

# JIOS

## Journal of Information and Organizational Sciences



**VOL. 50, NO. 1, PP. 1-262, JUNE 2026**

UDC 004:005:3

DOI: <https://doi.org/10.31341/jios>

## **Publisher**

University of Zagreb, Faculty of  
Organization and Informatics, Varaždin, Croatia

## **Editor-in-Chief**

Igor Balaban, University of Zagreb, Faculty of  
Organization and Informatics, Varaždin, Croatia

## **Editorial Secretary**

Goran Hajdin, University of Zagreb, Faculty of  
Organization and Informatics, Varaždin, Croatia

## **Editorial Board**

Alen Lovrenčić, University of Zagreb, Faculty  
of Organization and Informatics, Croatia  
Bart Rienties, The Open University, UK  
Blaženka Divjak, University of Zagreb, Faculty  
of Organization and Informatics, Croatia  
Daniel Rodriguez, University of Alcalá, Spain  
Diana Šimić, University of Zagreb, Faculty of  
Organization and Informatics, Croatia  
Dragutin Kermek, University of Zagreb, Faculty  
of Organization and Informatics, Croatia  
Ivan Luković, University of Belgrade – Faculty  
of Organizational Sciences, Serbia  
Ján Paralič, Technical University of Košice,  
Slovakia  
Krzysztof Podlaski, University of Lodz, Faculty  
of Physics and Applied Informatics, Department  
of Intelligent Systems, Poland  
Maria José Escalona Cuaresma,  
University of Seville, Spain  
Markus Schatten, University of Zagreb, Faculty  
of Organization and Informatics, Croatia  
Matej Černe, School of Economics and Business  
University of Ljubljana, Slovenia  
Michael Lang, University of Galway, Ireland  
Michal Kvet, University of Žilina, Slovakia

## **Publishing Board**

Darko Grabar, University of Zagreb, Faculty of  
Organization and Informatics, Croatia  
Mladen Konecki, University of Zagreb, Faculty of  
Organization and Informatics, Croatia  
Bernarda Kos, University of Zagreb, Faculty of  
Organization and Informatics, Croatia  
Ladislav Cvetko, University of Zagreb, Faculty of  
Organization and Informatics, Croatia

## **For Publisher**

Marina Klačmer Čalopa, University of Zagreb, Faculty  
of Organization and Informatics, Varaždin, Croatia

## **Advisory Board**

Jasminka Dobša, University of Zagreb,  
Faculty of Organization and Informatics, Croatia  
Neven Vrček, University of Zagreb,  
Faculty of Organization and Informatics, Croatia  
Dijana Oreški, University of Zagreb,  
Faculty of Organization and Informatics, Croatia  
Ivan Malbašić, University of Zagreb,  
Faculty of Organization and Informatics, Croatia

Mike Joy, University of Warwick, UK  
Mirjana Kljajić Borštnar, University of Maribor,  
Faculty of Organizational Sciences, Slovenia  
Nina Begicevic Redjep, University of Zagreb,  
Faculty of Organization and Informatics, Croatia  
Petra Grd, University of Zagreb, Faculty of  
Organization and Informatics, Croatia  
Roman Povalej, Police Academy of  
Lower Saxony, Germany  
Slobodan Beliga, University of Rijeka, Faculty of  
Informatics and Digital Technologies, Croatia  
Tomislav Hernaus, University of Zagreb,  
Faculty of Economics and Business, Croatia  
Veljko Jeremic, University of Belgrade,  
Faculty of Organizational Sciences, Serbia  
Vicente Julian, Universitat Politècnica  
de València, Spain  
Vjeran Strahonja, University of Zagreb, Faculty  
of Organization and Informatics, Croatia  
William Steingartner, Technical University of Košice,  
Faculty of Electrical Engineering  
and Informatics, Slovakia  
Zlatko Stapić, University of Zagreb Faculty  
of Organization and Informatics, Croatia

## **Senior communication and dissemination officer**

Josipa Bađari, University of Zagreb, Faculty of  
Organization and Informatics, Croatia

## CONTENTS

|                                                                                                                                                                                                                                                                            |                |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| <b>Mateja Gorenc and Janez Likozar:</b><br><b>Optimization of Batch Processing Through Advanced Management in Public Institution X</b><br>Original Scientific Paper                                                                                                        | <b>1-15</b>    |
| <b>Ivana Petkovski and Petar Vranić:</b><br><b>Evaluation of Digital Development Indexes Using MEREC and Hybrid MCDM Methods</b><br>Original Scientific Paper                                                                                                              | <b>17-43</b>   |
| <b>Raghad Baker Sadiq, Nur Fazidah Elias and Hazura Mohamed:</b><br><b>The Meta-analysis based Modified Unified Theory of Acceptance and Use of Technology Model: A Systematic Review</b><br>Original Scientific Paper                                                     | <b>45-64</b>   |
| <b>Shofia Amin, Akhmad Habibi, Siti Hajar Halili, Hamdy Abdullah, Muthia Shahnaz, Rafiqi Rafiqi and Nela Safelia:</b><br><b>Factors Affecting Behavior, Perceived Impact of AI on Work Engagement, and AI Application: Indonesian Context</b><br>Original Scientific Paper | <b>65-82</b>   |
| <b>Matija Novak and Marija Maček:</b><br><b>AI Tools in Programming Education: Student Perspectives and Usage Trends</b><br>Original Scientific Paper                                                                                                                      | <b>83-104</b>  |
| <b>Yudhy Setyo Purwanto, Rahmat Gernowo and Dinar Mutiara Kusumo Nugraheni:</b><br><b>User Experience in Mobile Language Learning Applications (MLLA): Insight from Mixed-Methods Triangulation Study</b><br>Original Scientific Paper                                     | <b>105-127</b> |
| <b>Raghad Baker Sadiq, Nur Fazidah Elias and Hazura Mohamed:</b><br><b>Factors Affecting Lecturers' Intention to Adopt Artificial Intelligence Tools for Assessing Students' Performance in Malaysian Universities</b><br>Original Scientific Paper                        | <b>129-155</b> |
| <b>Matea Cvjetković:</b><br><b>Why Communication Dimensions Matter: Change Readiness Under Workplace Stressors In SMEs</b><br>Original Scientific Paper                                                                                                                    | <b>157-175</b> |
| <b>Vu Nguyen Huynh Anh, Nguyen Ton Nu Khoa and Vy Nguyen Dinh Nhat:</b><br><b>Institutional and Resource-Based Drivers of E-Commerce Adoption in Relationship-Embedded Wholesale Markets: A Business Model Component Analysis</b><br>Original Scientific Paper             | <b>177-193</b> |
| <b>Cemil Ceylan, Caner Tacoglu and Gozde Koc:</b><br><b>Consultant Competencies and ERP Project Success: The Mediating Role of ERP Implementation Critical Success Factors</b><br>Original Scientific Paper                                                                | <b>195-209</b> |
| <b>Candraningrat, Nimas A. Pratiwi and Azmil Chusnaini:</b><br><b>Smart Agriculture in Southeast Asia: Digitization of Smallholders through Organizational Commitment and Collective Ambition</b><br>Original Scientific Paper                                             | <b>211-232</b> |
| <b>Rosalina, Noor Lees Ismail, Genta Sahuri and Joseph Tedja Nugraha Wibawa:</b><br><b>Embedding Retrieval-Augmented Generation into SME Customer Communication Work Practices</b><br>Original Scientific Paper                                                            | <b>233-246</b> |
| <b>Sabrina Nasrin Saima, Abdus Salam, Mir Md. Kawsur and K. M. Imtiaz-Ud-Din:</b><br><b>An Advanced Stacked Ensemble Framework for Enhanced Software Bug Predictions</b><br>Original Scientific Paper                                                                      | <b>247-262</b> |

## Erratum to: Smart Agriculture in Southeast Asia: Digitization of Smallholders through Organizational Commitment and Collective Ambition

Journal of Information and Organizational Sciences, Vol. 50, No. 1 (2026), pp. 211-232  
<https://doi.org/10.31341/jios.50.1.11>

In the originally published version of the article, the institutional affiliation of the third author, **Azmil Chusnaini**, was stated incorrectly as:

**Department of Management, Universitas Nahdlatul Ulama, Surabaya, Indonesia**

The correct affiliation is:

**Department of Management, Universitas Nahdlatul Ulama Surabaya, Surabaya, Indonesia**

The correction concerns the official name of the institution only and does not represent a change in the author's institutional affiliation. The author's affiliation with Universitas Nahdlatul Ulama Surabaya at the time of submission and acceptance of the article has been formally confirmed by the institution.

The original article remains unchanged. This erratum constitutes the formal correction to the published record.

### The Editorial Office

*Journal of Information and Organizational Sciences (JIOS)*

# Optimization of Batch Processing Through Advanced Management in Public Institution X

Mateja Gorenc<sup>1\*</sup> and Janez Likozar<sup>2</sup>

<sup>1</sup>International School for Business and Social Studies, Celje, Slovenia

<sup>2</sup>B2 College of business sciences, Ljubljana, Slovenia

\*Correspondence: [mateja.gorenc@vspv.si](mailto:mateja.gorenc@vspv.si)

## PAPER INFO

### *Paper history:*

Received 09 September 2025

Accepted 30 December 2025

### *Citation:*

Gorenc, M. & Likozar, J. (2026). Optimization of Batch Processing Through Advanced Management in Public Institution X. In Journal of Information and Organizational Sciences, vol. 50, no. 1, pp. 1-15

### *Copyright:*

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

The article discusses the optimization of batch processing in a public institution through the implementation of a proprietary centralized batch management system. The research is based on the analysis of processing log entries from 2018 to 2023 and the implementation of a proprietary information interface developed in C# and connected to an Oracle database. The study highlights the importance of operator roles, structured work orders, and socio-technical alignment between technology and organizational processes. The analysis confirms that advanced planning significantly reduces processing time, thereby improving operational efficiency. However, the impact on overall process success is limited, as reliability appears to depend on additional organizational and infrastructural factors.

**Keywords:** batch processing, process optimization, log entries, automation, information systems

## 1. Introduction

In the modern information environment of public sector organizations, managing large volumes of data has become central to ensuring operational efficiency and the reliability of business processes. Particularly important are so-called batch processes, which form the backbone of the information infrastructure in many institutions where large datasets must be processed routinely and regularly. These are automated or manual computing procedures executed within predefined timeframes or triggered by specific events, with their success often depending on a complex network of technical, logical, and temporal dependencies.

Public institutions frequently face challenges such as fragmented execution of processes, limited centralized control, outdated workflows, and insufficient documentation, all of which increase operational risks and reduce efficiency. Although digitalization has improved many administrative functions, batch management in the public sector remains less advanced compared to private-sector environments. This gap is reinforced by differences in organizational constraints: public institutions typically operate within rigid legal frameworks, face resource limitations, and rely on legacy information systems that are not easily adaptable to modern optimization approaches. These factors influence not only the feasibility but also the impact of structured batch scheduling.

Despite the extensive research in financial and manufacturing industries, considerably less is known about how structured, centralized scheduling affects efficiency and reliability in public-sector systems. The novelty of this research lies in examining a real large-scale institutional environment over a six-year period

and evaluating the effects of an internally developed centralized scheduling tool. The study combines socio-technical, contingency, and optimization perspectives to provide empirical evidence on how planning, dependency monitoring, and operator support influence batch process duration and success.

The goals of this research are threefold. First, we aim to quantify the impact of centralized and structured planning of batch processing on execution time and success rates in a real public-sector environment. Second, we seek to identify which aspects of planning and monitoring, such as dependency handling, automation of triggering, and operator oversight, contribute most to observed improvements. Third, we explore how insights obtained from log data can inform the design of future analytical tools and decision-support mechanisms. For IT managers and system administrators, the expected value of this work lies in concrete evidence on how a centralized scheduling tool can reduce delays and highlight remaining reliability issues. For institutional leadership and policymakers, the results offer a basis for investment decisions related to information infrastructure and human resources. For operators and technical staff, the study clarifies how structured work orders and clearer responsibilities can reduce operational uncertainty and error risk.

The research problem addressed is thus the following: *to what extent can centralized and structured management of batch processes improve duration and reliability in a public-sector institution, and what contextual factors determine the effectiveness of such interventions?*

Unlike studies that propose novel optimization algorithms or fully autonomous scheduling mechanisms, this paper does not aim to introduce a new scheduling paradigm. Its contribution lies instead in empirically evaluating the actual effects and limitations of centralized, rule-based batch planning in a real public-sector environment. By explicitly documenting both efficiency gains and their boundaries, the study seeks to provide a realistic and evidence-based understanding of centralized batch management, avoiding overly optimistic assumptions about automation as a universal solution.

The structure of the paper is as follows. Section 1 introduces the context, motivation, and research problem, and outlines the key challenges associated with batch processing in public-sector information systems. Section 2 presents the characteristics of batch processing and its relevance in the information systems of public institutions. Section 3 provides the theoretical framework of the study, which includes socio-technical approaches, the theory of constraints, and contingency theory. Section 4 formulates the research hypotheses. Section 5 describes the methodology, the analysed data, and the proprietary system for centralized batch management. Section 6 presents the results of the statistical analysis and hypothesis testing. Section 7 offers recommendations for improving existing processes. Section 8 concludes with the main findings and suggests possible directions for future research.

## 2. Batch Processing

Batch processing is a computational method in which a collection of tasks or application programs is executed sequentially without the need for direct user interaction. These tasks are typically carried out at specific time intervals, such as overnight or during periods of low system load, in order to optimize the use of computing resources and minimize disruptions to regular system operations. Batch processing thus represents the execution phase of an application program that operates based on predefined and clearly specified instructions (Thomas, 2024).

In financial institutions, batch processes often involve complex chains of tasks that handle large volumes of transactional data. For example, in payment processing, a batch workflow may include steps such as collecting and validating transaction data, posting to the general ledger, generating reports, and creating customer statements. Each step in this chain depends on the successful completion of the previous one, highlighting the importance of precise planning and control over the entire process (Khalfe, 2024).

Additionally, it is crucial to ensure that all data is properly stored and protected from loss or damage. This includes regularly creating data backups, especially before any changes or updates are made, to allow system restoration to a previous state in case of errors or failed processes. Doing so maintains data integrity and ensures the accuracy of outcomes, which is essential for user trust and regulatory compliance (Nicole, 2024).

Batch processing plays a critical role in the information systems of large organizations for several reasons:

- **Resource Optimization:** Batch processing enables the use of low-demand system periods to execute resource-intensive tasks, thereby avoiding system overload during peak hours and improving overall resource utilization efficiency (McHugh, 2024).
- **Automation of Repetitive Tasks:** By automating repetitive tasks, batch processing reduces the need for manual intervention, resulting in more consistent and reliable execution. This enhances productivity and minimizes the risk of human error (Talend, 2024).

- **Processing Large Volumes of Data:** Designed for the efficient handling of massive datasets, batch processing is particularly important for organizations dealing with extensive data, such as financial institutions, the healthcare sector, and government agencies. It allows for the processing and analysis of large data sets without straining system resources during peak periods (McHugh, 2024).
- **Reliability and Accuracy:** Batch processes ensure a high level of reliability and accuracy by executing tasks in a systematic and predictable manner, which is vital in business environments where data precision is critical (IBM, 2024).
- **Increased Operational Efficiency:** With centralized management and monitoring of batch processes, organizations can enhance operational efficiency. This includes tracking performance, optimizing processes, and responding swiftly to errors or failures. Centralized control enables better task organization and ensures alignment with business requirements (McHugh, 2024).
- **Integration with Various Systems:** Batch processing supports integration with different information systems and databases, simplifying complex business workflows and enhancing connectivity across multiple applications (McHugh, 2024).

Batch processing is a cornerstone of modern information systems, enabling automation, optimization, and reliable execution of numerous routine and data-intensive tasks. Through this, organizations achieve greater efficiency, reduced costs, and improved accuracy—factors essential for competitiveness and success in today's market (McHugh, 2024).

At the institution, we perform various types of batch processing:

- **Planned Processing:** These are pre-arranged processes executed according to a defined processing schedule.
- **Scheduled Processing:** These processes are not part of the main schedule but are automatically executed according to a predefined timetable, every day of the year, including weekends and holidays. Such processes ensure the continuous execution of critical operations, contributing to uninterrupted system functionality and compliance with operational requirements.
- **On-Demand Processing:** These are initiated at the request of the client, based on specific needs.
- **One-Time Processing:** These are processes that need to be executed only once, usually coordinated with a programmer or analyst.

The study builds on several theoretical perspectives that explain how structured planning and centralized management influence the performance of information systems. From the perspective of information systems theory, the socio-technical approach emphasizes the interdependence between technological solutions and organizational practices. In this view, batch processing efficiency cannot be achieved by technical improvements alone but requires alignment with organizational structures such as operator roles, work orders, and standardized procedures.

### 3. Theoretical Framework

The theoretical foundation of this study is based on socio-technical systems theory (STS), which emphasizes that the effectiveness of information systems depends on the alignment between technical solutions and organizational as well as human factors. As early as Trist and Bamforth (1951), it was demonstrated that technological improvements alone do not lead to greater efficiency if they are not simultaneously adapted to work practices and the social dynamics of the organization. Mumford (2006) highlighted the importance of participatory design, which fosters greater acceptance of information systems, while Alter (2015), through the development of Work System Theory, extended STS by offering a more holistic understanding of the relationships between people, processes, and technology. In the context of batch processing in public institutions, this means that centralized process scheduling requires both technical and organizational adaptations—from the standardization of procedures to the training of operators.

Another important theoretical foundation is contingency theory, which asserts that there is no universally optimal organizational design; rather, the effectiveness of specific practices depends on the given context. Lawrence and Lorsch (1967) demonstrated that the success of organizational structures arises from their adaptation to environmental complexity, while Donaldson (2001) expanded this perspective within a broader understanding of organizational contingencies. In the field of operations management, Sousa and Voss (2008) provided empirical evidence that structured planning yields the greatest benefits under conditions of high interdependence and uncertainty. More recently, scholars such as Zheng (2025) have developed adaptive scheduling models that respond dynamically to unexpected disruptions in real time. For public institutions, where the environment is often unpredictable and systemic constraints are significant, such an approach is particularly relevant.

The next body of theory derives from process optimization and the Theory of Constraints (TOC). Goldratt and Cox (1984) demonstrated that an organization's effectiveness is determined primarily by its ability to identify and manage bottlenecks. Within the field of workflow management, Van der Aalst and Van Hee (2004) developed models for analysing and optimizing task dependencies, while Méndez et al. (2006) showed through a comprehensive review of methods that systematic scheduling of batch processes significantly improves throughput and reduces processing times. More recent research (e.g., Feng et al., 2024; Zheng, 2025) has expanded this approach by incorporating multi-objective goals such as energy efficiency and cost reduction, indicating that process optimization extends beyond purely technical improvements to encompass broader aspects of sustainable organizational performance.

An additional perspective is offered by Structuration Theory, developed by Giddens (1984). This theory is based on the assumption that structures shape individual behaviour, while at the same time individuals reproduce and transform those structures through their actions. Orlikowski (1992) demonstrated that technology in organizations does not exist as a neutral tool but instead acquires meaning through its enactment in practice. In public institutions, this implies that the introduction of centralized batch scheduling represents not only a technical innovation but also a reorganization of work practices that evolves through the interaction of employees with the system.

Equally important is the Social Construction of Technology (SCOT) theory, developed by Pinch and Bijker (1987). According to their view, technology does not operate in a deterministic manner; rather, its meaning and function emerge through the interpretations and negotiations of different social groups. In the case of public institutions, this suggests that policymakers, information system developers, administrators, and end-users may each perceive and evaluate centralized scheduling in different ways, which in turn influences its actual effectiveness. This perspective explains why the introduction of centralized scheduling may successfully reduce processing times but does not necessarily increase process reliability, as this outcome also depends on the social acceptance and integration of the technology into everyday work practices.

#### 4. Hypotheses

In this study, the notion of “planning the execution of batch processing” refers to the set of activities and mechanisms introduced with the centralized batch management system. Concretely, this includes: defining and maintaining a centralized schedule of batch jobs; configuring temporal conditions (time windows, execution frequency) and logical dependencies (precedence constraints, prerequisite checks); linking batch jobs to standardized work orders and documentation; and enabling operators to monitor and adjust execution via a dedicated interface. Planning, as operationalized in our analysis, is therefore not an abstract concept but a socio-technical configuration in which temporal rules, dependency management, and organizational procedures are explicitly encoded into the system.

The objective of this research is to empirically test three hypotheses related to the impact of processing duration on success rates, the influence of planning on duration, and the influence of planning on the success of batch processes. Through quantitative analysis of real processing log data from the period 2018–2023, the article aims to demonstrate opportunities for process optimization in data-intensive environments and offer practical recommendations for improving information infrastructure within the public sector.

##### *H1: The Duration of Batch Processing Affects Its Success.*

Hypothesis H1 is based on several theories and empirical observations confirming that processing duration is a key indicator of process stability, resource availability, and execution success. Authors such as McHugh, Shiff, and Gaur emphasize that prolonged processing often reflects structural or systemic deficiencies that reduce system reliability.

According to McHugh (2024), processing time is a direct indicator of system efficiency, as longer durations often lead to bottlenecks and system overload. While each process does not require the same execution time, unusually long durations frequently signal failures in the processing chain, especially when involving complex technical dependencies or high system loads.

Shiff (2024) emphasizes that in sequentially dependent batch processes, timing synchronization is a critical factor. If one task takes too long, it disrupts the entire schedule, increasing the risk of failure for subsequent tasks, particularly those that are time-sensitive or depend on data from earlier stages. Such delays are not merely postponements; they can lead to invalid calculations or missing data, ultimately resulting in failed processing.

Mastering Batch Processing (2024) highlights that in systems with complex technical or logical dependencies, processes often execute more slowly due to waiting for resources or confirmations. When these

dependencies are not properly aligned, task durations increase, which can cause downstream process failures. This not only extends the total system time but also raises the likelihood of cascading errors.

Gaur (2024) stresses the importance of log data analysis for identifying patterns between duration and failure. In practice, processes that exceeded the average execution time by more than 25% had nearly twice the likelihood of failure, indicating that duration is a reliable risk indicator. This directly supports the theoretical foundation of Hypothesis H1.

### *H2: Planning the Execution of Batch Processing Reduces Processing Time*

Hypothesis H2 is based on the assumption that structured, pre-defined planning of batch processes enables better resource distribution, eliminates unnecessary delays, and ensures an optimal execution sequence. As a result, it shortens the overall processing time. This link between planning and processing duration has theoretical roots in workflow management, system optimization, and job scheduling theory.

Shorten (2024) emphasizes that effective scheduling is a key factor in reducing processing time in data-intensive environments. In systems lacking clear temporal and sequential planning, delays often occur due to resource contention, task conflicts, and duplicated procedures—all of which directly extend execution time. In contrast, dynamic planning and intelligent task orchestration allow for adjusting task order based on priorities, current system load, and historical performance.

Gaur (2024) highlights that systems like Apache Airflow and IBM Tivoli Workload Scheduler have demonstrably reduced process durations by enabling: dependency recognition, automatic scheduling based on optimal sequencing, and parallel execution where logical constraints allow. This shows that planning is not merely an administrative task but an active optimization mechanism that directly influences the temporal efficiency of processes.

nOps (2024) further points out that one of the most common causes of extended processing times is an unoptimized task sequence. When a system cannot predict resource availability or identify independent tasks, it often defaults to unnecessary serial execution, where parallelism would be possible. By introducing planning mechanisms that consider resource availability, time windows, and critical paths, bottlenecks can be eliminated, leading to significantly shorter execution times.

### *H3: Planning the Execution of Batch Processing Affects Processing Success*

Hypothesis H3 is based on the assumption that structured and deliberate planning of batch executions positively influences their success by enabling better system preparation, conflict resolution, error reduction, and improved availability of data and resources. Numerous studies support the idea that process organization significantly contributes to the reliability and quality of processing outcomes.

According to Johnson (2024), planning in data processing systems is not merely a temporal function but also a qualitative mechanism for ensuring stability and predictability. When the execution order is clearly defined, dependencies are resolved in advance, and the system is aware of the time windows during which tasks are executed, errors related to missing input data, resource access conflicts, or human mistakes during manual triggering are greatly reduced. Precise planning thus fosters higher process discipline, increasing the likelihood of successful completion.

Talend (2024) states that planning also introduces procedural standardization, meaning that processes are executed under consistent rules, at predictable times, and with the appropriate resource allocation. This consistency significantly lowers the risk of errors stemming from improvisation or undocumented interventions. From a quality management perspective, planning is therefore a prerequisite for success.

Nicole (2024) emphasizes that in systems where processing is initiated without planning, the probability of unexpected events, such as interruptions, unresolved dependencies, or system overload, increases. These conditions heighten the risk of failure. In contrast, planning introduces predictability, enabling systems to reserve resources appropriately and resolve potential conflicts, thus contributing to higher processing success rates.

Maronde (2024) notes that the success of processing is not solely dependent on planning, but also on factors such as operator competence, input data quality, and infrastructure reliability. Nevertheless, planning helps reduce errors caused by organizational deficiencies, functioning as a stabilizing factor even in complex environments.

While H1–H3 capture baseline expectations derived from the literature regarding duration, planning, and success, the empirical rejection of H1 and H3 suggests that other factors may be more decisive for processing success. Instead of formulating additional a priori hypotheses, we therefore complement our analysis with exploratory research questions for future work, such as:

- Which non-temporal factors (e.g., data quality, infrastructure stability, operator interventions) best explain variations in processing success?

- How do different configurations of planning (e.g., stricter dependency rules vs. more flexible scheduling) affect trade-offs between efficiency and robustness?

These questions extend the current hypothesis set and provide a roadmap for subsequent studies.

Although planning did not significantly increase processing success, the results highlight the importance of operator oversight and structured work orders. Unsuccessful cases were more frequently associated with inadequate input data, manual handling errors, or unrecorded dependencies—factors unrelated to system scheduling. This indicates that technological improvements must be complemented by organizational adjustments, including clearer documentation, operator training, and standardized work order procedures, consistent with socio-technical systems theory.

## 5. Methodology

Methodologically, the research was structured into four main phases. In the first phase, the proprietary centralized batch management system was designed and implemented, encapsulating the planning mechanisms described in Section 4. In the second phase, log data from 2018 to 2023 were extracted, cleaned, and segmented into pre-implementation and post-implementation periods, enabling a natural comparison between “non-planned” and “planned” execution regimes. The third phase consisted of descriptive analysis of key variables, processing duration, success rates, and processing types, providing a quantitative overview of the institutional workload. In the fourth phase, we conducted inferential statistical tests to evaluate Hypotheses H1–H3. In this way, each hypothesis is directly linked to a specific operationalization in the data and to a clearly defined analytical step.

The development phase of the research focused on the implementation of centralized batch process management. For this purpose, a dedicated user interface was developed in the C# programming language, enabling the management, monitoring, and triggering of batch processes. The system was tightly integrated with the existing Oracle database, allowing real-time access to metadata and processing logs.

The main functionalities of the developed system include:

- automatic dependency checks between processes,
- triggering based on temporal and logical conditions,
- visualization of schedules and tasks, and
- alerting in the event of processing failures.

To evaluate system performance, a database of batch processing logs from 2018 to 2023 was used. The dataset included the following elements:

- start and end timestamps of each process,
- process identifier,
- execution outcome (successful/unsuccessful),
- processing type (manual/automatic),
- associated work order.

Before analysis, the data had to be cleaned, removing missing, inconsistent, and duplicate records. The data were then segmented into two periods: before and after the implementation of the new system. The system was implemented in early 2020.

The proprietary batch management system does not implement adaptive or self-optimizing scheduling algorithms. Instead, it operates as a centralized, rule-based system that enforces predefined temporal schedules, logical dependencies, and execution conditions. Its primary function is to replace fragmented manual coordination with a unified control mechanism, ensuring consistent execution order, explicit dependency checks, and improved visibility for operators. The efficiency gains observed in the analysis therefore stem from reduced manual intervention, clearer sequencing of tasks, and the elimination of ad hoc execution practices rather than from dynamic optimization or predictive decision-making.

This study has several limitations that must be considered when interpreting the results. The research was conducted within a single public institution, which means that the findings cannot be directly generalized to all public organizations. The dataset used is based primarily on technical process logs, which provide precise measurements of efficiency but do not capture softer factors such as user satisfaction, employee competencies, or organizational culture. Methodologically, the study relies mainly on a quantitative approach, which limits the depth of contextual understanding that could have been enriched through qualitative methods. Furthermore, the analysis covers a specific time period, implying that subsequent technological or organizational changes may affect the replicability of the results. The analysis covers the period from 2018 to 2023, with 2018–2019 representing the pre-implementation period and 2020–2023 the post-implementation period.

6. Results

Log Data Analysis

The log data serve as a critical component for monitoring and analyzing batch processes at the institution. This system records all essential information, including process start and end times, execution outcomes, and processing results. These logs provide not only a clear overview of process performance but also valuable insights for identifying errors and enhancing procedures in the future.

The analysis of the institution’s log data also reveals a significant increase in the number of batch processes, particularly those executed on a scheduled basis. The data clearly show a year-over-year growth in the volume of automatically executed processes, while the number of manual processes has remained relatively stable.

| Year | Scheduled Processing | Manual Processing |
|------|----------------------|-------------------|
| 2018 | 41.562               | 18.570            |
| 2019 | 48.420               | 19.250            |
| 2020 | 53.516               | 17.557            |
| 2021 | 135.326              | 18.130            |
| 2022 | 195.464              | 16.978            |
| 2023 | 312.054              | 17.689            |

Table 1. Growth in scheduled and manual processing (2018-2023).

Table 1 presents the number of scheduled and manual batch processes from 2018 to 2023. The analysis of the data shows a clear and significant increase in scheduled processing over the observed period, while manual processing remained relatively stable with minor fluctuations. Between 2018 and 2019, scheduled processing increased by approximately 16.49%, whereas manual processing rose by only 3.66%. In 2020, scheduled processing continued to grow by 10.52%, while manual processing decreased by 8.79%. The most significant increase in scheduled processing occurred between 2020 and 2021, with a 152.82% rise in the number of executions. In contrast, manual processing increased by just 3.27%. In 2022, scheduled processing grew again by 44.43%, while manual processes declined by 6.36%. From 2022 to 2023, scheduled processes surged once more by 59.66%, while manual processes showed a slight increase of 4.19%. These figures clearly indicate a continuous upward trend in scheduled processing, particularly after 2020, reflecting greater automation and centralization of batch execution. In contrast, manual processing remained relatively stable, with only modest changes over time.

Analysis of Conditions and Dependencies

Conditions and dependencies are critical elements in the execution of batch processes, as they directly influence the correctness, sequence, and success of individual tasks. Every batch process depends on specific conditions that must be met to ensure a proper and orderly execution. These conditions may include time constraints, the status of preceding processes, the availability of input data, or predefined parameters that must be fulfilled before the process can start.

In the institution’s system, these conditions are explicitly recorded in a dedicated column, where they are clearly defined for each individual process. This field contains information about the prerequisites that must be met prior to execution. Often, these dependencies relate to previously completed processes or the status of system resources—such as file availability, access to data sources, or other processing requirements.

For manual processes, where an operator oversees execution, these conditions are especially critical. The operator must verify that all prerequisites are fulfilled before initiating a task. For instance, if a process is dependent on the outcome of another task, it is essential that the operator confirms the successful completion of the preceding process. Failure to meet these conditions can result in errors or incorrect outputs.

| Processing ID | Conditions   | Pre-Processing | Pre-Processing Performance | Status |
|---------------|--------------|----------------|----------------------------|--------|
| 1001          | Conditions A | 1000           | Successful                 | OK     |
| 1002          | Conditions B | 1001           | Failed                     | STOP   |

Table 2. Conditions and dependencies between processes.

Table 2 illustrates how individual batch processes are interconnected and how the success of preceding processes affects the overall workflow. Each process is subject to certain conditions and is dependent on the successful completion of a prior process. If the previous process is successful, the next one proceeds; otherwise, the sequence is halted.

Table Columns:

- Processing ID: Unique identifier for each batch process.
- Conditions: Specific requirements that must be fulfilled before execution.
- Pre-Processing: The process that must successfully complete beforehand.
- Pre-Processing Performance: Indicates whether the prior process succeeded.
- Status: Whether the current process can proceed ("OK") or is blocked ("STOP").

In automated processing, these conditions are verified through system rules and schedules. Automated condition checking ensures that processes are only triggered when all necessary criteria are met. The system evaluates: whether previous tasks have completed successfully, if all required input data is available, and whether any time constraints are satisfied.

Process Dependencies play a vital role because many batch jobs are linked in chains or sequences. A successful process execution is often a prerequisite for the next one to begin. Dependency analysis enables identification of critical control points, where potential failures or delays could cascade through the entire batch sequence. This makes understanding dependencies essential for effective planning and optimization of task execution.

At the institution, dependencies are either temporal or logical. Temporal dependencies relate to when tasks must run, at specific intervals or in a defined order. It is crucial that scheduling prevents resource conflicts (e.g., database or network contention). Logical dependencies are tied to the outcome of previous processes. A task may only proceed once the prior process has completed successfully and the necessary data is prepared.

Managing these dependencies and conditions is essential for the smooth operation of the batch processing system. Analysing dependencies and preconditions allows the system to: accurately determine execution order, prevent errors caused by improper sequencing, and reduce the risk of delays and system faults.

By systematically analysing conditions and dependencies, organizations can significantly improve the management of batch processing, leading to greater reliability, efficiency, and overall system performance.

Before proceeding with hypothesis testing, a descriptive analysis of the data must be conducted to gain a foundational understanding of the sample characteristics. This step involves summarizing key statistical indicators that provide insights into the structure and variability of the data.

The following metrics will be calculated:

- Means (Averages) – to determine the central tendency of processing times, success rates, and other key variables.
- Standard Deviations – to assess the variability and dispersion within the dataset.
- Distributions – to observe the frequency and patterns of specific values (e.g., successful vs. unsuccessful executions, scheduled vs. manual processing).

This initial overview will help identify trends, outliers, and potential anomalies, and will provide the necessary context for a more accurate and meaningful hypothesis testing process.

|                         | Duration of processing execution in minutes on a work order |
|-------------------------|-------------------------------------------------------------|
| Number (of work orders) | 1522                                                        |
| Mean                    | 161,64717                                                   |
| Standard deviation      | 80,21125                                                    |
| Minimum                 | 23                                                          |
| Maximum                 | 473                                                         |
| 25 %                    | 80                                                          |
| 50 %                    | 184                                                         |
| 75 %                    | 221                                                         |

**Table 3.** Descriptive statistics on processing duration on work orders.

Table 3 shows data on the duration of batch processing in minutes per work order, with a total of 1,522 work orders analysed. The average processing duration is 161.65 minutes, indicating that most processes take slightly more than two and a half hours. The standard deviation is 80.21 minutes, which means that the duration of processing can vary significantly between individual tasks. The shortest process lasted 23 minutes, while the longest took 473 minutes. Table 3 also presents the quartiles of processing duration. One quarter of

the processes lasted 80 minutes or less, half of the processes were completed in 184 minutes or less, and one quarter of the processes lasted more than 221 minutes.

|                         | Work order performance |
|-------------------------|------------------------|
| Number (of work orders) | 1522                   |
| Number of successful    | 1406                   |
| Number of unsuccessful  | 116                    |
| Minimum                 | 0 - Successful         |
| Maximum                 | 1 – Unsuccessful       |

**Table 4.** Descriptive statistics on the success of work order processing.

Table 4 shows the success rate of work orders, with a total of 1,522 orders analysed. Out of these, 1,406 were successfully executed, representing approximately 92.38% of all orders, while 116 orders failed, accounting for about 7.62%. The minimum success value is 0, meaning that in a certain period all orders were unsuccessful, while the maximum value is 1, meaning that in a certain period all orders were successfully executed.

|                         | Duration of processing execution in minutes on a work order |
|-------------------------|-------------------------------------------------------------|
| Number (of work orders) | 976                                                         |
| Mean                    | 14,46926                                                    |
| Standard deviation      | 22,09812                                                    |
| Minimum                 | 0                                                           |
| Maximum                 | 148                                                         |
| 25 %                    | 3                                                           |
| 50 %                    | 6                                                           |
| 75 %                    | 15                                                          |

**Table 5.** Descriptive statistics on the duration of selected treatments from 2018 to 2023.

Table 5 presents a descriptive analysis of two key variables related to the selected processes: processing duration and processing success. The analysis is based on a sample of 976 selected processes. The average processing duration is 14.47 minutes, with a standard deviation of 22.10 minutes, indicating high variability in the duration of individual processes. The shortest process lasted 60 seconds, while the longest process lasted 148 minutes. One quarter of all processes lasted less than 3 minutes, half of the processes were completed in less than 6 minutes, and one quarter lasted more than 15 minutes. The results show that the majority of processes were successful and that the duration of individual executions varied considerably, with the average time being slightly under 15 minutes.

|                        | Treatment performance | Treatment performance selected |
|------------------------|-----------------------|--------------------------------|
|                        | Frequency             | Percentage                     |
| Number of treatments   | 976                   | 100 %                          |
| Number of successful   | 975                   | 99,9 %                         |
| Number of unsuccessful | 1                     | 0,01 %                         |

**Table 6.** Descriptive statistics on the success of selected treatments from 2018 to 2023.

The table 6 presents the success rate of the selected processes, with a total of 976 processes analysed. Out of these, 975 were successful, which means that the vast majority of processes were completed successfully. Only one process failed, indicating a very high success rate.

*H1: The duration of batch processing affects its success.*

In this analysis, we tested Hypothesis H1, which claims that the execution time of batch processes influences their success. By comparing the time frames of process execution with their outcomes, we examined whether there is a statistically significant relationship between processing duration and success.

| Performance      | Mean<br>(in minutes) | Standard deviation | N    | ANOVA<br>F- statistics | Sig.    |
|------------------|----------------------|--------------------|------|------------------------|---------|
| Successful (0)   | 162,21692            | 79,32931           | 1406 | 0,93073                | 0,33482 |
| Unsuccessful (1) | 154,74137            | 90,31482           | 116  |                        |         |

**Table 7.** Comparison of duration of successful and unsuccessful treatments (2018-2023).

Table 7 presents a comparison of the average duration between successful and unsuccessful processes. The average duration of successful processes is 162.22 minutes with a standard deviation of 79.33 minutes, while unsuccessful processes last on average 154.74 minutes, with a standard deviation of 90.31 minutes. The number of successful processes was 1,406, while there were 116 unsuccessful processes. The results of the ANOVA analysis show an F-statistic of 0.93073 and a p-value of 0.33482, which indicates that there is no statistically significant difference between the duration of successful and unsuccessful processes. Based on the statistical analysis, we reject H1, as the duration of batch processes does not affect their success.

The rejection of Hypotheses 1 processing time is one of the key predictors of success in data-intensive systems; Gaur (2024), who claims that longer durations are often associated with a higher number of errors due to bottlenecks; and Shiff (2024), who highlights the cascading effect of delays leading to errors in dependent processes.

In the specific case of the analysed public institution, we may conclude that the quality of process execution was likely more dependent on other factors, such as infrastructure stability, input data quality, or operator supervision, rather than on the duration of the tasks themselves.

*H2: Planning the Execution of Batch Processing Reduces Processing Time.*

Hypothesis H2 is based on the assumption that structured planning of batch processing has a significant impact on time efficiency. This thesis is supported by several authors, such as Shorten (2024), Gaur (2024), and nOps (2024), who emphasize that well-scheduled processing tasks enable better utilization of resources, prevent waiting times, and eliminate unnecessary temporal overlaps within complex data workflows.

| Period      | Mean<br>(in minutes) | Standard deviation | N    | ANOVA<br>F- statistics | Sig.  |
|-------------|----------------------|--------------------|------|------------------------|-------|
| Before 2020 | 207,85317            | 60,75913           | 504  | 299,04597              | 0,000 |
| After 2020  | 138,77111            | 78,83241           | 1018 |                        |       |

**Table 8.** Comparison of processing times before and after 2020.

Processes were grouped into two periods: before 2020 and after 2020. The transition year 2020 is included in the post-implementation dataset, as proprietary centralized batch management system was introduced early that year. The table presents a comparison of the average duration of batch processes before 2020 and after the introduction of improved planning. Before 2020, the average duration of processes was 207.85 minutes with a standard deviation of 60.76 minutes, indicating longer but more consistent execution times. After 2020, the average duration decreased to 138.77 minutes, reflecting a reduction of approximately 69 minutes per process. The number of processes before 2020 was 504, while after 2020 it increased to 1,018. The results of the ANOVA analysis show an F-statistic of 299.05 and a p-value less than 0.5, which confirms that Hypothesis H2 can be accepted. We can conclude that planning of batch processes has a statistically significant effect on reducing execution time. After the introduction of advanced planning, the average execution time decreased by 69 minutes, confirming that careful process planning plays a crucial role in improving efficiency and accelerating execution.

These results are consistent with the findings of: Shorten (2024), who notes that real-time intelligent planning leads to shorter execution times in data systems. nOps (2024), who highlights that more than 40% of processing delays are caused by non-optimal task sequencing. Gaur (2024), who confirms that dynamic dependency management increases system throughput.

In this research, the results confirm that advanced planning has a key impact on reducing process duration, as it enables better scheduling control, timely activation of dependent tasks, and optimal coordination in highly complex systems. This is particularly important in public institutions, where process stability and time synchronization directly affect the quality of services.

*H3: Planning the Execution of Batch Processing Affects Processing Success.*

In this analysis, we tested Hypothesis H3, which states that planning of batch processes affects their success. By comparing process success before and after the implementation of planned processes, we analysed whether the introduced changes led to fewer errors, improved timeliness of execution, and greater user satisfaction.

| Period      | The proportion of successful |
|-------------|------------------------------|
| Before 2020 | 92,06349                     |
| After 2020  | 92,53438                     |

**Table 9.** Comparison of successful processing before and after implementation of planning.

The table presents a comparison of the share of successful processes before 2020 and after the implementation of improved planning. Before 2020, the share of successful processes was 92.06%. After 2020, this share slightly increased to 92.53%. Although the difference is small, the data indicate a slight improvement in success following the implementation of planned processes. Based on these results and the minimal difference in success rates, it can be concluded that planning did not have a significant impact on process success. On the basis of statistical analyses, Hypothesis H3 is rejected, meaning that the planning of batch processing execution does not significantly influence process success.

Taken together, the results for H2 and H3 suggest an important asymmetry in the effects of planning as operationalized in this study. On the one hand, planning, implemented through centralized scheduling, explicit temporal rules, and dependency management, has a strong and statistically significant effect on execution time (H2), confirming that such mechanisms are effective levers for improving temporal efficiency. On the other hand, the same planning configuration has only a negligible effect on success rates (H3). This indicates that “planning the execution”, as we define it here, primarily addresses when and in which order tasks are executed, but does not sufficiently intervene in other determinants of success, such as input data quality, infrastructure reliability, and exceptional event handling. In other words, the planning regime is effective in smoothing the temporal dimension of processing but only weakly influences the deeper socio-technical sources of reliability.

This finding suggests that the effect of planning on success is less pronounced than its effect on duration, which aligns with the conclusions of Johnson (2024), who argues that planning influences efficiency more strongly than it does execution quality, except in highly complex or critical workflows. Similarly, Maronde (2024) points out that organizational changes, such as the introduction of planning, may have limited impact on success if the main causes of failures lie elsewhere, for instance, in data quality, infrastructure robustness, or unpredictable system outages.

The results of this study confirm that centralized and structured scheduling of batch processes significantly reduces execution time and thereby increases the efficiency of information systems in a public institution. This finding advances existing knowledge by providing empirical evidence consistent with process optimization and workflow management theories, which emphasize the importance of eliminating bottlenecks and improving task sequencing. However, the study also shows that scheduling has only a limited impact on overall process success, indicating that factors such as infrastructure robustness, data quality, and operator competence remain equally important. This nuance extends previous research, which often assumes a stronger link between planning and reliability, by demonstrating that in the context of public institutions, efficiency gains do not automatically translate into higher levels of success.

Although the finding that centralized scheduling reduces execution time may appear intuitive, its empirical confirmation in a public-sector environment is non-trivial. Many studies implicitly assume that automation and planning also improve reliability; however, the rejection of H1 and H3 provides empirical evidence against this assumption. The results demonstrate that, in complex institutional settings, efficiency gains achieved through centralized scheduling do not automatically translate into higher success rates. This negative finding is theoretically relevant, as it challenges simplified narratives of automation-driven reliability and underscores the need to distinguish between temporal efficiency and execution robustness.

By critically reflecting on these findings, the study contributes to the broader theoretical debate in the literature on organizations and information systems. Specifically, it demonstrates that socio-technical alignment is essential: improvements in efficiency require not only technological solutions but also organizational adaptation, training, and standardized procedures. The rejection of Hypotheses 1 and 3 highlights the need to revisit assumptions within contingency theory and structuration theory, emphasizing that process success in highly complex and interdependent environments depends on more than scheduling mechanisms alone. Concretely, our findings suggest at least three directions for such refinement. First, future contingency-theoretic work should explicitly model infrastructure robustness, data quality, and regulatory

constraints as contextual variables that moderate the impact of planning mechanisms, using comparative designs across institutions with different technical and organizational profiles. Second, structuration-inspired studies could investigate how operators appropriate or circumvent centralized scheduling tools in practice, for instance through qualitative case studies or mixed-method designs that combine log analysis with interviews and observations. Third, empirical criteria for evaluating success should be broadened beyond binary execution outcomes to include measures such as recovery time after failures, stability across peak-load periods, and the effort required from human operators. These approaches would allow a more fine-grained assessment of when and how centralized scheduling creates value, and where complementary interventions are necessary.

Beyond the specific institutional case, the results have broader implications for other public-sector and organizational contexts facing similar digitalization challenges. Many government agencies and large organizations operate with legacy systems characterized by high interdependencies and limited resources. The evidence presented here suggests that centralized scheduling can be an effective intervention to reduce delays and increase efficiency in such contexts, even if it does not fully resolve issues of reliability. Thus, the study provides practical insights for policymakers and practitioners while also offering a framework that can be generalized to a wide range of organizations seeking to modernize their information infrastructures through structured process management.

The results of the analysis clearly show that non-temporal factors play a decisive role in explaining variations in processing success. Failures most commonly arise from issues such as incomplete or inconsistent input data, infrastructure instability, database locks, and fluctuations in system load, factors that are independent of execution timing or scheduling quality. Operator-related aspects, including manual overrides, mis-sequenced task initiation, or deviations from documented procedures, also significantly influence outcomes. Furthermore, processes embedded in long or fragile dependency chains are particularly vulnerable, as any upstream interruption propagates across subsequent tasks regardless of the underlying planning regime. These findings demonstrate that the determinants of processing success lie primarily in data quality, infrastructure robustness, operator competence, and the structural characteristics of dependency networks, which explains why planning mechanisms substantially improved temporal efficiency (H2) but did not meaningfully affect overall success rates (H3).

The analysis also highlights that different configurations of planning introduce varying trade-offs between efficiency and robustness. Stricter planning configurations, such as rigid dependency enforcement, comprehensive validation of prerequisites, and narrow execution windows, enhance robustness by restricting execution to conditions that minimize the risk of failure. However, these constraints reduce flexibility, limit parallelism, and may increase overall processing time. In contrast, more flexible configurations allow broader time windows, dynamic triggering, and greater concurrency, which improves throughput but increases exposure to resource contention, synchronization problems, and reduced operator oversight. A hybrid approach, in which critical or failure-sensitive processes follow stringent rules while routine tasks operate under more flexible conditions, offers a balanced solution. This indicates that the design of planning mechanisms matters as much as their presence: planning must be tailored to the technical, organizational, and operational characteristics of the environment if institutions are to achieve both efficiency gains and improved reliability.

## 7. Suggestions for Improvements

The proposed recommendations are directly derived from the empirical findings and the theoretical framework adopted in this study. Since the analysis shows that failures are primarily associated with non-temporal factors, such as data quality, infrastructure instability, and operator interventions, rather than with scheduling mechanisms, the suggested improvements deliberately focus on enhancing informational transparency, analytical support, and organizational coordination rather than on further tightening execution schedules.

Based on the analysis conducted and the review of current processes, the following improvements are recommended for the Institution:

1. **Enhancement of Log Records:**

Currently, log records mainly capture timestamps (start and end of processing) and execution success. While this provides basic monitoring, it would be beneficial to extend records with more detailed information. Standardized and richer logs that include all necessary process parameters would allow deeper insights and advanced analytics, such as machine learning and artificial intelligence. This would make the system more modern and prepared for future technological trends, while also supporting further process optimization. In particular, logs should include

information on resource usage (memory), error codes, dependency identifiers, data input validation results, operator actions, queue wait times, and system load at execution time. These fields would enable predictive analytics and support the identification of structural failure patterns.

2. **Standardization of Records:**

To ensure more effective use of data in the future, log records should be standardized. This would make analysis consistent regardless of the type or scope of processing. Standardization would simplify historical process analysis, help predict potential issues, and support more accurate planning of future processes.

3. **Development of Analytical Tools:**

To fully leverage past process data, tools should be developed to predict process success and duration based on historical records. This would allow more precise planning and reduce the likelihood of future errors. To fully leverage past process data, the institution should develop analytical tools that provide early-warning signals and decision support based on historical patterns. For example, dashboards could visualize distributions of processing times and failure rates by job type, time of day, or dependency chain, enabling operators to detect emerging bottlenecks. Statistical or machine-learning models could estimate the probability of failure for an upcoming execution based on features such as past duration, recent infrastructure incidents, or the length of the dependency chain. Such tools would not replace operators but would support them with concrete risk scores and alerts, allowing them to prioritize monitoring and preventive interventions for the most vulnerable processes.

4. **Use of Artificial Intelligence and Machine Learning:**

Artificial intelligence and machine learning can further enhance batch management by learning complex patterns that are difficult to capture with simple rules. For instance, anomaly detection algorithms could automatically flag executions whose duration or resource usage deviates significantly from historical norms, even if they have not yet failed. Sequence models could identify combinations of jobs and temporal conditions that are systematically associated with later disruptions. Recommendation systems could propose more robust scheduling configurations (e.g., alternative time windows or parallelization options) based on past successes. To make such approaches feasible, the institution must first ensure standardized and high-quality logs, as outlined in points 1 and 2, and then iteratively validate AI-generated insights with operators before automating any actions.

5. **Centralized Management and Planning:**

Analysis results have shown that centralized management and careful planning significantly reduce operational delays and improve process performance. With an even more systematic approach to planning, processes could be further optimized, reducing the risk of failures.

Log records from past processes have great potential for improving decision-making in proprietary centralized batch management systems. However, technical and content upgrades are necessary to unlock this potential. Unified and more detailed records would enable the use of advanced technologies, such as AI and machine learning, to optimize processes, making the system more aligned with modern trends and better equipped to predict and resolve issues in a timely manner.

Nevertheless, challenges remain, such as incomplete or inconsistent data in the current logs, which limits forecasting and optimization capabilities. For effective use of this data, consistency and accuracy must be ensured, along with robust security and protection mechanisms. Moving forward, investments should focus on improving log quality and developing tools that enable efficient forecasting and process optimization.

## 8. Conclusion

The research has shown that the introduction of advanced, centralized management significantly contributed to the reduction of batch processing duration in the analysed public institution. After the implementation of the centralized planning system, processing times decreased noticeably, confirming that careful planning and centralized process management have a positive impact on operational efficiency. At the same time, the results did not indicate a statistically significant effect on overall success rates, as the proportion of failed processes remained almost unchanged before and after the implementation. This asymmetry suggests that temporal optimization and reliability are shaped by partly different mechanisms.

This study does not claim to offer a breakthrough in automated batch scheduling or dynamic optimization. Instead, its contribution lies in providing a realistic and empirically grounded assessment of centralized batch management in a public-sector context. The findings show that while centralized, rule-based planning is an effective tool for reducing execution time, it has limited influence on processing success, which

depends on broader socio-technical conditions. By explicitly documenting these limitations, the paper contributes to a more balanced understanding of digital transformation in the public sector and cautions against overstating the capabilities of automation without corresponding organizational and infrastructural investments.

From a practical standpoint, the findings offer several benefits for stakeholders. For IT managers and system architects, the study provides quantitative evidence that investments in centralized scheduling and dependency management can yield substantial time savings, even in environments with legacy systems and limited resources. For operators, the results highlight where planning helps, by reducing manual interventions and clarifying execution sequences, and where it does not, namely in preventing failures caused by data quality issues or infrastructure incidents. For policymakers and organizational leaders, the study clarifies that digitalization efforts focused solely on automation and scheduling will improve efficiency but will not, on their own, guarantee higher reliability or service quality.

The results reported in this paper can be developed further in several directions. First, future work should extend the analysis beyond a single institution and compare public organizations with different technical infrastructures and governance models; to test how contextual factors moderate the effects observed here. Second, incorporating qualitative data, such as interviews with operators, administrators, and users, would allow a deeper understanding of how centralized scheduling is appropriated in practice and how informal workarounds influence outcomes. Third, the proposed improvements in logging and analytical tools could be implemented and evaluated in a pilot setting, for example by testing whether predictive models and anomaly detection reduce the number or impact of failures over time.

In a broader perspective, the study contributes to the ongoing discussion on digital transformation in the public sector. It demonstrates that centralized scheduling is a powerful instrument for improving efficiency, but not a universal remedy for all reliability problems. Achieving sustainably reliable batch processing requires a combination of technological enhancements (centralized management, better logging, analytical tools), organizational measures (clearer work orders, training, process standardization), and human resource development. Only by jointly addressing these technical and organizational dimensions can public institutions fully realize the potential of advanced batch management and move towards more resilient, data-intensive service provision.

## 9. References

- Chia, L. (2024). Batch Processing Analysis in Large Institutions: SWOT Perspective. *Journal of Information Systems*, 33(2), 123–135.
- Donaldson, L. (2001). *The contingency theory of organizations*. Sage Publications.
- Feng, Y., Zhang, H., & Sun, X. (2024). Energy-efficient batch scheduling with makespan and electricity cost minimization. *Mathematics*, 12(3), 376.
- Gaur, V. (2024). Efficient Workflow Orchestration in Public Systems. *Systems Engineering Review*, 40(1), 45–63.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press.
- Goldratt, E. M., & Cox, J. (1984). *The goal: A process of ongoing improvement*. North River Press.
- IBM. (2024). *Modern Batch Processing Systems and Job Control Language*. IBM Knowledge Center. <https://www.ibm.com/docs>
- Johnson, T. (2024). Cloud-Based Optimization of Batch Workflows. *Proceedings of the IEEE CloudTech Conference*, 2024.
- Khalfe, R. (2024). Transaction Data Flows and Batch Dependencies in Financial Institutions. *FinTech Data Journal*, 29(3), 188–201.
- Lawrence, P. R., & Lorsch, J. W. (1967). *Organization and environment: Managing differentiation and integration*. Harvard Business School Press.
- Maronde, P. (2024). Organizational Change and IT System Implementation. *Public Sector Informatics Review*, 15(1), 59–74.
- Mastering Batch Processing. (2024). *Principles of Batch Job Scheduling and Execution*. <https://www.masteringbatch.org>

- McHugh, D. (2024). Automation and Resource Optimization in Data-Intensive Institutions. *Information Management Quarterly*, 12(4), 212–230.
- Méndez, C. A., Cerdá, J., Grossmann, I. E., Harjunkski, I., & Fahl, M. (2006). State-of-the-art review of optimization methods for short-term scheduling of batch processes. *Computers & Chemical Engineering*, 30(6–7), 913–946.
- Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*, 16(4), 317–342.
- Nicole, B. (2024). Data Integrity and Backup Strategies for Critical Systems. *Journal of Data Safety*, 6(2), 88–97.
- nOps. (2024). Error Resilience and Dependency Mapping in Batch Workflows. White Paper Series. <https://www.nops.io/resources>
- Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science*, 3(3), 398–427.
- Pinch, T. J., & Bijker, W. E. (1987). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. V W. E. Bijker, T. P. Hughes & T. J. Pinch (ur.), *The social construction of technological systems* (17–50). MIT Press.
- Roth, S. (2023). Sociotechnical perspectives in Industry 4.0: Aligning technology and organizational practices. *Sustainability*, 15(23), 16438.
- Shiff, M. (2024). Dependency Management in Job Scheduling Systems. *ACM Transactions on Information Systems*, 41(1), 1–29.
- Shorten, C. (2024). Predictive Batch Processing using Real-Time Analytics. *DataOps Engineering Reports*, 18(2), 98–115.
- Sousa, R., & Voss, C. A. (2008). Contingency research in operations management practices. *Journal of Operations Management*, 26(6), 697–713.
- Talend. (2024). Data Pipeline Orchestration for Hybrid Environments. <https://www.talend.com/resources>
- Thomas, E. (2024). Foundations of Batch Processing and System Optimization. *IT Operations Handbook*, Vol. 2, 70–91.
- Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4(1), 3–38.
- Van der Aalst, W. M. P., & Van Hee, K. M. (2004). *Workflow management: Models, methods, and systems*. MIT Press.
- Zheng, S. (2025). Bi-objective scheduling of non-identical batch machines considering energy efficiency. *Annals of Operations Research*, 346(3), 145–167.



## Evaluation of Digital Development Indexes Using MEREC and Hybrid MCDM Methods

Ivana Petkovski<sup>1\*</sup> and Petar Vranić<sup>1</sup>

<sup>1</sup>Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

\*Correspondence: [ivana993@turing.mi.sanu.ac.rs](mailto:ivana993@turing.mi.sanu.ac.rs)

### PAPER INFO

#### Paper history:

Received 27 August 2025

Accepted 06 January 2026

#### Citation:

Petkovski, I. & Vranić, P. (2026). Evaluation of Digital Development Indexes Using MEREC and Hybrid MCDM Methods. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 17-43

#### Copyright:

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

### ABSTRACT

The use of relevant and structured instruments for measuring digital development is essential for policy-making in digitalization. The aim of the research is to compare structural adequacy of the global digital development indexes by means of multicriteria decision-making (MCDM). Theoretical contribution is to develop an evaluation framework and propose a novel methodological integration. Nine criteria were used to quantify six indexes: the Network Readiness Index (NRI), the E-Government Development Index (EGDI), the Digital Economy and Society Index (DESI), the ICT Development Index (IDI), the IMD World Digital Competitiveness Ranking (IMD) and the Global Digital Index (GDI). The criteria's objective weights were evaluated using the Method based on the Removal Effects of Criteria (MEREC) and the weights alterations effect was considered using the Shannon entropy method. The final prioritization was consolidated using five MCDMs scores: Combined Compromise Solution (CoCoSo), Measurement Alternatives and Ranking according to the Compromise Solution (MARCOS), Additive Ratio Assessment (ARAS), Combinative Distance-based ASsessment (CODAS) and Evaluation based on Distance from Average Solution (EDAS). Practical contribution and originality are presented by proposing first time evaluation framework of digital development indexes based on a recently proposed MEREC and selecting the most appropriate index (NRI) in a neutral MCDM context.

**Keywords:** Digital development, measurement frameworks, global digital indexes, multicriteria decision-making (MCDM), Method based on the Removal Effects of Criteria (MEREC), Combined Compromise Solution (CoCoSo), Measurement Alternatives and Ranking according to the Compromise Solution (MARCOS), Additive Ratio Assessment (ARAS), Combinative Distance-based ASsessment (CODAS), Evaluation based on Distance from Average Solution (EDAS)

## 1. Introduction

Providing a reliable and methodologically transparent measure of digital development of countries, despite a narrower or broader specialized scope, is a key tool for shaping the policy of digital progress (Bánhidi & Dobos, 2024). The existing literature offers a wide array of instruments for measuring digital development in the form of global indexes. The developed indexes (indices) were proposed by experts from world-renowned institutions in the field of technology such as Portulans Institute (PI), The Organization for Economic Co-operation and Development (OECD) and Huawei Technologies among others (Martínez et al., 2022; Huawei and IDC, 2024; Portulans Institute, 2024). Their availability greatly facilitates timely decision-making about

digitalization matters. However, there are significant variations between them. The most frequent differences concern geographical coverage, alongside more significant ones such as the structure of indicators and the choice of ranking methodology (Siqueira et al., 2019; Sidorov & Senchenko, 2020). Such variability can lead to heterogeneity among the scores of various indexes and provide different rankings among the same countries or regions depending on the selected index. While this could offer a second opinion on the digitalization issue, it could also compromise the entire decision-making process and prioritization for policymakers or investors since their decisions often rely on a specific index and its rank list.

Special branch of indexes is focused on digital development like the European Commission's Digital Economy and Society Index (DESI), or the United Nations E-Government Development Index (EGDI). These indices have become strong steering tools that shape governments in diagnosing gaps, setting agendas, and justifies digital reforms. By reducing complex socio-technical systems into comparable metrics, they became a platform for inter-state benchmarking and “soft” competition, that can influence agenda setting and resource allocation (Davis et al., 2012; EC, n.d.; EC, 2024). In the European Union DESI has transitioned from an independent scorecard toward formal monitoring by integrating itself with the Digital Decade Policy Program as its monitoring mechanism. It transitions gradually from a “critical evaluation” scorecard to a formal monitoring functioning. DESI trajectories and annual assessments via States of the Digital Decade, concerning progress towards 2030 targets and country-specific recommendations issued to member states, allow for the raised interfacing of indicator movement with meaningful policy guidance and cooperative governance cycles (EC, 2024; 2025). In order to align national with regional development, such metrics are often internalized by national governments in their strategies. For instance, in 2017 Portugal launched “National Digital Skills Initiative e.2030, (Portugal INCoDe.2030)”, an integrated public policy initiative, which aims to promote population's digital skills. According to INCoDe.2030, the plan is to consolidate the country's DESI standing, aligning investments in inclusion, skills, and research with the pillars of the scoreboard. Previous DESI country documentation also recorded the launch of INCoDe.2030 as a flagship strategy response to measured gaps (EC, 2018; International Trade Administration, 2024). Greece's “Digital Transformation Strategy” and later “Digital Decade” roadmap similarly placed reforms on the agenda because of low DESI rankings - an example of how rankings can increase attention and EU Recovery funding on connectivity, skills, and digitalization of public services (HRMDG, 2019). Another example beyond Europe is for instance rapid improvements in UN EGDI standings promoted in the Saudi Vision 2030 documents and Digital Government Authority communications as evidence of reform momentum, thus using index movement as legitimacy for investment in digital identity, portals and service redesign (DGA, 2022; Vision 2030, 2025).

Several authors have used digital development indexes in different areas. Jovanović et al. (2018) and Imran et al. (2022) use DESI to evaluate sustainable development of European Union countries while Adams & Paul (2023) use EGDI with the same purpose however on the African countries sample. Sofrankova et al. (2025) employ DESI to investigate the relationship between population digital skills and digital development. Ishnazarov et al. (2021) and Magoutas et al. (2024) discussed the ICT index in the context of digital economy progress while Oloyede et al. (2023) performed a similar analysis while limiting the sample towards developing countries. On the other side, Sagarik (2023) studies the use of the IMD index in assessing national digital competitiveness. Kolat & Ünver (2025) propose the use of the ICT index to reflect the question of cybersecurity in terms of economic development. On the other side, Fernández-Portillo et al. (2020) use DESI to examine the impact of digital progress on the trend of economic growth while Tokmergenova & Dobos (2024) use NRI to provide technological clusters among the countries.

Besides its practical use, index structure also matters: for instance, the World Bank Digital Adoption Index (DAI) opened itself as an explicit tool to policy-makers in the design of context-enabled strategies across people, business, and government while the Network Readiness Index (NRI) calls itself a “compass” for governments navigating digital transformations-a deliberate theory of use encouraging policy uptake (World Bank, 2016; Portulans Institute, 2024). These examples illustrate to how indicators have been moving into the policy processes through benchmarking (to identify strengths/weaknesses), conditionality, and cooperation (to align with regional targets and support from donors), and signaling (to attract investment and talent through reputational gains). However, following the advice from science, these opportunities bring risks with them: 1) metric-driven governance could narrow the attention to what is measured, 2) may induce isomorphic “teaching to the test”, and 3) may end up hiding the contextual realities in situations with lower capacity. Responsible usage must come with combining quantitative dashboards with qualitative diagnostics and inclusive deliberation (Davis, et al., 2012a; Merry, 2016).

Gerpott & Ahmadi (2015) emphasize that when selecting an adequate index of digital development for their study, all existing indexes failed to be developed as measures that can be connected with the economic or social development of a country. Gerpott & Ahmadi (2015) state that most of the indexes have structural gaps such as relying on subjective weighting of the component's importance. Therefore, a research gap exists in determining the most appropriate index in a neutral context at the macro level. At this point, a systematic

framework for evaluating digital development index is missing. The multi-criteria decision-making methodology emerges as an adequate choice for solving this problem. In this case, the problem is to prioritize the Network Readiness Index (NRI), the E-Government Development Index (EGDI), the Digital Economy and Society Index (DESI), the ICT Development Index (IDI), the IMD World Digital Competitiveness Ranking (IMD) and the Global Digital Index (GDI) that serve as alternatives. The prioritization of the alternatives is performed across multiple criteria. The similar examples are given in studies by Stević et al. (2020), Huy et al. (2022) and Dudić et al. (2024). In addition to the possibility of applying a larger number of MCDM methods in the prioritization context, the applied data normalization method also plays an important role (Dudic et al., 2024). The proposed MCDM methodological framework in this paper provides an objective and stable approach to decision-making about the choice of instruments for measuring digital progress with the simultaneous integration of different, often conflicting criteria.

The research objective of this study is to critically evaluate the structural adequacy of global digital development indexes in order to determine their methodological validity, broadness and relevance by means of multi-criteria analysis. The main idea is to compare the indexes based on their characteristics and prove model robustness. This paper does not state that researchers or other decision-makers should exclusively use one index over the all others. It simply tries to develop a comprehensive framework for easier decision-making about the adequate index when various context-neutral criteria are considered such as geographical coverage, data normalization, multidimensional structure or frequent revision and others. Since this represents a general perspective, the combination of criteria and alternatives will change in respect to the decision-making context and goals.

The main goal will be achieved through four phases:

1. Identification of alternatives (indexes) and evaluation criteria to form a decision-making matrix and determine the objective criteria importance weights with MEREC methodology,
2. Index ranking using five multi-criteria methods to check the reliability of hybrid prioritization models,
3. Aggregation of ranks and evaluation of similarities and differences of the obtained ranks.
4. Evaluation of rank robustness with alternative weighting method.

The scientific contribution of the research is based on the analysis of the index structural adequacy or in other words - index structural characteristics, which improves the measurement methodologies and the proposal of standardized approaches when developing new indices. The results of the prioritization assist in the decision-making process of selecting the most suitable index through the evaluation of index advantages and limitations. The proposed methodological framework simplifies the choice between several indexes. In this paper, the MCDM context is neutral and does not reflect any specific decision-maker objective. It only serves to compare existing indexes. Within a practical use of this framework, the decision-maker should set a specific goal and priorities and afterwards assess the considered indexes. It supports the fact that the proposed model is flexible and can be adjusted to the user's needs. As a limitation of the study the authors agree that the index measurement precision and quality in the form of scores are out of the scope of this paper. This limitation is, for example, a research subject of authors Miškuřová et al. (2025) who argue about the compliance between different ranking outcomes in digital competitiveness.

## 2. Literature Review

### 2.1. Digital development indexes

Although all digital indexes measure the digital progress of the world's economies, they are developed by different organizations and built on different methodologies.

The Network Readiness Index (NRI) is developed by the Portulans Institute in collaboration with partners such as Oxford University. The European Commission's Joint Research Centre - Competence Centre on Composite Indicators and Scoreboards audit NRI statistically to secure methodological thoroughness. The NRI (2024) is a composite that covers four main pillars: technology, people, governance, and impact, each with three sub-pillars. It comprises 54 indicators. For the construction of the index, primary objective data are used from renowned international sources. The index applies standard (min-max) normalization and employs equal weighting at each aggregation level. It tackles the digital divide by including both high- and low-income countries. The NRI specifically examines how countries utilize ICTs to enhance their competitiveness and overall development. It undergoes rigorous revision to suit the methodological updates and data availability. It is published annually. In the 2024 report, the index covers 133 countries, capturing a variety of economic development (95% of global GDP) (Portulans Institute, 2024).

The E-Government Development Index (EGDI) is published by the United Nations Department of Economic and Social Affairs (UN DESA). It is featured in its e-Government Survey. It focuses on three main pillars: provision of online services, telecommunication connectivity, and human capital, and thirteen sub-pillars with 191 indicators, if one counts every Online Service Index checklist item as an indicator. The survey reflects constructive improvements in the EGDI methodology, deriving from the lessons learned from previous editions. Each of the pillars is weighted equally and normalized by the Z-score method. The UNDESA irregularly revise EDGI indicators in respect to data availability and survey updates. Despite a stable core model, evolving technology and understanding shift the metrics' meanings, making the Survey a flexible comparative framework rather than a fixed linear path. It is developed on global survey frameworks and operationalized through consistent data sources, and it is released biannually. The report from 2024 covers 193 UN member states (UNDESA, 2022).

The ICT Development Index (IDI) is developed by the International Telecommunication Union (ITU). It is based on two main pillars: universal connectivity and meaningful connectivity, considering 11 indicators. The index values are normalized to a 0-100 scale by the min-max method. Methodology revisions are conducted roughly every four years following the ITU plenary processes. The index was suspended in 2018 and, after a methodology revision, was approved again in 2023 and a new edition is published in 2024. It regularly covers approximately 193 ITU member countries, but in 2024, 170 countries are considered, covering a wide spectrum of economies (ITU, 2024).

The Global Digitalization Index (GDI) is developed jointly by Huawei Technologies in collaboration with the International Data Corporation (IDC). It is built on four pillars: connectivity, digital platforms/foundation, sustainability/green energy, policy and ecosystem, and forty-two indicators normalized relative to defined targets. The data used are mostly objective data from bodies like ITU, WEF, UN, and World Bank acquired via Huawei and IDC measure frameworks. The index is published each year, with regular additions of indicators and country coverage to keep pace with digital transformation trends. It looked at seventy-seven countries constituting 93% of global GDP and 80% of world population at all development levels (Huawei and IDC, 2024).

The Digital Economy and Society Index (DESI) is developed by the European Commission to monitor the digital development of all EU Member States, covering diverse development levels. It is constructed of four equally weighted pillars: human capital, connectivity, technology integration and digital public services aggregated via simple weighted averages. It is organized through ten sub-pillars and a total of 32 indicators. The database is derived from Eurostat sources. DESI is press released annually and it is periodically updated (European Commission, 2022).

IMD World Digital Competitiveness Ranking (IMD) is an index that accumulates 59 criteria, where 38 belong to the hard data indicators and 21 survey responses. The structure is based on three main pillars: knowledge, technology and future readiness, assessed through 9 sub-pillars. Indicators are allocated balanced weights and normalized on scale from 0 to 100. IMD is published annually, with regular methodological updates that capture the digital transformation trend. It assesses 67 economies, covering a wide economic development range (IMD, 2024).

The Digital Government Index (DGI) is developed by the OECD. The DGI is constructed around six equally weighted pillar dimensions: digital by design, data-driven public sector, government as a platform, open by default, user-driven, and proactiveness, and thirty sub-pillars, evaluated through ninety-four questions covering each of the pillars. Data was drawn from a survey taken among senior government officials; thus, it considers a moderate degree of subjectivity. The DGI is published annually through the Government's at glance reports. Occasional changes are also being applied to the survey instrument and frameworks as well. The assessment includes 33 OECD member states, plus four accession countries (and sometimes partner regions), representing a range of economic contexts (OECD, 2023).

The Digital Adoption Index (DAI) is developed by World Bank. Data was published just once, in 2016, by the World Development Report of the World Bank, and has never been updated since. No further revisions have been made; hence, the data are now static. It covered roughly 180 countries across the whole economic spectrum. DAI is composed of three equally weighted sub-indices (people, government, business), each set of indicators built from objective supply-side data (World Bank, 2016).

Although similar in terms of general scopes, the indices differ significantly. Global-level annual composites such as NRI, EGDI, IDI, and GDI cover many economies based on a mix of infrastructure and skills, and governance are best for spotting gaps within the groups of income. Still, an equal rating of importance of components and normalization (min-max or z-scores) can all but mask crucial deficits, while OSI-style checklists tend to measure the service's availability and not its actual use or quality. Region or club-specific instruments (DESI for the EU; OECD's DGI) provide insights into public-sector capability but, by design, narrow variation-useful for benchmarking peers, less so for spotlighting extreme disparities. Survey-heavy indices (DGI; IMD's mix of hard data and executive opinions) can inject perception bias in a way that disfavors

lower-income contexts. Methodology-wise, frequent revisions keep pace with tech but hinder comparability over time. In principle, an index linking access (connectivity, affordability, device quality) with capability (skills, data readiness) and outcomes (usage, impact) across wide sets of countries with as little subjective input as possible shall paint the clearest picture of the divide. There still needs to be subnational, gender, and affordability indices, which remain blind spots of most national-level dashboards (Sidorov & Senchenko, 2020).

As Davis et al. (2012) commented, various international rankings essentially attempt to simplify complex institutional and organizational profiles into a comparable metrics, to the extent digestible for media, politicians, or financial experts and investors. This way ranking systems can influence both external perception as well as definition of internal, national priorities. In this form, composite institutional characteristics or country scores are translated to understandable signals that can inform the allocation of investments, by explaining socio-economic and institutional qualities through one standardized criterion; i.e., position on the ranking list can be turned into extent of foreign interests and investments (Groh et al., 2023).

The development of global indexes for measuring digital progress in the literature is addressed through multiple aspects that indicate their advantages but also numerous limitations.

## 2.2. Composite index considerations

Martínez et al. (2022) emphasize that the dimensionality and structure of the index are of crucial importance in the policy-making process because they contribute to a deeper understanding of the context and facilitate decision-making. This argumentation is supplemented by Lnenicka et al. (2022), who point out that the most common way of regulating index results is to alter the weight values of indicators or dimensions. When this strategy fails to yield the expected results, the structure itself must be changed. Such changes can lead to consistency problems making historical comparisons unfeasible. This challenge further complicates the position of developing countries, which are slower to implement new methodologies and lag behind developed countries in the preliminary reports.

The second important dimension of debate relates to the reliability and availability of datasets, as well as the index's universality. Sidorov and Senchenko (2020) show that common indexes such as NRI, EGDI and IDI do not offer universality in measurement because there is no complete database as a stable basis for their calculation. The authors propose an index derived from available data with the minimal presence of subjectivity in the calculations. Oloyede et al. (2023) offer a comparable view by discussing that the absence of a universal definition of the digitalization prevents the creation of comprehensive measuring instruments. The authors emphasize that databases from developing countries are often incomplete and inadequate, which makes analysis difficult. As one solution, Siqueira et al. (2019) use a dataset collected from a single source to develop a compatible index of digital development in enterprises. However, this approach has its limitations, since measuring once a year for a long period of time can be difficult because of the necessity to change indicators.

The third area of discussion concerns the way of assigning importance to certain indicators. Bánhidi and Dobos (2024) point out that assigning equal importance to indicators or dimensions, which are defined before data collection, is insufficient because it does not reflect the real informative value of the data. Instead, they propose that the significance weights should be based on the statistical characteristics of the set, which would increase the objectivity and precision of the measurement. These findings confirmed Denissova et al. (2025) who point to the problem of using outdated indicators that in the contemporary conditions fail to illuminate the level of digital development. The authors argue that such indicators do not allow measuring the disparities between urban and rural areas nor provide an assessment of important issues such as cyber security. As an alternative, it is proposed to introduce more modern measures, such as the level of automation. Further measures imply the simultaneous combination of subjective and objective indicators, which would ensure greater precision of the results and adaptability to the specific context.

Giovanni et al. (2005) note in their guide for creating an index in 10 steps that the accuracy of the data and their collection method are taken into account during the formation. In this regard, a controversy is placed on the surveying data because the correctness of the sample size and limitations must be analyzed. An additional dilemma is related to the data normalization because different methods yield different results. Those results can show sensitivity under extreme values or alteration of indicator weights. Then the principle of assigning weighting factors that directly affect the outcome is discussed. In that case, the authors suggest comparing the alignment of the obtained results with existing indices or indicators that measure the same or similar phenomenon. While Jiménez-Fernández et al. (2022) discuss the use of mathematical methods that enable the comparison of results within the index in a transparent manner. Greco et al. (2019) point out that the lack of differentiation in weights leads to confusion in the results. Certain indicators that are known to

have a direct impact, while proven in the literature, can be neglected. Such an approach loses the informative value and introduces subjectivity because it is a planned developer choice. This effect is especially clear in the case when the value of collinear indicators is doubled, which is also agreed upon by the authors Jiménez-Fernández et al. (2022).

3. Research framework

The problem of selecting an index with the most appropriate structure is recognized in the literature (Gerpott & Ahmadi, 2015). Even though attempts to discuss and compare digital development indexes exist, the majority of relative studies such as Oloyede et al. (2023) or Chaonan et al. (2024) are limited to literature reviews. Buturac (2024) provides a qualitative assessment of the methodology, limitations and advantages. The study of Bai et al. (2024) argues the structure and multidimensionality of recently developed indexes and supports the role of MCDMs in developing novel indices. In practice, modern studies such as Chen & Wu (2022) or Liang & Tan (2024) rely on the MCDM framework for generating a composite index. However, none of these studies use or develop an evaluation framework in the form of a tool for prioritizing existing indexes. Therefore, this paper addresses the identified gap among studies and evaluates digitalization indexes based on multi-criteria analysis. The list of mostly used and evaluated indexes is presented in Table 1.

| Index                                     | Abbreviation | Publisher                                                         | First release | Latest release |
|-------------------------------------------|--------------|-------------------------------------------------------------------|---------------|----------------|
| Network Readiness Index                   | NRI          | Portulans Institute (PI)                                          | 2002          | 2024           |
| E-Government Development Index            | EGDI         | United Nations (UN)                                               | 2001          | 2024           |
| The Digital Economy and Society Index     | DESI         | European Commission (EC)                                          | 2014          | 2024           |
| The ICT Development Index                 | IDI          | International Telecommunication Union (ITU)                       | 2009          | 2025           |
| Digital Government Index                  | DGI          | The Organisation for Economic Co-operation and Development (OECD) | 2019          | 2025           |
| Digital Adoption Index                    | DAI          | World Bank Group                                                  | 2014          | 2016           |
| IMD World Digital Competitiveness Ranking | IMD          | IMD World Competitiveness Center                                  | 2017          | 2025           |
| Global Digitalization Index               | GDI          | Huawei Technologies                                               | 2024          | 2024           |

Table 1. The common indexes for measuring digital development.

The research was divided into four phases, as illustrated in Figure 1:

**Phase one.** The evaluation of the criteria that express the structural adequacy of the most popular and latest global indexes of digital development is presented according to the established criteria in Table 2. The evaluation was performed using a recently reported MCDM method MEREC, adding to the objectivity of the analysis.

**Phase two.** The ranking of the digitalization indexes is performed. The ranking of indexes was realized using several MCDM methods with different methodological approaches to justify the stability of the results. In this procedure, newer methods in the literature introduced in the last decade such as CoCoSo, MARCOS, ARAS, CODAS and EDAS were used. These methods are used to show the ranking's reliability with the methods that have been confirmed in the last few years in the scientific literature.

**Phase three.** The result of the ranking presents an aggregated prioritization of indexes according to their structural adequacy and relevance. The overall prioritization was determined using the voting method, average absolute ranking differences and Spearman's correlation rank.

**Phase four.** The ranking's stability via the alternative weighting method is investigated. The authors use Shannon entropy to evaluate whether weight pounders have the capacity to modify rankings and how.

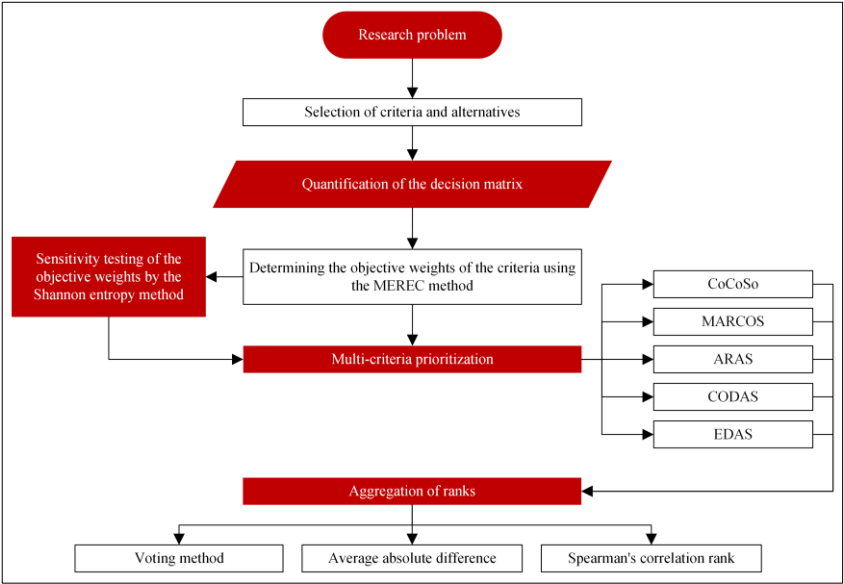


Figure 1. Research flow.

3.1. Multi-criteria decision-making framework

The preliminary review of the reports of world-renowned institutions in the field of digital development determined a large number of indexes, the basic information of which is shown in Table 1. The majority of reports are released for the year 2024, which was taken as the reference point. The DAI index and DGI index were eliminated from further research due to the lack of published report for 2024.

In this illustrative example the structural adequacy of the digitalization measurement indexes is evaluated based on a diverse criterion. The basic criteria with their definitions and qualitative and quantitative scales are shown in Table 2.

| No | Criterion                              | Criterion definition                                                     | Qualitative scale                               | Quantitative scale |
|----|----------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------|--------------------|
| 1  | Geographic coverage (GC)               | Number of countries included in the index.                               | High coverage > 100 countries                   | 3                  |
|    |                                        |                                                                          | Medium coverage of 50 - 100 countries           | 2                  |
|    |                                        |                                                                          | Low coverage < 50 countries                     | 1                  |
| 2  | Economic diversity of the sample (EDS) | Representation of countries of different levels of economic development. | Developed and developing countries              | 3                  |
|    |                                        |                                                                          | Developing countries                            | 2                  |
|    |                                        |                                                                          | Developed countries                             | 1                  |
| 3  | Number of indicators (NI)              | The total number of indicators that make up the index.                   | Large number of indicators > 20 indicators      | 3                  |
|    |                                        |                                                                          | Average number of indicators 10 - 20 indicators | 2                  |
|    |                                        |                                                                          | Small number of indicators < 10 indicators      | 1                  |
| 4  | Multidimensionality of the index (MI)  | The number of dimensions that make up the index.                         | > 2 dimensions                                  | 3                  |
|    |                                        |                                                                          | 2 dimensions                                    | 2                  |
|    |                                        |                                                                          | 1 dimension                                     | 1                  |

|    |                                      |                                                                     |                                                          |   |
|----|--------------------------------------|---------------------------------------------------------------------|----------------------------------------------------------|---|
| 5  | Coverage of the digital divide (CDD) | The coverage of the digital division that makes up the index.       | 3 dimensions of the digital divide                       | 3 |
|    |                                      |                                                                     | 2 dimensions of the digital divide                       | 2 |
|    |                                      |                                                                     | 1 dimension of the digital divide                        | 1 |
| 6  | Measurement approach (MA)            | The degree of subjectivity of the data on which the index is based. | Objective measurements (government's official reports)   | 3 |
|    |                                      |                                                                     | Combined measurements                                    | 2 |
|    |                                      |                                                                     | Subjective measurements (expert assessments and surveys) | 1 |
| 7  | Data source (DS)                     | Data origin.                                                        | International and official sources                       | 3 |
|    |                                      |                                                                     | Combined sources                                         | 2 |
|    |                                      |                                                                     | Private research and surveys                             | 1 |
| 8  | Normalization methodology (NM)       | Application of some of the standard data normalization methods.     | Standardized                                             | 3 |
|    |                                      |                                                                     | Partially documented                                     | 2 |
|    |                                      |                                                                     | Non-existent                                             | 1 |
| 9  | Weighting method (WM)                | Approach to determination of indicator weights.                     | Objective method                                         | 3 |
|    |                                      |                                                                     | Subjective method                                        | 2 |
|    |                                      |                                                                     | No weighting was applied                                 | 1 |
| 10 | Publication frequency (PF)           | Frequency of public release of index scores.                        | Regular publication = 1 year                             | 3 |
|    |                                      |                                                                     | Periodic publication = 2 year                            | 2 |
|    |                                      |                                                                     | Rare publication $\geq 3$ years                          | 1 |
| 11 | Indicator update frequency (IUF)     | Frequency of index indicator revision.                              | Periodic revision $\geq 2$ years                         | 3 |
|    |                                      |                                                                     | Regular revision = 1 per year                            | 2 |
|    |                                      |                                                                     | No revision                                              | 1 |

Table 2. Common criteria for evaluating digitalization indexes.

**Geographic coverage** - the number of countries included in the assessment of digital development provides significant information about the applicability of the index outcome to different territorial contexts. Broader coverage allows for regional or global cluster analysis (Oloyede et al., 2023) and policy formulation targeting marginalized or digitally less developed areas (Giovanni et al., 2005).

**Economic diversity of the sample** - enables the evaluation of countries that are in different stages of economic development. In this way, digital progress is also observed through the influence of the economic disparities within the sample (Siqueira et al., 2019; Oloyede et al., 2023). This informs policymakers about the emerging economies' issues regarding further digital development and guides the allocation of financial resources toward lagging countries.

**Number of indicators** - affects the equilibrium of the results because an excessive number can burden the structure of the index and reduce the precision of measurement due to excessive variability. An insufficient quantity of data simplifies the structure and may lead to inaccurate results due to a lack of useful information (European Union & Joint Research Centre, 2008). A high quantity of indicators can mislead policymakers when developing strategies with excessive information.

**Multidimensionality of the index** - fluctuates according to the number of areas associated with digital progress (Sidorov & Senchenko, 2020; Martínez et al., 2022). If the index refers to a specific field, then it is specialized for measuring digital progress within that area, while covering a larger number of fields, the index remains valid for measuring digital progress at the macro-level. This is important if policymakers plan interventions across multiple sectors.

**Coverage of the digital divide** - certain aspects of the digital divide are often analyzed in the literature, which according to the standard definition of the author Van Dijk (2005) includes three levels. The level of access, the level of use and skills, and the level of outcomes of the use of digital technology are very popular in contemporary literature and comprehensively depict the digital progress of a country. Ignoring some

aspects of the digital divide can present a distorted picture of digital progress (Denissova et al., 2025). This information helps policymakers in developing strategies for mitigating digital inequality.

**Measurement approach** - measurements of indicators of digital progress through the examination of experts' opinions or by conducting surveys introduce subjectivity in the formation of the index because the ratings are a representation of attitudes and biases of individuals (Sidorov & Senchenko, 2020). By using indicators that are physically measured, objectivity is added to the measurement process. A more reliable source of information is provided by objectively measured indicators. The use of subjective measures can present a disordered picture of digital development if an inadequate population is sampled.

**Data source** - well-known data sources have strong credibility and accuracy of information, while less prominent sources can be compromised due to lack of transparency in data collection (Oloyede et al., 2023). All official government strategies and action plans rely on formal reports so their accuracy cannot be questioned.

**Normalization methodology** - the raw collected data that form indexes are frequently expressed in different units of measure. To facilitate a precise comparison, it is necessary to normalize the collected data as proposed by Sidorov and Senchenko (2020). Different normalization techniques lead to diversity as stated by Greco et al. (2019). Therefore, policymakers need to be aware of these differences in results to correctly interpret and utilize them.

**Weighting method** - enables the allocation of the importance of indicators or dimensions in accordance with the impact on the assessment of digital progress (Sidorov & Senchenko, 2020). Different indicators provide a smaller or larger informative value to the formation of the index, so by giving weights, the influence of less significant indicators on the final score of the index is mitigated and the influence of important indicators is strengthened. This step contributes to balanced index results. For developing formal policies, important areas of digital development must be highlighted to prevent less important features from limiting the index accuracy (Giovanni et al., 2005).

**Publication frequency** - the time interval for the press release report with the updated index results (Siqueira et al., 2019; Lnenicka et al., 2022). Continuous publication allows policymakers to analyze the dynamics.

**Indicator update frequency** - the time frame for periodic adjustment of the structure and indicators of the index in line with trends in information society. This indicator points out to an emerging challenge that is the inconsistency of results produced by the same index after its revision (Lnenicka et al., 2022), a topic that is rare mentioned in the literature. By updating the index's structure, policymakers can follow upcoming trends in the area and update action plans and strategies accordingly.

The proposed framework is flexible in several terms. The study uses a three-point interval scale to maintain consistency in evaluating diverse criteria and to overcome potential subjectivity. Broader scales, such as five-point or seven-point scales, can provide detailed data granularity; however, in that case, experts' evaluations must be taken into account to justify derived values. Therefore, the three-point scale introduces simplicity into the decision-making process and allows transparent comparisons. It is based on the assumption of equal conceptual distances. Another important fact is that in this case, all criteria are considered as beneficial meaning that a higher evaluation is desirable. This rule can also be a subject of change depending on the decision-making context and decision-makers preferences. Decision-makers have the flexibility to add or remove any criteria and alternative (index) or to adjust the importance of each attribute that serves as a criterion. For instance, a wide geographical coverage may not always be an important question in selecting an adequate index. Alternatively, a public frequency may be irrelevant in the case when the countries' digital development is evaluated for one year.

The collinearity problem between distinct criteria was avoided due to clear definitions of the concepts they measure. Even if some of the criteria might look related, they contextually measure different aspects. For example, the number of indicators is diverse compared to the multidimensionality of the index because a higher number of indicators does not assure that different dimensions are being measured. In the same way, the regular publishing does not confirm that index structure is being revised to adjust to the specific digitalization trends. As a result, it may not be updated.

### 3.2. Research methodology

The applied research methodology followed a sequence consisting of the implementation of the MEREC method for calculating objective criteria weights and various MCDMs (CoCoSo, MARCOS, ARAS, CODAS and EDAS) for prioritizing global digital development indexes as alternatives.

The first step in all MCDM methods is to create a decision-making matrix (X) that is described as follows (Keshavarz-Ghorabae et al., 2021):

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}, i = 1, \dots, m; \quad j = 1, \dots, n, \quad (1)$$

where the element  $x_{ij}$  is the value of alternative  $i$  according to the  $j$ -th criterion. In this study, six alternatives ( $m=6$ ) in the form of indexes (NRI, EGDI, DESI, IDI, IMD and GDI) are evaluated using nine diverse criteria (GC, NI, MI, CDD, MA, DS, NM, PF and IUF) ( $n=9$ ). Decision matrix  $X$  reflects the input data into the MCDM methods while ranking lists are the outcome. The criteria can be divided into beneficial and non-beneficial (cost) type of criteria if they add or reduce to the value of prioritization goal. The importance of each criterion is described by an objective weight coefficient  $w_j$  where  $w$  is the weight value of the  $j$ -th criterion. The sum of all criteria weight coefficients is limited to 1 and expressed in equation (2).

$$\sum_{j=1}^n w_j = 1 \quad (2)$$

The initial weight coefficients are calculated using the MEREC method. Method inputs are defined as a decision-making matrix described in equation (1) and the outputs are the objective weights of the criteria. Additionally, the Shannon entropy method is implemented as a complementary method for sensitivity testing for weight alterations. This method has already been used in the area of digital development (Brodný & Tutak, 2015).

### 3.2.1. Objective weighting methods

**The Method based on the Removal Effects of Criteria (MEREC)** is a multi-criteria approach for objectively determining weighted values of criteria that is selected in this study. The method was developed in 2021 by the authors, Keshavarz-Ghorabae et al. (2021). The MEREC method varies from other methods for determining criteria weights because it does not rely on the criteria's variability (Ecer & Aycin, 2023). The application is based on the impact that the removal of each criterion individually has on the outcome of the alternatives (Kou et al., 2025). If the deviations of the value of the alternatives are greater upon eliminating a certain criterion, that criterion will assume greater importance (Štilić et al., 2024). Recent literature presents fuzzy form and hybrid MEREC with various multi-criteria methods (Abdelaal et al., 2024; Chaurasiya & Jain, 2024; Štilić et al., 2024; Kou et al., 2025). The basic steps in MEREC are presented as following (1-6):

1. Forming the decision-making matrix. The matrix is described as  $X$  expressed by equation (1).
2. Normalizing the decision-making matrix. The normalization is performed linearly, depending on the type of criterion (beneficial/non-beneficial) as shown in equation (3) (Chaurasiya & Jain, 2024).

$$n_{ij}^x = \begin{cases} \frac{\min x_{ij}}{x_{ij}} & \text{if } \text{beneficial} \\ \frac{x_{ij}}{\max x_{ij}} & \text{if } \text{non-beneficial} \end{cases} \quad (3)$$

3. Calculating the total value performance. The total value performances ( $S_i$ ) are calculated for each alternative according to the entire set of criteria with all criteria assigned equal significance as presented in equation (4) (Keshavarz-Ghorabae et al., 2021).

$$S_i = \ln \left( 1 + \left( \frac{1}{n} \sum_j |\ln(n_{ij}^x)| \right) \right) \quad (4)$$

4. Eliminating criteria. The next step follows by removing individual criteria and after each elimination, the outcome of alternatives without that criterion ( $S'_{ij}$ ) is calculated.

$$S'_{ij} = \ln \left( 1 + \left( \frac{1}{n} \sum_{k, k \neq j} |\ln(n_{ik}^x)| \right) \right) \quad (5)$$

5. Measuring deviations. Following the criteria elimination procedure, the value of absolute deviation ( $E_j$ ) is calculated separately for each criterion, indicating the variations in the alternative's outcome. This process is described in the equation (6).

$$E_j = \sum_i |S'_{ij} - S_i| \quad (6)$$

6. Calculating objective weights. The final objective weight coefficients are given by the equation (7) that expresses the elimination effect.

$$w_j = \frac{E_j}{\sum_k E_k} \quad (7)$$

A greater deviation resulting from the elimination of a criterion corresponds to a higher weight coefficient  $w_j$  assigned to the  $j$ -th criterion (Keshavarz-Ghorabae et al., 2021).

**Shannon entropy** is a method developed by Shannon (1948). Its concept belongs to the information theory and represents a measure of diversity among the information (Ali et al., 2023). It could also be referred to as a measure of uncertainty (Ali et al., 2023). If the distribution among the data is higher so the value of entropy will increase therefore the observed criterion will achieve a lower weight coefficient (Saraiva, 2023).

The elementary steps in implementing the Shannon entropy are given as follows (Shannon, 1948):

1. Forming the decision-making matrix. The matrix is described as matrix  $X$  presented by equation (1).
2. Normalizing the decision-making matrix. The formula for calculating the normalized values ( $p_{ij}$ ) is given as a ratio of element  $x_{ij}$  and a sum of the elements that belong to the same criteria.

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \quad (8)$$

3. Calculating the entropy.

$$E_j = -k \sum_{i=1}^m p_{ij} \ln(p_{ij}) \quad (9)$$

where  $k$  is a positive constant derived from a reciprocal natural logarithm of the number of alternatives ( $m$ ).

$$k = \frac{1}{\ln(m)} \quad (10)$$

4. Calculate the final objective weights. The procedure for determining the objective weights of each criterion is described in equation (11).

$$w_j = \frac{1 - E_j}{\sum_{j=1}^n (1 - E_j)} \quad (11)$$

### 3.2.2. Multi-criteria prioritization methods

**The Combined Compromise Solution (CoCoSo)** is a contemporary multi-criteria method that integrates simple additive weighting (SAW) and exponentially weighted product (EWP) (Dwivedi & Sharma, 2022). It was introduced to the public by Yazdani et al. (2019). The application procedure consists of multiple steps described below (1-5):

1. Forming the decision-making matrix. The matrix is described in matrix  $X$  expressed by equation (1).

2. Normalizing the decision-making matrix. The next step is the normalization of the decision-making matrix depending on the type of criteria as in equation (12) (Khan & Haleem, 2021).

$$r_{ij} = \begin{cases} \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} & \text{if } \textit{beneficial} \\ \frac{\max_i x_{ij} - x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} & \text{if } \textit{non-beneficial} \end{cases} \quad (12)$$

3. Summing the weighted values. The subsequent step involves summing the weighted values of each alternative ( $S_i$ ) presented by the equation (13) (Yazdani et al., 2019).

$$S_i = \sum_{j=1}^n (w_j r_{ij}) \quad (13)$$

4. Calculating the weighted product. The multiplication of the weighted values' product ( $P_i$ ) is implemented in equation (14), aligning with the EWP method (Yazdani et al., 2019).

$$P_i = \sum_{j=1}^n (r_{ij})^{w_j} \quad (14)$$

In the next step, three aggregate values (equations 15-17) are calculated as a combination of previously derived matrices. These values form the elements of the final aggregation according to which the alternatives are ranked (Yazdani et al., 2019).

$$k_{ia} = \frac{P_i + S_i}{\sum_{i=1}^m (P_i + S_i)} \quad (15)$$

$$k_{ib} = \frac{S_i}{\min_i S_i} + \frac{P_i}{\min_i P_i} \quad (16)$$

$$k_{ic} = \frac{\lambda(S_i) + (1-\lambda)(P_i)}{\left( \lambda \max_i S_i + (1-\lambda) \max_i P_i \right)} \quad (17)$$

where  $\lambda$  is a value between 0 and 1. In this case, the authors use the value of 0.5 that is typically used in such studies (Khan & Haleem, 2021).

5. Calculating the final rank values.

$$k_i = (k_{ia} k_{ib} k_{ic})^{\frac{1}{3}} + \frac{1}{3} (k_{ia} + k_{ib} + k_{ic}) \quad (18)$$

The multiple aggregation of the sum ( $k_{ia}$ ) and weights ( $k_{ib}$ ) and then the compromise solution ( $k_{ic}$ ) mitigates the effects that extreme values can have in the case of the application of only individual aggregation ( $k_i$ ), which could significantly distort the ranking results (Khan & Haleem, 2021). The advantage of this method is that the ranking is not sensitive to fluctuations in the criteria's value.

**Measurement Alternatives and Ranking according to Compromise Solution (MARCOS)** represents a novel multi-criteria decision-making method developed by Stević et al. (2020). The method involves determining the ideal (*ai*) and anti-ideal (*aai*) solution according to which the optimal solution or alternative is derived (Stanković et al., 2020; El-Araby, 2023). The calculation consists of the following several steps (1-7):

1. Forming the decision-making matrix. The initial decision-making matrix is described as matrix  $X$  expressed by equation (1).

2. Finding ideal and anti-ideal alternatives. In the case of beneficial criteria, the ideal is the maximum value and the anti-ideal is the minimum value in the decision-making matrix  $X$ . For cost criteria, the rule is in the opposite direction.

3. Normalizing the decision-making matrix. This step is carried out by dividing the individual value with the ideal or anti-ideal value in relation to the nature of the criterion as in equation (19) (Stević et al., 2020).

$$n_{ij} = \begin{cases} \frac{x_{ij}}{x_{ai}} & \text{if } \textit{beneficial} \\ \frac{x_{ai}}{x_{ij}} & \text{if } \textit{non-beneficial} \end{cases} \quad (19)$$

4. Weighting the normalized decision-making matrix. Subsequently, the normalized values are multiplied with the weighting coefficients ( $w_j$ ) that are calculated using another MCDM method, which is employed to calculate the significance of the criteria (Stanković et al., 2020). This is described in the equation (20):

$$v_{ij} = n_{ij} \times w_j \quad (20)$$

5. Calculating the utility coefficients. Afterwards, the utility coefficients for the ideal ( $K_i^+$ ) and anti-ideal ( $K_i^-$ ) solutions are determined as in equations (21) and (22). The  $S_i$  value is the sum of  $v_{ij}$  for each alternative, while  $S_{ai}$  are the summed ideal values of the ideal alternative and  $S_{aai}$  are the summed anti-ideal values of the anti-ideal alternative.

$$K_i^- = \frac{S_i}{S_{aai}} \quad (21)$$

$$K_i^+ = \frac{S_i}{S_{ai}} \quad (22)$$

6. Calculating the utility functions. The detailed formulas for calculating utility positive  $f(K_i^+)$  and negative  $f(K_i^-)$  functions are described subsequently in (23) and (24).

$$f(K_i^-) = \frac{K_i^+}{K_i^+ + K_i^-} \quad (23)$$

$$f(K_i^+) = \frac{K_i^-}{K_i^+ + K_i^-} \quad (24)$$

7. Calculating the total utility function. The total utility function serves as the foundation for ranking, where the higher value means a better-ranked alternative (Duc Trung, 2022). This coefficient value  $f(K_i)$  integrates the previous two values as demonstrated in equations (23) and (24) (Stević et al., 2020).

$$f(K_i) = \frac{K_i^- + K_i^+}{1 + \frac{1 - f(K_i^+)}{f(K_i^+)} + \frac{1 - f(K_i^-)}{f(K_i^-)}} \quad (25)$$

The MARCOS is robust when applying different methods for calculating criteria's weight coefficients, such as equal weighting, rank sum weighting, rank order centroid and entropy weighting (Nguyen et al., 2023). Another advantage is its compliance with other MCDMs such as simple additive weighting and preference selection index (Huy et al., 2022). However, the main drawback of this method is its susceptibility to outliers, because the calculations of utility coefficients are based on the minimum and maximum values of the initial decision matrix.

**The Additive Ratio Assessment (ARAS)** method is classified in the group of multi-criteria decision-making methods and was developed in 2010 by Zavadskas and Turskis (2010). The main advantage of the ARAS method is easy implementation in several steps and a comprehensible model (Zavadskas et al., 2010). The ARAS procedure includes the following steps (1-6):

1. Forming the decision-making matrix. The original decision-making matrix is expressed as matrix  $X$  in equation (1).

2. Finding the optimal value. If the criteria are beneficial the optimal value ( $x_{0j}$ ) is the maximum value of the criteria, in opposite if the criteria are non-beneficial the optimal value ( $x_{0j}^*$ ) is the minimum value of the criteria. Therefore, a hypothetical alternative with ideal values is formed as a base for comparison (Zavadskas & Turskis, 2010).

3. Normalizing the original decision matrix. The normalization of the decision matrix is presented by formula (26) that varies on the type of criteria.

$$\bar{x}_{ij} = \begin{cases} \frac{x_{ij}}{\sum_{i=0}^m x_{ij}} & \text{if } \textit{beneficial} \\ x_{ij} = \frac{1}{x_{ij}^*}; x_{ij} = \frac{x_{ij}}{\sum_{i=0}^m x_{ij}} & \text{if } \textit{non-beneficial} \end{cases} \quad (26)$$

4. Weighting the normalized decision-making matrix. The normalized values of the decision matrix are weighted ( $\hat{x}_{ij}$ ) by the equation (27).

$$\hat{x}_{ij} = \bar{x}_{ij} \times w_j \quad (27)$$

5. Calculating the optimality function. The aggregate weighted values of each alternative represent the optimality function ( $S_i$ ) described in formula (28) (Turskis & Zavadskas, 2010). The same procedure is applied to find the optimal alternative ( $S_0$ ).

$$S_i = \sum_{j=1}^n \hat{x}_{ij}, i = \bar{0}, \dots, \bar{m} \quad (28)$$

6. Calculating the utility degree. By comparing the value of the optimality function ( $S_i$ ) with the value of the optimal alternative ( $S_0$ ), a ranking is made (Zavadskas & Turskis, 2010).

$$K_i = \frac{S_i}{S_0}, i = \bar{0}, \dots, \bar{m} \quad (29)$$

The higher value of the utility degree means a better position in the ranking list.

The main disadvantage of this method is the possible existence of extreme values within one of the alternatives that can distort the ranking values.

**The Combinative Distance-based Assessment (CODAS)** is a modern multi-criteria decision-making method proposed by Ghorabae et al. (2016). The core component of the CODAS method is the ranking of alternatives according to the distance measurement of individual alternatives in relation to the worst-case scenario (Kumari & Acherjee, 2022). The first step is identical to most MCDM methods and it includes forming a decision-making matrix. The basic steps in CODAS are described as follows (1-7):

1. Forming the decision-making matrix. The preliminary decision-making matrix is expressed as matrix  $X$  in equation (1).

2. Normalizing the original decision matrix. The formula is presented in equation (30) and differs for beneficial and non-beneficial criteria (Ghorabae et al., 2016):

$$n_{ij} = \begin{cases} \frac{x_{ij}}{\max_i x_{ij}} & \text{if } \textit{beneficial} \\ \frac{\min_i x_{ij}}{x_{ij}} & \text{if } \textit{non-beneficial} \end{cases} \quad (30)$$

3. Weighting the normalized decision-making matrix. The formula for the weighting step is described in equation (31).

$$r_{ij} = w_j \times n_{ij} \quad (31)$$

4. Defining the negative-ideal solution. The negative-ideal value ( $n_{sj}$ ) is equal to the minimum weighted value ( $r_{ij}$ ) of alternative  $i$  to the  $j$ -th criterion from step 3.

5. Calculating distances from the negative-ideal solution. The distance measurement is performed using the Euclidean distance ( $E_i$ ) and the taxicab distance ( $T_i$ ), which contributes to the consistency of the rankings (Tüysüz & Kahraman, 2020). The derived weighted values are used to calculate the distance of each

alternative from the negative ideal solution defined in step 4 (Ghorabae et al., 2016). The calculations in equations (32) and (33) are done on weighted values  $r_{ij}$  and negative-ideal value  $n_{sj}$ .

$$E_i = \sqrt{\sum_{j=1}^m (r_{ij} - ns_j)^2} \quad (32)$$

$$T_i = \sum_{j=1}^m |r_{ij} - ns_j| \quad (33)$$

6. Calculating the relative assessment score. By aggregating the two distance measures, we obtain the relative assessment scores (Ghorabae et al., 2016). The formula for obtaining the value is presented in equation (34).

$$h_{ik} = (E_i - E_k) + (\Psi(E_i - E_k) \times (T_i - T_k)), k = 1, \dots, n \quad (34)$$

where  $\Psi$  is determined by a threshold value of  $\tau$  that is in this case expressed as 0.02 (Kumari & Acherjee, 2022). The distance between the alternatives is measured by the Euclidean distance ( $E_i$ ) and the taxicab distance ( $T_i$ ) if it satisfies the following rule:

$$\Psi(x) = \begin{cases} 1 & \text{if } |x| \geq \tau \\ 0 & \text{if } |x| < \tau \end{cases} \quad (35)$$

7. Calculating the assessment score. By summing the relative assessment scores, the final scores are obtained, according to which the alternatives are ranked.

$$H_i = \sum_{k=1}^n h_{ik} \quad (36)$$

The alternative with a higher assessment score ( $H_i$ ) attains a more favorable position in the ranking list since its distance from the negative ideal solution is the largest (Kumari & Acherjee, 2022).

**Evaluation based on Distance from Average Solution (EDAS)** is a multi-criteria method for ranking alternatives presented by Keshavarz Ghorabae et al. (2015). The essence of the methodology consists in quantifying the distance of the alternatives from the mean (Alinezhad & Khalili, 2019). In this way, the EDAS method demonstrates the relative superiority or inferiority of an alternative compared to the average solution. The initial steps in the application of the method involve the formation of the decision matrix and the calculation of the average values of the criteria. The required steps in EDAS are presented as follows (1-8):

1. Forming the decision-making matrix. The starting decision-making matrix is presented as matrix  $X$  in equation (1).

2. Calculating the average values. The equation (37) for the arithmetic mean is used to calculate the average values ( $AV_j$ ) that serve as a reference value for comparison (Torkayesh et al., 2023).

$$AV_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (37)$$

3. Calculating the positive and negative distance from the average. The next step involves measuring the positive ( $PDA_{ij}$ ) and negative ( $NDA_{ij}$ ) distance of the alternatives in relation to the average solution (Keshavarz Ghorabae et al., 2015). In this step, the nature of the criteria (benefit/cost) is taken into account. Therefore, equations (38) and (39) are used in case of beneficial criteria, while equations (40) and (41) are used for non-beneficial criteria.

$$PDA_{ij} = \frac{\max(0, x_{ij} - AV_j)}{AV_j} \quad (38)$$

$$NDA_{ij} = \frac{\max(0, AV_j - x_{ij})}{AV_j} \quad (39)$$

$$PDA_{ij} = \frac{\max(0, AV_j - x_{ij})}{AV_j} \quad (40)$$

$$NDA_{ij} = \frac{\max(0, x_{ij} - AV_j)}{AV_j} \quad (41)$$

4. Weighting the normalized decision-making matrix. In the next step, the weighting procedure is performed and the weighted positive ( $SP_i$ ) and weighted negative ( $SN_i$ ) distances of the alternatives are calculated (Keshavarz Ghorabae et al., 2015).

$$SP_i = \sum_{j=1}^n w_j \times PDA_{ij} \quad (42)$$

$$SN_i = \sum_{j=1}^n w_j \times NDA_{ij} \quad (43)$$

5. Normalizing the weighted positive and negative distances. The procedure is done using equations (44) and (45).

$$NSP_i = \frac{SP_i}{\max_i(SP_i)} \quad (44)$$

$$NSN_i = 1 - \frac{SN_i}{\max_i(SN_i)} \quad (45)$$

6. By aggregating normalized positive ( $NSP_i$ ) and normalized negative ( $NSN_i$ ) distances in equation (46), the values of the alternatives are derived, according to which the ranking is made (Keshavarz Ghorabae et al., 2015).

$$AS_i = \frac{1}{2}(NSP_i - NSN_i) \quad (46)$$

A higher value of the alternative ( $AS_i$ ) brings a better position in the ranking list (Torkayesh et al., 2023). The advantage of the EDAS method is that it mitigates the effects of extreme values on the ranking process by taking average values into account (Alinezhad & Khalili, 2019). Hybrid EDAS models integrated with the regression approach have proven to be effective in scenarios where the number of alternatives changes and an urgent decision need to be made, since this approach shortens the process of assigning ranks (Trung et al., 2023).

#### 4. Research results

Table 3 presents the quantitative assessment of each index according to the given criteria. The EDS criterion, which refers to the economic diversity of the countries included in the indexes, yields no discriminatory power across all alternatives, so it was eliminated from the decision-making process. It means that all Indexes are considering both developed and developing countries. Similarly, the WM criterion, which determines the method of calculating the weight of the components inside the index structure, was omitted from the study. All observed indices were assigned equal weights to groups of criteria thereby falling into the category of subjective weighting approaches. The theoretical explanation for implementing this type of weighting method is not transparent in the reports. Among the remaining criteria, the greatest diversity in ratings appeared in terms of the geographical coverage of the countries (GC) included in the index and the representation of the digital divide (CDD) within the index's indicator framework. NRI, EGDI and DESI cover the physical access to the digital technologies (first level), skills to use them (second level) and benefits that are acquired by individuals or groups from them (third level) within their indicators. The IDI index covers only the physical access to digital technologies (first level), while IMD and GDI cover physical access and skills to use digital

technologies (first and second levels). The remaining criteria express the theoretical and practical scope of the indexes.

| Criterion                        | Label | Alternative |      |      |     |     |     |
|----------------------------------|-------|-------------|------|------|-----|-----|-----|
|                                  |       | NRI         | EGDI | DESI | IDI | IMD | GDI |
| Geographic coverage              | GC    | 3           | 3    | 1    | 3   | 2   | 2   |
| Economic diversity of the sample | EDS   | 3           | 3    | 3    | 3   | 3   | 3   |
| Number of indicators             | NI    | 3           | 3    | 3    | 1   | 3   | 3   |
| Multidimensionality of the index | MI    | 3           | 3    | 3    | 1   | 3   | 3   |
| Coverage of the digital divide   | CDD   | 3           | 3    | 3    | 1   | 2   | 2   |
| Measurement approach             | MA    | 2           | 2    | 2    | 3   | 2   | 3   |
| Data source                      | DS    | 2           | 2    | 2    | 3   | 2   | 3   |
| Normalization methodology        | NM    | 3           | 3    | 3    | 3   | 1   | 1   |
| Weighting method                 | WM    | 2           | 2    | 2    | 2   | 2   | 2   |
| Publication frequency            | PF    | 3           | 2    | 2    | 3   | 3   | 3   |
| Indicator update frequency       | IUF   | 3           | 3    | 3    | 3   | 3   | 1   |

Table 3. Evaluation matrix with a quantitative rating scale.

4.1. Comparative analysis of the MEREC-MCDM models

By applying the objective approach of eliminating criteria the calculation of weight values was carried out using the MEREC method. The proposed methodology is sensitive to outliers however using the scale 1-3 this problem is minimized. All criteria are marked as beneficial because higher values enhance a higher index rating and their significance is reported in Table 4. The greatest impact on the subsequent ranking of the index are structural modifications. Following the digitalization trends ensures the timely availability of high-quality information, which is crucial for the decision-making process. Then follow the quantity of indicators within the structure of the index and the scope of the measurement area. More comprehensive indexes with a larger number of indicators are effective for ranking. The important dimensions were previously confirmed by Oloyede et al. (2023) and Denissova et al. (2025).

| Criterion                          | GC    | NI    | MI    | CDD   | MA    | DS    | NM    | PF    | IUF   |
|------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Alternative                        |       |       |       |       |       |       |       |       |       |
| Type of criteria                   | max   | max   | max   | max   | max   | max   | max   | max   | max   |
| MEREC weight coefficient ( $w_i$ ) | 0.140 | 0.164 | 0.164 | 0.138 | 0.026 | 0.026 | 0.130 | 0.049 | 0.165 |

Table 4. Objective weight coefficients using MEREC method.

Index ranking was conducted under the scope of five MCDM methods to test the validity and reliability of the results. The ranking results are presented in Table 5.

| Method | CoCoSo |      | MARCOS |      | ARAS  |      | CODAS  |      | EDAS  |      |
|--------|--------|------|--------|------|-------|------|--------|------|-------|------|
| Index  | Score  | Rank | Score  | Rank | Score | Rank | Score  | Rank | Score | Rank |
| NRI    | 1.955  | 1    | 0.895  | 1    | 0.982 | 1    | 0.260  | 1    | 0.987 | 1    |
| EGDI   | 1.742  | 2    | 0.880  | 2    | 0.967 | 2    | 0.258  | 2    | 0.949 | 2    |
| DESI   | 1.452  | 6    | 0.795  | 3    | 0.868 | 3    | -0.303 | 6    | 0.699 | 3    |
| IDI    | 1.479  | 5    | 0.628  | 6    | 0.695 | 6    | 0.066  | 3    | 0.322 | 5    |
| IMD    | 1.544  | 4    | 0.732  | 4    | 0.792 | 4    | -0.104 | 4    | 0.469 | 4    |
| GDI    | 1.616  | 3    | 0.647  | 5    | 0.706 | 5    | -0.175 | 5    | 0.284 | 6    |

Table 5. Final ranking scores using different MEREC-MCDM techniques.

The applied MCDM methods validate the NRI index as the first-ranked and the most favorable for the assessment of digital development, while the EGDI index is positioned second in terms of structural adequacy. The findings indicate the consistency of the decision-making process without deviations in the ranks. The difference in the scores of these two indices is negligible compared to the other lower-ranked indices, suggesting minor differences in terms of the quality and information they provide. The NRI index achieves a limited territorial scope (133 countries), with an annual release frequency. With these advantages, NRI provides an opportunity for comparative analysis and continuous monitoring of changes. The structure of the NRI index is routinely updated to modernize and refine the index according to the dynamics of the digital landscape. In contrast, EGDI covers a broader number of countries (193 countries) but is published biennially, which leaves the possibility that short-term fluctuations in digital flows go unnoticed. Both indexes, NRI and EGDI, in their structure contain a large number of indicators that can be connected to the analysis of the digital divide at all three levels, offering a comprehensive instrument for measuring digital development. The reports transparently present the data normalization methodology and cite credible sources of information. These two indices are rated as the most relevant in decision-making.

Additional confirmation of the stability of the MCDM ranking results is the consistency of the ranking of the IMD index, which took the fourth position across all MCDM methods. IMD can be classified as a medium-index with structural adequacy. The average territorial coverage of 67 countries provides a limited picture of digital development. IMD contains a large number of indicators, but according to it, the third level of the digital division, which refers to the measurement of the effects of the use of digital technology is lacking. The index fails to provide an explanation of the method for data normalization and information on whether the normalization procedure was implemented.

The remaining three indices DESI, IDI and GDI are characterized by fluctuations in the rank from the third to the sixth position, reflecting their sensitivity to the choice of ranking method. Although it has a larger number of indicators, adequate data sources and a robust framework, DESI is constrained in its geographical scope because it focuses on EU countries. This causes inferior ranking according to the CoCoSo and CODAS methods. In addition, the DESI report is published every second year, which affects the lack of a regular annual review of the digital development status. IDI and GDI have a smaller number of indicators in their structure, and according to them, the digital divide cannot be analyzed in terms of possession, use and benefit from digital technology, unlike DESI. According to the given weighting values of the criteria, the number of indicators and the coverage of the digital divide are crucial in the process of evaluating alternatives, which illustrate both criteria by which the IDI is rated weaker. The report on the GDI index lacks information on data normalization and does not provide any guidance on how to update the structure. It is a recently established index, which reduces its reliability and credibility. The difference in the ranking of the IDI and GDI indices arises only from the frequency of structural updates because the GDI has not been subject to changes until now.

In order to perform a more complex comparison of alternatives according to considered MCDM methods, the voting method was applied. The results of the overall ranking were calculated using a methodology according to which the ranks represent ratings and the index with the lowest summed value of the votes is ranked first, followed by the others in ascending order. The results are reported in Table 6.

| Criterion    | Alternative |      |      |     |     |     |
|--------------|-------------|------|------|-----|-----|-----|
|              | NRI         | EGDI | DESI | IDI | IMD | GDI |
| Voting score | 5           | 10   | 21   | 25  | 20  | 24  |
| Rank         | 1           | 2    | 4    | 6   | 3   | 5   |

**Table 6.** Aggregated ranking using voting approach.

A comparative analysis of differences and similarities in the application of MCDM methods is shown in Table 7. Author Dua (2024) suggests the use of Spearman's correlation rank for analyzing relationship among various MCDM ranking scores. Absolute concordance in ranks is present in the pair of MARCOS and ARAS. Despite differing normalization techniques (the MARCOS method performs normalization in relation to the ideal and anti-ideal solution while the ARAS method performs normalization according to the share of the alternative in relation to the sum of the scores of all alternatives), the obtained values have a similar proportionality that affects the allocation of the same ranks. The MARCOS and ARAS methods show a partial concordance in the ranks with the EDAS method evidenced by a Spearman correlation coefficient of 0.94. The EDAS method is similar to the MARCOS method in the normalization of data according to positive and negative deviations from the average value, while MARCOS considers ideal and anti-ideal values. Conversely,

the ARAS methodology relies on the average values of the alternatives according to the given criteria however the final ranking is performed comparing the alternative values with the ideal hypothetical alternative values. These three measures are related as they offer insight into the divergence of index values from ideal values.

The second category of CoCoSo and CODAS methods analyzes the distance and difference between indices. Spearman's correlation coefficient of ranks according to these methods is 0.77. Both methods share a comparable principle on the data normalization but a different procedure for ranking the alternatives. These differences can be evidenced by the high value of the average absolute difference in ranks between them, which equals 0.67.

MARCOS, ARAS and EDAS methods were chosen to rank the structural adequacy of the index based on a reference point that can be the average, ideal or anti-ideal value of the index rating according to the given criteria. Therefore, these methods yield more consistent findings. In contrast, CoCoSo and CODAS are more sensitive to extreme values due to deviation in values between alternatives but provide more accurate proof of the difference between similar alternatives.

| Method I | Method II | Average absolute difference in ranks | Spearman's correlation rank |
|----------|-----------|--------------------------------------|-----------------------------|
| CoCoSo   | MARCOS    | 1.00                                 | 0.60                        |
| CoCoSo   | ARAS      | 1.00                                 | 0.60                        |
| CoCoSo   | CODAS     | 0.67                                 | 0.77                        |
| CoCoSo   | EDAS      | 1.00                                 | 0.49                        |
| MARCOS   | ARAS      | 0.00                                 | 1.00                        |
| MARCOS   | CODAS     | 1.00                                 | 0.49                        |
| MARCOS   | EDAS      | 0.33                                 | 0.94                        |
| ARAS     | CODAS     | 1.00                                 | 0.49                        |
| ARAS     | EDAS      | 0.33                                 | 0.94                        |
| CODAS    | EDAS      | 1.00                                 | 0.60                        |

**Table 7.** Pairwise similarity and difference matrix.

#### 4.2. Effect of diverse weighting method

There is a partial convergence in the values when comparing the Shannon Entropy weight coefficients with the previously calculated values by the MEREC method (Table 8). This variance mostly derives from the weights of territorial coverage (MEREC: 0.140, Entropy: 0.139), coverage of the digital divide (MEREC: 0.138, Entropy: 0.139) and publication report frequency (MEREC: 0.049, Entropy: 0.040). The entropy approach prefers indices with the officially published normalization methodology ( $w_{NM} = 0.229$ ). The positive value of the correlation coefficient of 0.707 demonstrates the high degree of methods compatibility.

| Criterion                                    | GC    | NI    | MI    | CDD   | MA    | DS    | NM    | PF    | IUF   |
|----------------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Alternative                                  | GC    | NI    | MI    | CDD   | MA    | DS    | NM    | PF    | IUF   |
| Type of criteria                             | max   | max   | max   | max   | max   | max   | max   | max   | max   |
| Shannon Entropy weight coefficient ( $w_i$ ) | 0.139 | 0.119 | 0.119 | 0.139 | 0.048 | 0.048 | 0.229 | 0.040 | 0.119 |

**Table 8.** Objective weight coefficients using Shannon Entropy method.

The weight coefficients from Table 8 were used to test the rank stability with MCDM methods. The results are demonstrated in Figure 2.

A noticeable evidence from the graphic illustration proves the stability of the NRI and EGDI indices. Their consistency remains regardless of the chosen weight calculation method and ranking method. These two indices receive high marks that do not vary in relation to the criteria. In addition, the DESI index, as a relatively weaker rating, provides stability in the ranks when considering different methods for calculating the weighted importance of the criteria. An obvious fluctuation in ranks is present in the IDI and IMD indices, which are evaluated with different grades ranging from 1 to 3. The evidence demonstrates that the entropy

method in its hybrid form with CoCoSo, MARCOS, ARAS and EDAS is sensitive to index performance grades. These models are particularly useful when it is necessary to emphasize the difference between alternatives. The uniformity of ranks when altering the method for calculating weighting coefficients is validated by the coefficients of Spearman's rank correlation analysis (CoCoSo:  $\rho=0.829$ , MARCOS:  $\rho=0.829$ , ARAS:  $\rho=0.829$ , CODAS:  $\rho=1.000$ , EDAS:  $\rho=0.943$ ).

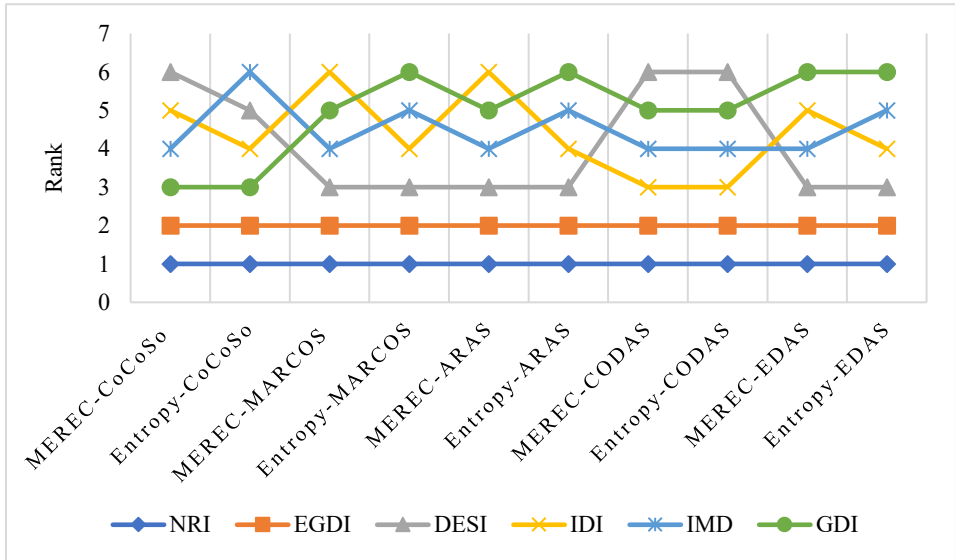


**Figure 2.** Stability of MCDM ranks in relation to changing the method for determining the weights to Shannon Entropy.

5. Discussion

This paper analyzes eight internationally recognized indexes for measuring digital development (NRI, EGDI, DESI, DGI, DAI, IDI, IMD and GDI) to determine which aligns more closely with the applied evaluation framework. Initial analyses removed the DAI and DGI indexes from comparison due to the lack of reports for the reference year 2024.

An integrated evaluation framework was developed using eleven different criteria for assessing six alternatives. With the hybrid approach of MEREC and other MCDM methods, namely CoCoSo, MARCOS, ARAS, CODAS and EDAS, five MCDM models were formed for the purpose of prioritization. The decision on methods was made to ensure that each provided a different normalization process and diversity of methods to find optimal solution. Ranking based on the MARCOS and ARAS methods indicated a complete positional overlap. The stability of the model is confirmed by the ranking results for the first (NRI), second (EGDI) and fourth (IMD) positions that have the same alternatives across all MCDM models. NRI and EGDI were characterized by strong and consistent ratings according to the highest importance attributes while IMD showed variability in score. The remaining alternatives (DESI, IDI, GDI) have recorded multiple deviations from the average values. The results remain stable even in the case when a different weighting method is applied (Shannon Entropy).

A comparative analysis of methodological similarities and differences conducted by the absolute difference between ranks and the Spearman correlation rank confirmed that reference-based methods (MARCOS, ARAS, EDAS) yield more consistent results than distance-based methods (CoCoSo, CODAS). The latter were found to be more sensitive to outliers, making them preferable when the decision-making process involves identifying exact differences between alternatives.

The proposed theoretical framework identified three attributes such as update frequency, the quantity of indicators and multidimensionality of the index as essential in the decision-making process. The combination of these three attributes captures both the theoretical and practical scope of digitalization within a dynamic environment. Overly narrow indicator structures that include one or two aspects of digital progress risk excluding important domains of development. In reverse, an overly broad multidimensional structure can

reduce the impact of key features due to the information overload. The most suitable are indices that depict the digital progress via a balanced set of dimensions and a structure that is periodically revised to align with social and economic priorities.

NRI as a leading index achieves high performance in the regularity of updating the index structure ( $w_{IUF}=0.165$ ) since control and changes are performed periodically. Frequent modifications contribute to the inconsistency of ranks and create a rather large gap in comparative analyses of historical data. The NRI report for 2024 has undergone minor changes compared to the previous 2023 report. Indicators such as tertiary enrollment, gross domestic expenditure on research and development by the business sector, high-tech manufacturing and exports have been removed (NRI report, 2023). However, more innovative metrics such as the number of venture capital deals invested in AI and public cloud computing market scale were introduced instead (Portulans Institute, 2023). In 2017 the publication of IDI index was paused due to problems with available data. A new report issued in 2023 provided an entirely new concept, so a comparative analysis was limited (International Telecommunication Union, 2025). This problem was also addressed by Lnenicka et al. (2022) showing that it is rarely discussed in contemporary literature. An additional problem emerges from the reverse scenario due to the lack of revision, because trends in the field of information technologies cannot be captured if we constantly measure the same indicators.

In the case of digital progress, an additional subject for further development is monitoring the digital divide. If the index lacks a structure for assessing digital inequality in terms of access, use and social and economic benefits from technology, then it is incomplete. A simple assessment of the use of digital technology fails to provide information on whether the application of these technologies has facilitated or hindered an activity. NRI is characterized by a large number of individual indicators ( $w_{NI}=0.164$ ). It also includes a large number of dimensions ( $w_{MI}=0.164$ ) through the integration of information technologies across fields of society, the business sector and public administration. The index deals with questions of digital trust and inclusion as important ethical and privacy terms in future development. Unlike the NRI index, the IDI index focuses on the accessibility and scale of these technologies but does not analyze their impact on other critical domains. Consequently, the IDI index is rated as less comprehensive because its structure is formed on the basis of a minimal 11 indicators. The obtained results according to the number of indicators align with the recommendations of the European Union & Joint Research Center (2008), who indicate that effective indexes are neither too simplistic nor overly complex. The IDI index is regarded as narrowly specialized aimed at measuring the share of information technology users while ignoring trends such as the side effects on the economy and society. In this sense, the IDI index does not deal with the digital divide ( $w_{CDD}=0.138$ ) because there is no data on the consequences of technological use, the development of e-government or digital inclusion. The GDI and IMD are presumed to effectively measure the development of digitalization at the macro level, but these indices also omit measurements that include issues of unequal territorial or gender access to technology for individuals. A classical example of an index that measures the digital divide adopted in this paper is the DESI index. This index relies on indicators of access and use of information technologies (e.g., mobile and fixed broadband coverage), measures skills for use (internet user skills and advanced skills and development) as well as the use of information technologies at the individual level (e-Government users). Similarly to DESI, the NRI index addresses the issue of equal use of information technologies by devoting the third sub-pillar, namely inclusion, to the main governance pillar. The findings were confirmed in the study by Denissova et al. (2025), who point out the need to include all three levels of the digital divide to form a realistic index. In terms of the technical aspect, IMD and GDI ignore valuable information about whether normalization ( $w_{NM}=0.130$ ) was carried out on the raw data, thereby reducing their significance. Authors Sidorov and Senchenko (2020) confirm that the absence of normalization questions the precision of the obtained results. NRI and EGDI index provide complete insight into the core methodology, which makes these indices transparent and desirable for professional use. The IMD and GDI indices, in addition to the DESI index, analyze a limited number of countries ( $w_{GC}=0.140$ ). This limitation prevents a global or regional analysis of digital development. For example, the DESI index has EU countries in its focus, while the IMD and GDI, although they have a global focus, omit technology hubs in Asia such as Israel and Hong Kong. Such situations can affect the analysis of regional differences and omit the influence of large technological players on the global market, as proved by Oloyede et al. (2023).

## 6. Conclusion

The proposed framework for selecting the most appropriate index aims to support decision-makers such as managers and policymakers by guiding them in choosing a transparent, consistent and comprehensive measurement instrument. The paper's additional value is reflected in the topic preference, as academic scholars have not previously addressed this issue using the proposed methodological and theoretical

framework. This analysis highlights an important observation that not every global index tackles emerging challenges. Comparative analysis raises the awareness of social differences such as access to and benefits derived from the new technology. This allows government officials to formulate investment policies adjusted to the active inclusion of digitally marginalized populations in ongoing digitalization processes. Another important topic that many indices ignore is digital trust and privacy, which are essential for further digital development. Transparency in the data collection and ranking is equally important. Practical recommendations include using an index that offers broad geographical coverage with frequent updates to mitigate bias and track trends. During the ranking process, not all indicators are of equal importance. Therefore, it is desirable to include objectively calculated weights in the index structure. For example, in this study, the frequency of index updates plays a more significant role in its selection than the data collection method itself. If the index is not adapted to the environment, data can be collected both subjectively and objectively but will not reflect the real picture of digital progress. Its structure lacks important indicators for that development phase and will not provide information on emerging trends. Such guidelines contribute to improving public trust in strategic planning and strengthening the initiative for digital progress in the long term.

The theoretical contribution is reflected in the newly developed hybrid model based on multi-criteria methods that allows for the assessment of the digitalization index as an analytical tool. Deciding on an appropriate index is based on criteria with varying importance, as not all criteria provide equally important information. This approach offers a theoretically supported framework as an alternative to the traditional arbitrary choice of indices and provides further insight into the issues of measuring digital development.

The main drawback of the study is reflected in the small alternative dataset. The focus of the study was on well-known and recognized global indices that are common sources of data and information for decision-makers. Therefore, the indices that were developed as part of the academic research projects and scholarly research were intentionally left out. This limitation leads to the question of relying on public documents as the primary source of information. The reliability and transparency of the dataset cannot be subjected to verification. The authors recognize another limitation in the form of using an ordinal scale for assessing indexes. The 1-3 ordinal scale was used to introduce simplicity into the multi-criteria analysis for qualitative criteria. However, some of these criteria could be quantified using alternative scales. Further research could include debate on this topic.

## Acknowledgement

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia through Mathematical Institute of the Serbian Academy of Sciences and Arts.

## References

- Abdelaal, R. M., Makki, A. A., Al-Madi, E. M., & Qhadi, A. M. (2024). Prioritizing Strategic Objectives and Projects in Higher Education Institutions: A New Hybrid Fuzzy MEREC-G-TOPSIS Approach. *IEEE Access*, 12, 89735-89753. <https://doi.org/10.1109/access.2024.3419701/mm6>
- Ali, A., Anam, S., & Ahmed, M. M. (2023). Shannon entropy in artificial intelligence and its applications based on information theory. *Journal of Applied and Emerging Sciences*, 13(1), 9-17. <http://dx.doi.org/10.36785/jaes.131549>
- Alinezhad, A., & Khalili, J. (2019). EDAS method. In *New methods and applications in multiple attribute decision making (MADM)* (pp. 149-155). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-15009-9\\_21](https://doi.org/10.1007/978-3-030-15009-9_21)
- Bai, C., Shiang, L. E., & Chua, S. Y. (2024). A Literature Review of Digital Economy Measurement Methods and Indexes. *International journal of academic research in business & social sciences*, 14(7), 1430-1440. <http://dx.doi.org/10.6007/IJARBS/v14-i7/21740>
- Bánhidi, Z., & Dobos, I. (2024). Measuring digital development: ranking using data envelopment analysis (DEA) and network readiness index (NRI). *Central European Journal of Operations Research*, 32(4), 1089-1108. <https://doi.org/10.1007/s10100-024-00919-y>
- Brodny, J., & Tutak, M. (2021). Assessing the level of digital maturity of enterprises in the Central and Eastern European countries using the MCDM and Shannon's entropy methods. *Plos one*, 16(7), e0253965. [https://doi.org/10.1371/journal.pone.0253965\\_](https://doi.org/10.1371/journal.pone.0253965_)

- Buturac, G. (2024). Methodological development of digital economy indices: a review and agenda for future research. *Radni materijali EIZ-a*, (1), 4-43.
- Chaonan, C., Jalil, N. A., & Rambeli, N. (2024). A literature review on the measurement methods of digital economy. *International Journal of Academic Reserach in Economics and Management Sciences*, 13(1). <https://doi.org/10.6007/ijarems/v13-i1/21072>
- Chaurasiya, R., & Jain, D. (2024). IoT adoption for smart cities waste management using Pythagorean fuzzy MEREC-SWARA-ARAS method. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 94(5), 533-548. <https://doi.org/10.1007/s10100-024-00919-y>
- Chen, W., & Wu, Y. (2022). Does intellectual property protection stimulate digital economy development?. *Journal of Applied Economics*, 25(1), 723-730. <https://doi.org/10.1080/15140326.2022.2045846>
- Davis, K. E., Kingsbury, B., & Merry, S. E. (2012). Indicators as a technology of global governance. *Law and Society Review*, 46(1), 71–104. <https://doi.org/10.1111/j.1540-5893.2012.00473.x>
- Davis, K., Fisher, A., Kingsbury, B., & Engle Merry, S. E. (2012a). *Governance by Indicators: Global Power through Quantification and Rankings*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199658244.001.0001>
- Denissova, O., Konurbayeva, Z., Kulisz, M., Yussubaliyeva, M., & Suieubayeva, S. (2025). Measuring the Digital Economy in Kazakhstan: From Global Indices to a Contextual Composite Index (IDED). *Economies*, 13(8), 225. <https://doi.org/10.3390/economies13080225>
- DGA - Digital Government Authority. (2022, October 5). Saudi Arabia achieves the best historical result and advances 12 places globally in E-Government Development Index [Webpage]. Digital Government Authority. <https://dga.gov.sa/en/node/394>
- Dua, T. V. (2024). Optimal selection for machining processes using the PSI-R-PIV method. *Applied Engineering Letters*, 9(3), 132-145. <https://doi.org/10.46793/aeletters.2024.9.3.2>
- Duc Trung, D. (2022). Multi-criteria decision making under the MARCOS method and the weighting methods: applied to milling, grinding and turning processes. *Manufacturing Review*, 9, 3. <https://doi.org/10.1051/mfreview/2022003>
- Dudic, B., Nguyen, N. T., & Ašonja, A. (2024). Data normalization for root assessment methodology. *International Journal of Industrial Engineering and Management*, 15(2), 156-168. <https://doi.org/10.24867/IJIEM-2024-2-354>
- Dudić, B., Van Duc, D., Son, N. H., & Ašonja, A. (2024). Comparison of MCDM methods effectiveness in the selection of plastic injection molding machines. *Teknomekanik*, 7(1), 1-19. <https://doi.org/10.24036/teknomekanik.v7i1.29272>
- Dwivedi, P. P., & Sharma, D. K. (2022). Application of Shannon entropy and CoCoSo methods in selection of the most appropriate engineering sustainability components. *Cleaner Materials*, 5, 100118. <https://doi.org/10.1016/j.clema.2022.100118>
- EC - European Commission. (2024, August 7). The Digital Economy and Society Index (DESI). Shaping Europe's digital future. <https://digital-strategy.ec.europa.eu/en/policies/desi>
- EC - European Commission. (2025, June 16). 2025 State of the Digital Decade package. Shaping Europe's digital future. <https://digital-strategy.ec.europa.eu/en/policies/2025-state-digital-decade-package>
- EC - European Commission. (n.d.). [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en)
- EC- European Commission. (2018). Digital Economy and Society Index (DESI) 2018: Country report — Portugal [PDF]. European Commission. [https://ec.europa.eu/information\\_society/newsroom/image/document/2018-20/pt-desi\\_2018-country\\_profile\\_eng\\_B440E073-A50F-CF68-82F6A8FB53D31DE5\\_52232.pdf](https://ec.europa.eu/information_society/newsroom/image/document/2018-20/pt-desi_2018-country_profile_eng_B440E073-A50F-CF68-82F6A8FB53D31DE5_52232.pdf)
- Ecer, F., & Aycin, E. (2023). Novel comprehensive MEREC weighting-based score aggregation model for measuring innovation performance: The case of G7 countries. *Informatica*, 34(1), 53-83. <https://doi.org/10.15388/22-infor494>
- El-Araby, A. (2023). The utilization of MARCOS method for different engineering applications: a comparative study. *International journal of research in industrial engineering*, 12(2), 155-164. <https://doi.org/10.22105/riej.2023.395104.1379>

- European Commission. (2022). Digital Economy and Society Index (DESI) 2022: Methodological Note. <https://digital-strategy.ec.europa.eu/en/library/desi-2022-methodological-note>
- European Union, & Joint Research Centre. (2008). Handbook on constructing composite indicators: methodology and user guide. OECD publishing.
- Fernández-Portillo, A., Almodóvar-González, M., & Hernández-Mogollón, R. (2020). Impact of ICT development on economic growth. A study of OECD European union countries. *Technology in Society*, 63, 101420. <https://doi.org/10.1016/j.techsoc.2020.101420>
- Gerpott, T. J., & Ahmadi, N. (2015). Composite indices for the evaluation of a country's information technology development level: Extensions of the IDI of the ITU. *Technological Forecasting and Social Change*, 98, 174-185. <https://doi.org/10.1016/j.techfore.2015.03.012>
- Ghorabae, M. K., Zavadskas, E. K., Turskis, Z., & Antucheviciene, J. (2016). A new combinative distance-based assessment (CODAS) method for multi-criteria decision-making. *Economic Computation and Economic Cybernetics Studies and Research*, 50(3), 25-44.
- Giovanni, E., Hoffmann, A., Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Handbook on constructing composite indicators: Methodology and user guide. OECD statistics working papers, 2005/3, OECD Publishing. <https://doi.org/10.1787/533411815016>
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social indicators research*, 141(1), 61-94. <https://doi.org/10.1007/s11205-017-1832-9>
- Groh, A., Liechtenstein, H., & Lieser, K. (2023). The Venture Capital and Private Equity Country Attractiveness Index 2023 (Annual report). IESE Business School. <https://blog.iese.edu/vcpeindex/files/2023/11/report2023.pdf> IESE Blog Network
- HRMDG - Hellenic Republic, Ministry of Digital Governance. (2019, July). Digital Transformation Strategy 2019–2023 [PDF]. [https://digitalstrategy.gov.gr/website/static/website/assets/Digital\\_Transformation\\_Strategy\\_2019.pdf](https://digitalstrategy.gov.gr/website/static/website/assets/Digital_Transformation_Strategy_2019.pdf)
- Huawei & IDC. (2024). Global Digitalization Index 2024. Huawei Technologies Co., Ltd. <https://www.huawei.com/en/gdi>
- Huy, T. V., Quyet, N. Q., Binh, V. H., Hoang, T. M., Tien, N. T., Anh, L. T., ... & Trung, D. D. (2022). Multi-criteria decision-making for electric bicycle selection. *Advanced Engineering Letters*, 1(4), 126-135. <https://doi.org/10.46793/adeletters.2022.1.4.2>
- IMD - IMD World Competitiveness Center. (2024). IMD World Digital Competitiveness Ranking 2024. Institute for Management Development. <https://www.imd.org/centers/world-competitiveness-center/rankings/world-digital-competitiveness>
- Imran, M., Liu, X., Wang, R., Saud, S., Zhao, Y., & Khan, M. J. (2022). The influence of digital economy and society index on sustainable development indicators: The case of European union. *Sustainability*, 14(18), 11130. <https://doi.org/10.3390/su141811130>
- International Telecommunication Union. (2025). History of the ICT Development Index (IDI). <https://www.itu.int/en/ITU-D/Statistics/Pages/IDI/history.aspx>
- International Trade Administration. (2024, January 25). Portugal – Information and Communications Technology. In *Portugal Country Commercial Guide*. U.S. Department of Commerce. <https://www.trade.gov/country-commercial-guides/portugal-information-and-communications-technology>
- Ishnazarov, A., Kasimova, N., Tosheva, S., & Isaeva, A. (2021, December). ICT and economic growth: evidence from cross-country growth modeling. In *Proceedings of the 5th International Conference on Future Networks and Distributed Systems* (pp. 668-671). <https://doi.org/10.1145/3508072.3508204>
- ITU - International Telecommunication Union. (2024). Measuring Digital Development: ICT Development Index 2024. [https://www.itu.int/hub/publication/D-IND-ICT\\_MDD-2024-4/](https://www.itu.int/hub/publication/D-IND-ICT_MDD-2024-4/)
- Jiménez-Fernández, E., Sánchez, A., & Ortega-Pérez, M. (2022). Dealing with weighting scheme in composite indicators: An unsupervised distance-machine learning proposal for quantitative data. *Socio-Economic Planning Sciences*, 83, 101339. <https://doi.org/10.1016/j.seps.2022.101339>
- Jovanović, M., Dlačić, J., & Okanović, M. (2018). Digitalization and society's sustainable development–Measures and implications. *Zbornik radova Ekonomskog fakulteta u Rijeci: časopis za ekonomsku teoriju i praksu*, 36(2), 905-928. <https://doi.org/10.18045/zbefri.2018.2.905>

- Keshavarz Ghorabae, M., Zavadskas, E. K., Olfat, L., & Turskis, Z. (2015). Multi-criteria inventory classification using a new method of evaluation based on distance from average solution (EDAS). *Informatica*, 26(3), 435-451. <https://doi.org/10.15388/informatica.2015.57>
- Keshavarz-Ghorabae, M., Amiri, M., Zavadskas, E. K., Turskis, Z., & Antucheviciene, J. (2021). Determination of objective weights using a new method based on the removal effects of criteria (MEREC). *Symmetry*, 13(4), 525. <https://doi.org/10.3390/sym13040525>
- Khan, S., & Haleem, A. (2021). Investigation of circular economy practices in the context of emerging economies: a CoCoSo approach. *International Journal of Sustainable Engineering*, 14(3), 357-367. <https://doi.org/10.1080/19397038.2020.1871442>
- Kolat, Ş. D., & Ünver, G. N. (2025). The impact of economic growth on cybersecurity: Evidence from ICT development index. *International Journal of Eurasia Social Sciences/Uluslararası Avrasya Sosyal Bilimler Dergisi*, 16(61). <https://doi.org/10.70736/ijoes.1711>
- Kou, G., Aydın, F. B., Eti, S., Yüksel, S., & Dinçer, H. (2025). A novel Spherical fuzzy MEREC-ARAS decision framework for optimizing innovation-driven investment strategies in energy systems. *Energy Strategy Reviews*, 59, 101771. <https://doi.org/10.1016/j.esr.2025.101771>
- Kumari, A., & Acherjee, B. (2022). Selection of non-conventional machining process using CRITIC-CODAS method. *Materials Today: Proceedings*, 56, 66-71. <https://doi.org/10.1016/j.matpr.2021.12.152>
- Liang, S., & Tan, Q. (2024). Can the digital economy accelerates China's export technology upgrading? Based on the perspective of export technology complexity. *Technological Forecasting and Social Change*, 199, 123052. <https://doi.org/10.1016/j.techfore.2023.123052>
- Lnenicka, M., Luterek, M., & Nikiforova, A. (2022). Benchmarking open data efforts through indices and rankings: Assessing development and contexts of use. *Telematics and informatics*, 66, 101745. <https://doi.org/10.1016/j.tele.2021.101745>
- Magoutas, A. I., Chaideftou, M., Skandali, D., & Chountalas, P. T. (2024). Digital progression and economic growth: Analyzing the impact of ICT advancements on the GDP of European Union countries. *Economies*, 12(3), 63. <https://doi.org/10.3390/economies12030063>
- Martínez, Y., Clastornik, J., & Campos, O. A. (2022, October). eGovernment Measuring Frameworks: A comparative analysis of different indexes based on their taxonomy: A comparative analysis of different indexes based on their taxonomy. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance* (pp. 397-403). <https://doi.org/10.1145/3560107.3560311>
- Merry, S. E. (2016). *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking* (Chicago Series in Law and Society). University of Chicago Press. [https://press.uchicago.edu/dam/ucp/books/pdf/course\\_intro/978-0-226-26128-7\\_course\\_intro.pdf](https://press.uchicago.edu/dam/ucp/books/pdf/course_intro/978-0-226-26128-7_course_intro.pdf) <https://doi.org/10.7208/chicago/9780226261317.001.0001>
- Miškufová, M., Košíková, M., Vašaničová, P., & Kiseľáková, D. (2025). Digitalization and Artificial Intelligence: A Comparative Study of Indices on Digital Competitiveness. *Information*, 16(4), 286. <https://doi.org/10.3390/info16040286>
- Nguyen, X. T., Ašonja, A., & Do, D. T. (2023). Enhancing handheld polishing machine selection: An integrated approach of marcos methods and weight determination techniques. *Appl. Eng. Lett. J. Eng. Appl. Sci*, 8, 131-138. <https://doi.org/10.18485/aeletters.2023.8.3.5>
- OECD - Organisation for Economic Co-operation and Development (2023). *Digital Government Index 2023*. In *Government at a Glance 2023*. OECD Publishing. <https://www.oecd.org/gov/government-at-a-glance-22214399.htm>
- Oloyede, A. A., Faruk, N., Noma, N., Tebepah, E., & Nwaulune, A. K. (2023). Measuring the impact of the digital economy in developing countries: A systematic review and meta-analysis. *Heliyon*, 9(7). <https://doi.org/10.2139/ssrn.4106167>
- Portulans Institute. (2023). *Network Readiness Index 2023*. [networkreadinessindex.org/reports/data/2023/nri-2023.pdf](https://networkreadinessindex.org/reports/data/2023/nri-2023.pdf)
- Portulans Institute. (2024). *Network Readiness Index 2024*. <https://networkreadinessindex.org/nri-2024-edition-press-release>

- Sagarik, D. (2023). Enhancing digital competitiveness through the lens of digital government among Asian economies. *International Journal of Public Administration in the Digital Age (IJPADA)*, 10(1), 1-11. <https://doi.org/10.4018/ijpada.326122>
- Saraiva, P. (2023). On Shannon entropy and its applications. *Kuwait Journal of Science*, 50(3), 194-199. <https://doi.org/10.1016/j.kjs.2023.05.004>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sidorov, A., & Senchenko, P. (2020). Regional digital economy: Assessment of development levels. *Mathematics*, 8(12), 2143. <https://doi.org/10.3390/math8122143>
- Siqueira, E. S., de Souza, C. A., & Barbosa, A. F. (2019). Using a digital divide index among enterprises in the context of public policies in Brazil. *The Electronic Journal of Information Systems in Developing Countries*, 85(3), e12088. <https://doi.org/10.1002/isd2.12088>
- Sofrankova, B., Sira, E., Horvathova, J., & Mokrisova, M. (2025). Digital Skills and Digital Transformation Performance in the EU-27: A DESI-Based Nonparametric and Panel Data Study. *Economies*, 13(11), 315. <https://doi.org/10.3390/economies13110315>
- Stanković, M., Stević, Ž., Das, D. K., Subotić, M., & Pamučar, D. (2020). A new fuzzy MARCOS method for road traffic risk analysis. *Mathematics*, 8(3), 457. <https://doi.org/10.3390/math8030457>
- Stević, Ž., Pamučar, D., Puška, A., & Chatterjee, P. (2020). Sustainable supplier selection in healthcare industries using a new MCDM method: Measurement of alternatives and ranking according to COMpromise solution (MARCOS). *Computers & industrial engineering*, 140, 106231. <https://doi.org/10.1016/j.cie.2019.106231>
- Štilić, A., Puška, A., Božanić, D., & Đurić, A. (2024). Ranking European Countries Using Hybrid MEREC-MARCOS MCDA Based on Travel and Tourism Development Index. *Tourism: An International Interdisciplinary Journal*, 72(4), 592-608. <https://doi.org/10.37741/t.72.4.6>
- Tokmergenova, M., & Dobos, I. (2024). Analysis of the Network Readiness Index (NRI) Using Multivariate Statistics. *Periodica Polytechnica Social and Management Sciences*, 32(1), 28-36. <https://doi.org/10.3311/ppso.20548>
- Torkayesh, A. E., Deveci, M., Karagoz, S., & Antucheviciene, J. (2023). A state-of-the-art survey of evaluation based on distance from average solution (EDAS): Developments and applications. *Expert systems with applications*, 221, 119724. <https://doi.org/10.1016/j.eswa.2023.119724>
- Trung, D. D., Truong, N. X., Dung, H. T., & Ašonja, A. (2023, December). Combining DOE and EDAS methods for multi-criteria decision making. In *International Conference on Organization and Technology of Maintenance* (pp. 210-227). Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-51494-4\\_19](https://doi.org/10.1007/978-3-031-51494-4_19)
- Turskis, Z., & Zavadskas, E. K. (2010). A novel method for multiple criteria analysis: grey additive ratio assessment (ARAS-G) method. *Informatica*, 21(4), 597-610. <https://doi.org/10.15388/informatica.2010.307>
- Tüysüz, N., & Kahraman, C. (2020). CODAS method using Z-fuzzy numbers. *Journal of Intelligent & Fuzzy Systems*, 38(2), 1649-1662. <https://doi.org/10.3233/jifs-182733>
- UNDESA - United Nations Department of Economic and Social Affairs (2022). *E-Government Survey 2022: The Future of Digital Government*. United Nations. <https://publicadministration.un.org/egovkb>
- United Nations (2024). *UN E-Government Survey 2024*. <https://desapublications.un.org/publications/un-e-government-survey-2024>
- van Dijk, J. A. G. M. (2005). *The deepening divide: Inequality in the information society*. SAGE Publications. <https://doi.org/10.4135/9781452229812.n7>
- Vision 2030. (2025, April 25). *Annual Report 2024* [PDF]. <https://www.vision2030.gov.sa/media/25042025ev1/En-Annual%20Report-Vision2030-2024.pdf>
- World Bank. (2016). *World Development Report 2016: Digital Dividends*. World Bank Publications. <https://www.worldbank.org/en/publication/wdr2016/Digital-Adoption-Index>
- Yazdani, M., Zarate, P., Kazimieras Zavadskas, E., & Turskis, Z. (2019). A combined compromise solution (CoCoSo) method for multi-criteria decision-making problems. *Management decision*, 57(9), 2501-2519. <https://doi.org/10.1108/md-05-2017-0458>

- Zavadskas, E. K., & Turskis, Z. (2010). A new additive ratio assessment (ARAS) method in multicriteria decision-making. *Technological and economic development of economy*, 16(2), 159-172. <https://doi.org/10.3846/tede.2010.10>
- Zavadskas, E. K., Turskis, Z., & Vilutiene, T. (2010). Multiple criteria analysis of foundation instalment alternatives by applying Additive Ratio Assessment (ARAS) method. *Archives of civil and mechanical engineering*, 10(3), 123-141. [https://doi.org/10.1016/s1644-9665\(12\)60141-1](https://doi.org/10.1016/s1644-9665(12)60141-1)



# The Meta-analysis based Modified Unified Theory of Acceptance and Use of Technology Model: A Systematic Review

Raghad Baker Sadiq<sup>1,2\*</sup>, Nur Fazidah Elias<sup>1</sup> and Hazura Mohamed<sup>1</sup>

<sup>1</sup>Faculty of Information Science and Technology, The National University of Malaysia, Selangor, Malaysia

<sup>2</sup>College of Business Informatics, University of Information Technology and Communications, Baghdad, Iraq

\*Correspondence: [bakiermajid@yahoo.com](mailto:bakiermajid@yahoo.com)

## PAPER INFO

### *Paper history:*

Received 25 September 2025

Accepted 17 December 2025

### *Citation:*

Sadiq, R. B., Elias, N. F. & Mohamed, H. (2026). The Meta-analysis based Modified Unified Theory of Acceptance and Use of Technology Model: A Systematic Review. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 45-64

### *Copyright:*

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

This study reviews state-of-the-art research on the application of the meta-analysis based Modified Unified Theory of Acceptance and Use of Technology model (Meta-UTAUT) to provide a comprehensive understanding of how the model has been applied across diverse contexts. From 1,881 citations of the original model, 36 Scopus-and Web of Science-indexed studies were analysed. Findings reveal that mobile payment is the most examined system, with financial technology as the predominant domain of application. Across studies, attitude consistently emerged as a strong determinant of behavioural intention and use behaviour, underscoring its central role in technology adoption. Behavioural intention and facilitating conditions were the most influential predictors of use behaviour, while performance expectancy and effort expectancy contributed substantially to shaping attitude. Many studies extended the Meta-UTAUT framework by incorporating external variables to enrich explanations of attitude, intention, and behaviour, deepening theoretical understanding of technology adoption. This review highlights recurring methodological limitations, including single-subject sampling, cross-sectional designs, and sampling methods, indicating the need for more rigorous approaches to strengthen theoretical refinement and improve generalisability. This study is the first to offer an extensive synthesis of Meta-UTAUT applications, providing valuable implications for researchers and guiding future inquiry toward more rigorous and contextually diverse investigations.

**Keywords:** Meta-UTAUT, Technology acceptance, Technology adoption, Methodological analysis, External variables, Weight analysis

## 1. Introduction

Numerous competing theories and models have been developed to better understand the acceptance and use of Information Systems/Information Technology (IS/IT). These frameworks provide valuable insights by examining contextual, technological, and personal factors that influence adoption decisions (Hawash et al., 2021; Wook et al., 2017). Their importance became evident as researchers and organizations increasingly recognized that technology implementations often fell short of expectations (Compeau & Higgins, 1995). Consequently, understanding these influencing factors has become a central concern for both management and IS/IT researchers, as it enables the development of more effective adoption strategies that leverage key motivators, address potential barriers, and ultimately enhance the success of technology adoption initiatives (Mukred et al., 2024).

The Unified Theory of Acceptance and Use of Technology model (UTAUT), introduced by Venkatesh et al. (2003), is a widely recognized framework that was developed based on eight prominent theories to provide a unified perspective on technology acceptance and address the limitations of earlier models. UTAUT includes four exogenous constructs, which are performance expectancy (PE), effort expectancy (EE), social influence (SI), and facilitating conditions (FC). PE, EE, and SI directly influence behavioural intention (BI), which in turn affects use behaviour (UB). Meanwhile, FC has a direct impact on UB. Additionally, its relationships are moderated by gender, age, experience, and the voluntariness of use.

UTAUT has been widely applied at both organizational and individual levels (Alsibhawi et al., 2023) due to its validity, reliability, and accuracy in predicting technology acceptance and adoption (Taherdoost, 2018). Its robustness stems from its parsimonious structure and high explanatory power (i.e., 70% of the variance in BI and 50% in UB) (Nooria & Jailanib, 2022), compared with other models, such as the Technology Acceptance Model (TAM) and the Theory of Planned Behaviour Model (TPB) (Mukred et al., 2019). In addition, UTAUT offers a more comprehensive model compared to earlier models such as TAM, which primarily focuses on technological and personal factors while overlooking the role of social influences (Gunasinghe et al., 2019). Nonetheless, UTAUT has faced criticism for proposing relationships that may not be universally applicable, omitting potentially important linkages, and excluding key factors necessary for understanding IS/IT acceptance and use (Dwivedi et al., 2019). To address these limitations, researchers have introduced new versions of UTAUT. For Instance, Venkatesh et al. (2012) proposed Unified Theory of Acceptance and Use of Technology 2 (UTAUT2), which incorporates three additional factors that directly influence BI. This extension improved the model's explanatory power, increasing the variance explained to 74% for BI and 52% for UB, compared to the original UTAUT. Farooq et al. (2017) introduced Unified Theory of Acceptance and Use of Technology 3 (UTAUT3), an extension of UTAUT2 that incorporates a new factor, increasing the variance in UB to 68.2%. Recently, Dwivedi et al. (2019) proposed the meta-analysis based modified Unified Theory of Acceptance and Use of Technology model (Meta-UTAUT). This revised model incorporates attitude as a mediator between independent and dependent factors to address some of the shortcomings of the original UTAUT. However, despite growing interest in the model, no study has systematically synthesized or critically examined the empirical applications of the Meta-UTAUT framework, particularly in terms of its performance and methodological approaches, and contextual deployment. This gap limits a holistic understanding of the model's potential, practical relevance, and theoretical boundaries. To address this issue, the present study systematically reviews the emerging body of literature that applies Meta-UTAUT to examine individual IT/IS acceptance and adoption. Specifically, this review evaluates the model's effectiveness and validity, synthesizes key findings and limitations, and outlines potential directions for future research.

The remainder of this paper is structured as follows: Section 2 provides a brief overview of the UTAUT and Meta-UTAUT models. Section 3 describes the methodology employed, followed by Section 4, which presents the findings. Section 5 offers a discussion of the results. Finally, Section 6 outlines the conclusions, future work, and limitations.

## 2. Literature Review

Venkatesh et al. (2003) introduced the Unified Theory of Acceptance and Use of Technology (UTAUT) to consolidate fragmented perspectives on technology acceptance and address limitations of earlier theoretical models. The model was developed through a systematic comparison of similarities and differences among eight prominent theories and models derived from information systems, psychology, and sociology, namely: the Theory of Reasoned Action (TRA), the Theory of Planned Behaviour (TPB), the Technology Acceptance Model (TAM), the Motivational Model (MM), Social Cognitive Theory (SCT), the Model of PC Utilization (MPCU), the combined TAM-TPB model (C-TAM-TPB), and the Diffusion of Innovations Theory (DOI).

The original UTAUT framework comprises four core exogenous constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions, as shown in Figure 1. In addition, the model incorporates four moderators, gender, age, experience, and voluntariness of use, to enhance predictive accuracy (Venkatesh et al., 2003). UTAUT has demonstrated strong explanatory power, accounting for approximately 70% of the variance in behavioural intention (Venkatesh et al., 2003) and nearly 50% of the variance in technology use (Venkatesh et al., 2012), representing a substantial improvement compared with earlier models (Mukred et al., 2019).

UTAUT is widely recognised as a leading framework in IS/IT research due to its strong theoretical grounding, substantial explanatory power, and comprehensive structure. However, scholars have noted that some of the model's proposed relationships may not be universally applicable and that it omits potentially important constructs relevant to specific contexts (Dwivedi et al., 2019). Furthermore, many studies modify

the original UTAUT by altering constructs or removing moderators, often citing concerns regarding its complexity (Dwivedi et al., 2020).

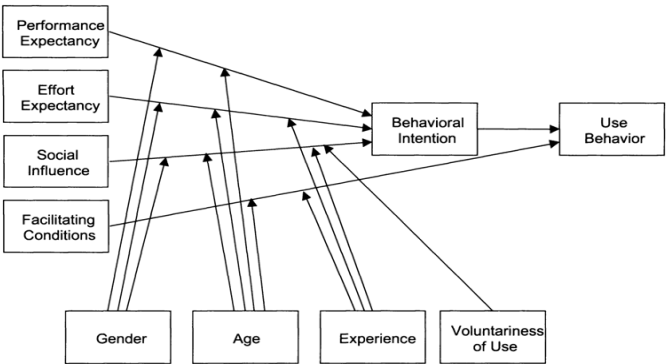


**Figure 1.** The unified theory of acceptance and use of technology (UTAUT) (Venkatesh et al., 2003)

Dwivedi et al. (2019) conducted a meta-analysis of 162 UTAUT-based empirical studies to develop an enhanced version of the model, known as the Meta-UTAUT model. The Meta-UTAUT model retains the four core exogenous constructs and the two endogenous constructs of the original UTAUT. Dwivedi et al. (2019) extended the original model by introducing attitude (ATT) as a key mediating factor, as depicted in Figure 2.

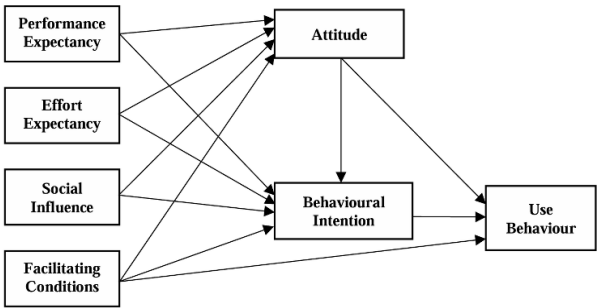


**Figure 2.** The meta-based modified model of the unified theory of acceptance and use of technology (Meta-UTAUT) (Dwivedi et al., 2019)

According to Venkatesh et al. (2003), PE refers to the degree to which an individual believes that using a particular system would improve his/her job performance. EE is defined as the degree of ease associated with using a system. SI is the degree to which a user perceives that important others believe he/she should use the new system. FC is defined as the degree to which an individual perceives that adequate organizational and technical resources are available to enable the effective use of a system. ATT has defined an individual's positive or negative feelings regarding performing a target behaviour (Davis, 1989). The four exogenous constructs directly influence ATT, in addition to their impact on BI. Additionally, ATT directly affects BI and UB. Moreover, the model excluded moderators, since they are not always applicable across different contexts (Dwivedi et al., 2019). Incorporating ATT into the theoretical model significantly enhances its explanatory power, as it increases the variance explained in BI from 38% to 45% and in UB from 21% to 27%, without and with the inclusion of ATT, respectively (Dwivedi et al., 2019).

Although Meta-UTAUT has received increasing attention within IS/IT research, a comprehensive synthesis of its application, performance, and contextual relevance remains limited. To date, only two reviews have examined the model. Dwivedi et al. (2020) conducted a systematic review that classified studies citing the original Meta-UTAUT article in Scopus, while Gulia and Singh (2024) carried out a bibliometric analysis of Meta-UTAUT publications, examining publication years, countries, subject areas, authors, and citation counts. While these studies provide valuable insights into citation patterns and the growing interest in Meta-UTAUT, they do not offer a detailed assessment of how the model has been operationalised in empirical

research, its effectiveness in predicting user attitudes, behavioural intentions, and technology use, or the methodological and contextual challenges encountered across various domains. As a result, understanding the model’s empirical robustness and practical applicability remains limited. Addressing this gap is essential, as a comprehensive review can identify the strengths and limitations of Meta-UTAUT and provide guidance for its refinement and adaptation in different IS/IT contexts. Therefore, this study systematically examines the application of Meta-UTAUT, evaluates its predictive performance, and synthesises current evidence regarding its effectiveness, limitations, and potential areas for improvement. The study is guided by the following research questions:

- RQ1:** What is the distribution of countries from which research participants in Meta-UTAUT studies were recruited?
- RQ2:** What systems and domains have the Meta-UTAUT model been applied to?
- RQ3:** What research methodologies were employed by studies using Meta-UTAUT?
- RQ4:** What are the most frequently applied models in conjunction with Meta-UTAUT?
- RQ5:** What are the results of hypothesis testing for internal variables in Meta-UTAUT studies?
- RQ6:** What external variables have been integrated with Meta-UTAUT?
- RQ7:** What are the major limitations of Meta-UTAUT studies?

3. Methodology

This systematic review was conducted in accordance with the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009). This protocol is recognized for its conciseness and standardized workflow, making it widely adopted across various studies, including in IS research (Maryati et al., 2020). The following steps describe the workflow of the review and are illustrated in Figure 3.

- 1. Identification:** All studies citing the original article (Dwivedi et al., 2019) up until 20 February 2025 were retrieved from Scopus and Web of Science (WoS). These trusted academic sources ensure broad and current research coverage. This approach was chosen to ensure comprehensive coverage of relevant research, as Meta-UTAUT initially lacked a unique name. For instance, some studies (Elprama et al., 2020; Kannan et al., 2018; Udang et al., 2023) referred to it as the “revised UTAUT”, while others (Sangeeta, 2021) labelled it the “modified UTAUT”. Some (Huseynov & Özkan Yildirim, 2019; Wiafe et al., 2020) had no specific name for it. Search focus was solely on journal articles and conference papers written in English. Resultantly, out of 2026 records, a total of 1881 studies were retrieved from Scopus (1040) and WoS (841). Metadata was managed using Microsoft Excel.
- 2. Screening:** In the first screening, 733 duplicates were removed, leaving 1148 studies. By applying the inclusion and exclusion criteria reviewed (see Table 1), titles, abstracts, and keywords were then reviewed. Subsequently, different types of review papers, studies focusing on system design, development and implementation, and experimental research were excluded during this phase, leading to the removal of 115 studies and keeping 1033 studies for more detailed examination.

| Inclusion                                                                                                   | Exclusion                                                                                                                                                         |
|-------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Studies that:<br>Proposed, developed, and/or validated a model utilizing Meta-UTAUT as its theoretical lens | Studies that:<br>Proposed, developed, and/or validated a model by utilizing theoretical models other than Meta-UTAUT or without relying on any foundational model |
| Were written in English                                                                                     | Suggested Meta-UTAUT as a model for future use                                                                                                                    |
|                                                                                                             | Meta-UTAUT was utilized either to support research findings or to justify causal relationships within their conceptual frameworks                                 |
|                                                                                                             | Focused on reviewing technology adoption and acceptance models                                                                                                    |

Table 1. Inclusion and exclusion criteria were used in this study

- 3. Eligibility:** A total of 37 studies met the inclusion criteria. The 996 studies that were excluded used other models, such as TAM, the Task-Technology Fit (TTF) model, UTAUT, the Artificial Intelligence Device Use Acceptance (AIDUA) model, and TPB. A second round of eligibility screening was

conducted by thoroughly reviewing the full texts of the previously identified studies to determine those that explicitly employed the Meta-UTAUT model as a foundational framework. In conclusion, the final sample of this stage yielded 36 papers, as a duplicate model was introduced in two studies.

4. **Included:** The final selection for this review comprises 36 studies, all of which met the criteria and were deemed appropriate for the analysis.

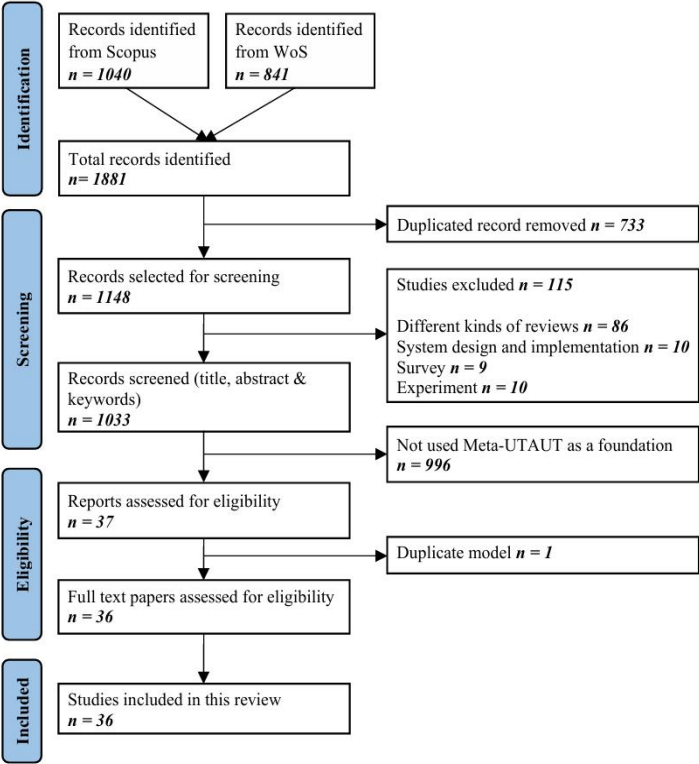


Figure 3. The flowchart of the systematic review (Moher et al., 2009)

4. Findings

The analysis shows that all the studies empirically validated their models, except for Sarker et al. (2020). The validation process crucially confirms the validity of the model, consistency, and relevance to the research problem (Creswell & Creswell, 2018). This analysis reviews Meta-UTAUT publications from 2019 to February 2025, showing growth since its debut. The peak was in 2022 (10 studies), followed by 2024 (nine studies) and 2023 (seven studies), as shown in Figure 4.

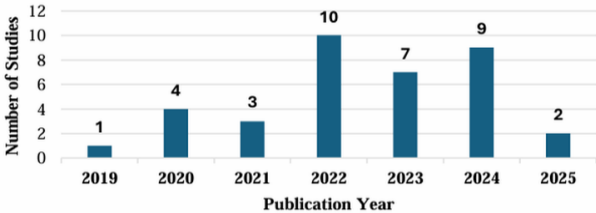


Figure 4. Distribution of studies per year

4.1. Sources of primary data by country

Primary data analysis was based on 33 studies, excluding one without empirical testing (Sarker et al., 2020) and two without country specification. Findings show Meta-UTAUT research spans 22 countries, as indicated in Table 2. The majority of the reviewed studies were conducted within a single-country context, with only two adopting a multi-country approach. India accounted for the largest proportion, contributing eight studies, followed by Indonesia with five and Vietnam with three. Other countries were represented minimally, appearing only once or twice in the dataset.

| Country                                                | Studies                                                                                                                                                                              |
|--------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| India (8)                                              | (Balakrishnan et al., 2022; Chatterjee et al., 2021; HM et al., 2024; Patil et al., 2020; Razi-ur-Rahim et al., 2024; Sangeeta, 2021; Tamilmani et al., 2022; Upadhyay et al., 2022) |
| Indonesia (5)                                          | (Alfarizi & Sari, 2022; Hakim et al., 2023; Hermanto et al., 2022; Kurniawan et al., 2023; Nur & Gosal, 2021)                                                                        |
| Vietnam (3)                                            | (Nguyen Trong et al., 2022; Nguyen, 2024; Pham & Nguyet, 2023)                                                                                                                       |
| Bangladesh (2)                                         | (Hossain et al., 2019; Sarker et al., 2025)                                                                                                                                          |
| Ghana (2)                                              | (Issaka et al., 2024; Wiafe et al., 2020)                                                                                                                                            |
| Gulf Countries (1)                                     | (Alkhowaiter, 2022)                                                                                                                                                                  |
| Australia (1)                                          | (Roy et al., 2021)                                                                                                                                                                   |
| Kenya (1)                                              | (Soehnchen et al., 2023)                                                                                                                                                             |
| Kenya, Nigeria, and South Africa (1)                   | (Olamijuwon & Odimegwu, 2022)                                                                                                                                                        |
| Malaysia, Indonesia, Thailand, and the Philippines (1) | (Atiqah et al., 2024)                                                                                                                                                                |
| Mauritius (1)                                          | (Gunnoo et al., 2023)                                                                                                                                                                |
| Netherlands (1)                                        | (Polyportis & Pahos, 2025)                                                                                                                                                           |
| Pakistan (1)                                           | (Hassaan & Yaseen, 2024)                                                                                                                                                             |
| Saudi Arabia (1)                                       | (Abed, 2024)                                                                                                                                                                         |
| Spain (1)                                              | (García de Blanes Sebastián et al., 2024)                                                                                                                                            |
| Taiwan (1)                                             | (Yang et al., 2022)                                                                                                                                                                  |
| Thailand (1)                                           | (Udang et al., 2023)                                                                                                                                                                 |
| USA (1)                                                | (Wu et al., 2024)                                                                                                                                                                    |
| N/A (3)                                                | (Elprama et al., 2020; Sarker et al., 2020; Wanner et al., 2022)                                                                                                                     |

Table 2. Countries used for primary data collection

4.2. IS research systems and domain examined

Table 3 summarizes the systems examined in the studies and their domains. The results reveal that the mobile payment systems were the most frequently analyzed in studies that applied the Meta-UTAUT model across various domains, especially in finance. Additionally, chatbots appeared in three studies across different sectors. On the other hand, financial technology (Fintech) emerged as the top domain with nine studies, followed by educational technology (EdTech), with five studies. E-health and marketing have been examined in four studies each.

| Domain                     | System examined                                                                                                            |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Financial technology (9)   | Digital banking (Alfarizi & Sari, 2022; Kurniawan et al., 2023)                                                            |
|                            | Mobile payment (Alkhowaiter, 2022; Gunnoo et al., 2023; Hassaan & Yaseen, 2024; Patil et al., 2020; Upadhyay et al., 2022) |
|                            | Unified payments interface (Razi-ur-Rahim et al., 2024)                                                                    |
|                            | E-wallet (Hakim et al., 2023)                                                                                              |
| Educational technology (5) | ChatGPT (Polyportis & Pahos, 2025)                                                                                         |
|                            | Mobile-Assisted Language Learning (MALL) (Nguyen, 2024)                                                                    |
|                            | Online classes (Udang et al., 2023)                                                                                        |

|                                |                                                                     |
|--------------------------------|---------------------------------------------------------------------|
|                                | Online teaching (Atiqah et al., 2024; Sangeeta, 2021)               |
| E-health (4)                   | Electronic health record (EHR) (Hossain et al., 2019)               |
|                                | Sexual health education digital tool (Soehnchen et al., 2023)       |
|                                | Sexuality information on social media (Olamijuwon & Odimegwu, 2022) |
|                                | Telemedicine (Wu et al., 2024)                                      |
| Marketing (4)                  | Smart retailing (Roy et al., 2021)                                  |
|                                | AI-customer relationship management (Chatterjee et al., 2021)       |
|                                | Social commerce (Sarker et al., 2020; Sarker et al., 2025)          |
| E-government (3)               | AI-chatbots (Abed, 2024)                                            |
|                                | Online business registration service (Nguyen Trong et al., 2022)    |
|                                | Online tax return reporting (Hermanto et al., 2022)                 |
| Industry (2)                   | Exoskeletons (Elprama et al., 2020)                                 |
|                                | Intelligent industrial maintenance systems (Wanner et al., 2022)    |
| E-service (1)                  | Chatbot-based services (Balakrishnan et al., 2022)                  |
| Agricultural technology (1)    | Sustainable intensification practices (Issaka et al., 2024)         |
| Construction (1)               | E-procurement (HM et al., 2024)                                     |
| E-commerce (1)                 | Mobile payment in online shopping (Nur & Gosal, 2021)               |
| Journalism and media (1)       | Blockchain (Pham & Nguyet, 2023)                                    |
| Maritime industry (1)          | A multicarrier booking and shipping (Wiafe et al., 2020)            |
| Mobile technology (1)          | Smartphone (Yang et al., 2022)                                      |
| Sustainable transportation (1) | Electric vehicles (García de Blanes Sebastián et al., 2024)         |
| Tourism and hospitality (1)    | Airbnb (Tamilmani et al., 2022)                                     |

Table 3. Systems examined and their domains applied Meta-UTAUT

4.3. Theoretical foundation

Among the 36 studies reviewed, 27 employed the Meta-UTAUT model independently, without integrating it with other established frameworks, as detailed in Table 4. However, some studies combined Meta-UTAUT with original UTAUT or its extensions. Five studies integrated it with the original UTAUT, one study combined it with both UTAUT and UTAUT2, and another study incorporated it with UTAUT2. Additionally, one study integrated Meta-UTAUT with models such as TAM plus, the Theory of Reasoned Action (TRA), TPB, UTAUT, and UTAUT2. Expanding Meta-UTAUT through such integrations can provide deeper insights and help address its limitations in studying adoption behaviours and emerging technological environments.

| Theory/model                                      | Studies                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Meta-UTAUT (27)                                   | (Abed, 2024; Alkhowaiter, 2022; Balakrishnan et al., 2022; Chatterjee et al., 2021; García de Blanes Sebastián et al., 2024; Hakim et al., 2023; Hassaan & Yaseen, 2024; Hermanto et al., 2022; HM et al., 2024; Issaka et al., 2024; Nguyen Trong et al., 2022; Nguyen, 2024; Nur & Gosal, 2021; Olamijuwon & Odimegwu, 2022; Patil et al., 2020; Polyportis & Pahos, 2025; Razi-ur-Rahim et al., 2024; Roy et al., 2021; Sangeeta, 2021; Sarker et al., 2020; Sarker et al., 2025; Soehnchen et al., 2023; Tamilmani et al., 2022; Udang et al., 2023; Upadhyay et al., 2022; Wiafe et al., 2020; Yang et al., 2022) |
| UTAUT and Meta-UTAUT (5)                          | (Atiqah et al., 2024; Elprama et al., 2020; Hossain et al., 2019; Pham & Nguyet, 2023; Wanner et al., 2022)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Meta-UTAUT and UTAUT2 (1)                         | (Kurniawan et al., 2023)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| TAM, TRA, TPB, UTAUT, UTAUT 2, and Meta-UTAUT (1) | (Gunnnoo et al., 2023)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

|                                     |                         |
|-------------------------------------|-------------------------|
| UTAUT, Meta-UTAUT, and TAM plus (1) | (Alfarizi & Sari, 2022) |
| UTAUT, UTAUT2 and Meta-UTAUT (1)    | (Wu et al., 2024)       |

Table 4. Theories/models applied with Meta-UTAUT

4.4. Methodological analysis

The findings indicate that 32 studies relied solely on a quantitative survey research method. In contrast, one study (Issaka et al., 2024) adopted a mixed-method approach by conducting a focus group followed by a survey. While another combined a survey with the System Usability Scale (SUS) (Soehnchen et al., 2023), and a third collected both qualitative and quantitative data within a survey (Udang et al., 2023). Additionally, one study (Sarker et al., 2020) developed a conceptual model based on a literature review without conducting empirical research. Therefore, the total number of studies analysed in this section is 35.

Regarding the analysis methods and tools, Table 5 reveals that the majority of the studies employed Structural Equation Modelling (SEM), with PLS-SEM (n = 18) and CB-SEM (n = 15) being the most commonly used statistical analysis methods, while one study utilized ESEM. Additionally, one study used the multinomial logit model, which is used for predicting categorical dependent variables when there are more than two unordered outcomes (Nguyen Trong et al., 2022). In terms of analysis tools, the findings in Table 6 indicate that SPSS (15 studies) and SmartPLS (14 studies) are the most used analysis tools.

| Analysis method                                              | No |
|--------------------------------------------------------------|----|
| Partial Least Squares Structural Equation Modeling (PLS-SEM) | 18 |
| Covariance-Based Structural Equation Modelling (CB-SEM)      | 15 |
| Exploratory Structural Equation Model (ESEM)                 | 1  |
| T-test                                                       | 3  |
| Multinomial Logit Model                                      | 1  |
| ANOVA                                                        | 1  |
| N/A                                                          | 1  |

Table 5. Data analysis methods used in the selected studies

| Analysis tool                                      | No |
|----------------------------------------------------|----|
| Statistical Package for the Social Sciences (SPSS) | 15 |
| SmartPLS                                           | 14 |
| AMOS                                               | 9  |
| R                                                  | 3  |
| STATA                                              | 1  |

Table 6. Data analysis tools used in the selected studies

4.5. Analysis of internal variables

This review evaluates the effectiveness and applicability of the Meta-UTAUT model’s internal variables across various domains, drawing on findings of 35 empirical studies. Following the guidance of Jeyaraj et al. (2006), a statistical analysis was conducted to estimate the relative weight of each predictor.

$$Weight\ of\ predictors = \frac{Total\ no.\ of\ significant\ relationships}{Total\ no.\ of\ relationships\ examined}$$

The results show that none of the studies confirmed all of the Meta-UTAUT relationships, as explained in Table 7. This is because some studies did not test all proposed relationships, whereas others expanded the Meta-UTAUT model by introducing external variables or integrating it with other models. Additionally, one study (Olamijuwon & Odimegwu, 2022) used BI more than once, slightly increasing the result counts to 36. Moreover, another study (Alfarizi & Sari, 2022) adopted the Meta-UTAUT model but did not examine any of the relationships among the variables specified in the original model.

| Author                                                                                                                                                                                                          | PE-ATT | PE-BI | EE-ATT | EE-BI | SI-ATT | SI-BI | FC-ATT | FC-BI   | FC-UB | ATT-BI | ATT-UB | BI-UB |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-------|--------|-------|--------|-------|--------|---------|-------|--------|--------|-------|
| Razi-ur-Rahim et al. (2024)                                                                                                                                                                                     | Yes    | X     | Yes    | X     | X      | Yes   | X      | Yes     | X     | Yes    | X      | Yes   |
| Soehnchen et al. (2023)                                                                                                                                                                                         | X      | Yes   | X      | No    | X      | No    | X      | Reverse | X     | Yes    | X      | X     |
| Kurniawan et al. (2023)                                                                                                                                                                                         | Yes    | No    | Yes    | No    | X      | Yes   | X      | No      | X     | Yes    | X      | Yes   |
| García de Blanes Sebastián et al. (2024)                                                                                                                                                                        | No     | No    | No     | Yes   | No     | No    | No     | No      | X     | Yes    | X      | X     |
| Yang et al. (2022)                                                                                                                                                                                              | Yes    | Yes   | Yes    | Yes   | Yes    | Yes   | Yes    | No      | X     | Yes    | X      | X     |
| Chatterjee et al. (2021)                                                                                                                                                                                        | Yes    | Yes   | Yes    | Yes   | X      | X     | Yes    | Yes     | Yes   | Yes    | Yes    | Yes   |
| Upadhyay et al. (2022)                                                                                                                                                                                          | Yes    | X     | Yes    | X     | X      | No    | X      | Yes     | X     | Yes    | X      | Yes   |
| Pham and Nguyet (2023)                                                                                                                                                                                          | X      | No    | X      | Yes   | X      | X     | X      | Yes     | X     | X      | X      | X     |
| Nguyen Trong et al. (2022)                                                                                                                                                                                      | X      | X     | X      | X     | X      | X     | X      | X       | Yes   | X      | No     | X     |
| Sarker et al. (2025)                                                                                                                                                                                            | Yes    | X     | Yes    | X     | X      | X     | X      | X       | Yes   | X      | Yes    | X     |
| Nguyen (2024)                                                                                                                                                                                                   | X      | Yes   | X      | Yes   | Yes    | X     | X      | X       | X     | Yes    | X      | Yes   |
| Udang et al. (2023)                                                                                                                                                                                             | Yes    | No    | Yes    | Yes   | No     | Yes   | Yes    | No      | X     | No     | X      | X     |
| Sangeeta (2021)                                                                                                                                                                                                 | Yes    | Yes   | Yes    | No    | No     | Yes   | Yes    | Yes     | Yes   | Yes    | Yes    | Yes   |
| Hassaan and Yaseen (2024)                                                                                                                                                                                       | Yes    | Yes   | No     | Yes   | No     | Yes   | Yes    | Yes     | X     | Yes    | X      | Yes   |
| Wiafe et al. (2020)                                                                                                                                                                                             | Yes    | Yes   | Yes    | No    | X      | No    | X      | Yes     | Yes   | Yes    | X      | No    |
| Tamilmani et al. (2022)                                                                                                                                                                                         | No     | Yes   | Yes    | X     | Yes    | X     | Yes    | X       | X     | Yes    | X      | X     |
| Alfarizi and Sari (2022)                                                                                                                                                                                        | X      | X     | X      | X     | X      | X     | X      | X       | X     | X      | X      | X     |
| Hossain et al. (2019)                                                                                                                                                                                           | X      | No    | X      | No    | X      | Yes   | X      | Yes     | Yes   | X      | X      | Yes   |
| Nur and Gosal (2021)                                                                                                                                                                                            | No     | No    | Yes    | Yes   | No     | No    | No     | No      | X     | Yes    | X      | X     |
| Roy et al. (2021)                                                                                                                                                                                               | Yes    | X     | No     | X     | Yes    | X     | Yes    | X       | X     | X      | Yes    | X     |
| Atiqah et al. (2024)                                                                                                                                                                                            | Yes    | X     | Yes    | X     | Yes    | X     | No     | X       | X     | Yes    | X      | X     |
| Issaka et al. (2024)                                                                                                                                                                                            | X      | Yes   | X      | No    | X      | Yes   | X      | Yes     | Yes   | Yes    | X      | Yes   |
| Olamijuwon and Odimegwu (2022) Use intention                                                                                                                                                                    | Yes    | Yes   | No     | Yes   | Yes    | Yes   | X      | Yes     | X     | Yes    | X      | X     |
| Olamijuwon and Odimegwu (2022) Interact intention                                                                                                                                                               | Yes    | Yes   | No     | Yes   | Yes    | Yes   | X      | Yes     | X     | No     | X      | X     |
| Elprama et al. (2020)                                                                                                                                                                                           | Yes    | Yes   | Yes    | Yes   | No     | Yes   | Yes    | No      | X     | Yes    | X      | X     |
| Hermanto et al. (2022)                                                                                                                                                                                          | Yes    | X     | Yes    | X     | X      | No    | X      | Yes     | X     | Yes    | X      | Yes   |
| HM et al. (2024)                                                                                                                                                                                                | No     | Yes   | Yes    | Yes   | Yes    | Yes   | Yes    | Yes     | X     | Yes    | X      | X     |
| Hakim et al. (2023)                                                                                                                                                                                             | Yes    | Yes   | Yes    | No    | Yes    | No    | X      | X       | Yes   | Yes    | X      | Yes   |
| Wanner et al. (2022)                                                                                                                                                                                            | Yes    | Yes   | Yes    | No    | X      | X     | X      | X       | X     | Yes    | X      | X     |
| Gunnoo et al. (2023)                                                                                                                                                                                            | X      | X     | X      | X     | X      | X     | X      | X       | X     | X      | Yes    | Yes   |
| Balakrishnan et al. (2022)                                                                                                                                                                                      | Yes    | Yes   | Yes    | Yes   | Yes    | No    | Yes    | Yes     | X     | Yes    | X      | X     |
| Patil et al. (2020)                                                                                                                                                                                             | Yes    | X     | Yes    | X     | X      | Yes   | X      | Yes     | X     | Yes    | X      | Yes   |
| Polyportis and Pahos (2025)                                                                                                                                                                                     | Yes    | Yes   | Yes    | No    | X      | Yes   | X      | Yes     | No    | Yes    | Yes    | Yes   |
| Abed (2024)                                                                                                                                                                                                     | Yes    | X     | Yes    | X     | X      | No    | X      | No      | X     | Yes    | X      | Yes   |
| Alkhowaiter (2022)                                                                                                                                                                                              | Yes    | Yes   | Yes    | Yes   | Yes    | Yes   | Yes    | Yes     | X     | Yes    | X      | Yes   |
| Wu et al. (2024)                                                                                                                                                                                                | Yes    | Yes   | Yes    | No    | Yes    | Yes   | No     | Yes     | X     | Yes    | X      | X     |
| Notes: Yes: A significant and positive relationship was found.; No: A non-significant relationship was found; X: The relationship was not examined; Reverse: A significant but negative relationship was found. |        |       |        |       |        |       |        |         |       |        |        |       |

Table 7. Findings from hypothesis testing of Meta-UTAUT internal hypotheses

Among the relationships examined, ATT-BI was the most frequently tested, being reported in 29 studies, as shown in Table 8. This underscores the role of ATT in explaining BI. This was followed by the relationships PE-ATT and EE-ATT, each investigated in 28 studies, and FC-BI was reported in 26 studies. PE-BI and SI-BI were each investigated in 25 studies. Meanwhile, EE-BI was examined in 24 studies. These findings reflect the influential roles of PE and EE in shaping ATT, as well as the impact of PE, EE, SI, and FC on BI. Other relationships examined include SI-ATT in 18 studies, BI-UB in 17 studies, and FC-ATT in 15 studies, which makes each play a comparatively less substantial role. FC-UB and ATT-UB had the lowest testing frequencies, examined in nine and seven studies, respectively. This limited examination suggests a research gap concerning these relationships. Overall, BI remains the most frequently studied outcome in Meta-UTAUT research.

The findings indicate that ATT-BI was reported as significant in 27 studies, representing the highest frequency among all examined relationships. It is followed by PE-ATT (24 studies), EE-ATT (23 studies), PE-

BI (19 studies), and both SI-BI and BI-UB, each examined in 16 studies. In contrast, other relationships yielded less significant results, including EE-BI (14 studies), SI-ATT (12 studies), FC-ATT (11 studies), FC-UB (eight studies), and ATT-UB (six studies).

| Hypotheses                             | PE-ATT | PE-BI | EE-ATT | EE-BI | SI-ATT | SI-BI | FC-ATT | FC-BI | FC-UB | ATT-BI | ATT-UB | BI-UB |
|----------------------------------------|--------|-------|--------|-------|--------|-------|--------|-------|-------|--------|--------|-------|
| Significant relationship               | 24     | 19    | 23     | 14    | 12     | 16    | 11     | 18    | 8     | 27     | 6      | 16    |
| Non-significant relationship           | 4      | 6     | 5      | 10    | 6      | 9     | 4      | 7     | 1     | 2      | 1      | 1     |
| Negative relationship                  | 0      | 0     | 0      | 0     | 0      | 0     | 0      | 1     | 0     | 0      | 0      | 0     |
| Not tested                             | 8      | 11    | 8      | 12    | 18     | 11    | 21     | 10    | 27    | 7      | 29     | 19    |
| Total                                  | 36     | 36    | 36     | 36    | 36     | 36    | 36     | 36    | 36    | 36     | 36     | 36    |
| Total no. of relationships examined    | 28     | 25    | 28     | 24    | 18     | 25    | 15     | 26    | 9     | 29     | 7      | 17    |
| Total no. of significant relationships | 24     | 19    | 23     | 14    | 12     | 16    | 11     | 19    | 8     | 27     | 6      | 16    |
| Weight of predictors                   | 0.86   | 0.76  | 0.82   | 0.58  | 0.67   | 0.64  | 0.73   | 0.73  | 0.89  | 0.93   | 0.86   | 0.94  |

Table 8. Summary of internal relationship analysis in Meta-UTAUT

According to Jeyaraj et al. (2006), an independent variable is considered a strong predictor if its weight is at least 0.8. Among the internal relationships examined, six relationships out of 12 (PE-ATT, EE-ATT, FC-UB, ATT-BI, ATT-UB, and BI-UB) met this threshold, with ATT-BI demonstrating the highest weight significance. The remaining relationships played comparatively less significant roles. Figure 5 illustrates the weight significance among Meta-UTAUT constructs, where the thickness of the connecting lines represents the strength of the relationships.

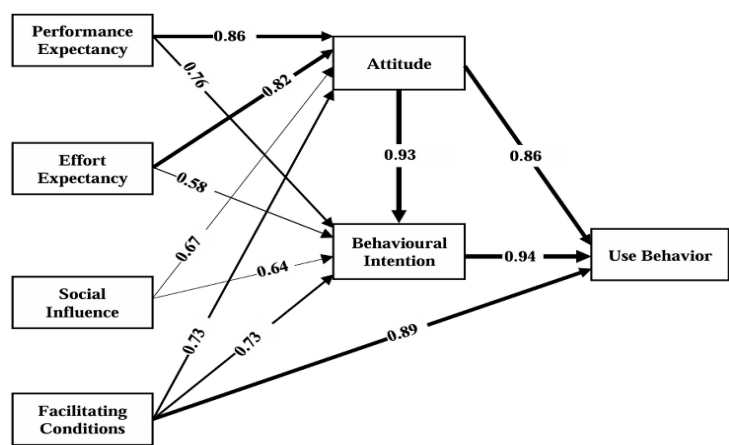


Figure 5. Weight significance of internal relationships in Meta-UTAUT

4.6. Analysis of external variables

Various external variables have been systematically incorporated into the Meta-UTAUT model. However, eight studies (Elprama et al., 2020; HM et al., 2024; Issaka et al., 2024; Olamijuwon & Odimegwu, 2022; Sangeeta, 2021; Soehnnchen et al., 2023; Udang et al., 2023; Yang et al., 2022) relied solely on the model’s core variables without adding external variables. Figure 6 provides a detailed mapping of how these external variables are integrated into the Meta-UTAUT model, highlighting their relationships with the core variables.

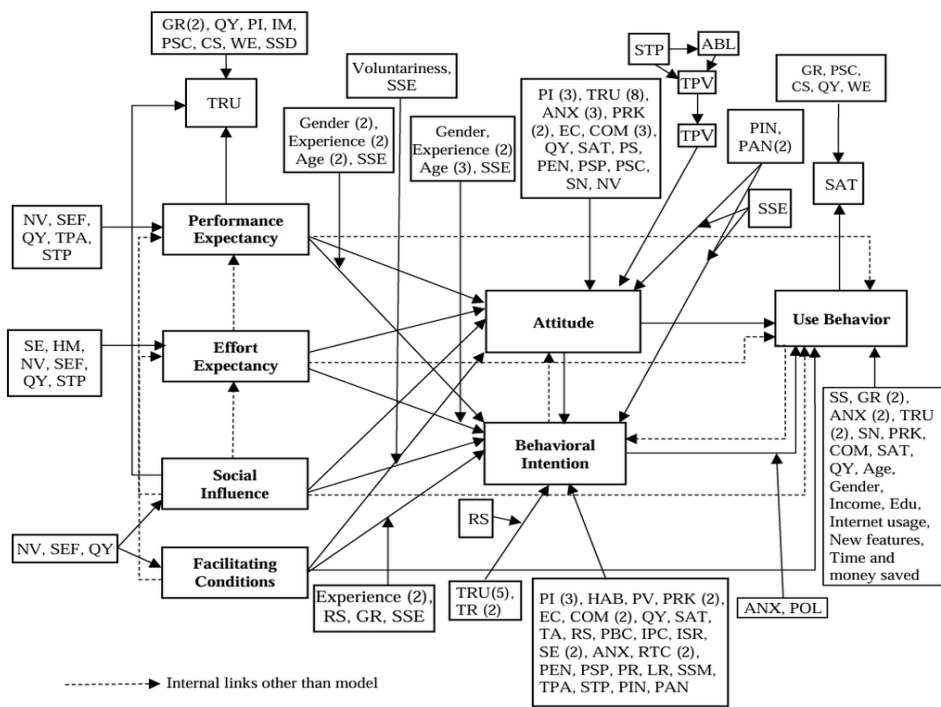


Figure 6. External variables incorporated with Meta-UTAUT

Table 9 summarizes 43 external variables that have been incorporated into the Meta-UTAUT model, along with their definitions and frequency of usage. The analysis indicates that the majority of external variables incorporated into the Meta-UTAUT framework were individual-level constructs, comprising 22 distinct factors that primarily reflect users’ personal perceptions, psychological characteristics, and behavioural tendencies. Among these, trust was the most frequently examined variable, appearing in 13 studies, followed by anxiety (seven studies) and personal innovativeness (six studies). A comparison with Williams et al. (2015) reveals that constructs such as self-efficacy, trust, anxiety, personal innovativeness, and perceived risk have been frequently examined across UTAUT studies. In contrast, technology/system-related factors encompassed 11 distinct variables, with system/service quality appearing in four studies, compatibility in three, and perceived anthropomorphism in two. Institutional/regulatory factors included five variables, with grievance redressal examined in five studies. Similarly, contextual/environmental factors comprised five variables; however, most of these were reported in only a single study, such as price value.

| Variable                    | Definition                                                                                                             | Referred articles                                                                                                                                                                                                                                                                                   |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Personal/Individual Factors |                                                                                                                        |                                                                                                                                                                                                                                                                                                     |
| Trust (TRU) (13)            | The extent to which a party believes that another party will fulfil their expectations in performing a specific action | (Alfarizi & Sari, 2022; Alkhowaiter, 2022; Gunnoo et al., 2023; Hakim et al., 2023; Hermanto et al., 2022; Kurniawan et al., 2023; Patil et al., 2020; Pham & Nguyet, 2023; Polyportis & Pahos, 2025; Razi-ur-Rahim et al., 2024; Sarker et al., 2020; Sarker et al., 2025; Tamilmani et al., 2022) |

|                                       |                                                                                                                                                                          |                                                                                                                                                                 |
|---------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Anxiety (ANX) (7)                     | An unpleasant emotional state that is often accompanied by subjective feelings of tension, apprehension, and worry                                                       | (Hakim et al., 2023; Hermanto et al., 2022; Nguyen Trong et al., 2022; Patil et al., 2020; Razi-ur-Rahim et al., 2024; Sarker et al., 2025; Wiafe et al., 2020) |
| Personal innovativeness (PI) (6)      | An individual would like to try out new technologies                                                                                                                     | (Alfarizi & Sari, 2022; Hossain et al., 2019; Nur & Gosal, 2021; Patil et al., 2020; Razi-ur-Rahim et al., 2024; Sarker et al., 2025)                           |
| Self-efficacy (SE) (4)                | An individual believes in their ability to perform a particular behaviour                                                                                                | (Balakrishnan et al., 2022; Tamilmani et al., 2022; Upadhyay et al., 2022; Wiafe et al., 2020)                                                                  |
| Satisfaction (SAT) (4)                | An individual's expectations regarding a product, service, or system are met or exceeded, resulting in an emotional or cognitive response based on their experience      | (Abed, 2024; Alfarizi & Sari, 2022; Chatterjee et al., 2021; Nguyen Trong et al., 2022)                                                                         |
| Perceived risk (PRK) (3)              | A combination of uncertainty and the perceived seriousness of potential outcomes is involved                                                                             | (García de Blanes Sebastián et al., 2024; Gunnoo et al., 2023; Sarker et al., 2020)                                                                             |
| Technology readiness (TR) (2)         | An organization possesses the necessary infrastructure and resources to enable its employees to adopt and effectively utilize new technology in the workplace            | (Atiqah et al., 2024; Pham & Nguyen, 2023)                                                                                                                      |
| Novelty (NV) (2)                      | The degree to which technology is perceived as distinctive due to its innovativeness                                                                                     | (Polyportis & Pahos, 2025; Roy et al., 2021)                                                                                                                    |
| Perceived security (PSC) (2)          | An individual's belief that technology is safe to use and adequately protects their personal and sensitive information from unauthorized access                          | (Alfarizi & Sari, 2022; Nur & Gosal, 2021)                                                                                                                      |
| Resistance to change (RTC) (2)        | An individual's or organization's reluctance or opposition to adopting new ideas, behaviours, or processes                                                               | (Hossain et al., 2019; Nur & Gosal, 2021)                                                                                                                       |
| Technology affinity (TA) (1)          | An individual's tendency to accept and adopt new technology                                                                                                              | (Pham & Nguyet, 2023)                                                                                                                                           |
| Habit (HAB) (1)                       | The manner in which users utilize a specific technology in their daily lives                                                                                             | (Kurniawan et al., 2023)                                                                                                                                        |
| Perceived severity (PS) (1)           | An individual believes they would experience serious consequences as a result of a specific threat or adverse event                                                      | (Upadhyay et al., 2022)                                                                                                                                         |
| Perceived behaviour control (PBC) (1) | An individual's belief about how easy or difficult it is to perform a specific behaviour, shaped by their past experiences, perceived obstacles, and necessary resources | (Nguyen, 2024)                                                                                                                                                  |
| Hedonic motivation (HM) (1)           | The degree of pleasure, enjoyment, or fun an individual experiences when using a particular technology                                                                   | (Tamilmani et al., 2022)                                                                                                                                        |
| Perceived enjoyment (PEN) (1)         | The extent to which an individual finds using technology intrinsically enjoyable, regardless of any performance-related outcomes.                                        | (Nur & Gosal, 2021)                                                                                                                                             |
| Pedagogical readiness (PR) (1)        | The familiarity with effective teaching methods                                                                                                                          | (Atiqah et al., 2024)                                                                                                                                           |
| Lifestyle readiness (LR) (1)          | The ability of an individual to work remotely and the flexibility                                                                                                        | (Atiqah et al., 2024)                                                                                                                                           |

|                                             |                                                                                                                                                                                      |                                                                                                             |
|---------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| Trust propensity (TPA) (1)                  | An individual's general tendency or inclination to trust others                                                                                                                      | (Wanner et al., 2022)                                                                                       |
| Ability beliefs (ABL)                       | The perceived competence and knowledge base of the system in performing a given task                                                                                                 | (Wanner et al., 2022)                                                                                       |
| Subjective norm (SN) (1)                    | An individual's perception of social pressure to perform or not a particular behaviour                                                                                               | (Gunnnoo et al., 2023)                                                                                      |
| Islamic religiosity (ISR) (1)               | An individual practices and adheres to the teachings, principles, and values of Islam                                                                                                | (Alkhowaiter, 2022)                                                                                         |
| Technology/System Factors                   |                                                                                                                                                                                      |                                                                                                             |
| System/service quality (QY) (4)             | An individual's perception of how effectively a system meets the needs of its users                                                                                                  | (Alfarizi & Sari, 2022; Chatterjee et al., 2021; Nguyen Trong et al., 2022; Roy et al., 2021)               |
| Compatibility (COM) (3)                     | The degree to which an innovation is perceived as aligned with an individual's work style, beliefs, values, experiences, and needs                                                   | (Chatterjee et al., 2021; Gunnnoo et al., 2023; Nur & Gosal, 2021)                                          |
| Perceived anthropomorphism (PAN) (2)        | An individual's belief that non-human entities possess human-like characteristics                                                                                                    | (Balakrishnan et al., 2022; Polyportis & Pahos, 2025)                                                       |
| Perceived intelligence (PIN) (1)            | An individual's belief that technology demonstrates human-like cognitive abilities                                                                                                   | (Balakrishnan et al., 2022)                                                                                 |
| Cloud service (CS) (1)                      | Internet-based computing solutions that allow users to access, store, and process data on remote servers rather than on local devices                                                | (Alfarizi & Sari, 2022)                                                                                     |
| Cloud service (CS) (1)                      | Internet-based computing solutions that allow users to access, store, and process data on remote servers rather than on local devices                                                | (Alfarizi & Sari, 2022)                                                                                     |
| Work efficiency (WE) (1)                    | The greatest number of tasks and labour accomplished with the least amount of time and effort                                                                                        | (Alfarizi & Sari, 2022)                                                                                     |
| System sophistication and display (SSD) (1) | The perceived level of user interface quality of a system and how it presents information and interacts with users                                                                   | (Alfarizi & Sari, 2022)                                                                                     |
| Technology System Bad Issue (TSBI) (1)      | User-perceived negative aspects or failures of a technology system that raise concerns about the system                                                                              | (Alfarizi & Sari, 2022)                                                                                     |
| System effectiveness (SEF) (1)              | The extent to which individuals perceive a system as successfully achieving its intended goals or functions                                                                          | (Roy et al., 2021)                                                                                          |
| Service smartness (SSM) (1)                 | An individual's belief that a system is intelligent                                                                                                                                  | (Hakim et al., 2023)                                                                                        |
| System transparency (STP) (1)               | The extent to which a technology openly conveys its functionality, decision-making logic, and data handling practices, thereby enabling users to comprehend and trust its operations | (Wanner et al., 2022)                                                                                       |
| Institutional/Regulatory Factors            |                                                                                                                                                                                      |                                                                                                             |
| Grievance redressal (GR) (5)                | The process through which individuals or groups can formally raise complaints or dissatisfaction regarding a service, product, or action and seek resolution or compensation         | (Alfarizi & Sari, 2022; Hakim et al., 2023; Hermanto et al., 2022; Patil et al., 2020; Sarker et al., 2025) |
| Institutional privacy concerns (IPC) (1)    | Individuals' apprehensions about the handling and protection of their personal or sensitive information by institutions                                                              | (Hassaan & Yaseen, 2024)                                                                                    |
| Institutional source reliability (ISR) (1)  | The discrepancy between trustworthiness and credibility is attributed to information provided by institutions and the actual experiences of users                                    | (Hassaan & Yaseen, 2024)                                                                                    |

|                                     |                                                                                                                                                                                                                                                 |                                           |
|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| Regulatory support (RS) (1)         | The set of rules, policies, and legal frameworks established by government bodies or regulatory authorities to promote the implementation, adoption, and operation of specific technologies, practices, or systems within an industry or sector | (Pham & Nguyet, 2023)                     |
| Institutional policy (POL) (1)      | A formal set of rules and procedures established by an institution to regulate behaviour and operations within its domain, ensuring alignment with its values and objectives                                                                    | (Polyportis & Pahos, 2025)                |
| Contextual/Environmental Factors    |                                                                                                                                                                                                                                                 |                                           |
| Price value (PV) (1)                | The extent to which perceived costs influence the adoption or usage of a particular technology                                                                                                                                                  | (Kurniawan et al., 2023)                  |
| Environmental concerns (EC) (1)     | Individuals are aware of ongoing environmental degradation                                                                                                                                                                                      | (García de Blanes Sebastián et al., 2024) |
| Social support (SS) (1)             | The assistance, encouragement, and resources an individual receives from their social network to help them cope with challenges and stressors                                                                                                   | (Sarker et al., 2025)                     |
| Image (IM) (1)                      | An individual believes that using innovation will enhance their status in their social system                                                                                                                                                   | (Alfarizi & Sari, 2022)                   |
| Perceived social pressure (PSP) (1) | An individual's response to the expectations, opinions, or behaviours of others concerning engagement in a specific behaviour                                                                                                                   | (Nur & Gosal, 2021)                       |

**Table 9.** Summary of external variables in prior studies

#### 4.7. The major limitations of Meta-UTAUT studies

The major limitations identified in prior studies (22 studies) reflect a narrow focus on specific subjects, such as particular communities, regions, countries, cities, organizations, educational institutions, individuals, or demographic groups (e.g., age, gender, or language). Additionally, seven studies highlighted issues with sampling methods, notably the use of non-probability techniques (e.g., purposive and convenience sampling), which introduced bias and constrained the generalizability of their findings. Moreover, an equal proportion of studies (seven) employed a cross-sectional design, which may constrain the ability to observe changes in technology usage over time and, in turn, may not accurately capture long-term behavioural patterns. Other reported limitations include the use of small sample sizes (six studies), exclusive reliance on quantitative methods instead of mixed-method approaches (five studies), lack of investigation into actual usage behavior (five studies), reliance on online surveys (two studies), voluntary participation (two studies), limited availability of related research for comparison (either on the topic or specifically applying the Meta-UTAUT model) (two studies), language bias in the questionnaire (one study), and absence of mediation analysis (one study).

### 5. Discussion

This review offers an in-depth analysis of studies employing the Meta-UTAUT model as a theoretical foundation. The findings reveal increasing research interest in the model, peaking in 2022 with the publication of 10 studies. This surge can be attributed to the rapid expansion and adoption of digital technologies, such as mobile payments and AI-driven services, trends that were further accelerated by the COVID-19 pandemic (Upadhyay et al., 2022). Furthermore, the increasing number of publications during this period likely underscored the model's relevance and applicability, thereby stimulating greater scholarly interest. As a result, this upward trend underscores its growing recognition as a robust framework for examining technology adoption and acceptance across diverse contexts. However, its theoretical application remains relatively limited, highlighting the need for broader use across varied technologies to further validate and strengthen its effectiveness.

The analysis of 33 studies reveals that research has been conducted in only 22 countries, with India emerging as the primary location for data collection. This prominence can be attributed to the rapid proliferation of digital technologies, including mobile payments, fintech platforms, and AI-enabled services, in emerging economies, where adoption continues to accelerate and is projected to expand substantially in the near future (Panetta et al., 2025). These regions also face distinct challenges such as varying levels of

digital literacy, infrastructure development, and regulatory maturity, making extended adoption models like Meta-UTAUT particularly relevant for understanding user behaviour, thereby increasing the need for exploratory theoretical studies (Adel, 2024). However, this highlights the importance of achieving broader geographical representation in future research. Mobile payment systems and the FinTech sector were the most frequently examined technologies and domains in studies employing the Meta-UTAUT model. FinTech continues to advance rapidly, fueled by ongoing innovation and strong support from both governments and businesses to promote cashless transactions (Hakim et al., 2023). This shift was particularly intensified in the post-COVID-19 era, leading to the widespread real-world applicability of mobile payment systems (Upadhyay et al., 2022). This has led to a marked increase in scholarly attention to user adoption behaviours of these systems (Hakim et al., 2023).

In terms of theoretical foundation, the Meta-UTAUT model has predominantly been applied in isolation in the reviewed studies. This preference is likely due to the incorporation of the individual attribute "attitude" as a mediating factor, which strengthens the model's explanatory power in capturing user acceptance behaviour compared to the original UTAUT, which does not account for personal characteristics (Kurniawan et al., 2023; Upadhyay et al., 2022; Yang et al., 2022). In addition, Hakim et al. (2023) and Kurniawan et al. (2023) reported that the model offers greater simplicity and flexibility than UTAUT, particularly due to the removal of moderators that may not be applicable across diverse research contexts. Its relative novelty also makes it well-suited for examining individual acceptance in the IS/IT domain, further contributing to the establishment of its validity (Kurniawan et al., 2023). Nevertheless, the Meta-UTAUT model has frequently been integrated with UTAUT and UTAUT2, and, to a lesser extent, with models such as TAM, TRA, and TPB. These integrations seek to address limitations within the Meta-UTAUT model and to provide a more comprehensive understanding of user behaviour by leveraging the complementary strengths of multiple theoretical perspectives. Integration of Meta-UTAUT with UTAUT2 or with other models, including TAM, TPB, and TRA, enhances its explanatory power, addresses theoretical limitations, and enables a more comprehensive understanding of technology usage behaviour, particularly in consumer contexts (Gunnool et al., 2023; Kurniawan et al., 2023). Moreover, incorporating moderators such as age and experience, which are included in UTAUT and UTAUT2 but omitted in Meta-UTAUT, enhances the model's ability to capture contextual variation and addresses criticisms that Meta-UTAUT alone may inadequately reflect behavioural differences across demographic and experiential segments (Wu et al., 2024). In synthesis, these integrations suggest that while Meta-UTAUT offers flexibility and explanatory breadth, integrated frameworks are often necessary in specific contexts to provide richer, context-sensitive insights into technology adoption across diverse user populations.

Regarding research methodologies, the majority of studies employed a quantitative design through questionnaire surveys. Quantitative methods are primarily used to examine relationships between variables and analyze causal effects (Creswell & Creswell, 2018). Additionally, questionnaire surveys are widely used in quantitative research due to their cost-effectiveness and efficiency in gathering large datasets from diverse samples and locations (Ranganathan & Caduff, 2023). However, quantitative surveys do not capture the underlying reasons behind participants' responses and often lack the depth to explain why certain findings occur. Therefore, to address this limitation, future research could integrate qualitative methods, such as interviews, alongside quantitative surveys. This mixed-method approach would provide richer, more nuanced insights into user behaviour, enabling a more comprehensive understanding of the factors influencing technology adoption (Creswell & Creswell, 2018).

The primary analytical method employed in the reviewed studies was SEM for its effectiveness in modelling latent variables (Creswell & Creswell, 2018). Among the SEM approaches, PLS-SEM is the most frequently utilized approach. This growing preference for PLS-SEM in recent research is largely due to its superior ability to handle complex models, accommodate smaller sample sizes, and manage non-normal data distributions, which are limitations often associated with CB-SEM (Hair et al., 2018).

Our weight analysis revealed that six of the 12 internal relationships (ATT-UB, BI-UB, ATT-BI, FC-UB, PE-ATT, and EE-ATT) met the threshold established by Jeyaraj et al. (2006) and were thus classified as the strongest predictors. It has confirmed the central role of ATT in adopting and using technology, particularly during the early stages of the adoption process (Balakrishnan et al., 2022; Patil et al., 2020; Upadhyay et al., 2022). However, the results also underscore the need for further examination of the remaining variables to evaluate their potential to meet the criteria for inclusion in the strongest predictor category.

The diverse range of external variables examined across the reviewed studies indicates that most focus on their influence on attitude, behavioural intention, and behaviour. This aligns with the core purpose of the Meta-UTAUT model, which is designed to evaluate attitude, intention, and usage behaviour. Additionally, Williams et al. (2015) demonstrated that these individual constructs are frequently examined within UTAUT-based studies, further reinforcing their importance in predicting adoption outcomes. Notably, individual-type

external variables predominate (22 factors), underscoring that research has primarily focused on personal perceptions, psychological characteristics, and behavioural tendencies as key drivers of user acceptance. This emphasis aligns with the central premise of technology adoption theories, which highlight the critical role of individual cognition and affective factors in shaping behavioural intention (Venkatesh et al., 2003; Venkatesh et al., 2012). Trust and anxiety are among the most frequently examined external variables in Meta-UTAUT models, reflecting deep and fundamental issues rather than merely recurring variables. The prominence of trust stems from the pervasive perceived risks inherent in contemporary digital systems, including security, privacy, institutional, and financial risks (Gunnool et al., 2023). Incorporating trust into Meta-UTAUT frameworks highlights users' need for assurance not only regarding a technology's functionality and usefulness but also its security, reliability, and overall integrity (Tamilmani et al., 2022). Thus, trust becomes a prerequisite for adoption, particularly as digital technologies such as e-wallets, AI assistants, smart home systems, mobile payments, and digital banking become deeply embedded in everyday life (Hakim et al., 2023). Conversely, rapid digitalisation has intensified emotional barriers alongside cognitive assessments, which explains the increased scholarly focus on anxiety as a critical determinant within Meta-UTAUT-based adoption studies. Engagement with new and emerging technologies often provokes emotional responses such as fear, tension, and negative perceptions, manifesting in concerns about data loss, system errors, or misuse (Venkatesh, 2000). These anxieties are heightened when users perceive limited control over technological processes, prompting them to view technological events as potential threats (Rana et al., 2016). Importantly, anxiety persists even among experienced users, as ongoing apprehension and fear continue to influence behavioural responses, ultimately increasing the likelihood of resistance or discontinuation of use (Patil et al., 2020), particularly in contexts where technology adoption is voluntary (Mayr et al., 2023).

The limitations acknowledged by the studies in our review mainly centre on issues related to data collection. Common concerns include reliance on specific populations, lack of longitudinal research, the frequent use of non-probability sampling methods, small sample sizes, limited use of mixed-method approaches, and insufficient focus on actual use behaviour. Although Dwivedi et al. (2019) did not empirically examine the mediating effect of attitude, they underscored its significance. Similarly, most studies have not tested this mediating relationship, which constrains a full understanding of how attitude operates within the Meta-UTAUT model. The absence of such testing limits insight into the causal pathways through which attitude influences behavioural intention and use behaviour, reducing the accuracy of recommendations for technology adoption strategies. These observations highlight significant opportunities for future research aimed at addressing these limitations, thereby contributing to more rigorous and robust studies.

## 6. Conclusions, Future Work, And Limitations

This study is the first systematic review to present a comprehensive analysis of 36 studies that have cited the original Meta-UTAUT model (Dwivedi et al., 2019) up to February 2025. The growing adoption of the Meta-UTAUT model in IS/IT research underscores its relevance for understanding technology adoption and user behaviour. Nevertheless, despite its extensive citation, relatively few studies have empirically implemented the model (Dwivedi et al., 2020). A thorough synthesis of how Meta-UTAUT has been operationalized across diverse contexts, its predictive capabilities, and the methodological and contextual challenges encountered is still lacking. This gap constrains a full appreciation of the model's practical applicability, theoretical rigor, and areas requiring further refinement. Accordingly, this study systematically examines the empirical applications of Meta-UTAUT, assesses its effectiveness in predicting user attitudes, behavioural intentions, and technology use, and highlights both methodological limitations and avenues for future research. The findings are organized into six principal categories: the geographical distribution of research, domains of application and types of systems investigated, theoretical frameworks adopted, methodological approaches employed, analysis of internal and external variables, and identification of key limitations.

Our analysis identifies several promising avenues for future inquiry, as well as areas that remain underexplored. Although the original Meta-UTAUT study has been cited 1881 times, only 36 studies have applied the model, indicating its limited adoption as a theoretical framework. This presents substantial opportunities for further development in the field. In particular, there is significant potential for researchers in unrepresented countries and various technologies to conduct original, contextually grounded studies that examine cultural and environmental influences on Meta-UTAUT applications. The model continues to be refined and adapted, offering avenues for future research to incorporate new variables, integrate complementary theoretical frameworks, and explore alternative relationships among its core constructs across various contexts.

The limitations commonly acknowledged in the existing literature highlight several promising avenues for future research. Addressing these gaps may require the use of more rigorous and diverse study designs,

such as examining actual system usage, adopting probability sampling methods, utilizing larger and more representative sample sizes, and conducting longitudinal investigations to capture changes over time. Additionally, integrating mixed-methods approaches, rather than relying solely on quantitative data, would further enrich understanding. Equally importantly, testing the mediating effect of attitude is vital to achieve complete and theoretically sound explanations of technology adoption.

This study has certain limitations that should be acknowledged. The literature search was restricted to studies that cited the original Meta-UTAUT article within the Scopus and Web of Science (WoS) databases. However, expanding to other databases or publications like book chapters or dissertations could significantly enhance the comprehensiveness and depth of future reviews.

## References

- Abed, S. S. S. (2024). Understanding the determinants of using government AI-chatbots by citizens in Saudi Arabia. *International Journal of Electronic Government Research*, 20(1), 1-20. <https://doi.org/10.4018/IJEGR.349733>
- Adel, N. (2024). The impact of digital literacy and technology adoption on financial inclusion in Africa, Asia, and Latin America. *Heliyon*, 10(24), e40951. <https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e40951>
- Alfarizi, M., & Sari, R. K. (2022). Indonesian digital bank adoption defence determination post-privacy data leakage issues. IWBIS, Depok, Indonesia.
- Alkhowaiter, W. A. (2022). Use and behavioural intention of m-payment in GCC countries: Extending meta-UTAUT with trust and Islamic religiosity. *Journal of Innovation and Knowledge*, 7(4), 100240. <https://doi.org/https://doi.org/10.1016/j.jik.2022.100240>
- Alsibhawi, I. A. A., Yahaya, J. B., & Mohamed, H. B. (2023). Business intelligence adoption for small and medium enterprises: Conceptual framework. *Applied Sciences*, 13(7), 4121. <https://www.mdpi.com/2076-3417/13/7/4121>
- Atiqah, S. N., Hanafiah, M. H., Ismail, H., & Kandappan, B. (2024). Preparing instructors to transition to online distance learning: A pandemic panacea? *Innoeduca*, 10(1), 5-28. <https://doi.org/https://doi.org/10.24310/ijtei.101.2024.16820>
- Balakrishnan, J., Abed, S. S., & Jones, P. (2022). The role of meta-UTAUT factors, perceived anthropomorphism, perceived intelligence, and social self-efficacy in chatbot-based services? *Technological Forecasting and Social Change*, 180, 121692. <https://doi.org/https://doi.org/10.1016/j.techfore.2022.121692>
- Chatterjee, S., Rana, N. P., Khorana, S., Mikalef, P., & Sharma, A. (2021). Assessing organizational users' intentions and behavior to AI integrated CRM systems: A Meta-UTAUT approach. *Information Systems Frontiers*, 25(4), 1299–1313. <https://doi.org/https://doi.org/10.1007/s10796-021-10181-1>
- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 19(2), 189-211. <https://doi.org/https://doi.org/10.2307/249688>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications. [https://spada.uns.ac.id/pluginfile.php/510378/mod\\_resource/content/1/creswell.pdf](https://spada.uns.ac.id/pluginfile.php/510378/mod_resource/content/1/creswell.pdf)
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. <https://doi.org/https://doi.org/10.2307/249008>
- Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers*, 21(3), 719–734. <https://doi.org/https://doi.org/10.1007/s10796-017-9774-y>
- Dwivedi, Y. K., Rana, N. P., Tamilmani, K., & Raman, R. (2020). A meta-analysis based modified unified theory of acceptance and use of technology (meta-UTAUT): A review of emerging literature. *Current Opinion in Psychology*, 36, 13-18. <https://doi.org/https://doi.org/10.1016/j.copsyc.2020.03.008>
- Elprama, S. A., Vannieuwenhuyze, J. T. A., De Bock, S., Vanderborght, B., De Pauw, K., Meeusen, R., & Jacobs, A. (2020). Social processes: What determines industrial workers' intention to use exoskeletons? *Hum Factors*, 62(3), 337-350. <https://doi.org/https://doi.org/10.1177/0018720819889534>

- Farooq, M. S., Salam, M., Jaafar, N., Fayolle, A., Ayupp, K., Radovic-Markovic, M., & Sajid, A. (2017). Acceptance and use of lecture capture system (LCS) in executive business studies. *Interactive Technology and Smart Education*, 14(4), 329-348. <https://doi.org/https://doi.org/10.1108/ITSE-06-2016-0015>
- García de Blanes Sebastián, M., Sarmiento Guede, J. R., Azuara Grande, A., & Juárez-Varón, D. (2024). Analysis of factors influencing attitude and intention to use electric vehicles for a sustainable future. *The Journal of Technology Transfer*, 49(4), 1347-1368. <https://doi.org/https://doi.org/10.1007/s10961-023-10046-6>
- Gulia, A., & Singh, L. (2024). A bibliometric analysis of the usage of 'Meta-UTAUT': An emerging IS/IT acceptance theory. *Global Journal of Enterprise Information System*, 16(2), 108-113. <https://doi.org/https://doi.org/10.18311/gjeis/2024>
- Gunasinghe, A., Abd Hamid, J., Khatibi, A., & Azam, S. F. (2019). Academicians' acceptance of online learning environments: A review of information system theories and models. *Global Journal of Computer Science and Technology: Information & Technology*, 19(1), 30-39.
- Gunnoo, L., Subadar, U. A., & Fauzel, S. (2023). The mediating effect of attitude on adoption of mobile payment services in small island developing states. The case of Mauritius. *Social Sciences & Humanities Open*, 8(1), 100724. <https://doi.org/https://doi.org/10.1016/j.ssaho.2023.100724>
- Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2018). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hakim, M. M., Afifah, A. N., & Aryotejo, G. (2023). The analysis of factors affecting behavioral intention and behavior usage of e-wallet using Meta-UTAUT model. *International Journal on Advanced Science, Engineering and Information Technology*, 13(2), 786-793. <https://doi.org/10.18517/ijaseit.13.2.18297>
- Hassaan, M., & Yaseen, A. (2024). Factors influencing customers' adoption of mobile payment in Pakistan: Application of the extended meta-UTAUT model. *Journal of Science and Technology Policy Management*. <https://doi.org/https://doi.org/10.1108/JSTPM-01-2024-0029>
- Hawash, B., Mokhtar, U. A., & Yusof, Z. M. (2021). Users' acceptance of an electronic record management system in the context of the oil and gas sector in Yemen: An application of ISSM-TAM. *International Journal of Management and Enterprise Development*, 20(1), 75-98. <https://doi.org/10.1504/ijmed.2021.113661>
- Hermanto, A. H., Windasari, N. A., & Purwanegara, M. S. (2022). Taxpayers' adoption of online tax return reporting: Extended meta-UTAUT model perspective. *Cogent Business & Management*, 9(1), 2110724. <https://doi.org/https://doi.org/10.1080/23311975.2022.2110724>
- HM, A. J., KB, A., & VR, H. (2024). Technology adoption in material procurement: An empirical study applying the UTAUT model among construction companies in India. *Global Business Review*, 1-20. <https://doi.org/https://doi.org/10.1177/09721509241244989>
- Hossain, A., Quaresma, R., & Rahman, H. (2019). Investigating factors influencing the physicians' adoption of electronic health record (EHR) in healthcare system of Bangladesh: An empirical study. *International Journal of Information Management*, 44, 76-87. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2018.09.016>
- Huseynov, F., & Özkan Yıldırım, S. (2019). Online consumer typologies and their shopping behaviors in B2C e-commerce platforms. *SAGE Open*, 9(2). <https://doi.org/https://doi.org/10.1177/2158244019854639>
- Issaka, Y. B., A., D. S., & Kranjac-Berisavljevic, G. (2024). Psycho-social factors influencing the adoption of sustainable intensification practices by smallholder rice farmers in Northern Ghana. *Cogent Food & Agriculture*, 10(1), 2345434. <https://doi.org/https://doi.org/10.1080/23311932.2024.2345434>
- Jeyaraj, A., Rottman, J. W., & Lacity, M. C. (2006). A review of the predictors, linkages, and biases in IT innovation adoption research. *Journal of Information Technology*, 21(1), 1-23. <https://doi.org/https://doi.org/10.1057/palgrave.jit.2000056>
- Kannan, R. J., Punithavathi, P., & Sambandam, N. (2018, Nov. 2018). Technology adoption models – adoption of ICT in educational institutions in India. In Proceedings of the 2018 World Engineering Education Forum-Global Engineering Deans Council (WEEF-GEDC), Albuquerque, NM, USA.
- Kurniawan, Y., Dumais, S. M., & Anwar, N. (2023). An empirical study on the factors affecting the usage of digital banking in Generation Z. *Journal of System and Management Sciences*, 13(5), 31-48.
- Maryati, I., Purwandari, B., Santoso, H. B., & Budi, I. (2020, Nov. 2020). Implementation strategies for adopting digital library research support services in academic libraries in Indonesia. In Proceedings of the

2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia.

Mayr, A., Stahmann, P., Nebel, M., & Janiesch, C. (2023). Unified theory of acceptance and use of technology (UTAUT) for intelligent process automation. *ICIS 2023 Proceedings*, Hyderabad, India.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7), e1000097. <https://doi.org/doi:10.1371/journal.pmed.1000097>

Mukred, M., Mokhtar, U. A., Hawash, B., AlSalman, H., & Zohaib, M. (2024). The adoption and use of learning analytics tools to improve decision making in higher learning institutions: An extension of technology acceptance model. *Heliyon*, 10(4), e26315. <https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e26315>

Mukred, M., Yusof, Z. M., Alotaibi, F. M., Mokhtar, U. A., & Fauzi, F. (2019). The key factors in adopting an electronic records management system (ERMS) in the educational sector: A utaut-based framework. *IEEE Access*, 7, 35963-35980. <https://doi.org/10.1109/ACCESS.2019.2904617>

Nguyen Trong, H., Dang, T. V., Nguyen, V., & Nguyen, T. T. (2022). Determinants of e-government service adoption: An empirical study for business registration in Southeast Vietnam. *Journal of Asian Public Policy*, 15(3), 453-468. <https://doi.org/https://doi.org/10.1080/17516234.2020.1805396>

Nguyen, T. T. (2024). Exploring the interaction of Meta-UTAUT in the mobile-assisted language learning environment for EFL learning. *Journal of Educators Online*, 21(4). <https://doi.org/https://doi.org/10.9743/jeo.2024.21.4.13>

Nooria, M. A., & Jailanib, M. N. (2022). Biometric authentication technology acceptance in mobile banking in iraq. *Neuro Quantology*, 20(7), 3004-3019. <https://doi.org/doi:10.14704/nq.2022.20.7.NQ33378>

Nur, T., & Gosal, G. A. (2021). Mobile payment usage in online shopping among Gen Z in the Jabodetabek Area: META-UTAUT approach. *ICIMTech*, Jakarta, Indonesia.

Olamijuwon, E., & Odimegwu, C. (2022). Sexuality education in the digital age: Modelling the predictors of acceptance and behavioural intention to access and interact with sexuality information on social media. *Sexuality Research and Social Policy*, 19(3), 1241-1254. <https://doi.org/10.1007/s13178-021-00619-1>

Panetta, I. C., Anjomrouz, E., Paiardini, P., & Leo, S. (2025). Digital by Default? A Critical Review of Age-Driven Inequalities in Payment Innovation. *Journal of Risk and Financial Management*, 18(6), 313. <https://www.mdpi.com/1911-8074/18/6/313>

Patil, P., Tamilmani, K., Rana, N. P., & Raghavan, V. (2020). Understanding consumer adoption of mobile payment in India: Extending Meta-UTAUT model with personal innovativeness, anxiety, trust, and grievance redressal. *International Journal of Information Management*, 54, 102144. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2020.102144>

Pham, C. T., & Nguyet, T. T. T. (2023). Determinants of blockchain adoption in news media platforms: A perspective from the Vietnamese press industry. *Heliyon*, 9(1), e12747. <https://doi.org/https://doi.org/10.1016/j.heliyon.2022.e12747>

Pham, L. T. T., & Nguyen, Y. T. H. (2023). An integrated framework approach to understanding Vietnamese people's intention to adopt smart home solutions. *Human Behavior and Emerging Technologies*, 2023(1), 8882543. <https://doi.org/https://doi.org/10.1155/2023/8882543>

Polyportis, A., & Pahos, N. (2025). Understanding students' adoption of the ChatGPT chatbot in higher education: The role of anthropomorphism, trust, design novelty and institutional policy. *Behaviour & Information Technology*, 44(2), 315-336. <https://doi.org/https://doi.org/10.1080/0144929X.2024.2317364>

Rana, N. P., Dwivedi, Y. K., Williams, M. D., & Weerakkody, V. (2016). Adoption of online public grievance redressal system in India: Toward developing a unified view. *Computers in Human Behavior*, 59, 265-282. <https://doi.org/https://doi.org/10.1016/j.chb.2016.02.019>

Ranganathan, P., & Caduff, C. (2023). Designing and validating a research questionnaire - Part 1. *Perspect Clin Res*, 14(3), 152-155. [https://doi.org/doi:10.4103/picr.picr\\_140\\_23](https://doi.org/doi:10.4103/picr.picr_140_23)

Razi-ur-Rahim, M., Rabbani, M. R., Uddin, F., & Shaikh, Z. H. (2024). Adoption of UPI among Indian users: Using extended meta-UTAUT model. *Digital Business*, 4(2), 100093. <https://doi.org/https://doi.org/10.1016/j.digbus.2024.100093>

- Roy, S. K., Singh, G., & Shabnam, S. (2021). Modelling customer engagement behaviour in smart retailing: Modelling customer engagement behaviour. *Australasian Journal of Information Systems*, 25. <https://doi.org/https://doi.org/10.3127/ajis.v25i0.2967>
- Sangeeta, U. T. (2021). Factors influencing adoption of online teaching by school teachers: A study during covid-19 pandemic. *Journal of Public Affairs*, 21(4), e2503. <https://doi.org/doi:10.1002/pa.2503>
- Sarker, P., Hughes, D. L., & Dwivedi, Y. K. (2020, May. 2020). Extension of META-UTAUT for examining consumer adoption of social commerce: Towards a conceptual model. In F. J. Martínez-López & S. D'Alessandro, *Advances in Digital Marketing and eCommerce (DMEC 2020)*, Cham.
- Sarker, P., Hughes, L., Malik, T., & Dwivedi, Y. K. (2025). Examining consumer adoption of social commerce: An extended META-UTAUT model. *Technological Forecasting and Social Change*, 212, 123956. <https://doi.org/https://doi.org/10.1016/j.techfore.2024.123956>
- Soehnchen, C., Weirauch, V., Schmook, R., Henningsen, M., & Meister, S. (2023). An acceptance analysis of a sexual health education digital tool in resource-poor regions of Kenya: An UTAUT based survey study. *BMC Womens Health*, 23(676). <https://doi.org/https://doi.org/10.1186/s12905-023-02839-6>
- Taherdoost, H. (2018). A review of technology acceptance and adoption models and theories. *Procedia Manufacturing*, 22, 960-967. <https://doi.org/https://doi.org/10.1016/j.promfg.2018.03.137>
- Tamilmani, K., Rana, N. P., Nunkoo, R., Raghavan, V., & Dwivedi, Y. K. (2022). Indian travellers' adoption of Airbnb platform. *Information Systems Frontiers*, 24(1), 77-96. <https://doi.org/https://doi.org/10.1007/s10796-020-10060-1>
- Udang, L. N., Shyong, A. L. S., & Zhou, F. (2023). Exploring UTAUT factors on lecturers' adoption of online classes during covid-19 pandemic. *Eurasian Journal of Educational Research*, 104, 243-266. <https://doi.org/10.14689/ejer.2023.104.014>
- Upadhyay, N., Upadhyay, S., Abed, S. S., & Dwivedi, Y. K. (2022). Consumer adoption of mobile payment services during COVID-19: Extending meta-UTAUT with perceived severity and self-efficacy. *International Journal of Bank Marketing*, 40(5), 960-991. <https://doi.org/https://doi.org/10.1108/IJBM-06-2021-0262>
- Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, 11(4), 342-365. <https://doi.org/https://doi.org/10.1287/isre.11.4.342.11872>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478. <https://doi.org/https://doi.org/10.2307/30036540>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157-178. <https://doi.org/https://doi.org/10.2307/41410412>
- Wanner, J., Herm, L.-V., Heinrich, K., & Janiesch, C. (2022). The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study. *Electronic Markets*, 32(4), 2079-2102. <https://doi.org/10.1007/s12525-022-00593-5>
- Wiafe, I., Koranteng, F. N., Tettey, T., Kastriku, F. A., & Abdulai, J.-D. (2020). Factors that affect acceptance and use of information systems within the maritime industry in developing countries: The case of Ghana. *Journal of Systems and Information Technology*, 22(1), 21-45. <https://doi.org/https://doi.org/10.1108/JSIT-06-2018-0091>
- Williams, M. D., Rana, N. P., & Dwivedi, Y. K. (2015). The unified theory of acceptance and use of technology (UTAUT): A literature review. *Journal of Enterprise Information Management* 28(3), 443-488. <https://doi.org/https://doi.org/10.1108/JEIM-09-2014-0088>
- Wook, M., Yusof, Z. M., & Nazri, M. Z. A. (2017). Educational data mining acceptance among undergraduate students. *Education and Information Technologies*, 22(3), 1195-1216. <https://doi.org/10.1007/s10639-016-9485-x>
- Wu, M. Y.-C., Wu, C. T.-L., Saleem, N., & Zhang, M. (2024). An assessment of the COVID-19 pandemic's impact on US physicians' acceptance of telemedicine: An exploratory study. *Journal of Health Management*, 26(1), 124-138. <https://doi.org/https://doi.org/10.1177/09720634231217027>
- Yang, C.-C., Li, C.-L., Yeh, T.-F., & Chang, Y.-C. (2022). Assessing older adults' intentions to use a smartphone: Using the meta-unified theory of the acceptance and use of technology. *International Journal of Environmental Research and Public Health*, 19(9), 5403. <https://doi.org/https://doi.org/10.3390/ijerph19095403>

## Factors Affecting Behavior, Perceived Impact of AI on Work Engagement, and AI Application: Indonesian Context

Shofia Amin<sup>1\*</sup>, Akhmad Habibi<sup>1,2,3</sup>, Siti Hajar Halili<sup>4</sup>, Hamdy Abdullah<sup>3</sup>, Muthia Shahnaz<sup>1</sup>, Rafiqi Rafiqi<sup>1</sup> and Nela Safelia<sup>1</sup>

<sup>1</sup>Faculty of Economics and Business, Universitas Jambi, Jambi, Indonesia

<sup>2</sup>Graduate School of Education, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea

<sup>3</sup>Faculty of Business and Management, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia

<sup>4</sup>Faculty of Education, Universiti Malaya, Kuala Lumpur, Malaysia

\*Correspondence: [akhmad.habibi@unj.ac.id](mailto:akhmad.habibi@unj.ac.id)

### PAPER INFO

#### *Paper history:*

Received 25 July 2025

Accepted 08 February 2026

#### *Citation:*

Amin, S., Habibi, A., Halili, S. H., Abdullah, H., Shahnaz, M., Rafiqi, R. & Safelia, N. (2026). Factors Affecting Behavior, Perceived Impact of AI on Work Engagement, and AI Application: Indonesian Context. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 65-82

#### *Copyright:*

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

### ABSTRACT

This study aimed to elaborate on the predictors of behavioral intention, actual behavior, perceived impact of AI on work engagement, and the perceived impact of AI on teaching and learning in Asian higher education institutions (HEIs). We extended the Unified Theory of Acceptance and Use of Technology (UTAUT) to test the relationships among these variables. In total, 516 lecturers from three different universities contributed to the dataset. Partial least squares structural equation modeling (PLS-SEM) was employed to analyze the respondents' data through both measurement and structural models. The findings indicate that actual behavior has a significant positive relationship with the perceived impact of AI on teaching and learning and the perceived impact of AI on work engagement. Behavioral intention is strongly correlated with actual behavior. Performance expectancy is the most prominent link to behavioral intention, followed by social influence and facilitating conditions. The results of the study underscore the importance of fostering a conducive environment through adequate facilitating conditions and aligning performance expectations to drive behavioral intention and perceived engagement with AI technologies in HEIs.

**Keywords:** Artificial intelligence; behavior; higher education; Indonesian students; perceived impact of AI; work engagement; teaching and learning

## 1. Introduction

AI plays a significant role in higher education, especially in changing how classes are taught. AI-powered techniques help teachers assess efficacy, organize materials, and enhance teaching. They can manage admin tasks and customize methods using chatbots, virtual assistants, and adaptive learning systems. Chatbots such as ChatGPT, Gemini, and Meta AI assist lecturers by answering questions, providing teaching advice, and managing administrative tasks like scheduling. AI also improves grading speed and gives thorough assignment feedback, allowing educators to concentrate more on student engagement and instructional design. AI algorithms can spot locations where instructional enhancements are needed and find trends in teaching data that assist teachers in delivering better courses. Additionally, AI improves communication and monitors progress, helping lecturers identify and resolve student issues for a more flexible, responsive learning environment (Ashraf, 2023; Deng & Lin, 2023; Lin et al., 2022). More creative and consequential applications

of AI in education are on the horizon and hold great potential for improving education quality, relevance, and efficiency through increased individualization and work engagement.

AI in Higher Education Institutions (HEIs) can boost perceived AI impact on lecturers' work engagement, influencing behavioral intentions and actions. Assessing factors affecting AI use, behavior, and its impact on engagement helps guide stakeholders on AI's importance. Many studies examine AI's role in enhancing learning. (An et al., 2023; Habibi, Muhaimin et al., 2023; Ragheb et al., 2022; Wu et al., 2022). However, a few studies reported the use of AI and factors affecting behavioral intention to use AI, behavior, perceived impact of AI on work engagement, and perceived effects of AI on teaching and learning perceived by university lecturers (Clifford & Perez, 2024). The current research expanded the UTAUT model, integrating the perceived impact of AI on work engagement and the perceived effects of AI on teaching and learning. Prior UTAUT studies (e.g., Venkatesh, 2003; Habibi et al., 2023) focused on performance and effort expectancy as adoption drivers, but this study shows how AI behavior sustains engagement by reducing routine tasks, increasing autonomy, and fostering meaningful pedagogical interactions—critical in high-workload Indonesian HEI context behavioral intention and behavior predict to boost lecturers' motivation and long-term adoption by reducing administrative load. This dual-path influence—utility to intention, engagement to sustained behavior—provides a new theoretical lens from transactional acceptance to revolutionary workplace results. These findings strengthen UTAUT's educational explanations and inform robust AI ecosystem implementation in education. Therefore, the current research aims to understand the factors by extending the unified theory of acceptance and use of technology (UTAUT) framework.

## **2. Literature Review**

The current study extends UTAUT to examine AI integration in HEIs. Technology adoption and use behavior can be predicted and explained using the UTAUT framework, which emphasizes performance expectancy, effort expectancy, social influence, and facilitating conditions. These constructs directly affect behavioral intention and behavior. In this study, the perceived impact of AI on work engagement and the perceived effects of AI on teaching and learning were hypothesized to correlate with behavioral intention and behavior. The involved constructs are linked to the classic UTAUT model to provide a nuanced view of technology adoption and motivating variables needed to sustain and improve AI in education. Thus, this model offers a robust theoretical foundation for HEIs to implement AI applications in education. Although the basic elements of UTAUT reliably forecast AI adoption in HEIs (Venkatesh et al., 2003; Habibi et al., 2023), the conflicting findings regarding effort expectancy necessitate examination. EE substantially influences intention among early adopters (Lin et al., 2022; Li et al., 2024) but declines among seasoned users (Grassini et al., 2024), indicating that skill maturation moderates its impact. Landmark assessments underscore AI's transformative potential (Holmes et al., 2019) while exposing inconsistent integration attributable to institutional deficiencies (Zawacki-Richter et al., 2019). Augmenting UTAUT with job engagement resolves this by correlating behavioral outcomes with enduring motivation.

### **2.1. Performance Expectancy**

Performance expectancy refers to the degree to which a person believes a technology will enhance their job performance (Venkatesh et al., 2003). UTAUT paradigm relies on this concept to explain technology adoption. In the current study, performance expectancy refers to the belief that AI will enhance academic and administrative efficiency in higher education (Venkatesh, 2022; Wu et al., 2022). When educational institutions implement new technology, performance expectancy becomes an important determinant of user engagement with new systems (Sebastián et al., 2022). Venkatesh et al. (2003) identified performance expectancy as one of UTAUT's four significant components, impacting users' technology use intentions. Prior studies have indicated that performance expectancy is a critical factor for teachers or students in using AI in education (Andrews et al., 2021; Sebastián et al., 2022; Habibi, Muhaimin et al., 2023; Lin et al., 2022; Venkatesh, 2022; Wu et al., 2022). For example, performance expectancy influences ChatGPT use among HEIs' students in Indonesia (Habibi, Muhaimin, et al., 2023). When users have reasonable performance expectancy in AI, the technology promises to improve work outcomes and simplify educational processes (Wu et al., 2022). These prior findings demonstrate the continuous importance of understanding the intention of users in using AI in HEIs:

### **2.2. Effort Expectancy**

Effort expectancy in this study measures how easily HEIs lecturers assess their capacity to adopt and apply AI technology in their daily work (Venkatesh et al., 2003). Users are more likely to adopt easy-to-use technology.

UTAUT suggests that effort expectancy predicts behavioral intention, especially early in technology adoption when users are still learning new systems (Alhwaiti, 2023; Grassini et al., 2024). Venkatesh et al.'s (2003) foundational research on technology acceptance emphasized the relevance of effort expectancy, or perceived ease of use, in influencing technology adoption. Recent studies confirmed AI is more likely to be embraced and integrated into teaching and learning activities in HEIs as rapid AI tool adoption is crucial for improving educational outcomes and operational efficiency (Alhwaiti, 2023; Grassini et al., 2024; Habibi, Muhaimin, et al., 2023; Li et al., 2024; Lin et al., 2022; Sebastián et al., 2022). Habibi et al. (2023), who studied the use of ChatGPT in higher education, found that effort expectancy was significantly correlated with the behavioral intention to use AI in education. On the other hand, effort expectancy was not significantly related to behavioral intention to use AI in HEIs (Grassini et al., 2024).

### 2.3. Social Influence

Social influence is defined as the extent to which an individual believes important individuals think they should use AI technology in HEIs (Venkatesh et al., 2003). Coworkers, superiors, and other influential people often shape impressions and decisions. Social influence refers to the perceived pressure or encouragement from colleagues, management, or the academic community to adopt and utilize AI technologies in higher education (Habibi, Muhaimin et al., 2023; Lin et al., 2022). Social influence can motivate technology adoption, especially when social pressure or group norms favor new technology. Venkatesh et al. (2003) noted that social influence is significant in social and professional settings. For example, if lecturers learn from their peers or leaders by using AI tools in their teaching or administration, they may be inspired to do the same. This social dynamic can speed up the adoption of AI in higher education, fostering institutional change and innovation. Prior studies have explored the essential role of social influence in shaping behavioral intention to use AI (Sebastián et al., 2022; Habibi, Muhaimin et al., 2023; Li et al., 2024; Lin et al., 2022). However, Grassini et al. (2024) found that social influence does not significantly affect university students' perceived use of AI in Norway.

### 2.4. Facilitating Conditions

Facilitating Conditions refer to the extent to which an individual perceives that the organizational and technological infrastructure supports technology (Venkatesh, 2022; Wu et al., 2022). Training, technical assistance, and enough resources are included. AI adoption and integration in HEIs requires technological resources, faculty and staff training, and management support. Even advanced technology needs solid support for adoption (Habibi, Muhaimin, et al., 2023; Li et al., 2024). Studies suggest that facilitating conditions are crucial for implementing technology, especially AI in education (Andrews et al., 2021; Cabrera-Sánchez et al., 2021; García de Blanes Sebastián et al., 2022; Grassini et al., 2024; Habibi, Muhaimin, et al., 2023; Li et al., 2024; Lin et al., 2022; Venkatesh, 2022; Wu et al., 2022). In the early years of UTAUT studies, Venkatesh et al. (2003) found that facilitating conditions directly impact technology adoption, especially in user-supporting organizations. Andrews et al. (2021) found that institutions (university libraries) with adequate resources and support boost AI use in education. Khan and Rafi (2021) also noted that faculty need ongoing technical support and professional development to cultivate a positive attitude toward emerging technologies such as AI, which, in turn, enhances their effectiveness in higher education (Clifford & Perez, 2024). One hypothesis was established regarding the role of facilitating conditions in behavioral intention to use IA in HEIs, as perceived by Indonesian university lecturers.

### 2.5. Behavioral Intention

Behavioral intention refers to a person's intention to use technology. Behavioral intention predicts technology use in UTAUT, enabling attitudes and perceptions to become actions. This study defines behavioral intention as the desire of Indonesian HEI lecturers to use AI. Behavioral intention's ability to predict technology adoption or behavior can determine the success of technology integration. Recent research confirms behavioral intention's importance in predicting behavior in educational settings (Andrews et al., 2021; Cabrera-Sánchez et al., 2021; Chatterjee & Bhattacharjee, 2020; Habibi, Muhaimin et al., 2023; Raffaghelli et al., 2022). Cabrera-Sánchez et al. (2021) found that behavioral intention highly predicted AI use in education. Chatterjee and Bhattacharjee (2020) found that excellent intentions boost educational professionals' AI-driven solution use. Supportive surroundings, training, and a clear explanation of AI technology benefits are needed to encourage positive behavior and successful integration of AI technology in education. In the current study, we extended the role of behavioral intention to predict perceived impact of AI on work engagement and AI use in HEIs, which has gained little scholarly attention (AlAjmi, 2022; Al-Takhayneh et al., 2022; Rahiman &

Kodikal, 2024). Behavioral intention, rooted in the TPB, predicts the perceived impact of AI on work engagement. AI's integration in HEIs enhances efficiency and personalized learning, potentially boosting engagement (Rahiman & Kodikal, 2024). Prior studies highlight how positive attitudes toward AI correlate with the increased perceived impact of AI on work engagement, particularly when users perceive AI as beneficial and supportive in their academic and administrative roles (AlAjmi, 2022; Rahiman & Kodikal, 2024). Three hypotheses were developed regarding the role of behavioral intention on behavior, the perceived impact of AI on work engagement, and the impact of AI on teaching and learning.

## **2.6. Behavior**

In this study, behavior refers to how lecturers use AI tools. UTAUT confirms that behavior directly results from behavioral intention, which encompasses AI-graded assignments, tailored learning, and administrative efficiency. Lecturers who use AI-driven tasks may feel more accomplished and efficient, which can boost their job satisfaction. The behavior of lecturers in embracing AI technology can be a major predictor of perceived impact of AI on work engagement and perceived impact of AI on teaching and learning (AlAjmi, 2022; Al-Takhayneh et al., 2022; Rahiman & Kodikal, 2024). Lecturers actively using AI in their education and administrative responsibilities improve their work processes and help the institution adopt AI (Takhayneh et al., 2022). Positive AI experiences lead to broad adoption across departments, fostering an environment where AI is fully integrated into education (AlAjmi, 2022). Behavior is crucial to the perceived impact of AI on work engagement and the successful perceived AI effect on teaching and learning.

## **2.7. Perceived impact of AI on work engagement and perceived impact of AI on teaching and learning**

The way AI affects employees' motivation, commitment, and emotional attachment to their work is reflected in the perceived impact of AI on work engagement (Kong et al., 2021). In this study, AI is being incorporated into more and more professional settings, changing how activities are completed and allowing for increased accuracy and efficiency for students. The decrease in routine and repetitive chores is one substantial effect. AI increases lecturers' happiness and engagement by automating repetitive tasks, allowing them to concentrate on more valuable and innovative work (Sreenivasan & Suresh, 2024). AI is reported to have a revolutionary effect on teaching and learning in HEIs, changing conventional teaching methods and enhancing student and teacher outcomes (Katsamakos et al., 2024). AI allows more individualized and effective educational experiences by introducing adaptive learning technologies, personalized learning paths, and enhanced analytics. AI enables customized learning by assessing each student's needs, pace, and progress through adaptable platforms. It automates tasks like grading, attendance, and report creation, freeing lecturers to focus on mentoring and curriculum development (Rane et al., 2023).

## **2.8. Hypotheses of the study**

With eight variables in the current study (Figure 1), we proposed 10 hypotheses to evaluate their relationships.

- H1: Performance expectancy correlates with behavioral intention.
- H2: Effort expectancy correlates with behavioral intention.
- H3: Social influence correlates with behavioral intention.
- H4: Facilitating conditions correlate with behavioral intention
- H5: Facilitating conditions correlate with behavior.
- H6: Behavioral intention correlates with behavior.
- H7: Behavioral intention correlates with perceived impact of AI on work engagement.
- H8: Behavioral intention correlates with perceived impact of AI on teaching and learning.
- H9: Behavior correlates with the perceived impact of AI on work engagement.
- H10: Behavior correlates with perceived impact of AI on teaching and learning.

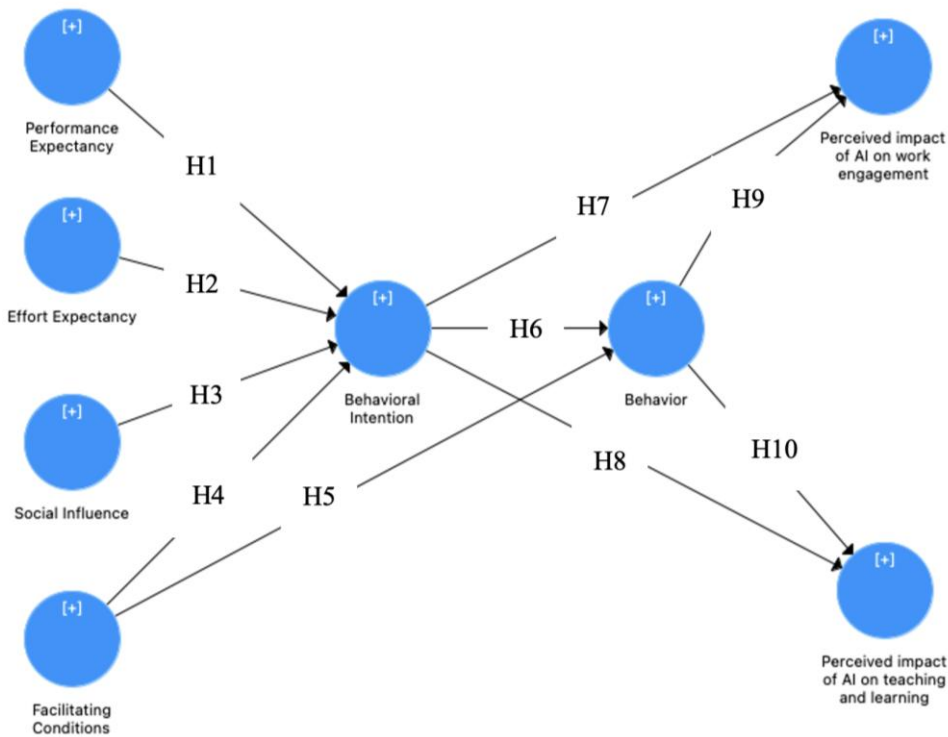


**Figure 1.** Proposed model: factors affecting behavior, perceived impact of AI on work engagement, and AI application

3. Methods

The survey included a Likert scale consisting of 27 items, each measured on a 5-point scale (Cabrera-Sánchez et al., 2021; Rahiman & Kodikal, 2024; Venkatesh, 2022). We requested five educational technology specialists to assess the items for content validity, and we revised them based on their suggestions. The remaining items (n. 24) in Appendix 1 were used for the pilot study and data collection after eliminating three questions due to the changes. The experts were selected based on their doctoral qualifications in educational technology, at least 10 years of experience in curriculum design and assessment, and prior publications in peer-reviewed journals on survey methodology. Validation involved independent assessments in which each expert rated items on a 4-point scale for relevance, clarity, and simplicity, and provided qualitative feedback via email surveys. No focus groups were used; assessments were administered individually to minimize bias. Inter-rater agreement was measured using Fleiss' Kappa ( $\kappa=0.82$ , indicating substantial agreement), while intra-rater reliability was not assessed as it was a one-time evaluation. Expert data-guided revisions: items with average scores below 3.0 or conflicting feedback were modified or removed, ensuring the content validity index (CVI) exceeded 0.80 for the final instrument.

3.1. Data Collection

Using a Google Forms survey, we conducted a pilot study with 51 lecturers to assess the model's reliability. Participants in the pilot study were 18 men and 33 women from a single Indonesian HEI. The pilot research participants were not included in the primary data collection to prevent bias in the work. Good dependability was indicated by a Cronbach's alpha of  $>.700$  (Brown, 2002). The reliability ranges from .773 to .897 (performance expectancy = .773, Effort expectancy = .810, social influence = .776, facilitating conditions = .821, facilitating conditions = .846, behavioral intention = .812, behavior = .757, perceived impact of AI on work engagement GPTU = .878, and perceived impact of AI on teaching and learning .897).

After calculating the first reliability test, we sent the survey to respondents at four Indonesian HEIs for the primary data collection. The questionnaire included a brief description, informed consent, demographic data, and the primary items. Ensuring the respondents employed AI in their activities was the first step in gathering data. We received 550 answers in all. Nonetheless, 516 responses—274 from women and 241 from men—were quantifiable. Ninety-seven lecturers had between one and ten years of professional experience, whereas 191 lecturers had between eleven and twenty years. In the meantime, 231 lecturers have over 20 years of experience in the classroom.

3.2. Data Analysis

We utilised the current SmartPLS statistical tool (SmartPLS 4.0) to test the hypotheses and theoretical framework. We used PLS-SEM for model assessment, measurement and structural models, because of its flexibility with sample sizes and non-normal data (Becker et al., 2023; Rigdon et al., 2017). The current study’s intended population comprises lecturers from four Indonesian HEIs who utilize AI in teaching and research. We checked for normality by examining the data’s skewness and kurtosis. The normality assumption was reviewed to determine if the data distribution met parametric test requirements, though PLS-SEM is robust to non-normality. This check is crucial for justifying the choice of analytical method, ensuring reliability in parameter estimates, and avoiding biased inferences in structural modeling. Presenting normality results enhances transparency, allowing readers to verify data characteristics and the appropriateness of PLS-SEM for handling any deviations from normality in the sample (Rigdon et al., 2017).

4. Results

4.1. Common Method Bias (CMB) and Normality of the Data

To enhance methodological rigor, we addressed potential biases thoroughly. Self-selection was alleviated by focusing on AI-utilizing lecturers from four distinct Indonesian HEIs, hence minimizing volunteer bias. Social desirability was mitigated by employing anonymous responses and neutral item language. The assessment of common method bias (CMB) was thorough; Harman’s single-factor test indicated a variance of 42.3% (below the 50% threshold), and full collinearity variance inflation factors (VIFs) ranged from 1.12 to 2.87 (below 3.3), demonstrating the absence of CMB (Kock, 2015). PLS-SEM was chosen over CB-SEM not only for its versatility but also for the exploratory enhancement of UTAUT with new dimensions (work engagement, teaching impact), non-normative data (skewness/kurtosis  $\pm 2$ ), and a predictive emphasis. These selections guarantee strong, generalizable insights into the dynamics of AI adoption (Abdulmuhsin et al., 2022; Churi et al., 2022).

We checked for normality by examining the data’s skewness and kurtosis; we also reported Mean and standar deviation (SD) (Table 1). The skewness and kurtosis values must be between -2 and +2 for the data to be normal. The skewness and kurtosis values fall within the range of -2 to +2 (Zhang et al., 2023). The lowest values of Skewness are reported to emerge on PAITL3 (-1.197), while the highest is reported for SI3 (.15). For Kurtosis, the values range from 1.932 (PAITL2) to -.921 (PAIWE1). Descriptive reports exhibit a sufficient value of PE1 Mean (4.250) and SD (.749) to EE4 (Mean = 3.472; SD = .961). To alleviate evaluation fear and minimize common-method bias, we included consent forms in the survey questionnaire that did not request personal identifiers. The CMB variation inflation factor (VIF) was determined; values of 4 or greater indicate common-method bias (Podsakoff et al., 2024).

| Item | Mean  | SD   | Kurtosis | Skewness |
|------|-------|------|----------|----------|
| PE1  | 4.25  | .749 | .004     | -.722    |
| PE2  | 4.293 | .668 | -.783    | -.418    |
| PE3  | 4.156 | .706 | -.983    | -.231    |
| EE1  | 4.181 | .67  | -.802    | -.229    |
| EE2  | 4.129 | .683 | -.867    | -.169    |
| EE3  | 3.944 | .72  | -.455    | -.164    |
| SI1  | 3.472 | .961 | -.196    | -.084    |
| SI2  | 3.484 | .977 | -.234    | -.06     |
| SI3  | 3.688 | .875 | -.949    | .15      |
| HM1  | 4.358 | .607 | -.663    | -.376    |
| HM2  | 4.11  | .808 | -.307    | -.576    |

|        |       |       |       |        |
|--------|-------|-------|-------|--------|
| HM3    | 3.942 | .968  | .278  | -.766  |
| FC1    | 4.175 | .701  | .087  | -.527  |
| FC2    | 3.963 | .771  | -.641 | -.216  |
| FC3    | 4.042 | .687  | -.88  | -.055  |
| HB1    | 3.289 | .984  | -.184 | .113   |
| HB2    | 3.351 | .993  | -.433 | -.05   |
| HB3    | 2.998 | 1.097 | -.617 | .135   |
| BI1    | 4.26  | .691  | .362  | -.677  |
| BI2    | 4.198 | .672  | -.816 | -.259  |
| BI3    | 3.909 | 1.019 | -.35  | -.662  |
| B1     | 3.655 | 1.032 | -.768 | -.294  |
| B2     | 3.89  | .932  | .475  | -.883  |
| B3     | 3.89  | .94   | .259  | -.784  |
| PAIWE1 | 3.611 | 1.061 | -.921 | -.198  |
| PAIWE2 | 4.012 | .871  | .539  | -.724  |
| PAIWE3 | 3.705 | 1.059 | -.87  | -.298  |
| PAITL1 | 3.984 | .952  | .593  | -.79   |
| PAITL2 | 4.121 | .811  | 1.923 | -1.073 |
| PAITL3 | 4.239 | .793  | 1.344 | -1.197 |

**Table 1.** Mean, SD, Kurtosis, and Skewness

## 4.2. Measurement Model

A detailed analysis of factors affecting behavior, behavioral intention, and AI's perceived impact on work engagement, teaching, and learning was conducted using PLS-SEM. This analysis employed a measurement model to assess the validity and reliability of the scale. Multiple items represent each construct, with factor loadings indicating the degree to which each item reflects the measure. Items PAITL1, PAITL2, and PAITL3 measure the Perceived impact of AI on teaching and learning construct, with loadings of .951, .960, and .948, respectively. The items' high loadings indicate potent indicators of the construct, encapsulating AI's role in education. Behavior and behavioral intention are also measured by items with high loadings, demonstrating reliability in capturing the intended dimensions. Other constructs (perceived impact of AI on work engagement, performance expectancy, effort expectancy, and facilitating conditions) have also informed a good loading value for each item, with values of greater than .700 (Habibi et al., 2024; Habibi, Sofyan, et al., 2023; Hair et al., 2022). Cronbach's alpha, rho\_A, composite reliability (CR), and average variance extracted (AVE) support the constructs' reliability. The structures' Cronbach's alpha values range from .678 to .949, with most above .700, indicating excellent internal consistency (Benitez et al., 2020; Henseler, 2017; Sarstedt et al., 2022; Schuberth et al., 2022). Dijkstra-Henseler's rho (rho\_A) is usually more robust than Cronbach's alpha and CR values. Rho\_A values (.763-.949) further support internal consistency, lying between alpha and CR. All values are above .700, confirming the constructs' reliability and ensuring that the items collectively measure the constructs consistently (Dijkstra & Henseler, 2015). In detail, the measurement model exhibits strong to excellent psychometric properties across all constructs. Behavior shows excellent internal consistency (alpha = .896, rho\_A = .898, CR = .935), while Behavioral intention demonstrates very good reliability (alpha = .890, rho\_A = .895, CR = .931). Effort expectancy possesses the lowest alpha (.678) but remains acceptable thanks to a solid rho\_A (.798) and CR (.854). Facilitating conditions has good reliability (alpha = .762, rho\_A = .763, CR = .862). The perceived impact of AI on teaching and learning achieves outstanding reliability (alpha = .949, rho\_A = .949, CR = .967), and both indicators of the perceived impact of AI on work engagement display excellent consistency (alpha = .926, rho\_A = .926, CR = .953). Performance Expectancy exhibits very strong reliability (alpha = .906, rho\_A = .907, CR = .941), and Social Influence provides good, acceptable values (alpha = .772, rho\_A = .785, CR = .868). The structures account for more variance than measurement error, as AVE values range from .677 to .908. This thorough examination confirms constructs are well-defined and accurately assessed, underpinning future analysis and interpretation (Table 2).

| Variable                                        | Alpha | rho_A | CR   | AVE  |
|-------------------------------------------------|-------|-------|------|------|
| Behavior                                        | .896  | .898  | .935 | .828 |
| Behavioral Intention                            | .890  | .895  | .931 | .819 |
| Effort Expectancy                               | .678  | .798  | .854 | .746 |
| Facilitating Conditions                         | .762  | .763  | .862 | .677 |
| Perceived impact of AI on teaching and learning | .949  | .949  | .967 | .908 |
| Perceived impact of AI on work engagement       | .926  | .926  | .953 | .871 |
| Performance Expectancy                          | .906  | .907  | .941 | .843 |
| Social Influence                                | .772  | .785  | .868 | .687 |

**Table 2.** Reliability and convergent validity: factors affecting behavior, perceived impact of AI on work engagement, and AI application

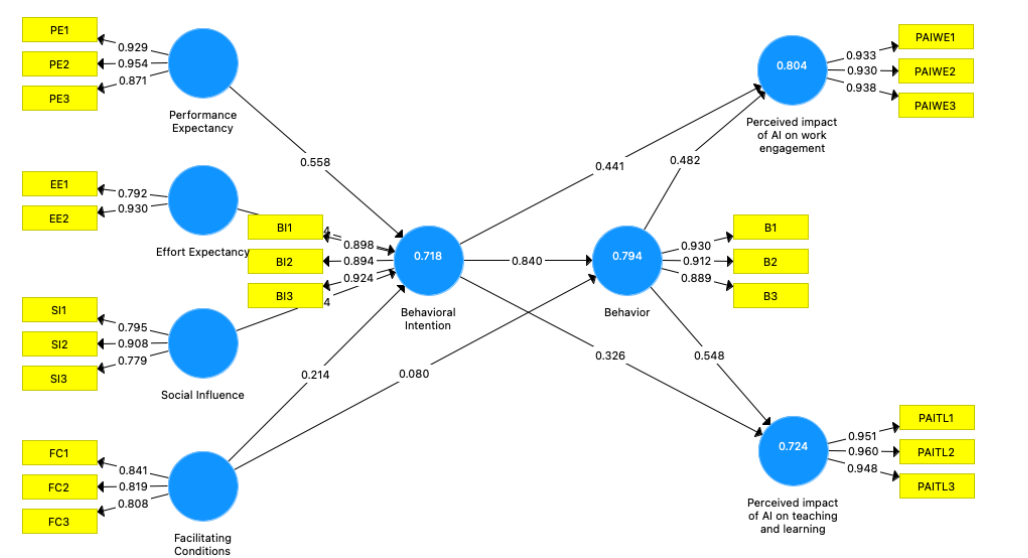


**Figure 2.** Measurement model: factors affecting behavior, perceived impact of AI on work engagement, and AI application

The Fornell-Larcker Criterion assesses discriminant validity; each construct's diagonal value, the square root of AVE, shows its variance relative to measurement error (Hamid et al., 2017). The perceived impact of AI on teaching and learning has a value of .953, indicating a high construct representation. Off-diagonal construct correlations should be lower than diagonal values to confirm discriminant validity. Behavior (.910), behavioral intention (.905), and perceived impact of AI on work engagement (.934) have high diagonal values, demonstrating internal consistency and validity. The discriminant validity of constructs like perceived impact of AI on teaching and learning with behavior (.838) and behavioral intention and lower correlations than diagonal values support perceived impact of AI on work engagement (.869). Effort expectancy (.864) and facilitating conditions (.823) similarly have good discriminant validity because their correlations with other dimensions are smaller than their AVE square roots. This analysis affirms each construct uniquely measures perceived AI impact on teaching and learning, providing a framework for understanding their relationships. Table 3 shows discriminant validity, supporting reliable educational research and implementation.

|                         | B    | BI   | EE   | FC   | PAITL | PAIWE | PE | SI |
|-------------------------|------|------|------|------|-------|-------|----|----|
| Behavior                | .910 |      |      |      |       |       |    |    |
| Behavioral Intention    | .888 | .905 |      |      |       |       |    |    |
| Effort Expectancy       | .557 | .590 | .864 |      |       |       |    |    |
| Facilitating Conditions | .593 | .612 | .727 | .822 |       |       |    |    |

|                                                 |      |      |      |      |      |      |      |      |
|-------------------------------------------------|------|------|------|------|------|------|------|------|
| Perceived impact of AI on teaching and learning | .837 | .813 | .563 | .640 | .953 |      |      |      |
| Perceived impact of AI on work engagement       | .873 | .869 | .553 | .558 | .792 | .933 |      |      |
| Performance Expectancy                          | .836 | .800 | .585 | .526 | .711 | .791 | .918 |      |
| Social Influence                                | .677 | .622 | .414 | .428 | .613 | .664 | .567 | .829 |

**Table 3.** Fornell-Larcker Criterion: factors affecting behavior, perceived impact of AI on work engagement, and AI application

The cross-loading analysis of the current study shows how well each item correlates with its concept relative to others, which is essential for discriminant validity (Hair et al., 2019). For example, PAITL1, PAITL2, and PAITL3 have significant loadings on the perceived impact of AI on teaching and learning construct (.951, .960, and .948), showing strong alignment compared to other loading values on other constructs. Items (B1, B2, and B3) also have high behavior construct loadings (.930, .912, and .889), confirming their significance better than others. The result shows that most items load stronger on their constructs than others, supporting the measurement model's discriminant validity (Table 4). This study guarantees each construct's distinctness and precise quantification, supporting AI research and implementation in higher education.

|        | <b>B</b>    | <b>BI</b>   | <b>EE</b>   | <b>FC</b>   | <b>PAITL</b> | <b>PAIWE</b> | <b>PE</b>   | <b>SI</b>   |
|--------|-------------|-------------|-------------|-------------|--------------|--------------|-------------|-------------|
| B1     | <b>.930</b> | .818        | .582        | .571        | .753         | .864         | .821        | .591        |
| B2     | <b>.912</b> | .820        | .482        | .545        | .786         | .802         | .818        | .612        |
| B3     | <b>.889</b> | .788        | .454        | .504        | .750         | .715         | .637        | .650        |
| BI1    | .781        | <b>.898</b> | .438        | .559        | .751         | .779         | .715        | .525        |
| BI2    | .729        | <b>.894</b> | .536        | .531        | .697         | .724         | .644        | .570        |
| BI3    | .893        | <b>.924</b> | .623        | .571        | .758         | .850         | .804        | .594        |
| EE1    | .289        | .367        | <b>.792</b> | .578        | .256         | .308         | .293        | .205        |
| EE2    | .610        | .610        | <b>.930</b> | .676        | .639         | .593         | .647        | .460        |
| FC1    | .453        | .442        | .601        | <b>.841</b> | .482         | .398         | .399        | .386        |
| FC2    | .474        | .520        | .607        | <b>.819</b> | .462         | .495         | .432        | .298        |
| FC3    | .530        | .538        | .588        | <b>.808</b> | .624         | .476         | .460        | .373        |
| PAITL1 | .780        | .777        | .532        | .640        | <b>.951</b>  | .753         | .692        | .627        |
| PAITL2 | .811        | .789        | .535        | .581        | <b>.960</b>  | .784         | .724        | .574        |
| PAITL3 | .804        | .758        | .544        | .612        | <b>.948</b>  | .729         | .616        | .555        |
| PAIWE1 | .843        | .809        | .555        | .500        | .681         | <b>.933</b>  | .763        | .593        |
| PAIWE2 | .801        | .833        | .519        | .489        | .760         | <b>.930</b>  | .743        | .596        |
| PAIWE3 | .803        | .792        | .473        | .575        | .782         | <b>.938</b>  | .709        | .675        |
| PE1    | .800        | .742        | .494        | .475        | .662         | .723         | <b>.929</b> | .542        |
| PE2    | .799        | .733        | .517        | .434        | .647         | .712         | <b>.954</b> | .518        |
| PE3    | .704        | .728        | .601        | .541        | .650         | .745         | <b>.871</b> | .504        |
| SI1    | .439        | .416        | .247        | .348        | .425         | .436         | .342        | <b>.795</b> |
| SI2    | .581        | .557        | .395        | .338        | .511         | .578         | .514        | <b>.908</b> |
| SI3    | .638        | .552        | .366        | .380        | .572         | .614         | .526        | <b>.779</b> |

**Table 4.** Cross-loading: factors affecting behavior, perceived impact of AI on work engagement, and AI application

The HTMT (Heterotrait-Monotrait) ratio of correlations, a crucial metric for evaluating discriminant validity, is shown in Table 5. HTMT compares average heterotrait to monotrait correlations to assess construct differences. Values below .85 indicate sufficient discriminant validity (or .90 for conceptually comparable items). Strong correlations between behavioral intention and important constructs, specifically performance expectancy (.924) and perceived impact of AI on teaching and learning (.908), highlight the significance of perceived utility and effectiveness in influencing adoption intentions. A substantial correlation between effort expectancy and behavioral intention (.720) and performance expectancy (.689) indicates that the ease of use of AI systems is significant. The moderate associations between facilitating conditions and behavioral

intention (.736) and perceived AI on teaching and learning (.746) highlight the importance of outside assistance and resources in boosting adoption intentions. With a strong correlation between behavioral intention (.953) and performance expectancy (.863), the perceived impact of AI on work engagement underscores the importance of AI's beneficial effects on engagement and performance expectations. Although less significant than direct performance and engagement elements, Social influence represents the influence of peer and societal perceptions and correlates reasonably with behavioral intention (.740) and performance expectancy (.664). Adequate discriminant validity is confirmed by HTMT results below the threshold, guaranteeing the measured constructs are unique and the model's linkages are strong and clear.

|                                                 | B    | BI   | EE   | FC   | PAITL | PAIW<br>E | PE   |
|-------------------------------------------------|------|------|------|------|-------|-----------|------|
| Behavior                                        |      |      |      |      |       |           |      |
| Behavioral Intention                            | .990 |      |      |      |       |           |      |
| Effort Expectancy                               | .661 | .720 |      |      |       |           |      |
| Facilitating Conditions                         | .713 | .736 | .903 |      |       |           |      |
| Perceived impact of AI on teaching and learning | .908 | .883 | .641 | .746 |       |           |      |
| Perceived impact of AI on work engagement       | .956 | .953 | .652 | .659 | .845  |           |      |
| Performance Expectancy                          | .924 | .886 | .689 | .629 | .766  | .863      |      |
| Social Influence                                | .802 | .740 | .517 | .558 | .708  | .774      | .664 |

**Table 5.** HTMT: factors affecting behavior, perceived impact of AI on work engagement, and AI application

4.3. Structural Model

Using PLS-SEM procedures, we analyzed the factors that affect behavioral intention, behavior, perceived impact of AI on work engagement, and perceived impact of AI on teaching and learning through bootstrapping the data (5000 subsamples). Performance expectancy was the strongest predictor of behavioral intention, with a path coefficient of .558, indicating that technology's benefits and usefulness greatly influence users' intention to use AI. A t-value of 16.701 and an effect size ( $f^2$ ) of .579 show that performance expectancy is statistically significant and strongly influences behavioral intention. However, effort expectancy, which evaluates the ease of use of technology, had no significant effect on behavioral intention, with a path coefficient of .024 and a t-value of .512. The finding shows that AI adoption in HEIs is not influenced by ease of use, potentially due to users' knowledge or comfort with the technology, making effort expectancy unimportant.

Social influence, another essential factor, moderately affected behavioral intention with a path coefficient of .204 and a t-value of 7.518. The report suggests that peer pressure, cultural norms, and colleagues' and leaders' influence influence instructional AI engagement goals. The moderate impact size of .097 shows that social influence drives behavioral intention more than social influence. Facilitating conditions, which examine technology resources and assistance, also had a substantial but smaller effect on behavioral intention (path coefficient = .214, t-value = 5.080) and behavior (path coefficient = .080, t-value = 2.856). The model's best direct link between behavioral intention and behavior was .840, indicating that once people want to use AI, they are likely to do so. Both behavior and behavioral intention had significant positive effects on perceived impact of AI on work engagement and perceived impact of AI on teaching and learning, demonstrating their interconnectedness and the central role of behavioral intention and behavior in educational AI integration as well as perceived impact of AI on work engagement (Table 6 and Figure 3).

| H  | Relationship                                    | $\beta$ | t value | p values | $f^2$  | Sign. |
|----|-------------------------------------------------|---------|---------|----------|--------|-------|
| H1 | Performance Expectancy -> Behavioral Intention  | .558    | 16.786  | .000     | .5796  | Yes   |
| H2 | Effort Expectancy -> Behavioral Intention       | .024    | .520    | .603     | .0008  | No    |
| H3 | Social influence -> Behavioral Intention        | .204    | 7.372   | .000     | .0969  | Yes   |
| H4 | Facilitating Conditions -> Behavioral Intention | .214    | 5.080   | .000     | .0725  | Yes   |
| H5 | Facilitating Conditions -> Behavior             | .080    | 2.856   | .004     | .0194  | Yes   |
| H6 | Behavioral intention -> Behavior                | .840    | 36.540  | .000     | 2.1383 | Yes   |

|     |                                                                            |      |        |      |       |     |
|-----|----------------------------------------------------------------------------|------|--------|------|-------|-----|
| H7  | Behavioral intention -><br>Perceived impact of AI on work engagement       | .441 | 11.458 | .000 | .2094 | Yes |
| H8  | Behavioral intention -><br>Perceived impact of AI on teaching and learning | .326 | 4.692  | .000 | .0811 | Yes |
| H9  | Behavior -><br>Perceived impact of AI on work engagement                   | .482 | 11.952 | .000 | .2494 | Yes |
| H10 | Behavior -><br>Perceived impact of AI on teaching and learning             | .548 | 8.090  | .000 | .2293 | Yes |

**Table 6.** Structural model: factors affecting behavior, perceived impact of AI on work engagement, and AI application

The current research also presents the  $R^2$  and  $Q^2$  values for various constructs related to AI applications in HEIs, providing insights into the model's predictive accuracy and relevance. The  $R^2$  value, or coefficient of determination, indicates the proportion of variance in the dependent variable that is predictable from the independent variables (Habibi et al., 2020; Hair et al., 2019). For the perceived impact of AI on teaching and learning, the  $R^2$  value is .724, suggesting that the model explains 72.4% of the variance, indicating a strong predictive power. Similarly, the behavior construct has an  $R^2$  of .794, meaning that 79.4% of its variance is accounted for, reflecting a robust model fit.

Behavioral intention shows an  $R^2$  of .718, while the perceived impact of AI on work engagement has an  $R^2$  of .804, indicating substantial explanatory power. The  $Q^2$  value, derived from the Stone-Geisser test, assesses the model's predictive relevance through a blindfolding procedure. A  $Q^2$  value greater than zero suggests the model has predictive relevance for a particular construct. The perceived impact of AI on teaching and learning has a  $Q^2$  of .653, indicating good predictive relevance. Behavior, with a  $Q^2$  of .652, also demonstrates strong predictive relevance (Table 7). Behavioral intention and perceived impact of AI on work engagement have  $Q^2$  values of .579 and .695, respectively, further supporting the model's predictive capabilities. Overall, the high  $R^2$  and  $Q^2$  values across these constructs suggest that the model is accurate and relevant in predicting the impact of perceived impact of AI on teaching and learning (Table 7). This analysis underscores the model's effectiveness in capturing the dynamics of AI integration, behavior, intention, and engagement, providing a solid foundation for further research and practical application in educational settings.

|                                                 | $R^2$ | $Q^2$ |
|-------------------------------------------------|-------|-------|
| Perceived impact of AI on teaching and learning | .724  | .653  |
| Behavior                                        | .794  | .652  |
| Behavioral intention                            | .718  | .579  |
| Perceived impact of AI on work engagement       | .804  | .695  |

**Table 7.**  $R^2$  and  $Q^2$

5. Discussion

The model's validation and reliability in exploring factors affecting behavioral intention, behavior, perceived impact of AI on work engagement, and perceived impact of AI on teaching and learning were rigorously established through Content Validity Index (CVI), a pilot study, and measurement model assessment. CVI obtained initial content validity. Five educational technology experts with doctorates, with over 10 years of curriculum design experience and survey methodology publications, independently evaluated the 27 survey items on a 4-point scale for relevance, clarity, and representativeness. Quantifying inter-rater agreement with Fleiss' Kappa ( $\kappa=0.82$ ) revealed significant consensus (Habibi et al., 2024). The instrument's alignment with the structures was confirmed by 24 items with an overall CVI greater than 0.80 after revising or eliminating items below 3.0 or with contradicting feedback. Following this, 50 Indonesian HEI lecturers who employ AI in teaching and research participated in a pilot project. It identified minor phrasing and response-pattern issues and tested the instrument's practicality. Initial internal consistency was shown by subscale Cronbach's alpha of 0.78–0.92. Final revisions improved comprehensibility without affecting fundamental validity. Finally, SmartPLS 4.0 was used to evaluate the measurement model via PLS-SEM. All indicators had outer

loadings above 0.70, confirming convergence. The Fornell-Larcker criterion and Heterotrait-Monotrait (HTMT) ratios below 0.85 confirmed reliability and discriminant validity. Composite reliability (CR) scores were over 0.80, and Average Variance Extracted (AVE) exceeded 0.50. Skewness and kurtosis validations supported the use of PLS-SEM with non-normal data (Becker et al., 2023; Rigdon et al., 2017). This comprehensive methodology maintained the model's robustness in capturing AI's implications for lecturers' intentions, behaviors, and perceived impacts, enabling hypothesis testing and theoretical insights.

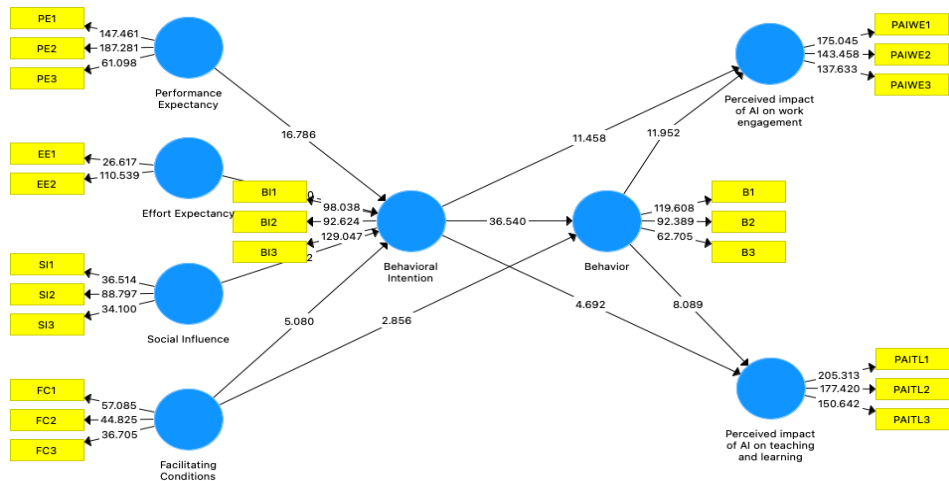


Figure 3. Structural model

The study found that performance expectancy positively affects behavioral intention with a path coefficient of .558, consistent with technology adoption research. In HEIs, users are more likely to adopt and employ AI technologies if they believe they will improve their academic or professional performance. UTAUT confirms that people are motivated by expected performance benefits from new technology (Venkatesh, 2022). Recent research has confirmed this, especially in educational contexts where AI's significance in improving learning results is becoming more apparent (Andrews et al., 2021; Habibi, Muhaimin et al., 2023; Lin et al., 2022; Sebastián et al., 2022; Venkatesh, 2022; Wu et al., 2022). Habibi, Muhaimin, et al. (2023) found that performance expectancy significantly influences university lecturers' propensity to employ AI-driven learning technologies. However, effort expectancy, which evaluates the ease of use of technology, had no significant effect on Behavioral Intention (path coefficient = .024). This suggests that HEIs users may be more concerned with the potential benefits of AI than the effort necessary to utilize AI technologies, especially if they believe the long-term performance benefits surpass any early effort-related hurdles, similar to a prior study (Grassini et al., 2024). These findings contradict previous studies on UTAUT and technology acceptance (Sebastián et al., 2022; Habibi, Muhaimin, et al., 2023; Li et al., 2024; Lin et al., 2022). For instance, Habibi et al. (2023) found effort expectancy to be a key determinant in technology or system uptakes. Growing familiarity with digital tools among educators may have reduced the perceived difficulty of adopting new technologies.

Social influence predicts behavioral intention with a path coefficient of .204. Peers, coworkers, and academic leaders can influence people's AI adoption aspirations. UTAUT model confirms that social impact strongly influences behavioral intention, especially in organizational settings where peer pressure and institutional norms are necessary (Venkatesh, 2022). Prior studies have found that social norms or influences and peer endorsements considerably increase educators' willingness to use AI (Grassini et al., 2024; Habibi, Muhaimin et al., 2023; Li et al., 2024; Lin et al., 2022; Sebastián et al., 2022). Creating a supportive culture surrounding AI use in HEIs, where leadership and peers endorse its use, may increase adoption rates. Facilitating conditions, such as technology resources and support, also significantly affected behavioral intention (path coefficient of .214) and behavior (.080). This finding supports past technology adoption research that facilitating environments are necessary but not sufficient to drive behavior (Andrews et al., 2021; Cabrera-Sánchez et al., 2021; García de Blanes Sebastián et al., 2022; Grassini et al., 2024; Habibi, Muhaimin, et al., 2023; Li et al., 2024; Lin et al., 2022; Venkatesh, 2022; Wu et al., 2022). Supportive settings improve technology adoption early on but decrease as users become more familiar with AI. The findings show that enough infrastructure, training, and technical support are necessary to establish an intention to utilize AI, but it has little effect on actual utilization.

Behavioral intention predicted behavior best in the study, with a path coefficient of .84. Intention is crucial to whether people will engage in an activity, according to the UTAUT model. Lecturers in Indonesian HEIs are more likely to implement AI if lecturers and students are strongly encouraged to use it. This is important in contexts where AI technologies are readily available and accessible, lowering utilization barriers. The literature shows that purpose strongly influences behavior (Andrews et al., 2021; Cabrera-Sánchez et al., 2021; Chatterjee & Bhattacharjee, 2020; Habibi, Muhaimin, et al., 2023; Raffaghelli et al., 2022). Habibi, Muhaimin, et al. (2023) found that strong intentions influence behavior in many circumstances, including technology adoption. Behavioral intention also significantly predicted perceived impact of AI on work engagement (path coefficient of .441) and perceived impact of AI on teaching and learning (path coefficient of .326), suggesting that motivated AI adopters work harder and use AI more in school. Job demands and resources influence perceived impact of AI on work engagement (AlAjmi, 2022; Al-Takhayneh et al., 2022; Rahiman & Kodikal, 2024). AI can improve the perceived impact of AI on work engagement and the perceived effect of AI on teaching and learning if users are motivated to use the technology in their daily activities.

The study concludes that behavior affects perceived impact of AI on work engagement (.482) and perceived impact of AI on teaching and learning (.548). The correlation shows that AI technologies improve the perceived impact of AI on work engagement and lead to more AI adoption in education. These findings are crucial to comprehending HEI AI adoption's long-term effects. The substantial association between behavior and perceived impact of AI on work engagement shows that using AI makes people more invested in their work (AlAjmi, 2022; Al-Takhayneh et al., 2022; Rahiman & Kodikal, 2024). Recent research (Rahiman & Kodikal, 2024) has examined how AI in education might boost the perceived impact of AI on work engagement by making tasks more efficient and fulfilling. Behavior's considerable impact on AI applications suggests that actual usage drives AI's widespread adoption in education. Once people use AI, they will discover benefits and apply it in various cases, leading to broader organizational adoption. Future research should explore how behavior mediates between intention and outcomes, and how different AI applications influence perceptions of AI's impact on work engagement and performance.

This study extends UTAUT by incorporating perceived AI impact on engagement and outcomes, creating a comprehensive framework linking technology acceptance with motivation and pedagogical effectiveness in higher education. Using the extended UTAUT shows that behavioral intention and behavior drive adoption and sustain lecturers' commitment by reducing administrative burdens and enabling meaningful instructional design. This reciprocal mechanism—where AI behavior reinforces engagement—offers a fresh theoretical viewpoint unseen in past UTAUT investigations in high-pressure Indonesian HEI contexts, where faculty often shoulder significant teaching, research, and service loads. The model exhibits better predictive power, explaining job engagement and teaching impact variance, surpassing standard UTAUT criteria (~50-60%). Moreover, experienced Indonesian lecturers' insignificant effort expectancy indicates a maturation effect, with performance expectancy dominating intention formulation as digital familiarity increases. This contextual nuance refines UTAUT's temporal dynamics and emphasizes stage-specific adoption models, adding theoretical depth beyond replication studies like Grassini et al. (2024). The results leave crucial holes that future research must fill to improve generalizability and practicality. Digital literacy, organisational culture, and institutional type (public vs. private, urban vs. rural) may moderate pathway strength. For example, inadequate digital literacy may increase effort expectancy in late adopters. Longitudinal designs are needed to determine if AI fatigue reduces engagement gains. Policymakers and HEI leaders should consider a strategic pivot as conducive conditions predict intention and behavior, suggesting infrastructure alone is insufficient. To avoid digital divides, investments must prioritize role-specific training, performance-focused communication (e.g., measurable time savings), and equity-driven deployment. If human-centered implementation protects teacher-student interactions, scalable customization improves education quality, graduate preparedness for Industry 4.0, and institutional ecosystems. The extended UTAUT provides a model for revolutionary, sustainable AI integration in higher education, especially in Indonesia.

## 6. Conclusion

This study sheds light on how Indonesian HEIs adopt and use AI technologies. Performance expectancy and behavioral intention were key factors, highlighting the importance of perceived benefits and strong intentions in AI adoption. Social influence and facilitating conditions greatly affected behavioral intention but had little effect on actual behavior, underscoring the complexity of technology adoption. The significant correlation between behavioral intention and behavior predicts that motivated people would engage more with AI, improving the perceived impact of AI on work engagement and broader AI use in education. These findings add to the literature on AI adoption in education and demonstrate the value of expanded UTAUT models in understanding technological acceptance dynamics.

However, this study includes shortcomings that should be addressed in future research. First, cross-sectional statistics only record a snapshot in time, limiting causal inferences. Longitudinal research would help determine how these linkages change and how AI adoption affects the perceived impact of AI on work engagement and education. This study focused on AI in HEIs; hence, its conclusions may not apply to other settings or technologies. Future research should encompass educational contexts and examine moderating variables like organizational culture and individual variations affecting AI acceptance and use. Future studies can overcome these constraints to better understand technology adoption in education and inform ways to integrate AI into teaching and learning.

## References

- Abdulmuhsin, A. A., Değirmenci, B., Efendioğlu, İ. H., & Durmaz, Y. (2022). The perception of COVID-19 and avoidance behavior in Turkey: the role of income level, gender and education. *International Journal of Emerging Markets*, 17(10). <https://doi.org/10.1108/IJOEM-11-2020-1308>
- AlAjmi, M. K. (2022). The impact of digital leadership on teachers' technology integration during the COVID-19 pandemic in Kuwait. *International Journal of Educational Research*, 112. <https://doi.org/10.1016/j.ijer.2022.101928>
- Alhwaiti, M. (2023). Acceptance of Artificial Intelligence Application in the Post-Covid Era and Its Impact on Faculty Members' Occupational Well-being and Teaching Self-efficacy: A Path Analysis Using the UTAUT 2 Model. *Applied Artificial Intelligence*, 37(1). <https://doi.org/10.1080/08839514.2023.2175110>
- Al-Takhayneh, S. K., Karaki, W., Hasan, R. A., Chang, B. L., Shaikh, J. M., & Kanwal, W. (2022). Teachers' psychological resistance to digital innovation in jordanian entrepreneurship and business schools: Moderation of teachers' psychology and attitude toward educational technologies. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1004078>
- An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2023). Modeling students' perceptions of artificial intelligence-assisted language learning. *Computer-Assisted Language Learning*. <https://doi.org/10.1080/09588221.2023.2246519>
- Andrews, J. E., Ward, H., & Yoon, J. (2021). UTAUT as a Model for Understanding Intention to Adopt AI and Related Technologies among Librarians. *The Journal of Academic Librarianship*, 47(6), 102437. <https://doi.org/10.1016/j.acalib.2021.102437>
- Ashraf, D. (2023). *Digital Finance and Artificial Intelligence: Islamic Finance Challenges and Prospects*. [https://doi.org/10.1007/978-3-031-13302-2\\_11](https://doi.org/10.1007/978-3-031-13302-2_11)
- Benitez, J., Henseler, J., Castillo, A., & Schubert, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information and Management*, 57(2). <https://doi.org/10.1016/j.im.2019.05.003>
- Brown, J. D. (2002). The Cronbach alpha reliability estimate. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 6(1).
- Cabrera-Sánchez, J. P., Villarejo-Ramos, Á. F., Liébana-Cabanillas, F., & Shaikh, A. A. (2021). Identifying relevant segments of AI applications adopters – Expanding the UTAUT2's variables. *Telematics and Informatics*, 58. <https://doi.org/10.1016/j.tele.2020.101529>
- Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: a quantitative analysis using structural equation modelling. *Education and Information Technologies*, 25(5). <https://doi.org/10.1007/s10639-020-10159-7>
- Churi, P., Pawar, A. V., & Abdulmuhsin, A. A. (2022). Perception of privacy issues and awareness in health-care knowledge management systems: empirical study in Indian health-care context. *International Journal of Organizational Analysis*, 30(5). <https://doi.org/10.1108/IJOA-11-2020-2486>
- Clifford, R., & Perez, L. (2024). AI in higher education: Faculty perspective towards artificial intelligence through UTAUT approach. *HCMCOUJS-Social Sciences*, 14(4). <https://doi.org/10.46223/HCMCOUJS>
- Deng, J., & Lin, Y. (2023). The Benefits and Challenges of ChatGPT: An Overview. *Frontiers in Computing and Intelligent Systems*, 2(2). <https://doi.org/10.54097/fcis.v2i2.4465>
- Dijkstra, T. K., & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics and Data Analysis*, 81. <https://doi.org/10.1016/j.csda.2014.07.008>

- Grassini, S., Aasen, M. L., & Møgelvang, A. (2024). Understanding University Students' Acceptance of ChatGPT: Insights from the UTAUT2 Model. *Applied Artificial Intelligence*, 38(1). <https://doi.org/10.1080/08839514.2024.2371168>
- Habibi, A., Muhaimin, M., Danibao, B. K., Wibowo, Y. G., Wahyuni, S., & Octavia, A. (2023). ChatGPT in higher education learning: Acceptance and use. *Computers and Education: Artificial Intelligence*, 5. <https://doi.org/10.1016/j.caeai.2023.100190>
- Habibi, A., Mukminin, A., & Sofyan, S. (2024). Access to the digital technology of urban and suburban vocational schools. *Education and Information Technologies*, 29(4). <https://doi.org/10.1007/s10639-023-12006-x>
- Habibi, A., Sofyan, S., & Mukminin, A. (2023). Factors affecting digital technology access in vocational education. *Scientific Reports*, 13(1), 5682. <https://doi.org/10.1038/s41598-023-32755-6>
- Habibi, A., Yusop, F. D., & Razak, R. A. (2020). The role of TPACK in affecting pre-service language teachers' ICT integration during teaching practices: Indonesian context. *Education and Information Technologies*, 25(3), 1929–1949. <https://doi.org/10.1007/s10639-019-10040-2>
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. In *European Business Review* (Vol. 31, Issue 1, pp. 2–24). <https://doi.org/10.1108/EBR-11-2018-0203>
- Hair, J., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *PLS-SEM Book: A Primer on PLS-SEM (3rd Ed.)*. Thousand Oaks: Sage.
- Hamid, M. R., Sami, W., & Mohmad Sidek, M. H. (2017). Discriminant Validity Assessment: Use of Fornell & Larcker criterion versus HTMT Criterion. *Journal of Physics: Conference Series*, 890(1). <https://doi.org/10.1088/1742-6596/890/1/012163>
- Henseler, J. (2017). *Partial Least Squares Path Modeling* (pp. 361–381). [https://doi.org/10.1007/978-3-319-53469-5\\_12](https://doi.org/10.1007/978-3-319-53469-5_12)
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education. Promise and Implications for Teaching and Learning*.
- Katsamakas, E., Pavlov, O. V., & Saklad, R. (2024). Artificial Intelligence and the Transformation of Higher Education Institutions: A Systems Approach. *Sustainability*, 16(14), 6118. <https://doi.org/10.3390/su16146118>
- Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of E-Collaboration*, 11(4), 1–10. <https://doi.org/10.4018/ijec.2015100101>
- Kong, H., Yuan, Y., Baruch, Y., Bu, N., Jiang, X., & Wang, K. (2021). Influences of artificial intelligence (AI) awareness on career competency and job burnout. *International Journal of Contemporary Hospitality Management*, 33(2). <https://doi.org/10.1108/IJCHM-07-2020-0789>
- Li, L., Peng, W., & Rheu, M. M. J. (2024). Factors Predicting Intentions of Adoption and Continued Use of Artificial Intelligence Chatbots for Mental Health: Examining the Role of UTAUT Model, Stigma, Privacy Concerns, and Artificial Intelligence Hesitancy. *Telemedicine and E-Health*, 30(3). <https://doi.org/10.1089/tmj.2023.0313>
- Lin, H. C., Ho, C. F., & Yang, H. (2022). Understanding adoption of artificial intelligence-enabled language e-learning system: An empirical study of UTAUT model. *International Journal of Mobile Learning and Organisation*, 16(1). <https://doi.org/10.1504/IJMLO.2022.119954>
- Podsakoff, P. M., Podsakoff, N. P., Williams, L. J., Huang, C., & Yang, J. (2024). Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix. In *Annual Review of Organizational Psychology and Organizational Behavior* (Vol. 11). <https://doi.org/10.1146/annurev-orgpsych-110721-040030>
- Raffaghelli, J. E., Rodríguez, M. E., Guerrero-Roldán, A. E., & Bañeres, D. (2022). Applying the UTAUT model to explain the students' acceptance of an early warning system in Higher Education. *Computers and Education*, 182. <https://doi.org/10.1016/j.compedu.2022.104468>
- Ragheb, M. A., Tantawi, P., Farouk, N., & Hatata, A. (2022). Investigating the acceptance of applying chat-bot (Artificial intelligence) technology among higher education students in Egypt. In *International Journal of Higher Education Management (IJHEM)* (Vol. 8). CBER. [www.cberuk.com](http://www.cberuk.com)
- Rahiman, H. U., & Kodikal, R. (2024). Revolutionizing education: Artificial intelligence empowered learning in higher education. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2023.2293431>

- Rane, N., Choudhary, S., & Rane, J. (2023). Education 4.0 and 5.0: Integrating Artificial Intelligence (AI) for Personalized and Adaptive Learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4638365>
- Sarstedt, M., Radomir, L., Moisescu, O. I., & Ringle, C. M. (2022). Latent class analysis in PLS-SEM: A review and recommendations for future applications. *Journal of Business Research*, 138. <https://doi.org/10.1016/j.jbusres.2021.08.051>
- Schuberth, F., Rademaker, M. E., & Henseler, J. (2022). Assessing the overall fit of composite models estimated by partial least squares path modeling. *European Journal of Marketing*. <https://doi.org/10.1108/EJM-08-2020-0586>
- Sebastián, M., Sarmiento Guede, J. R., & Antonovica, A. (2022). Application and extension of the UTAUT2 model for determining behavioral intention factors in use of the artificial intelligence virtual assistants. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.993935>
- Sreenivasan, A., & Suresh, M. (2024). Design thinking and artificial intelligence: A systematic literature review exploring synergies. *International Journal of Innovation Studies*, 8(3), 297–312. <https://doi.org/10.1016/j.ijis.2024.05.001>
- Venkatesh, V. (2022). Adoption and use of AI tools: a research agenda grounded in UTAUT. *Annals of Operations Research*, 308(1–2). <https://doi.org/10.1007/s10479-020-03918-9>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly: Management Information Systems*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Wu, W., Zhang, B., Li, S., & Liu, H. (2022). Exploring Factors of the Willingness to Accept AI-Assisted Learning Environments: An Empirical Investigation Based on the UTAUT Model and Perceived Risk Theory. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.870777>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? In *International Journal of Educational Technology in Higher Education* (Vol. 16, Issue 1). <https://doi.org/10.1186/s41239-019-0171-0>
- Zhang, D., Jia, J., Han, Z., Ouyang, H., Liu, J., & Han, X. (2023). An efficient uncertainty quantification and propagation method through skewness and kurtosis fitting region. *Structural and Multidisciplinary Optimization*, 66(2). <https://doi.org/10.1007/s00158-022-03481-w>

## APPENDIX 1

### Survey instrument

Umur

Jenis kelamin

Pengalaman mengajar

#### A. Performance Expectancy

1. Menggunakan alat AI meningkatkan kinerja pekerjaan saya.  
- Sangat Tidak Setuju [1] - Sangat Setuju [5]
2. Aplikasi AI meningkatkan efisiensi kerja saya.  
- Sangat Tidak Setuju [1] - Sangat Setuju [5]
3. Saya menemukan alat AI membantu mencapai tujuan kerja saya.  
- Sangat Tidak Setuju [1] - Sangat Setuju [5]

#### B. Effort Expectancy (Ekspektasi Usaha)

1. Mempelajari cara menggunakan aplikasi AI mudah bagi saya.  
- Sangat Tidak Setuju [1] - Sangat Setuju [5]
2. Saya merasa alat AI mudah digunakan dalam tugas sehari-hari saya.  
- Sangat Tidak Setuju [1] - Sangat Setuju [5]

3. Berinteraksi dengan aplikasi AI tidak memerlukan banyak usaha.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

#### C. Social Influence (Pengaruh Sosial)

1. Orang-orang yang pendapatnya saya hargai menginginkan saya menggunakan aplikasi AI.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

2. Rekan kerja saya berpikir saya harus menggunakan alat AI di tempat kerja.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

3. Di organisasi saya, menggunakan aplikasi AI dianggap penting.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

#### D. Facilitating Conditions (Kondisi yang Mendukung)

1. Saya memiliki sumber daya yang diperlukan untuk menggunakan aplikasi AI.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

2. Saya memiliki pengetahuan yang dibutuhkan untuk menggunakan alat AI secara efektif.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

3. Alat AI kompatibel dengan sistem lain yang saya gunakan.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

#### E. Niat

1. Saya berniat menggunakan AI dalam kegiatan pembelajaran saya di masa depan.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

2. Saya berencana untuk memanfaatkan teknologi AI untuk meningkatkan kualitas pengajaran saya.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

3. Saya akan secara aktif mencari cara untuk mengintegrasikan AI ke dalam metode pembelajaran saya.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

#### F. Perilaku

1. Saya sering menggunakan aplikasi AI dalam pekerjaan saya.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

2. Saya merekomendasikan alat AI kepada orang lain.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

3. Saya berencana untuk terus menggunakan aplikasi AI di masa depan.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

#### G. Keterlibatan Kerja

1. Alat AI membuat saya lebih terlibat dalam pekerjaan.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

2. Menggunakan aplikasi AI membuat pekerjaan saya lebih menarik.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

3. Alat AI meningkatkan komitmen saya terhadap pekerjaan.

- Sangat Tidak Setuju [1] - Sangat Setuju [5]

#### H. Aplikasi AI di Perguruan Tinggi

1. Aplikasi AI meningkatkan kualitas pengajaran di perguruan tinggi.  
- Sangat Tidak Setuju [1] - Sangat Setuju [5]
2. Alat AI memfasilitasi proses pembelajaran di perguruan tinggi.  
- Sangat Tidak Setuju [1] - Sangat Setuju [5]
3. Aplikasi AI sangat penting untuk inovasi di perguruan tinggi.  
- Sangat Tidak Setuju [1] - Sangat Setuju [5]

# AI Tools in Programming Education: Student Perspectives and Usage Trends

Matija Novak<sup>1\*</sup> and Marija Maček<sup>1</sup>

<sup>1</sup>Faculty of Organization and Informatics, University of Zagreb, Varaždin, Croatia

\*Correspondence: [matija.novak@foi.unizg.hr](mailto:matija.novak@foi.unizg.hr)

## PAPER INFO

### Paper history:

Received 24 July 2025

Accepted 08 February 2026

### Citation:

Novak, M. & Maček, M. (2026). AI Tools in Programming Education: Student Perspectives and Usage Trends. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 83-104

### Copyright:

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

This research investigates the expanding use of generative AI (Mahmoud et al., 2025) tools among computer programming students. Building on a 2023 study, a 2024 survey of 182 undergraduate and graduate students explored usage patterns, perceived effectiveness, and attitudes towards their academic use. Thematic analysis revealed a significant increase in AI adoption, with students frequently using tools for code generation, debugging, and explanation. While students value the speed and efficiency AI offers, they express significant concerns about output accuracy, over-reliance, and the impact on developing independent reasoning skills. Notably, attitudes shifted from favouring full allowance of AI tools to partial allowance for programming assignments, with distinct usage patterns observed across different study levels. These findings underscore the urgent need for educational institutions to establish clear guidelines. A balanced approach is required to integrate AI tools constructively while mitigating risks to academic integrity and deep learning.

**Keywords:** Artificial Intelligence, Student perspectives, Computer Programming

## 1. Introduction

Computer programming often presents a challenge for students (Bain & Barnes, 2014; Guzdial, 2010). Research indicates that many students rely on online resources to solve programming tasks, raising ethical concerns, especially in the context of academic integrity (Cosma & Joy, 2008; Novak, 2020). The ubiquitous availability of artificial intelligence (AI) tools, such as ChatGPT, has further amplified these concerns (Denecke et al., 2023). This trend reflects the expanding use of generative AI (Mahmoud et al., 2025), which has been met with a notably polarized public response characterized by a tension between technological excitement and concern over its impact on human skills. Unlike classic search engines, AI tools can generate complete solutions for programming tasks, without the need for code integration or major modification. This capability of AI tools calls into question the effectiveness of traditional plagiarism detection methods (Novak et al., 2019), as recent studies suggest that distinguishing AI-generated code remains challenging even for specialised tools (Ellis et al., 2024). While plagiarism detection is important, prevention is key. Understanding student perspectives on the use of AI tools is important for the prevention of academic dishonesty.

This study is situated within computing education (CE) research on academic integrity in programming assessment, where generative AI complicates both the detectability of unauthorised assistance and the interpretability of students' code as evidence of individual learning. Prior work has documented challenges with source-code plagiarism in programming coursework and the limits of detection-focused responses, motivating a shift towards more authentic assessment designs that foreground process evidence and

demonstration of understanding rather than relying solely on end-product similarity. (Cosma & Joy, 2008; Novak et al., 2019; Ellis et al., 2024; Prather et al., 2023; Smolansky et al., 2023)

Our contribution is threefold. First, using a parallel mixed-methods survey design, we provide a 2024 primary account of how students report using AI tools in programming courses, distinguishing common functional uses (e.g., debugging, code generation, and explanation) and the perceived trade-offs between efficiency and learning. Second, we report patterns by study level within the 2024 sample, showing that adoption for programming assignments varies across programmes. Third, by repeating the survey across two consecutive years, we provide an exploratory year-on-year comparison that indicates a statistically detectable (though modest) shift from “fully allow” towards “partially allow” positions on AI use in programming assignments. We interpret this shift cautiously given non-equivalent study groups and possible repeated respondents, but we use it to motivate concrete, course-level implications for assessment and AI-use policy.

The main goal of this research is to explore students' perspectives on the use of AI tools in solving programming tasks. The paper examines how and to what extent students use AI tools, and analyses their attitudes towards the permissibility of such use. Following this introduction, the paper proceeds with a review of the relevant literature on the use of AI tools in education. The subsequent sections describe the related work, research methods, present the results, and discuss their implications, including suggestions for future research considering faculty perspectives (Mah & Groß, 2024). The paper concludes with a summary of the key findings and answers to research questions.

## 2. Related Work

Research on AI in programming education encompasses studies on detection, student perspectives, and reintegration strategies (Denecke et al., 2023; Rajendran & Ramasamy, 2024). (Ellis et al., 2024) investigated the detectability of AI-generated code, finding it challenging even with AI detection tools and recommending that instructors focus on student awareness and assignment design. Surveys (Saari et al., 2024; Farhi et al., 2023) have explored student usage of AI tools, revealing substantial use for learning, searching, coding, and debugging, alongside diverse reasons for abstaining. These studies also highlight student perceptions of AI's potential and concerns regarding academic integrity (Farhi et al., 2023). (Smolansky et al., 2023) examined perspectives from both students and teachers, highlighting the need for assessment reform that foregrounds process, higher-order thinking, and authentic applications rather than relying on detectability alone. (Yilmaz & Yilmaz, 2023) explored the use of ChatGPT for programming assignments, noting benefits such as rapid feedback and debugging assistance, as well as drawbacks including potential dependency and inaccuracies. They advocate for AI integration within programming courses, coupled with instruction on verification, critical thinking, and ethical considerations; related work in introductory programming similarly frames ChatGPT as a teaching assistant whose value depends on guided use (Yilmaz & Yilmaz, 2023; Ahmed et al., 2024). Similarly, in (Ali et al., 2024), it is highlighted how AI can contribute to personalized learning by enhancing student engagement and improving educational outcomes.

In computing education (CE), the integrity question is increasingly framed as one of valid evidence of learning rather than detection alone: a submitted programme may be syntactically correct and “original” in similarity terms yet still provide weak evidence of individual reasoning when produced with substantial external assistance. This perspective builds on established work on source-code plagiarism and similarity detection in programming coursework (Cosma & Joy, 2008; Novak, 2020; Novak et al., 2019) and is sharpened by the limited reliability of AI-code detectability as a control strategy (Ellis et al., 2024). Consequently, computing education scholarship and practice increasingly emphasise assessment designs that make students' process and understanding visible, such as traceable development checkpoints via version-controlled repositories and brief oral or walkthrough-style evaluations that require students to explain, test, and adapt their submitted code, aligning closely with students' own concerns about accuracy, over-reliance, and weakened independent reasoning (Prather et al., 2023; Smolansky et al., 2023).

Publications examining students' perspectives on digital tools in programming courses and, more broadly, AI-enabled tools in higher education, report predominantly positive attitudes towards their adoption and perceived usefulness (Asgari et al., 2024; Aydemir & Seferoğlu, 2024; Balogh, 2024). Students recognise the potential of AI to enhance learning experiences, foster increased motivation, and ultimately improve learning outcomes (Aydemir & Seferoğlu, 2024). They express appreciation for AI-driven tools such as conversational agents (chatbots), adaptive learning platforms, and automated assessment methodologies (Aydemir & Seferoğlu, 2024; Balogh, 2024). However, persistent concerns remain regarding ethical implications for learning and assessment, including authorship and academic integrity (Ou et al., 2024), as well as data security and privacy (Balogh, 2024; Idroes et al., 2023). A significant minority of students, representing between 10.17% and 35.42% of those surveyed, express anxieties concerning the potential for

AI to negatively impact the long-term sustainability of higher education institutions (Okulich-Kazarin et al., 2024). Despite these reservations, students generally exhibit confidence in their perceived ability to learn effectively with and utilise AI-based tools (Balogh, 2024).

While this research centres on student viewpoints, a comprehensive understanding requires integrating faculty and institutional perspectives. Recent studies confirm that educators are actively grappling with how to adapt their pedagogy and assessments in response to GenAI. Educators internationally share concerns about academic integrity, the risk of student over-reliance, and the need for clear institutional policies (Khlaif et al., 2024). Furthermore, phenomenological studies reveal the lived experiences of educators integrating these tools, including those in challenging contexts like Palestinian higher education (Hamamra et al., 2025a). These faculty perspectives provide crucial context for the students' shift towards partial allowance observed in our data, suggesting a growing consensus among all stakeholders on the need for a balanced, critically-aware approach to AI integration. In response to these findings, researchers advocate for a nuanced and balanced approach to AI integration within educational contexts, one that prioritises not only technological innovation and advancement but also careful consideration of ethical imperatives and the implementation of human-centred practices in AI development and deployment (Balogh, 2024; Idroes et al., 2023).

Recent educator-focused studies further contextualise students' ambivalence about full versus partial allowance of GenAI. Qualitative accounts describe tensions between perceived efficiency gains and ethical or institutional uncertainty in adoption, including evidence from Palestinian higher education showing how local context shapes both use and restrictions (Hamamra et al., 2025a; Hamamra et al., 2025b), while work on educator technostress underscores the need for clear guidance that supports both learning and wellbeing (Khlaif et al., 2025). Complementary quantitative modelling using the extended Unified Theory of Acceptance and Use of Technology (UTAUT) identifies acceptance drivers such as performance expectancy, effort expectancy, social influence, and facilitating conditions (Ayyoub et al., 2025); these same constructs offer a useful interpretive frame for students' reported motivations and hesitations in our survey, although we did not measure UTAUT variables directly.

Integrating this model would allow for a deeper interpretation of why, for instance, graduate students (GL1) in our study show higher adoption rates for programming tasks (97.62%) compared to first-year undergraduates (76.48%). This difference could be attributed to higher performance expectancy (i.e., belief that AI will improve job performance) or greater exposure to facilitating conditions within their advanced coursework.

Building on prior student surveys in programming contexts (Farhi et al., 2023; Saari et al., 2024; Smolansky et al., 2023), this study contributes a 2024 primary mixed-methods snapshot that distinguishes common usage functions (e.g., generation, debugging, explanation) and examines patterns by study level within one institutional context. An exploratory consecutive-year comparison is then used to contextualise how student preferences may be shifting towards partial allowance and what this implies for programming-course policy and assessment design.

### 3. Methods

In a previous study (Novak & Maček, 2024), a survey was administered to students in Croatia across five courses in 2023. A notable limitation of the 2023 study was the uneven distribution of respondents between graduate students at their first (8 responses) and second year (7 responses) of study (GL2023) and undergraduate students (74 responses) at their third year of study (UL2023). Additionally, the majority of participants were enrolled in a single course, which constrained the ability to draw meaningful conclusions regarding differences between student groups. Despite these limitations, the study yielded relevant and insightful findings.

To build on the previous study, the survey was repeated in 2024 across three courses: one at the first-year undergraduate level (UL1), one at the second-year undergraduate level (UL2), and one at the first-year graduate level (GL1). A total of 187 responses were collected, of which 5 were deemed invalid, resulting in 182 valid responses (42 from GL1, 106 from UL2, and 34 from UL1). Overview of the study workflow is presented in Figure 1.

In 2024, 94 students were enrolled in the first-year undergraduate course (UL1), 195 in the second-year course (UL2), and 77 in the first-year graduate course (GL1), yielding response rates of 36.17% (34/94) for UL1, 56.36% (106/195) for UL2, and 54.55% (42/77) for GL1. For the 2023 survey, 85 students were enrolled in the third-year undergraduate course, yielding a response rate of 87.06% (74/85). For the 2023 graduate level courses this information is not available but the approximate number is around 100, which would make a response rate of ~14.00% (14/100).

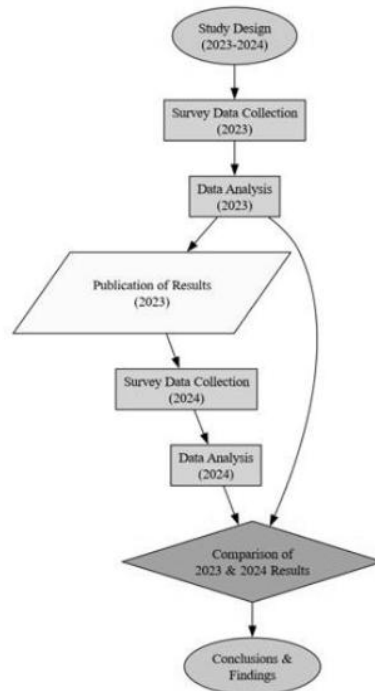


**Figure 1.** Study flowchart

The 2023 and 2024 surveys were administered to partially overlapping study programmes at the same institution. In 2023, the questionnaire was completed by students in third-year undergraduate and graduate level programming courses, whereas in 2024 it was administered in first-year and second-year undergraduate and first-year graduate programming courses. The survey was delivered online in the Croatian language using Google Forms, and links were shared with enrolled students through the institutional learning management system. Participation was voluntary and anonymous, and no identifiers were collected beyond self-reported level and programme of study, year of study, and current grade average. Because the 2023 and 2024 samples were drawn from different groups of students and some individuals may have participated in both surveys, the 2024 data are treated as the primary analytic sample. The 2023 results are used only for cautious, descriptive comparison rather than as a basis for formal longitudinal inference.

The questionnaire comprised background items (faculty, level and programme of study, year of study, and current grade average) and questions about students' use of AI tools for non-programming and programming assignments, the specific ways in which they used these tools, and their views on allowing or restricting AI in programming coursework. A final set of open-ended questions invited students to elaborate on their overall impressions of AI tools (advantages and disadvantages in the study context) and to add any further comments. An English translation of the full questionnaire, including response options, is provided in Appendix A. In both 2023 and 2024 the same questionnaire was used.

As already mentioned we did not collect detailed demographic information such as age or gender, nor did we include separate scales for prior programming or AI experience. This choice was made to keep the questionnaire concise and to preserve perceived anonymity, but it limits our ability to analyse differences by demographic subgroup or by formally measured prior experience.

To minimise the influence on students towards particular views, the questionnaire combined closed-ended items (e.g., yes/no questions, single-choice questions on AI allowance) with open-ended questions asking students to describe how they used AI tools and to explain their reasoning. Two binary items provide key anchors for subsequent analyses. "Used AI for assignments" refers to responses to a yes/no question asking whether students had used AI tools to solve non-programming coursework such as essays, seminar papers and other written tasks. "Used AI for programming assignments" refers to responses to a yes/no question asking whether they had used AI tools to solve programming assignments in the participating courses.

The questionnaire was purpose-designed for this study, drawing on the authors' experience of teaching programming and supervising student coursework. It was not subjected to formal psychometric validation, and we did not calculate reliability coefficients for multi-item scales. Consequently, the results should be interpreted as descriptive and context-specific, rather than as providing evidence for a validated measurement instrument.

The primary objective of the survey was to establish whether and how students used AI tools for programming assignments. A secondary objective was to understand students' perspectives on the advantages and disadvantages of such tools in the wider study context. These aims are reflected in the following research questions:

**RQ1:** What are students' perspectives on the use of AI tools in higher education for programming assignments?

**RQ1.1:** To what extent do students use AI tools for solving programming assignments?

**RQ1.2:** In what ways do students utilise AI tools for programming tasks, and how do they perceive their effectiveness?

**RQ1.3:** What are students' attitudes towards allowing or restricting AI tools in programming education?

**RQ2:** How do students' opinions on AI tools differ across different years of study within the same study population?

**RQ2.1:** Are there significant differences in AI tool adoption and perception between undergraduate and graduate students?

**RQ2.2:** How do students at different study levels (e.g., first-year vs. second-year undergraduate) differ in their use of AI for programming?

**RQ3:** In what ways have students' perspectives on AI-assisted programming changed between two consecutive years?

**RQ3.1:** Have students' perceptions of AI tools evolved over time, and if so, in what ways?

**RQ3.2:** Are there any trends in AI tool adoption between the two consecutive survey years?

## 1.1. Quantitative analysis

Quantitative analysis combined descriptive statistics with inferential tests. For both survey years, responses to closed-ended items were summarised using counts and percentages. The 2024 student sample was treated as the primary analytic sample because it is larger and better aligned with the current structure of the degree programmes. The 2023 responses were used only for exploratory comparison.

To examine associations between categorical variables (for example, use of AI tools versus study programme, and preferences for AI allowance versus study programme), we used Pearson's chi-square test where the expected cell counts satisfied standard assumptions. For  $2 \times 2$  contingency tables with adequate expected counts we report the chi-square test with Yates's continuity correction. Where one or more expected cell counts were below 5, we instead used Fisher's exact test. For key  $2 \times 2$  comparisons we report odds ratios (OR) with 95% confidence intervals (CI). For larger contingency tables we report Cramer's V with 95% confidence intervals as a measure of effect size. Statistical significance was assessed at  $\alpha = 0.05$ . When we conducted families of related tests on the same data (for example, pairwise comparisons between the three 2024 study programmes GL1, UL2 and UL1), we treated them as a single family and applied a Bonferroni correction to control the family-wise error rate, with adjusted threshold  $\alpha_{\text{adj}} = 0.05/3 \approx 0.0167$ . Given the non-equivalent student groups described above, year-on-year comparisons are reported as exploratory and interpreted descriptively rather than as formal longitudinal inference.

For single-response items, percentages use all valid responses to that item as the denominator. For multiple-response items (for example, lists of AI tools used or categories of AI usage), individual student responses could contribute to several categories. These items were coded as a set of binary variables (category present/absent), and percentages in the corresponding tables represent the proportion of students who mentioned each category at least once. As a result, column percentages may sum to more than 100%.

## 1.2. Qualitative analysis

Qualitative analysis focused on the 2024 survey. Open-ended responses were obtained for four core questions, covering (i) how students used AI tools for non-programming assignments (e.g. essays, seminar papers), (ii) how they used AI tools for programming assignments, (iii) their reasons for allowing or restricting AI use in programming coursework, and (iv) their overall impressions of the advantages and disadvantages of AI tools in the study context. A final, "anything else" question invited additional comments. Among 182 valid respondents, 24 (13.19%) wrote something in this field; most entries were brief statements such as "nothing

to add” or greetings. Only 8 students (4.40%) provided additional remarks that directly concerned AI tools or their use in studying, and these mainly reinforced themes already identified in the core open-ended questions.

We applied a descriptive thematic analysis of short open-ended responses at the semantic level. Our procedure drew on general guidance on conducting transparent and trustworthy thematic analysis (Nowell et al., 2017; Lochmiller, 2021), but was adapted to the brevity of survey answers. Two instructors who teach web programming on the participating degree programmes conducted the analysis. First, one researcher read all responses to a given question several times and generated initial inductive codes and categories that captured recurring ideas in the data. Coding was primarily data-driven in the sense that categories were derived from students’ own wording, but it was also guided by the study’s research questions about how AI tools are used, which uses are seen as educationally acceptable, and how students perceive impacts on learning, programming practice and assessment integrity. The same procedure was then applied to the remaining open-ended questions, with each question analysed separately so that categories remained specific to the question. Individual responses could be assigned to more than one category where students mentioned several distinct ideas.

The second researcher then reviewed the full set of coded responses and the proposed category labels for each question. Any disagreements or ambiguities were discussed until consensus was reached. Inter-coder reliability statistics were not calculated, as coding was treated as a collaborative interpretive process rather than a psychometric rating task.

All coding was carried out in a shared spreadsheet, with one row per student response and one column for each category, indicating whether the category was present in that response. This provided a simple audit trail of coding decisions. We then used this spreadsheet to summarise the final categories as descriptive themes and to quantify their frequencies (allowing for multiple codes per response). These summaries are reported in the relevant tables for 2024. This quantification of codes allowed us to relate the qualitative themes to the closed-ended survey items on AI usage, perceived benefits and risks, and preferences for full versus partial allowance of AI tools in programming education. Given the limited length of most responses and the study’s aim of informing course-level AI-use policies, the analysis was deliberately descriptive rather than deeply interpretive. Short translated examples are used to illustrate typical patterns of student reasoning.

Quantitative and qualitative strands were analysed in parallel in a parallel mixed-methods approach. Closed-ended items provided estimates of the prevalence and distribution of AI use and attitudes across study levels, while the open-ended responses were used to explain and contextualise these patterns. Quantitative results are therefore treated as the primary evidence for the strength and direction of associations (e.g., differences between study programmes or years), whereas qualitative themes are used to illuminate underlying reasons and to illustrate how students articulate perceived benefits and risks. Where an interpretation is supported solely by one strand, we indicate this in the text.

### **1.3. Limitations and ethics considerations**

The study protocol, including the anonymous online questionnaire and data-handling procedures, was reviewed and approved by the institutional ethics committee of the authors’ home institution. Participation was voluntary, and students indicated consent by proceeding to complete the survey after reading the information statement on the first page. No identifying information (such as names, e-mail addresses or student IDs) was collected, and results are reported only in aggregated form.

This design has several limitations. The study was conducted in web programming courses at a single Croatian higher education institution, whose AI uptake, institutional policies and digital readiness may differ from other contexts, which limits the generalisability of the findings. The 2023 and 2024 samples are non-equivalent student groups, and anonymity means that some students may have responded in both surveys. Year-on-year contrasts are therefore treated as descriptive comparisons rather than formal longitudinal inferences.

All data are self-reported and may not fully reflect students’ actual behaviours, and the voluntary online format introduces potential self-selection bias, so reported adoption rates may overestimate AI use in the wider student population.

Finally, the questionnaire was developed specifically for this study and was not formally validated, which limits the extent to which our items can be treated as standardised measures or directly compared with instruments from other contexts. These constraints are taken into account in the interpretation of the Results and Discussion.

## 4. Results

This section presents the results of the survey in 2024 and the differences to the previously conducted survey in 2023. As noted in the Methods, the 2024 data are treated as the primary analytic sample, while the 2023 results are used for exploratory, descriptive comparison.

### 4.1. Which AI Tools are Students Using?

In the survey of 2024 a total of 187 responses were received. In response to the initial question about whether students used AI tools, only 5 indicated “no”. Since the primary focus of this research revolves around the usage of AI tools, the responses from these 5 students are excluded from the analysis. This leaves 182 responses which are used for analysis. In the 2023 survey, 84 students reported using at least one AI tool which are included in the comparative analyses.

Concerning the question which AI tools are used by students, among the 182 students in 2024 the majority mentioned using ChatGPT (181/182, 99.5%) as illustrated in Figure 2. Microsoft Copilot was reported by 100 students (54.9%), Bing Chat by 68 students (37.4%), DALL-E by 42 students (23.1%), BlackBox AI by 22 students (12.1%) and Gemini by 40 students (22.0%). Among the 84 students in 2023, ChatGPT was also the most frequently reported tool (62/84, 73.8%), followed by Microsoft Copilot (12 students, 14.3%), Bing Chat (8 students, 9.5%), DALL-E (1 student, 1.2%), BlackBox AI (3 students, 3.6%), Gemini (4 students, 4.8%), and other tools (9 students, 10.7%). Some tools that were mentioned by three or fewer students in 2024 were: Midjourney (2 responses), Ollama (1 response), Claude (1 response), ThorAI (1 response), Perplexity (1 response), Phind (1 response), Julius (1 response). In 2023 some other mentioned tools were: you.com, Gencraft, RunwayML, LLaMA, Forefront.ai, Amazon CodeWhisperer, and Codeium AI.

As this was a multiple-response item, percentages represent the proportion of students who reported using each tool at least once and therefore do not sum to 100%.

From Figure 2. it is evident that there is an increase in the reported use of AI tools among students. To assess changes in AI tool usage between the two years, we conducted separate  $2 \times 2$  comparisons for each individual tool (see Appendix B). These analyses, which treat each tool as a binary independent variable (user vs. non-user), show significant increases in the adoption of ChatGPT, Microsoft Copilot, Bing Chat, DALL-E, and Gemini (all  $p < 0.007$  after Bonferroni correction). This approach provides a statistically valid assessment of adoption trends while maintaining the independence of observations required for inferential testing. Given the differences in the student populations between years and the possibility that some individuals responded in both surveys, this year-on-year comparison is treated as exploratory and does not constitute formal longitudinal evidence of change.

At the level of individual tools, we conducted exploratory  $2 \times 2$  comparisons of use (yes/no) between 2023 and 2024 for each tool (Appendix B). These analyses showed that students in 2024 were substantially more likely to report using ChatGPT than in 2023 (OR = 64.23, 95% CI [8.48, 486.42]), and that reported use of Microsoft Copilot (OR = 7.32, 95% CI [3.72, 14.40]), Bing Chat (OR = 5.67, 95% CI [2.58, 12.46]), DALL-E (OR = 24.90, 95% CI [3.36, 184.29]) and Gemini (OR = 5.63, 95% CI [1.94, 16.32]) also increased. After applying a Bonferroni correction for seven comparisons, these differences remained statistically significant (adjusted  $\alpha \approx 0.007$ ). In contrast, differences for BlackBox AI (OR = 3.71, 95% CI [1.08, 12.77]) and the aggregated “Other” category (OR = 0.38, 95% CI [0.14, 1.03]) were not statistically significant after correction. Again, these per-tool comparisons should be interpreted cautiously as exploratory indicators of change rather than as formal longitudinal effects.

### 4.2. Using AI Tools for Solving Assignments in General

When the students were asked if they used AI tools to solve non-programming assignments in general (for example, essays, seminar papers and other written coursework), the majority answered affirmatively. The data are presented in Table 1.

In 2024, 82.97% indicated that they used AI tools for such assignments, while 17.03% reported not using them. Usage was high across all three course groups (GL1, UL2, UL1), with between roughly four out of five and nine out of ten students reporting use. In the 2023 survey, 90.48% reported using AI tools for assignments and 9.52% did not, again indicating high adoption in both the graduate (GL2023) and undergraduate (UL2023) groups.

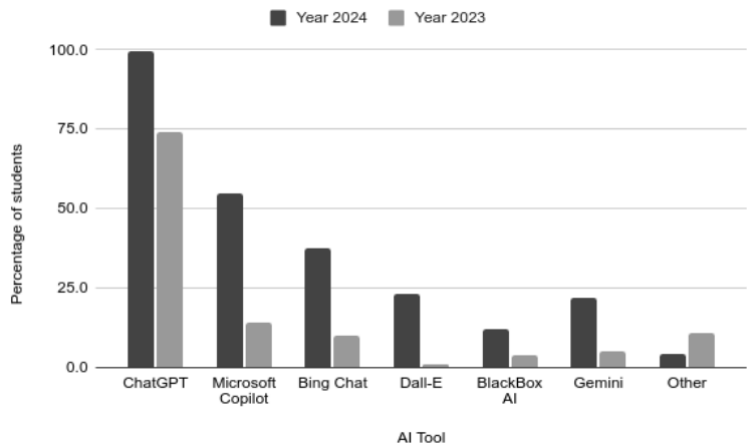


Figure 2. Percentage of students who used an AI tool

As an exploratory comparison, we examined whether overall use of AI tools for non-programming assignments differed between 2023 and 2024. A  $2 \times 2$  chi-square test with Yates’s continuity correction (year  $\times$  use vs. non-use) showed no statistically significant association,  $\chi^2(1) = 2.02$ ,  $p = 0.15$ . The corresponding effect size was small, Cramer’s  $V = 0.09$ , 95% CI [0.00, 0.21]. Expressed as an odds ratio, students in 2024 had somewhat lower odds of reporting AI use than students in 2023 ( $OR_{(2024 \text{ vs } 2023)} = 0.51$ , 95% CI [0.22, 1.17]), but the confidence interval includes 1, indicating that this apparent decrease is statistically uncertain.

To examine whether AI tool usage for non-programming assignments differed between study programmes within the same year, we conducted a chi-square test on the 2024 data comparing GL1, UL2 and UL1. The test indicated no statistically significant difference in usage between these groups,  $\chi^2(2) = 2.23$ ,  $p = 0.33$ . The associated effect size was small, Cramer’s  $V = 0.11$ , 95% CI [0.03, 0.25], consistent with only modest variation in usage rates across the three course groups. Overall, the quantitative data suggest that the use of AI tools for non-programming assignments was widespread and relatively similar across study programmes and between the two survey years.

While the quantitative data indicate that AI tool usage is relatively stable across different student groups and years. Further qualitative aspects were explored of AI adoption to understand factors influencing students’ decisions to use or avoid these tools. For the 2024 survey, open-ended responses about non-programming assignments were coded into five usage categories (Table 2). Because this was a multiple-response item, individual responses could be assigned to more than one category, and percentages represent the proportion of students in each course group whose response was coded with a given category.

|                                                                                                                                                                                                                                                                                                                                                             |                          | 2024 survey     |                  |                 |                    | 2023 survey        |                    |                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|-----------------|------------------|-----------------|--------------------|--------------------|--------------------|-------------------|
| Response                                                                                                                                                                                                                                                                                                                                                    | Course<br>(Participants) | GL1<br>(n = 42) | UL2<br>(n = 106) | UL1<br>(n = 34) | Total<br>(n = 182) | GL2023<br>(n = 14) | UL2023<br>(n = 70) | Total<br>(n = 84) |
| Did not use AI tools                                                                                                                                                                                                                                                                                                                                        | n                        | 4               | 20               | 7               | 31                 | 1                  | 7                  | 8                 |
|                                                                                                                                                                                                                                                                                                                                                             | %                        | 9.52%           | 18.87%           | 20.59%          | 17.03%             | 7.14%              | 10.00%             | 9.52%             |
| Used AI tools                                                                                                                                                                                                                                                                                                                                               | n                        | 38              | 86               | 27              | 151                | 13                 | 63                 | 76                |
|                                                                                                                                                                                                                                                                                                                                                             | %                        | 90.48%          | 81.13%           | 79.41%          | 82.97%             | 92.86%             | 90.00%             | 90.48%            |
| Note: This was a single-response item. Percentages show, for each course group and year, the proportion of students who reported using or not using AI tools.<br>2024: UL1 = first-year undergraduate; UL2 = second-year undergraduate; GL1 = first-year graduate.<br>2023: UL2023 = third-year undergraduate; GL2023 = first-year and second-year graduate |                          |                 |                  |                 |                    |                    |                    |                   |

Table 1. Usage of AI tools for solving non-programming assignments in 2024 and 2023

Across all three course groups in 2024, the most frequently coded use of AI tools was support with text composition for seminar papers, essays, task completion and structure generation, followed by content

correction, restructuring and translation, and information retrieval and verification. Idea generation and support with content analysis and comprehension were also common. Descriptively, graduate students (GL1) more often mentioned content correction and translation, whereas UL2 students more often mentioned text composition and information retrieval, which may point to subtle differences in how students at different stages of study integrate AI tools into their written coursework. The overall pattern of categories is, however, broadly similar across GL1, UL2 and UL1, and these patterns are exploratory and would need to be examined more rigorously in future work. Overall, the coded responses suggest that students at different study levels primarily draw on AI tools to help them draft, refine and make sense of written coursework, rather than to bypass it entirely.

These categories are clearly visible in students’ own descriptions. For idea generation, one student wrote that the tool “helped me to generate text, that is, to generate ideas that I could integrate into my assignments”. Others emphasised information retrieval and verification, using AI “to check whether information is correct”. Several students described AI as a way to polish or restructure their writing, for example by “generating example text or sentence beginnings that then needed to be corrected afterwards”. Students also reported relying on AI for drafting seminar papers and essays “to help me with writing seminar papers and essays, or to give me ideas about what to write”, and for making sense of complex topics, “to better understand the topic, to explain to me what it is about”.

| Categories                                                                                                                                                                                                                                                                                                                                                | Course<br>(Participant) | GL1<br>(n = 42) | UL2<br>(n = 106) | UL1<br>(n = 34) | Total 2024<br>(n = 182) |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|-----------------|------------------|-----------------|-------------------------|
| Idea generation                                                                                                                                                                                                                                                                                                                                           | n                       | 12              | 14               | 8               | 34                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 28.57%          | 13.21%           | 23.53%          | 18.68%                  |
| Information retrieval and<br>verification                                                                                                                                                                                                                                                                                                                 | n                       | 12              | 23               | 6               | 41                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 28.57%          | 21.70%           | 17.65%          | 22.53%                  |
| Content correction and<br>restructuring / text translation                                                                                                                                                                                                                                                                                                | n                       | 16              | 19               | 8               | 43                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 38.10%          | 17.92%           | 23.53%          | 23.63%                  |
| Text composition (seminar<br>papers, essays etc.) / task<br>completion / structure<br>generation                                                                                                                                                                                                                                                          | n                       | 10              | 37               | 9               | 56                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 23.81%          | 34.91%           | 26.47%          | 30.77%                  |
| Content analysis and<br>comprehension                                                                                                                                                                                                                                                                                                                     | n                       | 11              | 21               | 5               | 37                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 26.19%          | 19.81%           | 14.71%          | 20.33%                  |
| Note: This was a multi-response item where each participant’s open-ended response could be coded into more than one category. Percentages are the proportion of students in each course group whose open-ended response was coded with the given category.<br>UL1 = first-year undergraduate; UL2 = second-year undergraduate; GL1 = first-year graduate. |                         |                 |                  |                 |                         |

Table 2. AI tools usage for non-programming assignments in 2024

4.3. Using AI Tools for Solving Programming Assignments

When students were asked whether they used AI tools specifically for solving programming assignments, the majority answered affirmatively, although there was more variation than for non-programming assignments.

From the data (Table 3.) it is visible that there is an increase of using AI tools for programming assignments between 2023 and 2024. Also there seems to be a difference between the different levels of study programmes.

|                                                                                                                                                                                                                                                                                                                                                             |                          | 2024 survey     |                  |                 |                    | 2023 survey        |                    |                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|-----------------|------------------|-----------------|--------------------|--------------------|--------------------|-------------------|
| Response                                                                                                                                                                                                                                                                                                                                                    | Course<br>(Participants) | GL1<br>(n = 42) | UL2<br>(n = 106) | UL1<br>(n = 34) | Total<br>(n = 182) | GL2023<br>(n = 14) | UL2023<br>(n = 70) | Total<br>(n = 84) |
| Did not use AI tools                                                                                                                                                                                                                                                                                                                                        | n                        | 1               | 13               | 8               | 22                 | 6                  | 11                 | 17                |
|                                                                                                                                                                                                                                                                                                                                                             | %                        | 2.38%           | 12.26%           | 23.53%          | 12.09%             | 42.86%             | 15.71%             | 20.24%            |
| Used AI tools                                                                                                                                                                                                                                                                                                                                               | n                        | 41              | 93               | 26              | 160                | 8                  | 59                 | 67                |
|                                                                                                                                                                                                                                                                                                                                                             | %                        | 97.62%          | 87.74%           | 76.47%          | 87.91%             | 57.14%             | 84.29%             | 79.76%            |
| Note: This was a single-response item. Percentages show, for each course group and year, the proportion of students who reported using or not using AI tools.<br>2024: UL1 = first-year undergraduate; UL2 = second-year undergraduate; GL1 = first-year graduate.<br>2023: UL2023 = third-year undergraduate; GL2023 = first-year and second-year graduate |                          |                 |                  |                 |                    |                    |                    |                   |

Table 3. Usage of AI tools for solving programming assignments in 2023 and 2024

As an exploratory comparison between years, we examined whether overall use of AI tools for programming assignments differed between 2023 and 2024. A 2 × 2 chi-square test with Yates’s continuity correction (year × use vs non-use) showed no statistically significant association,  $\chi^2(1) = 2.43$ ,  $p = 0.12$ . The corresponding Cramer’s V was small,  $V = 0.10$ , 95% CI [0.00, 0.22]. Expressed as an odds ratio, students in 2024 had higher odds of reporting AI use than students in 2023 (OR = 1.85, 95% CI [0.92, 3.69]), but the confidence interval includes 1. Combined with the non-equivalent student groups across years and the possibility that some individuals responded in both surveys, these results suggest that any year-on-year difference in reported AI use for programming assignments is uncertain and should be interpreted descriptively.

We then examined whether, within 2024, the use of AI tools for programming assignments differed between study programmes. A chi-square test comparing GL1, UL2 and UL1 indicated a statistically significant association between programme and AI use,  $\chi^2(2) = 7.92$ ,  $p = 0.019$ , with a small-to-moderate effect size, Cramer’s V = 0.21, 95% CI [0.10, 0.36]. To explore which groups differed, we conducted three pairwise 2 × 2 comparisons using Fisher’s exact test (GL1 vs UL2, GL1 vs UL1, UL2 vs UL1), treating them as a single family with a Bonferroni-corrected threshold  $\alpha_{adj} \approx 0.05/3 \approx 0.017$ . The only pair that remained statistically significant after correction was GL1 versus UL1 ( $p = 0.009$ ), where the odds of reporting AI use were much higher for GL1 than for UL1 (OR = 12.62, 95% CI [1.49, 106.81]), consistent with the near-universal usage reported by GL1 students. For GL1 versus UL2, Fisher’s exact test yielded  $p = 0.11$  with an odds ratio of 5.73 (95% CI [0.73, 45.28]), and for UL2 versus UL1,  $p = 0.16$  with an odds ratio of 2.20 (95% CI [0.82, 5.88]). In both of these comparisons the confidence intervals include 1, indicating that any differences in usage between these groups are small or imprecisely estimated and do not meet the Bonferroni-adjusted criterion for statistical significance.

For the 2023 data, GL2023 ( $n = 14$ ) combines students from two graduate programmes and is therefore both small and heterogeneous. Given this and the exploratory role of the 2023 survey, we did not conduct formal within-year tests for 2023. Instead, the GL2023 and UL2023 percentages in Table 3 are reported descriptively and interpreted with caution. Notably, the 2024 UL2 usage rate (87.7%) is similar to that of UL2023 (84.3%), whereas the lower usage observed in GL2023 (57.1%) contrasts with the very high usage in GL1 (97.6%). However, because the graduate groups differ in size and composition across years, and because we lack true longitudinal tracking, these differences cannot be taken as evidence of a real shift in graduate behaviour.

Overall, the results indicate that AI tool use for programming assignments is high across study programmes, with particularly strong uptake among first-year graduate students in 2024. Descriptively, this may reflect greater readiness or opportunity among GL1 students to integrate AI support into programming coursework, for example because of their prior experience or the way AI is discussed in their classes. At the same time, first-year undergraduates (UL1) reported the lowest usage rate in 2024 (76.5%), and similar patterns of more cautious or uneven adoption among early-year students have been noted in previous work on AI in programming education (Yilmaz & Yilmaz, 2023). The absence of statistically robust differences between UL2 and UL1 within 2024, and the uncertainty around year-on-year changes, nonetheless suggests that patterns of AI adoption for programming tasks are broadly similar across neighbouring study levels, with some local variation that warrants further investigation in future work.

Table 4. shows that AI tools are widely used in programming assignments, with code generation, debugging, and code explanation being the most prominent use cases across all three course groups. Documentation writing was mentioned only by a small subset of students, and only in the GL1 group, while

“learning” (for example, using AI to understand concepts or review material) appeared across all levels. As this was a multiple-response item, individual responses could be coded into several categories, and the percentages reflect the proportion of students in each course group whose open-ended description of AI use was coded with each category. Descriptively, GL1 students more often mentioned code generation, debugging and documentation than the undergraduate groups, whereas UL1 students somewhat more frequently emphasised code explanation. We did not conduct formal statistical tests on these categories, so these differences should be interpreted cautiously as indicative patterns rather than definitive group contrasts.

Students’ comments illustrate how these categories played out in practice. For code generation, one student described using AI “to generate very specific pieces of code that require more detailed knowledge of the language syntax”. Debugging support was also prominent: “When I had certain errors at run time or during compilation, I tried to resolve them using AI tools.” Others focused on explanation and learning, for example asking AI “to write out and explain parts of the code that I did not know how to write myself”, or using it “to explain certain programming concepts, and for examples that illustrate those concepts”. Only a small subset of students mentioned documentation explicitly, describing AI as useful “for documenting the programmes I had written and for solving problems in the programme”.

| Categories                                                                                                                                                                                                                                                                                                                                                | Course<br>(Participant) | GL1<br>(n = 42) | UL2<br>(n = 106) | UL1<br>(n = 34) | Total 2024<br>(n = 182) |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|-----------------|------------------|-----------------|-------------------------|
| Code generation                                                                                                                                                                                                                                                                                                                                           | n                       | 22              | 42               | 11              | 76                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 52.38%          | 39.62%           | 32.35%          | 41.76%                  |
| Debugging                                                                                                                                                                                                                                                                                                                                                 | n                       | 19              | 41               | 7               | 68                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 45.24%          | 38.68%           | 20.59%          | 37.36%                  |
| Code explanation                                                                                                                                                                                                                                                                                                                                          | n                       | 10              | 34               | 12              | 58                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 23.81%          | 32.08%           | 35.29%          | 31.87%                  |
| Documentation writing                                                                                                                                                                                                                                                                                                                                     | n                       | 5               | 0                | 0               | 5                       |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 11.90%          | 0%               | 0%              | 2.75%                   |
| Learning                                                                                                                                                                                                                                                                                                                                                  | n                       | 10              | 20               | 7               | 38                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 23.81%          | 18.87%           | 20.59%          | 20.88%                  |
| Note: This was a multi-response item where each participant’s open-ended response could be coded into more than one category. Percentages are the proportion of students in each course group whose open-ended response was coded with the given category.<br>UL1 = first-year undergraduate; UL2 = second-year undergraduate; GL1 = first-year graduate. |                         |                 |                  |                 |                         |

Table 4. AI tools usage for programming assignments in 2024

This categorisation was developed inductively from student responses to reflect their self-described activities. While this student-centred approach differs from some existing taxonomies, our themes align with primary use cases reported in prior research. For example, our “Code generation” and “Debugging” categories correspond to the “Feature implementation” and “Software quality (Fixing)” tasks noted by (Keuning et al., 2024), as well as the common uses of “writing code” and “debugging” identified by (Hou et al., 2025) and (Prather et al., 2023). Our categorisation complements these works by offering a transparent scheme grounded directly in students’ own descriptions of their AI usage in an educational context.

4.4. Students Opinion about AI Allowance

Another interesting question was students’ opinions on whether AI tools should be allowed or forbidden for solving programming assignments. A comparison was made between the data from 2023 and 2024, presented in Table 5. In 2024, most students favoured partially allowing AI tools 57.1%, with a smaller proportion supporting full allowance 40.1% and very few favouring a complete ban 2.7%. In 2023, the pattern was more evenly split between partial allowance 41.7% and full allowance 57.1%, with only one student favouring a complete ban 1.2%.

To explore whether the overall distribution of opinions differed between the two survey years, we conducted a chi-square test on the three response categories (partially allow, allow, forbid) by year. The test indicated a statistically significant association,  $\chi^2(2) = 6.91$ ,  $p = 0.031$ . The corresponding effect size was

small, Cramer’s  $V = 0.16$ , 95% CI [0.06, 0.28], indicating a modest shift in the pattern of responses rather than a large change. Descriptively, this shift reflects an increased preference for partial allowance of AI tools in 2024 and a decrease in the proportion of students favouring full allowance. Given the non-equivalent student groups in 2023 and 2024 and the possibility that some individuals responded in both surveys, these year-on-year differences should be interpreted as exploratory and descriptive.

|                                                                                                                                                                                                                                                                                                                                                                        |                         | 2024 survey     |                 |                  |                    | 2023 survey        |                    |                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|-----------------|-----------------|------------------|--------------------|--------------------|--------------------|-------------------|
| Response                                                                                                                                                                                                                                                                                                                                                               | Course<br>(Participant) | GL1<br>(n = 42) | UL1<br>(n = 34) | UL2<br>(n = 106) | Total<br>(n = 182) | GL2023<br>(n = 14) | UL2023<br>(n = 70) | Total<br>(n = 84) |
| Partially<br>Allow                                                                                                                                                                                                                                                                                                                                                     | n                       | 26              | 21              | 57               | 104                | 7                  | 28                 | 35                |
|                                                                                                                                                                                                                                                                                                                                                                        | %                       | 61.90%          | 61.76%          | 53.77%           | 57.14%             | 50.00%             | 40.00%             | 41.67%            |
| Allow                                                                                                                                                                                                                                                                                                                                                                  | n                       | 16              | 10              | 47               | 73                 | 7                  | 41                 | 48                |
|                                                                                                                                                                                                                                                                                                                                                                        | %                       | 38.10%          | 29.41%          | 44.34%           | 40.11%             | 50.00%             | 58.57%             | 57.14%            |
| Forbid                                                                                                                                                                                                                                                                                                                                                                 | n                       | 0               | 3               | 2                | 5                  | 0                  | 1                  | 1                 |
|                                                                                                                                                                                                                                                                                                                                                                        | %                       | 0.0%            | 8.82%           | 1.89%            | 2.75%              | 0.0%               | 1.43%              | 1.19%             |
| Note: This is a single-response item. Percentages show, for each course group and year, the proportion of students who selected “allow”, “forbid”, or “partially allow”.<br>2024: UL1 = first-year undergraduate; UL2 = second-year undergraduate; GL1 = first-year graduate.<br>2023: UL2023 = third-year undergraduate; GL2023 = first-year and second-year graduate |                         |                 |                 |                  |                    |                    |                    |                   |

**Table 5.** Students responses regarding AI allowance for solving programming assignments in 2023 and 2024

Further, we examined whether attitudes in 2024 differed between the three study programmes. Because the “forbid” option was chosen by very few students, we compared only the “partially allow” and “allow” responses across GL1, UL1 and UL2. A chi-square test showed no statistically significant association between programme and attitude,  $\chi^2(2) = 1.87$ ,  $p = 0.39$ . The corresponding Cramer’s  $V$  was small,  $V = 0.10$ , 95% CI [0.03, 0.25]. Although UL2 students numerically showed the highest support for fully allowing AI tools and UL1 students were somewhat more likely to support a complete ban, these differences are small and statistically uncertain. Overall, the 2024 data suggest that support for partial allowance versus full allowance of AI tools is broadly similar across the three course groups.

Overall, these results indicate that a clear majority of students in both survey years favoured a partial allowance model for AI tools in programming assignments, in which AI can be used but not without constraints. This pattern suggests that many students recognise both the benefits of AI for supporting their work and the need for responsible, controlled use that maintains academic integrity and encourages them to engage meaningfully with programming tasks, rather than relying on AI tools as a complete substitute.

Next, we analysed the qualitative data, the results of which are summarised in Table 6. This was a multiple-response item, so individual open-ended responses could be coded into several categories. Descriptively, there is substantial overlap between the “Partially allow” and “Allow” groups: in both, many students emphasised that AI tools help them to program faster, and relatively few raised ethical concerns. At the same time, the pattern of categories suggests important differences in emphasis between these two positions.

In the “Partially allow” group, students more often framed AI as something that should support learning rather than replace it. A higher proportion of these students referred to AI as “good for learning” and stressed that “code should be understood by students” and that AI tools are acceptable for partial code generation but not for writing entire solutions. They also more frequently mentioned debugging support and the idea that AI should be forbidden for simple problems, where students are expected to develop core skills.

This cautious stance in the “Partially allow” group is well illustrated by one student’s reflection: “AI is a fine option for solving assignments only if it is not used as a complete substitute for learning the material.” Another argued that AI should be permitted but coupled with verification of understanding: “They should be allowed, but then there should be defences of assignments where students must show that they understand the code that has been written.” Students in this group often framed AI as acceptable for partial code generation: “AI tools should be allowed in the sense that they may be used to write only a small part of the assignment, and that this is explicitly stated”, and for debugging support: “AI should not write the entire code; it would also be good if the student can write code without AI for practice and creativity. However, debugging is something for which AI would be excellent.”

| Categories                                                                                                                                                                                                                                                                                                                          | Response<br>(Participant) | Partially allow<br>(n = 104) | Allow<br>(n = 73) | Forbid<br>(n = 5) |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|------------------------------|-------------------|-------------------|
| Good for learning                                                                                                                                                                                                                                                                                                                   | n                         | 29                           | 10                |                   |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 27.88%                       | 13.70%            |                   |
| AI tools are just help to program faster                                                                                                                                                                                                                                                                                            | n                         | 27                           | 20                |                   |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 25.96%                       | 27.40%            |                   |
| Code should be understood by students                                                                                                                                                                                                                                                                                               | n                         | 25                           | 10                |                   |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 24.04%                       | 13.70%            |                   |
| AI tools for partial code generation                                                                                                                                                                                                                                                                                                | n                         | 24                           | 6                 |                   |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 23.08%                       | 8.22%             |                   |
| Debugging                                                                                                                                                                                                                                                                                                                           | n                         | 14                           | 2                 |                   |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 13.46%                       | 2.74%             |                   |
| Forbid in simple problems                                                                                                                                                                                                                                                                                                           | n                         | 13                           | 2                 | 1                 |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 12.50%                       | 2.74%             | 20.00%            |
| Used in companies, this is the future                                                                                                                                                                                                                                                                                               | n                         | 12                           | 24                |                   |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 11.54%                       | 32.88%            |                   |
| Same as google                                                                                                                                                                                                                                                                                                                      | n                         | 5                            | 7                 |                   |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 4.81%                        | 9.59%             |                   |
| Ethics problems                                                                                                                                                                                                                                                                                                                     | n                         | 3                            | 2                 | 1                 |
|                                                                                                                                                                                                                                                                                                                                     | %                         | 2.88%                        | 2.74%             | 20.00%            |
| Forbid if the goal is learning                                                                                                                                                                                                                                                                                                      | n                         |                              |                   | 3                 |
|                                                                                                                                                                                                                                                                                                                                     | %                         |                              |                   | 60.00%            |
| Note: This was a multi-response item where each participant's open-ended response could be coded into more than one category.<br>For each response option ("allow", "forbid", "partially allow"), percentages show the proportion of students who selected that option whose open-ended response was coded with the given category. |                           |                              |                   |                   |

Table 6. Reasons regarding AI allowance for solving programming assignments in 2024

By contrast, students in the “Allow” group more often justified their position by appealing to external practice and future-oriented arguments. One student stated simply that AI tools “should be allowed because this is the future and we need to learn how to use these tools”, while another compared AI directly to existing online resources: “Honestly? I think it should be fully allowed, because what is the point of allowing help from StackOverflow and Google but not from AI?” This ‘same as Google’ framing positions AI as a natural extension of current search and forum-based help.

The “Forbid” group is small, but their responses are more tightly focused on learning and ethics. One student wrote that AI tools “should be banned for any type of assignment whose aim is to demonstrate knowledge”, while another warned that “they can sometimes help, but they can be very harmful for learning and are most often used for cheating”. These comments underline a stricter interpretation of academic integrity, particularly in tasks designed to evidence individual competence.

Overall, these patterns suggest that while both the “Partially allow” and “Allow” groups recognise the practical benefits of AI in programming, the partial-allowance position is more strongly tied to concerns about preserving learning opportunities and code understanding, whereas full allowance is more often justified in terms of real-world practice and the perceived inevitability of AI.

4.5. General Opinion of AI Tools

Regarding the advantages presented in Table 7., the data suggests nuances in how students at different stages perceive the benefits of AI tools. While faster problem-solving and clarification is a key advantage overall (46.70%), it has a slightly higher prevalence among graduate (GL1: 50%) and second-year undergraduate (UL2: 48.11%) students compared to first-years (UL1: 38.24%). This might reflect the increasing complexity

of tasks they face, making speed a more critical factor. Conversely, first-year undergraduates (UL1) stand out for their strong view of AI as more efficient than traditional search methods (52.94%), which is higher compared to UL2 (16.98%) and GL1 (33.33%). This could imply that students newer to higher education find the leap from conventional search tools to AI more impactful, perhaps having less established research habits. The minimal recognition of creativity fostering (1.10% overall) indicates students across levels primarily view these tools through a lens of productivity and problem-solving rather than generative exploration.

| Categories                                                                                                                                                                                                                                                                                                                                                | Course<br>(Participant) | GL1<br>(n = 42) | UL2<br>(n = 106) | UL1<br>(n = 34) | Total 2024<br>(n = 182) |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|-----------------|------------------|-----------------|-------------------------|
| <b>Faster problem-solving and solution clarification</b>                                                                                                                                                                                                                                                                                                  | n                       | 21              | 51               | 13              | 85                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 50.00%          | 48.11%           | 38.24%          | 46.70%                  |
| <b>More efficient than traditional search engines or online forums</b>                                                                                                                                                                                                                                                                                    | n                       | 14              | 18               | 18              | 50                      |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 33.33%          | 16.98%           | 52.94%          | 27.47%                  |
| <b>Creativity fostering</b>                                                                                                                                                                                                                                                                                                                               | n                       | 1               | 1                | 0               | 2                       |
|                                                                                                                                                                                                                                                                                                                                                           | %                       | 2.38%           | 0.94%            | 0%              | 1.10%                   |
| Note: This was a multi-response item where each participant's open-ended response could be coded into more than one category. Percentages are the proportion of students in each course group whose open-ended response was coded with the given category.<br>UL1 = first-year undergraduate; UL2 = second-year undergraduate; GL1 = first-year graduate. |                         |                 |                  |                 |                         |

**Table 7.** Advantages of AI tools in the study context in 2024

Analysis of disadvantages in Table 8. reveal varying levels of concern among the student groups. The risk of incorrect results is a major concern across all study levels (36.26% overall), with graduate students (GL1: 42.86%) expressing this slightly more often, possibly due to the higher stakes associated with accuracy in advanced tasks. More pronounced differences emerge regarding the impact on learning habits. Graduate students (GL1) show higher concern about developing an over-reliance on AI (23.8%) and a potential decline in independent reasoning (19.0%) compared to both first-year (UL1: 14.71% for both concerns) and especially second-year undergraduates (UL2: 12.26% and 11.32% respectively). This heightened concern among GL1 students might stem from a greater awareness of the necessity for deep, independent critical thinking at the graduate level and a recognition of how easily AI could potentially undermine its development. Second-year students (UL2) appear least concerned about these specific learning-related drawbacks. Ethical considerations were a minor theme overall (7.14%), with first-year students (UL1: 8.82%) mentioning them slightly more, perhaps reflecting an initial engagement with the academic integrity implications of these new tools.

The perceived advantages and disadvantages of AI tools also come through clearly in students' own words. For advantages, one student noted that AI is "useful if it is used for the right purposes... the advantages are faster problem-solving and more interesting solutions", while another described it as "a good replacement for Google and StackOverflow" that "makes work easier and faster" and "speeds up searching". A smaller group linked AI to creativity, commenting that it "increases students' creativity and efficiency, but there is also the possibility of plagiarism, which is negative". On the risk side, several students highlighted over-reliance and reduced independent thinking: "The advantage is saving time. The disadvantage is that relying too much on AI can sometimes be bad," and simply, "the disadvantage is a lack of one's own thinking." Concerns about incorrect outputs were equally direct: AI was seen as "a quick and simple way to get an answer to a given question, but sometimes those answers are incorrect." Together, these comments echo the survey categories by portraying AI as simultaneously time-saving and potentially misleading, and by tying ethical concerns directly to plagiarism and misuse.

| Categories                           | Course<br>(Participant) | GL1<br>(n = 42) | UL2<br>(n = 106) | UL1<br>(n = 34) | Total 2024<br>(n = 182) |
|--------------------------------------|-------------------------|-----------------|------------------|-----------------|-------------------------|
| <b>Over-reliance on AI tools</b>     | n                       | 10              | 13               | 5               | 28                      |
|                                      | %                       | 23.81%          | 12.26%           | 14.71%          | 15.38%                  |
| <b>Lack of independent reasoning</b> | n                       | 8               | 12               | 5               | 25                      |
|                                      | %                       | 19.05%          | 11.32%           | 14.71%          | 13.74%                  |

|                                                                                                                                                                                                                                                                                                                                                           |   |        |        |        |        |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|--------|--------|--------|--------|
| Incorrect results                                                                                                                                                                                                                                                                                                                                         | n | 18     | 35     | 13     | 66     |
|                                                                                                                                                                                                                                                                                                                                                           | % | 42.86% | 33.02% | 38.24% | 36.26% |
| Ethical considerations (misuse, plagiarism, human replacement)                                                                                                                                                                                                                                                                                            | n | 3      | 7      | 3      | 13     |
|                                                                                                                                                                                                                                                                                                                                                           | % | 7.14%  | 6.60%  | 8.82%  | 7.14%  |
| Note: This was a multi-response item where each participant's open-ended response could be coded into more than one category. Percentages are the proportion of students in each course group whose open-ended response was coded with the given category.<br>UL1 = first-year undergraduate; UL2 = second-year undergraduate; GL1 = first-year graduate. |   |        |        |        |        |

Table 8. Disadvantages of AI Tools in the study context in 2024

No formal significance tests were conducted for these multi-response categories, so the patterns reported here should be interpreted descriptively. Nonetheless, the combination of perceived advantages (speed, efficiency) and disadvantages (risk of inaccuracy, over-reliance, weaker independent reasoning, and ethical concerns) aligns with the broader shift towards partial allowance reported in Section 4.4. Students appear to value AI tools as supportive resources in programming and study more generally, while simultaneously recognising that unregulated or opaque use could compromise both learning outcomes and academic integrity.

5. Discussion

The use of AI-powered tools among students appears to have increased between 2023 and 2024, with a clear descriptive rise in the proportion of students reporting use of mainstream tools. In particular, reported use of ChatGPT rose from around three-quarters of respondents in 2023 to almost all respondents in 2024, with similar descriptive increases for Microsoft Copilot, Bing Chat and DALL-E. However, because the 2023 and 2024 samples were drawn from non-equivalent student groups and some individuals may have responded in both surveys, these year-on-year patterns should be interpreted as exploratory trends rather than as formal longitudinal evidence of change.

These findings highlight the growing role of AI in student education, with a clear preference for tools that integrate seamlessly into existing workflows and provide a high level of automation. Given the rapid development of AI models and their increasing adaptation to academic and professional needs, this trend is expected to continue. This consolidation around a small set of widely available tools also has practical implications for programming education, because course-level policies and assessment designs can realistically target a finite set of tools rather than an open-ended ecosystem.

Regarding usage of tools for solving assignments, the findings suggest that AI tools remain widely used for academic assignments. Although the proportion of students reporting AI use for non-programming assignments was somewhat lower in 2024 than in 2023, the year-on-year difference was not statistically significant and the associated effect size (Cramer's V) was small, with a confidence interval that includes zero. Similarly, within 2024 we found no statistically significant differences between programmes and only small effect sizes. Overall, these findings suggest that AI tool usage for non-programming assignments is broadly stable across study programmes and years, although the non-equivalent study groups limit the strength of year-on-year inferences.

For programming assignments, the 2024 data highlight meaningful variation in usage between study levels. First-year graduate students (GL1) in 2024 demonstrated the highest adoption rate, while first-year undergraduate students (UL1) had the lowest. These patterns are consistent with the idea that adoption of AI tools for programming varies by study level, with the most pronounced contrast between first-year graduate students (GL1) and first-year undergraduates (UL1).

The higher adoption rate among graduate students may reflect their greater familiarity with advanced tools, a stronger focus on efficiency in completing assignments, or better awareness of benefits of AI tools. On the other hand, the lower adoption rate among undergraduate students could be attributed to a lack of exposure, insufficient training, or a perception that such tools are not necessary for their level of study, similar to (Yilmaz & Yilmaz, 2023). The odds ratio for reporting AI use was much higher for GL1 than for UL1, although the confidence interval was wide due to the very small number of GL1 students who reported non-use, so this difference should be interpreted cautiously.

These results underscore the importance of targeted interventions that support transparent and critical use of AI tools, particularly among undergraduates. For example, workshops, tutorials or explicit integration

of AI into the curriculum can help students understand both the benefits and limitations of these tools. Additionally, further research is needed to explore the specific barriers faced by undergraduate students and to evaluate the effectiveness of potential interventions. Rather than simply encouraging more use among undergraduates, targeted interventions should help students at all levels to use AI tools transparently and critically. In programming education this points to assessment designs that foreground code understanding and provenance, for example, combining supervised labs with version-controlled repositories and brief code walkthroughs that require students to explain, adapt and test AI-assisted solutions.

Similarly, based on the question about allowing AI tools for programming assignments, there has been a shift in students' critical perspectives on AI tools. Consistent with the statistically detectable but modest shift from full towards partial allowance between 2023 and 2024, the qualitative analysis shows that the 'Allow' and 'Partially allow' groups articulate AI's benefits and risks in systematically different ways. Specifically, the percentage of students supporting the complete allowance of AI tools has decreased, while those advocating for partial allowance have increased between 2023 and 2024. While the majority of students across all courses in 2024 preferred a balanced approach, the lack of statistically significant differences between programmes suggests that viewpoints regarding AI allowance are broadly similar across neighbouring academic stages and study programmes.

However, the observed shift toward "partial allowance" should be interpreted with caution. While it may suggest a maturation of student perspectives over time, it could also be influenced by the different composition of the 2024 student population. Specifically, the inclusion of first-year undergraduate students (UL1), who reported more cautious adoption rates (76.5%), may introduce a "freshman mindset" characterized by stricter adherence to perceived academic rules compared to the more experienced third-year students surveyed in 2023. These student concerns regarding efficiency versus over-reliance mirror the broader societal polarization identified by Mahmoud et al. (2025), where public sentiment oscillates between excitement over AI's potential and fear of its long-term impact on human skills.

Nonetheless, from the qualitative data one can see that "Allow" and "Partially Allow" groups reflect distinct attitudes towards the use of AI tools in programming, but also show some overlap in terms of where they find value in these tools. The "Allow" group, with 73 students, represents a more permissive stance, where AI tools are fully embraced, especially for their ability to speed up programming processes (27.39%). This group seems to believe that AI tools can be fully integrated into the programming workflow, particularly in practical, real-world scenarios, such as in companies, where they are seen as a future technology (32.87%). However, the "Partially Allow" group, while similarly open to AI's utility, is more cautious and reserved, with 104 students supporting a limited usage model. This group tends to appreciate AI tools as helpful for certain tasks like partial code generation (23.07%) and debugging (13.46%), but is less inclined to fully trust or rely on them in all contexts, especially when the goal is learning or understanding code.

Where these two groups converge is in their shared recognition of AI tools' potential to improve programming speed and assist in tasks like debugging and partial code generation. Both groups see value in AI as a tool to help students become more efficient, yet they differ significantly in how much they are willing to allow AI to take over the programming process. While the "Allow" group supports AI as a future-forward, all-encompassing tool, the "Partially Allow" group emphasizes the importance of maintaining a balance between using AI as an assistant and ensuring students learn the fundamentals of programming. Our qualitative data clarifies that "partial allowance" is specifically envisioned as using AI for debugging support (13.46%) and partial code generation (23.08%), while strictly forbidding its use for solving entire assignments or simple foundational problems where core skill acquisition is the primary goal. This divergence highlights a central tension for programming educators: how to capitalise on AI's productivity benefits while preserving opportunities for students to practise algorithmic reasoning, debugging and code comprehension in ways that support academic integrity.

For instructional practice, these findings imply a necessary shift from outcome-based to process-based evaluation. Educators should explicitly integrate AI literacy into the curriculum, moving beyond mere permission or prohibition. Specifically, we suggest "collaborative debugging" sessions where instructors and students critique AI-generated code together. This approach not only addresses the accuracy concerns raised by 36.26% of our participants but also reinforces the "student-as-editor" role, which is essential for maintaining independent reasoning while utilising generative tools.

Operationalising a partial-allowance approach in programming assessment. Our results suggest that many students prefer partial allowance because they value GenAI's efficiency while recognising risks related to inaccurate outputs, over-reliance, and weakened independent reasoning. Rather than treating "balance" as a general principle, we translate it into a simple assessment framework that links (i) accuracy risks to explicit verification requirements (e.g., tests and justification of outputs), (ii) over-reliance to traceable process evidence and disclosure (e.g., staged development checkpoints and a brief statement of when AI was used and what was accepted or rejected), and (iii) weakened reasoning to short walkthrough-style checks in which

students must explain, test, and adapt their submitted code. This framing aligns with computing-education calls to refocus assessment on process evidence and demonstrated understanding, given that AI-generated programming work may be difficult to identify reliably and can weaken the evidential value of take-home artefacts (Prather et al., 2023; Smolansky et al., 2023; Ellis et al., 2024).

While this study provides valuable insights from a Croatian context, its focus on web programming students necessarily limits the direct generalisability of the findings. However, the core themes, such as the student-led shift towards partial AI allowance and concerns about over-reliance, resonate with broader trends observed in international studies. Large-scale surveys have documented prevailing attitudes and usage patterns across 20 countries (Prather et al., 2023), while specific studies in North America have tracked a dramatic increase in GenAI adoption, with the percentage of students who reported never using ChatGPT plummeting from 34.1% in 2023 to just 6.3% in 2024. (Hou et al., 2025). Similar trends of evolving usage have also been documented at large European universities (Keuning et al., 2024).

Future research should continue to explore these dynamics in diverse settings, including the Global South, where cultural and institutional factors may uniquely shape AI adoption, as demonstrated in studies from Palestinian universities that explore both student usage patterns and educator experiences (Hamamra et al., 2025a; Hamamra et al., 2025b). In addition, it would be beneficial to add more courses and different study programmes, and investigate things such as the role of course content, prior exposure to AI tools, or cultural factors within the programmes, to better understand how to promote responsible and effective use of AI tools in education. Pursuing these research directions can lead to a more comprehensive understanding of AI tools' evolving role in computer programming education, shaping more effective integration strategies.

### 5.1. Summary of findings

Regarding students' overall perspectives on using AI tools in higher education for programming assignments (RQ1), the findings revealed several key aspects.

*RQ1.1: To what extent do students use AI tools for solving programming assignments?* The study found high usage, with 87.92% of students reporting using AI for programming tasks in 2024, representing a slight (though not statistically significant and the effect size was small) increase from 79.76% in 2023. Almost all students surveyed had experimented with AI tools.

*RQ1.2: In what ways do students utilise AI tools for programming tasks, and how do they perceive their effectiveness?* Students primarily use AI for code generation (41.76%), debugging (37.36%), and code explanation (31.87%). They perceive these tools as effective for faster problem-solving and increasing efficiency compared to traditional methods, although concerns about accuracy are prevalent.

*RQ1.3: What are students' attitudes toward allowing or restricting AI tools in programming education?* There was a statistically detectable but modest shift from 2023 to 2024, with the majority preference moving from fully allowing AI (57.14% in 2023) to partially allowing it (57.14% in 2024). This indicates a developing preference for a balanced approach that leverages AI benefits while managing risks.

Overall, students generally hold a positive yet increasingly cautious perspective towards AI tools for programming, recognising their utility while also advocating for controlled usage and acknowledging potential drawbacks like inaccuracy and over-reliance (RQ1).

The study also examined how students' opinions on AI tools differ across different years of study (RQ2).

*RQ2.1: Are there significant differences in AI tool adoption and perception between undergraduate and graduate students?* Yes, statistically detectable differences were observed in 2024, primarily between first-year graduate and first-year undergraduate students, with a small-to-moderate effect size. First-year graduate students (GL1) showed higher adoption for programming tasks (97.62%) compared to first-year undergraduates (UL1: 76.48%). GL1 students also expressed greater concern regarding over-reliance and the potential negative impact on independent reasoning.

*RQ2.2: How do students at different study levels (e.g., first-year vs. second-year undergraduate) differ in their use of AI for programming?* Second-year undergraduates (UL2) used AI for programming more frequently (87.74%) than first-years (UL1: 76.48%), though this difference was not statistically significant. UL2 students also appeared least concerned about over-reliance and lack of independent reasoning compared to other groups.

Overall, students' opinions and usage patterns regarding AI tools do appear to differ by level of study, with the most pronounced differences between first-year graduate students and first-year undergraduate students in terms of adoption rates and concerns about learning impacts.

The research also investigated how students' perspectives on AI-assisted programming changed between two consecutive years (RQ3).

*RQ3.1: Have students' perceptions of AI tools evolved over time, and if so, in what ways?* Yes, perceptions appear to have evolved. The most notable change was the shift in attitude regarding allowance for programming assignments, with a move towards favouring partial allowance and restriction in 2024 compared to 2023.

*RQ3.2: Are there any trends in AI tool adoption between the two consecutive survey years?* Yes, there was a substantial descriptive increase in overall familiarity with and adoption of various AI tools, with specific tools like ChatGPT seeing major growth. The use specifically for programming assignments also trended upwards.

The descriptive trends suggest an evolution in students' AI usage and attitudes between 2023 and 2024, but the non-equivalent groups mean that these patterns should not be interpreted as definitive longitudinal change. The clearest year-on-year pattern is a statistically detectable, though modest, shift from fully allowing AI tools towards preferring partial allowance for programming assignments; even here, however, non-equivalent samples and possible repeated respondents warrant caution. Nonetheless, this research indicates faster integration of these tools into academic practice alongside a developing critical awareness regarding their appropriate use (RQ3).

## 6. Conclusion

This research explored the use of AI tools by students in computer programming education and their perspectives on this use, building upon previous work (Novak & Maček, 2024).

This study underscores the growing integration of AI tools into programming education. While students recognise the value of AI in enhancing efficiency and supporting learning, their perspectives have evolved towards a balanced, partial-allowance model. This model aligns with a view of AI as a supportive tool whose use must remain transparent and verifiable. As AI tools continue to evolve, it is crucial for educators to establish clear guidelines and interventions that promote responsible use, ensuring that students can use these tools effectively without compromising the development of critical thinking and problem-solving skills. Instead of a single, institution-wide policy, guidelines should be tailored to the course level and learning outcomes. For foundational programming courses, where the goal is skill acquisition, policies might restrict AI-generated solutions and require students to demonstrate independent reasoning through in-class work and brief viva voce code walkthroughs. For more advanced courses, where AI increasingly reflects industry practice, policies can be more permissive but should still require disclosure of AI use and traceable code provenance (e.g. via version-control histories).

To maintain academic integrity, assessments should place greater weight on the learning process and on students' ability to verify, critique and adapt AI-assisted code, for example through assignments that involve debugging or improving AI-generated solutions. Recognising that effective GenAI use is a skill, the curriculum should include instruction on prompt engineering and the critical evaluation of AI outputs. This empowers students to use AI as an effective co-pilot rather than an unreliable oracle, mitigating risks associated with inaccurate 'hallucinated' information.

## Acknowledgements

This work was supported by the Croatian Science Foundation under the project number IP-2025-02-4117.

## References

- Ahmed, Z., Shanto, S. S., & Jony, A. I. (2024). Potentiality of generative AI tools in higher education: Evaluating ChatGPT's viability as a teaching assistant for introductory programming courses. *STEM Education*, 4(3), 165–182. <https://doi.org/10.3934/steme.2024011>
- Ali, W., Alami, R., Alsmairat, M. A. K., & AlMasaeid, T. (2024). Consensus or Controversy: Examining AI's Impact on Academic Integrity, Student Learning, and Inclusivity Within Higher Education Environments. 2024 2nd International Conference on Cyber Resilience (ICCR), 01–05. <https://doi.org/10.1109/ICCR61006.2024.10532968>
- Asgari, M., Tsai, F.-C., Mannila, L., Strömbäck, F., & Sadique, K. M. (2024). Students' perspectives on using digital tools in programming courses. *Discover Education*, 3(1), 57. <https://doi.org/10.1007/s44217-024-00144-4>
- Aydemir, H., & Seferoğlu, S. S. (2024). Investigation of users' opinions, perceptions and attitudes towards the use of artificial intelligence in online learning environments. *European Journal of Open, Distance and E-Learning*, 26(s1), 70–92. <https://doi.org/10.2478/eurodl-2024-0007>

- Ayyoub, A., Khlaif, Z., Hamamra, B., Bensalem, E., Mitwally, M., Sanmugam, M., Fteiha, A., Joma, A., Bsharat, T. R. K., Eidah, B. A., & Khaldi, M. (2025). Drivers of Acceptance of Generative AI Through the Lens of the Extended Unified Theory of Acceptance and Use of Technology. *Human Behavior and Emerging Technologies*, 2025(1). <https://doi.org/10.1155/hbe2/6265087>
- Bain, G., & Barnes, I. (2014). Why is programming so hard to learn? In *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education* (p. 356). ACM. <https://doi.org/10.1145/2591708.2602675>
- Balogh, A. (2024). Educational Innovation of Using Artificial Intelligence in University Education: A Comprehensive Student Survey. *International Journal on Lifelong Education and Leadership*, 10(2), 55–65. <https://doi.org/10.25233/ijlel.1528746>
- Cosma, G., & Joy, M. (2008). Towards a Definition of Source-Code Plagiarism. *IEEE Transactions on Education*, 51(2), 195–200. [10.1109/TE.2007.906776](https://doi.org/10.1109/TE.2007.906776)
- Denecke, K., Glauser, R., & Reichenpfader, D. (2023). Assessing the Potential and Risks of AI-Based Tools in Higher Education: Results from an eSurvey and SWOT Analysis. *Trends in Higher Education*, 2(4), 667–688. <https://doi.org/10.3390/higheredu2040039>
- Ellis, M. E., Casey, K. M., & Hill, G. (2024). ChatGPT and python programming homework. *Decision Sciences Journal of Innovative Education*, 22(1), 4–15. <https://doi.org/10.1111/dsji.12306>
- Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F. F., Al-shami, S. A., & Slamene, R. (2023). Analyzing the students' views, concerns, and perceived ethics about chat GPT usage. *Computers and Education: Artificial Intelligence*, 5, 100180. <https://doi.org/10.1016/j.caeai.2023.100180>
- Guzdial, M. (2010). Why is it so hard to learn to program. In A. Oram & G. Wilson (Eds.), *Making Software: What Really Works, and Why We Believe It* (pp. 111–124). O'Reilly Media.
- Hou, I., Mettelle, S., Li, Z., Man, O., Zastudil, C., & MacNeil, S. (2024). The effects of generative AI on computing students' help-seeking preferences. In *Proceedings of the 26th Australasian Computing Education Conference (ACE '24)* (pp. 39–48). Association for Computing Machinery. <https://doi.org/10.1145/3636243.3636248>
- Hou, I., Nguyen, H. V., Man, O., & MacNeil, S. (2025). The evolving usage of GenAI by computing students. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2* (pp. 1438–1439). ACM. <https://doi.org/10.1145/3641555.3705266>
- Hamamra, B., Khlaif, Z. N., Mayaleh, A., & Baker, A. A. (2025a). A phenomenological examination of educators' experiences with AI integration in Palestinian higher education. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2025.2526435>
- Hamamra, B., Khlaif, Z. N., & Mayaleh, A. (2025b). Between promise and precarity: Palestinian university educators' experiences with AI integration in higher education. *Interactive Learning Environments*, 1–16. <https://doi.org/10.1080/10494820.2025.2547264>
- Idroes, G. M., Noviandy, T. R., Maulana, A., Irvanizam, I., Jalil, Z., Lenisoni, L., Lala, A., Abas, A. H., Tallei, T. E., & Idroes, R. (2023). Student Perspectives on the Role of Artificial Intelligence in Education: A Survey-Based Analysis. *Journal of Educational Management and Learning*, 1(1), 8–15. <https://doi.org/10.60084/jeml.v1i1.58>
- Keuning, H., Alpizar-Chacon, I., Lykourantzou, I., Beehler, L., Köppe, C., de Jong, I., & Sosnovsky, S. (2024). Students' perceptions and use of generative AI tools for programming across different computing courses [Preprint]. arXiv. <https://arxiv.org/abs/2410.06865>
- Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University Teachers' Views on the Adoption and Integration of Generative AI Tools for Student Assessment in Higher Education. *Education Sciences*, 14(10), 1090. <https://doi.org/10.3390/educsci14101090>
- Khlaif, Z. N., Hamamra, B., Bensalem, E., Mitwally, M. A., & Sanmugam, M. (2025). Factors influencing educators' technostress while using generative AI: A qualitative study. *Technology, Knowledge and Learning*, 1–26.
- Lochmiller, C. (2021). Conducting Thematic Analysis with Qualitative Data. *The Qualitative Report*. <https://doi.org/10.46743/2160-3715/2021.5008>

- Mah, D.-K., & Groß, N. (2024). Artificial intelligence in higher education: exploring faculty use, self-efficacy, distinct profiles, and professional development needs. *International Journal of Educational Technology in Higher Education*, 21(1), 58. <https://doi.org/10.1186/s41239-024-00490-1>
- Mahmoud, A. B., Kumar, V., & Spyropoulou, S. (2025). Identifying the Public's Beliefs About Generative Artificial Intelligence: A Big Data Approach. *IEEE Transactions on Engineering Management*, 72, 827–841. <https://doi.org/10.1109/TEM.2025.3534088>
- Novak, M. (2020). Effect of source-code preprocessing techniques on plagiarism detection accuracy in student programming assignments.
- Novak, M., Joy, M., & Kermek, D. (2019). Source-code Similarity Detection and Detection Tools Used in Academia. *ACM Transactions on Computing Education*, 19(3), 1–37. <https://doi.org/10.1145/3313290>
- Novak, M., & Maček, M. (2024). Student perspectives on using Artificial Intelligence tools for solving programming assignments. In N. Vrček & D. Korent (Eds.), *Proceedings of the 35th Central European Conference on Information and Intelligent Systems - CECIIS 2024* (pp. 351–363). University of Zagreb Faculty of Organization and Informatics.
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis. *International Journal of Qualitative Methods*, 16(1). <https://doi.org/10.1177/1609406917733847>
- Okulich-Kazarin, V., Artyukhov, A., Skowron, Ł., Artyukhova, N., & Wołowicz, T. (2024). Will AI Become a Threat to Higher Education Sustainability? A Study of Students' Views. *Sustainability*, 16(11), 4596. <https://doi.org/10.3390/su16114596>
- Ou, A. W., Stöhr, C., & Malmström, H. (2024). Academic communication with AI-powered language tools in higher education: From a post-humanist perspective. *System*, 121, 103225. <https://doi.org/10.1016/j.system.2024.103225>
- Prather, J., Denny, P., Leinonen, J., Becker, B. A., Albluwi, I., Craig, M., Keuning, H., Kiesler, N., Kohn, T., Luxton-Reilly, A., MacNeil, S., Petersen, A., Pettit, R., Reeves, B. N., & Savelka, J. (2023). The robots are here: Navigating the generative AI revolution in computing education. In *Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education (ITiCSE-WGR '23)* (pp. 108–159). Association for Computing Machinery. <https://doi.org/10.1145/3623762.3633499>
- Rajendran, V., & Ramasamy, R. K. (2024). Artificial Intelligence Integration in Programming Education: Implications for Pedagogy and Practice (pp. 197–206). [https://doi.org/10.1007/978-981-97-7794-5\\_16](https://doi.org/10.1007/978-981-97-7794-5_16)
- Saari, M., Rantanen, P., Nurminen, M., Kilamo, T., Systä, K., & Abrahamsson, P. (2024). Survey of AI Tool Usage in Programming Course: Early Observations (pp. 182–191). [https://doi.org/10.1007/978-3-031-48550-3\\_18](https://doi.org/10.1007/978-3-031-48550-3_18)
- Smolansky, A., Cram, A., Radulescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative AI on assessments in higher education. In *Proceedings of the Tenth ACM Conference on Learning @ Scale* (pp. 378–382). ACM. <https://doi.org/10.1145/3573051.3596191>
- Yilmaz, R., & Yilmaz, F. G. K. (2023). Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. *Computers in Human Behavior: Artificial Humans*, 1(2), 100005. <https://doi.org/10.1016/j.chbah.2023.100005>

## Appendix A: Questionnaire (English translation)

**Introductory text:** This survey is part of the study “*Students’ views on AI tools in the context of solving programming assignments*”. The survey is anonymous and the data will be reported in aggregated form only.

### Q1. Faculty *Single-response*.

Select or write the name of the faculty you are attending.

☐ University of ... ☐ Other: \_\_\_\_\_

### Q2. Level of study *Single-response*.

Select your current level of study.

☐ Undergraduate ☐ Graduate (Master’s)

**Q3. Study programme** *Open-ended.*

Please write the study programme (major) you are attending.

**Q4. Year of study** *Single-response.*

Select or state the year of study you are currently in.

- ☐ 1st year – undergraduate ☐ 2nd year – undergraduate ☐ 3rd year – undergraduate  
☐ 1st year – graduate (Master's) ☐ 2nd year – graduate (Master's)  
☐ Other: \_\_\_\_\_

**Q5. Average grade** *Open-ended (numeric).*

What is your current average grade on your study programme?

**Q6. Prior use of AI tools** *Single-response; required.*

Have you tried using AI tools such as ChatGPT, Microsoft Copilot, etc.?

- ☐ Yes ☐ No

**Q7. AI tools used** *Multiple-response.*

Please indicate which AI tools you have used.

- ☐ ChatGPT ☐ Bing Chat ☐ BlackBox AI ☐ Microsoft Copilot ☐ DALL-E ☐ Bard ☐ Other: \_\_\_\_\_

**Q8. Use of AI tools for non-programming assignments** *Single-response.*

Have you used AI tools for assignments that are **not** programming (e.g. seminar papers, essays, etc.)?

- ☐ Yes ☐ No

**Q9. How AI tools were used for non-programming assignments** *Open-ended.*

In what ways have you used AI tools for assignments that are **not** programming (e.g. seminar papers, essays, etc.)? Please describe in as much detail as you can.

**Q10. Use of AI tools for programming assignments** *Single-response.*

Have you used AI tools for solving programming assignments?

- ☐ Yes ☐ No

**Q11. How AI tools were used for programming assignments** *Open-ended.*

In what ways have you used AI tools for solving programming assignments? Please describe in as much detail as you can.

**Q12. General stance on allowing AI tools for programming assignments** *Single-response.*

Regarding programming assignments, do you think that AI tools should:

- ☐ Be fully allowed for solving assignments (please explain in the next question)  
☐ Be completely prohibited for solving assignments (please explain in the next question)  
☐ Be partially allowed and partially prohibited (please explain in the next question)

**Q13. Explanation of your stance on allowing or prohibiting AI tools** *Open-ended*

Do you think that AI tools should be allowed for solving programming assignments, or should they be

prohibited? Please explain your reasoning and, if possible, specify for which types of assignments they should or should not be used.

**Q14. Overall impression of AI tools in the study context** *Open-ended.*

What is your overall impression of these AI tools? What advantages and disadvantages do you see in the context of studying?

**Q15. Additional comments** *Open-ended.*

If there is anything else you would like to add, please write it here.

**Appendix B: Exploratory 2 × 2 comparisons of AI tool use between 2023 and 2024**

| AI tool           | 2023 users<br>n (%) | 2024 users<br>n (%) | OR<br>(2024 vs. 2023) | 95% CI for OR | p-value<br>(two-sided) |
|-------------------|---------------------|---------------------|-----------------------|---------------|------------------------|
| ChatGPT           | 62 (73.8%)          | 181 (99.5%)         | 64.23                 | 8.48 – 486.42 | < 0.001                |
| Microsoft Copilot | 12 (14.3%)          | 100 (54.9%)         | 7.32                  | 3.72 – 14.40  | < 0.001                |
| Bing Chat         | 8 (9.5%)            | 68 (37.4%)          | 5.67                  | 2.58 – 12.46  | < 0.001                |
| DALL-E            | 1 (1.2%)            | 42 (23.1%)          | 24.90                 | 3.36 – 184.29 | < 0.001                |
| BlackBox AI       | 3 (3.6%)            | 22 (12.1%)          | 3.71                  | 1.08 – 12.77  | 0.047                  |
| Gemini            | 4 (4.8%)            | 40 (22.0%)          | 5.63                  | 1.94 – 16.32  | < 0.001                |
| Other             | 9 (10.7%)           | 8 (4.4%)            | 0.38                  | 0.14 – 1.03   | 0.091                  |

$N_{(2023)} = 84$ ;  $N_{(2024)} = 182$ .

Percentages are the proportion of students in each year who reported using the tool at least once (multiple-response item; percentages do not sum to 100%).

Odds ratios (OR) > 1 indicate higher odds of reporting use in 2024 than in 2023; OR < 1 indicates lower odds.

p-values are based on 2 × 2 tests of association between year (2023 vs 2024) and use of each tool (yes/no), using Pearson’s chi-square test with Yates’s continuity correction.

Bonferroni-adjusted p-values are computed for a family of seven per-tool comparisons ( $\alpha_{adj} \approx 0.05/7 \approx 0.007$ ).

# User Experience in Mobile Language Learning Applications (MLLA): Insight from Mixed-Methods Triangulation Study

Yudhy Setyo Purwanto<sup>1,3\*</sup>, Rahmat Gernowo<sup>2</sup> and  
Dinar Mutiara Kusumo Nugraheni<sup>2</sup>

<sup>1</sup>Doctoral Program of Information Systems, School of Postgraduate Studies, Universitas Diponegoro, Semarang, Indonesia

<sup>2</sup>Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia

<sup>3</sup>Faculty of Energy Telematics, Institut Teknologi PLN, Jakarta, Indonesia

\*Correspondence: [yudhysp@students.undip.ac.id](mailto:yudhysp@students.undip.ac.id)

## PAPER INFO

### Paper history:

Received 11 September 2025

Accepted 19 February 2026

### Citation:

Purwanto, Y. S., Gernowo, R. & Nugraheni, D. M. K. (2026). User Experience in Mobile Language Learning Applications (MLLA): Insight from Mixed-Methods Triangulation Study. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 105-127

### Copyright:

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

Mobile Language Learning Applications (MLLAs) are gaining widespread use in higher education because of the flexible nature of practice opportunities with languages. The majority have easy-to-use interfaces, but few enable long-term engagement, retention, and productive learning. There is much empirical work that isolates usability or outcomes singularly without noting how user experience, motivation, and pedagogy intersect in the mobile environment. This study investigates MLLA user experience among university students in Indonesia using a mixed-methods design. Structural Equation Modelling (SEM) was used to investigate the inter-plays of usability, acceptance, engagement, and retention. Qualitative data provided information about user views and context-based limitations. Findings suggest that while overall MLLAs are usable and widely accepted, they tend to lack instructional intensity and intrinsic motivation support. Motivation and perceived usefulness significantly impact engagement, but redundant repetition, superficial individualization, and minimal interaction reduce retention. With reference to TPACK, FRAME, and SDT, the study highlights learner-centered design featuring effective pedagogy, social interaction, and adaptive functionality as being vital. To transcend surface gamification, MLLAs must support deep, long-term language learning. These results provide actionable recommendations for developers and instructors seeking to optimize mobile language learning by synergizing technology, pedagogy, and learner psychology.

**Keywords:** mobile language learning applications, user experience, mixed methods, TPACK- FRAME-SDT integration, SEM

## 1. Introduction

Mobile technology has changed access to and engagement with learning materials. In language acquisition, mobile-assisted language learning (MALL) applications have gained popularity since they are convenient, handy, and increasingly accepted by users. Among these, Mobile Language learning Applications (MLLA) is a subcategory that is enabled by the high smartphone penetration rate and the growing need for anytime-anywhere learning. These mobile applications seek to cater to learners with interactive tasks, multimedia learning, and on-demand language training that suit the habits and expectations of modern learners (Solihin, 2021). However, while such apps are on the rise in terms of availability, empirical studies are still limited

concerning the way users perceive MLLA as it relates to participation, ease of use, retention, and acceptance (Febriyanti et al., 2021).

Previous studies in MLLA have predominantly been focused on the general mobile learning benefits, such as convenience, accessibility, and instant feedback, and most of the existing literature has been focused on functionality, learning outcomes, or app performance, as opposed to how the users interact with the apps over time and what motivates their continued usage (Ekoç, 2021; Madhura Govind Date, 2020; Phuc & Nghi, 2023). This creates a gap in understanding the experiential aspects of mobile learning, especially how factors such as engagement, retention, usability, and acceptance work together to shape learners' long-term relationship with MLLAs. User experience (UX) has become increasingly important in educational settings because long-term usage, persistence, and perceived usefulness influence whether an app becomes part of a meaningful learning routine.

This study positions MLLAs as a set of tools with unique affordances such as personalization, contextual relevance, and continuous access, features that often differentiate them from traditional digital platforms. These affordances may play a significant role in shaping user satisfaction, frequent use, and perceived learning effectiveness. UX here, however, is complex and susceptible to the effects of numerous variables like individual user characteristics, environmental context, and technology design. To address this complexity, this study employs a mixed-method triangulation design that combines quantitative analysis of user survey data with qualitative data from student interviews.

The research is founded on three theory frameworks: Technological Pedagogical Content Knowledge, i.e., TPACK (Mishra & Koehler, 2006), which supports the confluence of pedagogy and technology; the Framework for the Rational Analysis of Mobile Education or FRAME (M. Koole, 2022; M. L. Koole, 2009), which emphasizes device, learner, and context interaction; and Self-Determination Theory or SDT (R. Ryan & Deci, 2017; R. M. Ryan & Deci, 2020), which emphasizes autonomy, competence, and relatedness motivation. These frameworks are used not only as descriptive models but as analytical lenses in interpreting the relationships among UX dimensions in MLLA use.

This study is guided by two refined research questions addressing (1) how users experience engagement, retention, usability, and acceptance when using MLLAs, and (2) how mixed-method triangulation can deepen the understanding of user interaction and satisfaction.

Through these questions, the study aims to expand the broader literature on user-centered design in educational technology, especially for location-aware systems. It is also offering worthwhile feedback for practice, pedagogy, and policy makers interested in enhancing language learning experience through the deployment of mobile technology in a context-aware manner.

## **2. Literature Review**

### **2.1. MALL and MLLA**

Mobile-Assisted Language Learning (MALL) entails the use of mobile technology to support and facilitate language learning. Development has been made possible by expanding smartphone penetration, making learning location-independent, self-paced, and integrated into daily routines. Early research into MALL highlighted accessibility and learner control advantages (Kukulska-Hulme & Shield, 2008; Loewen et al., 2019). Current studies have explored the ways in which mobile applications offer limited learning paths that aid in vocabulary building (Janebi Enayat et al., 2025), teaching grammar (Guerrero et al., 2010; Lin et al., 2020), pronunciation training (Dai & Wu, 2022), and communicative practice exercises (Hwang et al., 2024).

In this study, the term Mobile Language Learning Applications (MLLAs) is used to refer specifically to general-purpose language learning apps (e.g., Duolingo, ELSA Speak, Babbel, HelloTalk). However, critical examination of such platforms tends to identify a gap between the way the app was designed and the way it supports learning. For instance, (Kessler et al., 2023; Loewen et al., 2019) reported that while Duolingo strengthened vocabulary gains, it fell short in fostering communicative competence. Similarly, recent analysis (Al-Jamili et al., 2024; Sakkir & Syamsuddin, 2023; Sevara, 2024) note that many MLLAs rely heavily on repetitive gamification, limited contextual depth, and content that remains at a beginner level regardless of learner progress.

Some studies have looked to outcome-based metrics (like improved test results or task accomplishment rates) to evaluate such apps, but that only tells half the tale. Increasingly, researchers are attributing significance to user experience (UX) as a key factor in determining mobile learning effectiveness (Lubis et al., 2019; Okonkwo, 2024). Poor UX, especially engagement or app usability, has been found to lead to low motivation and high drop-out rates even when the instructional content of the app is of high quality. While

UX is essential, relatively little systematic, integrated research of MLLAs has examined how UX components interact and combine to affect lasting learning.

This research addresses that gap by examining how users interact with MLLA in authentic contexts through four key UX dimensions: engagement, retention, usability, and acceptance. Rather than focusing solely on learning outcomes, the study emphasizes the patterns of use, motivation, and adoption that determine whether MLLAs become sustainable learning tools.

## 2.2. User Experience in Educational Applications

User Experience (UX) in educational technology extends beyond interface design or usability. UX, to (Okonkwo, 2024), encompasses all the user's attitudes and emotions towards digital experience shaped by expectations, motivation, and context. UX in learning environments directly affects the ways learners experience, navigate, and engage with learning tools (Lubis et al., 2019; Royo & Laborda, 2018).

There are a variety of UX models that have been created in the past few years. In one such study by (Alshammare et al., 2025), UX is categorized into three dimensions, i.e., application, user, and context dimensions. Among its conclusions were efficiency, satisfaction, effectiveness, learnability, memorability, attractiveness, errors, and ease of use were the attributes that were most utilized. These are especially relevant to language learning applications, which must be good to look at, easy to use, and meaningful in the context of learners' personal or professional goals.

ISO 9241-210 (ISO 9241-210, 2010) also defines UX as “*a person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service,*” with emphasis on an overall framework that includes emotional, psychological, and situational factors. In the context of learning, UX is not merely a question of how students utilize the app, but how their utilization of the app affects their confidence, motivation, and persistence.

Earlier studies have examined individual determinants of UX in mobile learning. For example, (Kelly et al., 2023) studied usability and acceptance while learning through mobile devices and concluded that intuitive interaction and perfect error management enhanced learners' satisfaction. On the other hand, (M. Koole, 2022) examined engagement and established that adaptive tasks and instant feedback encouraged more cognitive engagement. However, these studies often examine UX dimensions in isolation. There remains limited research integrating these dimensions to explain how they collectively influence adoption and long-term learning behavior in MLLAs.

This work addresses this gap by focusing on four key aspects of user experience. The first is engagement, emotional and cognitive attachment to learning activities. The second is retention, users repeated use of the app and long-term adherence. Third is usability, which focuses on how easy the app is to use, how clear it is, and the quality of the app's interface overall. Lastly, the study considers acceptance, which reflects users' willingness to adopt the app and recommend it to others.

By analyzing these four dimensions together, this research presents UX as both an outcome of design and a predictor of educational persistence. It situates UX not merely as a design issue, but as a structural factor influencing learning success in mobile language environments.

## 2.3. Theoretical Foundations

In order to theorize the intersection of mobile learning, user motivation, and user experience, this study invokes three other frameworks: TPACK, FRAME, and Self-Determination Theory (SDT). However, in order to more fully understand the importance of UX in MLLA, this section also considers the broader theory of MALL and mobile app engagement literature.

This study invokes three theoretical frameworks for investigating user experience in mobile language learning applications. The first is Technological Pedagogical Content Knowledge (TPACK), designed by (Mishra & Koehler, 2006), which emphasizes that successful digital learning requires balanced integration of technology, pedagogy, and content. In MLLAs, misalignment between these components, such as content presented without pedagogical grounding or gamified interfaces that do not connect to learning goals, can undermine UX and retention. The second is the FRAME model of (M. L. Koole, 2009), which theorizes mobile learning as the intersection of device ease of use, social interaction, and learner profile. This is a very practical view since learners employ personal devices to pursue MLLAs in different social settings with different learning goals, factors largely determining user experience. The third is Self-Determination Theory which explains how autonomy, competence, and relatedness drive sustained motivation (Ryan & Deci, 2017). MLLAs that provide meaningful feedback, personalized challenges, or social connection opportunities are more likely to support intrinsic motivation and long-term use. In this study, TPACK informs the alignment between

content quality and usability, FRAME contextualizes engagement and retention through device and social interaction, and Self-Determination Theory explains acceptance and continued use through autonomy, competence, and relatedness.

In addition, the MALL theory places more attention on learner-centered design. Mobile learning is not merely about delivering content while on the move; it involves context-aware, affective, and behaviorally facilitating interaction (Kukulka-Hulme, 2009, 2021; Kukulka-Hulme & Viberg, 2018). UX becomes the channel through which pedagogical intentions are either reinforced or weakened.

Figure 1 summarizes the theoretical framework. The model integrates TPACK (alignment of technology, pedagogy, and content), the FRAME model (device, learner, and social context), and Self-Determination Theory (autonomy, competence, and relatedness) to explain how four user experience dimensions (engagement, retention, usability, and acceptance) influence learners' long-term use of mobile language learning applications.

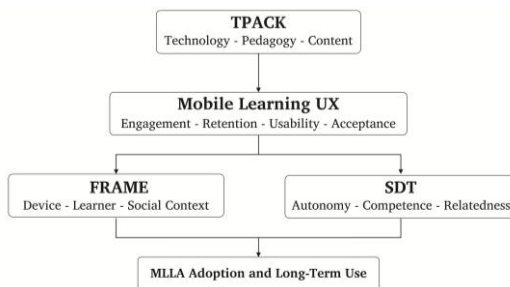


**Figure 1.** Theoretical Framework Overview

By integrating TPACK, FRAME, SDT, and MALL perspectives, this study positions UX as a central mechanism that connects design, motivation, and educational outcomes. This theoretical integration supports the study's focus on examining how engagement, retention, usability, and acceptance interact to shape learners' adoption of MLLAs.

### 3. Methodology

This study adopted a sequential explanatory mixed-methods design, in which quantitative data collection and analysis preceded the qualitative phase. The initial survey phase examined relationships among four user experience (UX) constructs: Engagement, Retention, Usability, and Acceptance, using a structural path model based on composite scores. The subsequent qualitative phase involved purposive interviews with selected participants to further interpret and contextualize the quantitative results. Rather than collecting both strands concurrently, the qualitative data were used to explain patterns observed in the survey findings, particularly where statistical relationships required deeper insight. Integration occurred at the interpretation stage, where qualitative themes were connected to the quantitative associations to provide a more comprehensive understanding of user experience in mobile language learning applications. This study follows Creswell's convergent mixed-methods triangulation design (Creswell & Clark, 2018), in which quantitative and qualitative data are collected in parallel and integrated during interpretation.

Participants were adult learners (18 years and above) who had used one or more MLLAs (such as Duolingo, Cake, or ELSA Speak) within the past six months. Convenience-purposive sampling (Andrade, 2021; Etikan, 2016) was used due to the researcher's limited access to a comprehensive student list. The survey link was distributed through WhatsApp groups (including class and student organizations) and university email networks.

A total of 266 responses were collected: 198 were active MLLA users and 68 were non-users. From the pool of MLLA users, 15 Interview participants were selected purposively based on survey responses to ensure that the qualitative phase captured diverse engagement and usability experiences.

Although the study was conducted within a single private university in Jakarta, the campus hosts students originating from various cities across Indonesia, representing diverse linguistic and educational backgrounds. This composition makes the sample contextually appropriate for the study's focus on Indonesian higher education learners. Conducting the study within one institution also ensures tighter control over demographic consistency and reduces institutional variability, which supports internal validity while remaining representative of typical urban Indonesian university environments.

Participation was voluntary, and informed consent was obtained from all respondents. Anonymity and confidentiality were maintained during data handling and reporting. The research protocol was reviewed and approved by the ethics board at the researcher's affiliated institution.

The survey instrument consisted of 28 questions: 4 demographic questions (age, gender, study program, and MLLA usage status), 4 Application-usage questions, and 20 items measuring UX aspects; five items each for engagement, retention, usability, and acceptance (Appendix A). These items were designed based on literature in user experience and mobile language learning and were informed by TPACK, FRAME, and SDT frameworks.

The questionnaire was pilot tested with 20 participants (18 MLLA users, 2 non-users). Validity and reliability testing (Appendix F2) revealed that two items (Q11-Usability; Q16-Acceptance) scored below the acceptable threshold ( $\alpha < 0.7$ ). These items were removed before full deployment to avoid reducing internal consistency. Removing low-performing items is a standard refinement step in survey development and improves the reliability and precision of subsequent analyses. The qualitative phase involved semi-structured interviews with the 15 selected MLLA users. The interviews were conducted via Microsoft Teams, using built-in recording and automatic transcription features. Transcripts were manually cleaned and checked before coding and analysis.

The quantitative structural analysis in this study was conducted using a single structural path model based on composite user experience (UX) scores. Specifically, Engagement, Retention, Usability, and Acceptance were computed as averaged scale scores and entered the model to examine their associative relationships. No additional structural equation models or extended UTAUT-type models were estimated beyond this four-construct configuration. Accordingly, the model should be understood as a structural path analysis on composite UX dimensions rather than as a multi-model or latent-variable SEM framework. Although the model comprises four composite UX dimensions, a structural modeling approach was selected instead of separate regression analyses to examine the system of relationships simultaneously within a single theoretically specified framework. This approach allows the estimation of direct and indirect associations among engagement, retention, usability, and acceptance in one coherent model, rather than testing isolated pairwise regressions. The structural path model therefore reflects the theoretically driven structure of UX dynamics while maintaining parsimony in model specification. Quantitative data were analyzed using Python in Jupyter Notebook (Fangohr et al., 2019; McKinney, 2017). Cronbach's alpha (Bonett & Wright, 2015; Tavakol & Dennick, 2011) was used to test internal consistency for each UX dimension. Descriptive statistics (mean, standard deviation, frequency) were used to summarize the data (Alabi & Bukola, 2023; Fisher, 2009; Marshall & Jonker, 2010), followed by regression testing to identify relationships among UX factors (Blatna, 2013; Freedman, 2007; Xin & Ren, 2016), correlation analysis using Structural Equation Modelling (SEM) (Blanchin et al., 2019; Elliott & Timulak, 2015; Kunnan, 1998). Because the data are cross-sectional, the SEM results indicate associative rather than causal relationships among engagement, retention, usability, and acceptance. Model fit indices including CFI, TLI, GFI, AGFI, and RMSEA were computed to evaluate overall SEM adequacy. K-Means clustering was used to segment users into groups based on UX patterns (Bock, 2007; Chen & Hoe, 2025; Oti et al., 2021; Priyatma et al., 2024). To ensure conceptual clarity, it should be noted that the structural analysis in this study involved only four empirically measured user experience constructs: Engagement, Retention, Usability, and Acceptance. These constructs were operationalized through validated questionnaire items and aggregated into composite scores prior to model estimation. No additional UTAUT2 constructs (e.g., perceived usefulness, effort expectancy, hedonic motivation, price value) were directly measured or statistically modeled in the structural analysis. References to UTAUT2, TPACK, FRAME, and Self-Determination Theory in later sections serve as interpretive frameworks to contextualize and explain the observed relationships among the four UX constructs, rather than as empirically estimated latent variables within the structural model. The purpose of applying a structural modeling approach was to examine the simultaneous relationships among multiple UX dimensions within a single coherent framework, rather than to compare alternative model specifications.

Qualitative data were analyzed using thematic analysis, following an inductive coding approach (Cooper, 2010; DiCicco-Bloom & Crabtree, 2006; Elhami, 2022; Robinson, 2023). Codes captured recurring patterns: motivation, satisfaction, challenges, and design expectations. The resulting themes were triangulated with quantitative findings to identify convergence, divergence, and complementarity across data sources, thereby fulfilling the mixed-method triangulation requirement.

Integration of quantitative and qualitative findings was conducted at the interpretation stage, consistent with a sequential explanatory mixed-methods design. Rather than merging datasets statistically, integration was achieved through side-by-side comparison and explanatory linking between quantitative patterns and qualitative themes.

Specifically, results from the structural path model and descriptive analyses were first examined to identify key relationships among engagement, retention, usability, and acceptance. The subsequent qualitative interview data were then used to explain, contextualize, and elaborate on these quantitative patterns, particularly where statistical results indicated moderate or ambiguous user experience outcomes.

To make this integration explicit, selected summary tables were structured as joint displays. These tables indicate whether findings derive from survey results, structural modeling, qualitative themes, or their combined interpretation. This approach enables transparent tracing of how qualitative insights inform the interpretation of quantitative relationships without implying that qualitative themes were statistically modeled.

## **4. Result**

### **4.1. Quantitative Results**

The quantitative results of 198 out of 266 Mobile Language Learning Application (MLLA) users are included in the section. Emphasis is on four key variables: engagement, retention, usability, and acceptance. These were measured with Likert-scale ratings and contrasted to discern general user trends, degrees of satisfaction, and points of improvement. The results provide a broad picture of the way in which students interact with MLLAs and serve as the basis for comparison with qualitative remarks below.

#### **4.1.1. General and Demographic Questions**

This section reports the findings from the general and demographic questions completed by all participants ( $n = 266$ ). The information collected involves age, gender, study program, and whether respondents have experience with mobile-based language learning apps (MLLAs). These variables set context for knowing the background of the participants and for comparing opinions between users and non-users, which is necessary for the following analysis of user experience and engagement behaviors.

In terms of experience, most of the participants (74.4%) or 198 respondents reported ever having used mobile-based language learning apps (MLLA). Conversely, 68 respondents (25.6%) reported never having tried using such apps. This difference provides a good comparison basis for evaluating user and non-user attitudes toward MLLAs and vindicates the UX (user experience) comparison analysis used in this study.

The age distribution among the respondents was diverse, with the ages ranging from 18 to over 23 years. The predominant ages were 20 years old (21.8%) and 22 years old (24.4%), followed by 19 years old (19.2%) and 21 years old (15.0%). Among the ages 18 and older than 23, the percentages were lower at 6.8% and 12.8%, respectively. This trend shows that most of the respondents were current university students at the 2<sup>nd</sup> to 6<sup>th</sup> semester, who are likely to start using mobile learning applications. And gender-wise, 165 respondents (62.0%) were male and 101 respondents (38.0%) were female. This proportion is a reasonable representative of the overall population of students in technology and engineering-related programs, which are the subjects of this study.

Respondents were from various academic programs, though the largest proportion was from the bachelor's in informatics engineering (39.5%) and bachelor's in electrical engineering (15.0%). Other academic programs with a significant number of respondents were bachelor's in information systems (11.3%) and bachelor's in mechanical engineering (8.6%). A few respondents were from vocational programs like the Diploma in Electrical Technology and Diploma in Mechanical Engineering, though in smaller numbers. This heterogeneity enriches the analysis by allowing comparisons across majors with different levels of exposure to technology and English-learning needs.

Meanwhile, among the 198 respondents who said they had utilized mobile language learning apps, most employed Duolingo (140 respondents), followed by ELSA Speak (46 respondents), and the remaining few employed other apps such as Babbel, Busuu, or domestic apps (12 respondents). This shows the dominance of two mainstream apps that rely on contrasting instructional designs: gamification vs. AI-driven pronunciation.

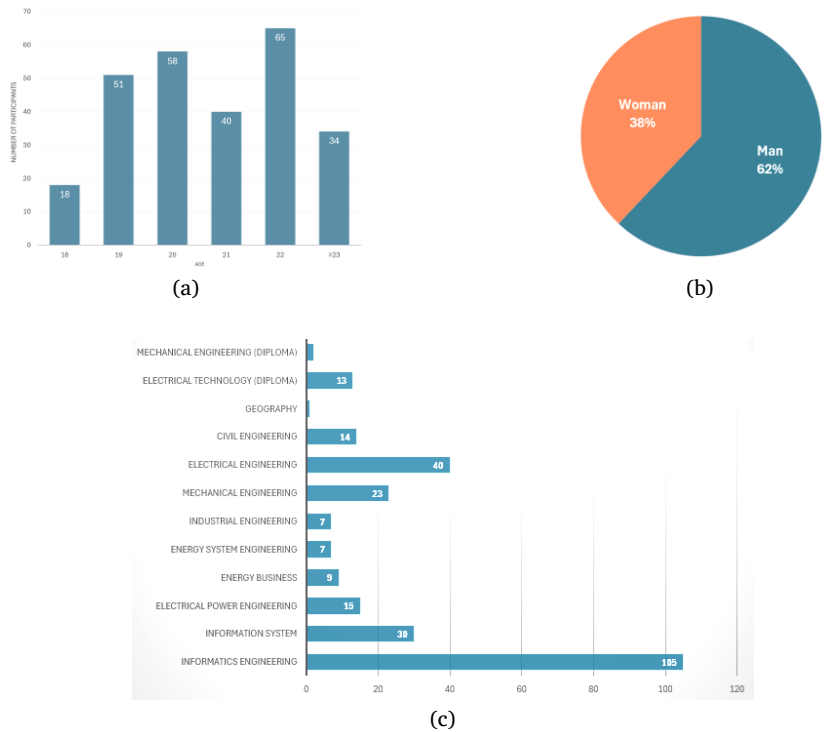


Figure 2. Demographic Questions Results, (a) age; (b) gender; (c) study programs

With respect to usage duration, the majority of the respondents had used the apps for over 6 months (48 respondents), while 45 respondents had used them between 1–2 months; the remaining respondents were evenly distributed across less than 1 month and 5–6 months. This indicates that some of the respondents had had long-term exposure to MLLAs, which can provide richer data on areas of user experience (UX).

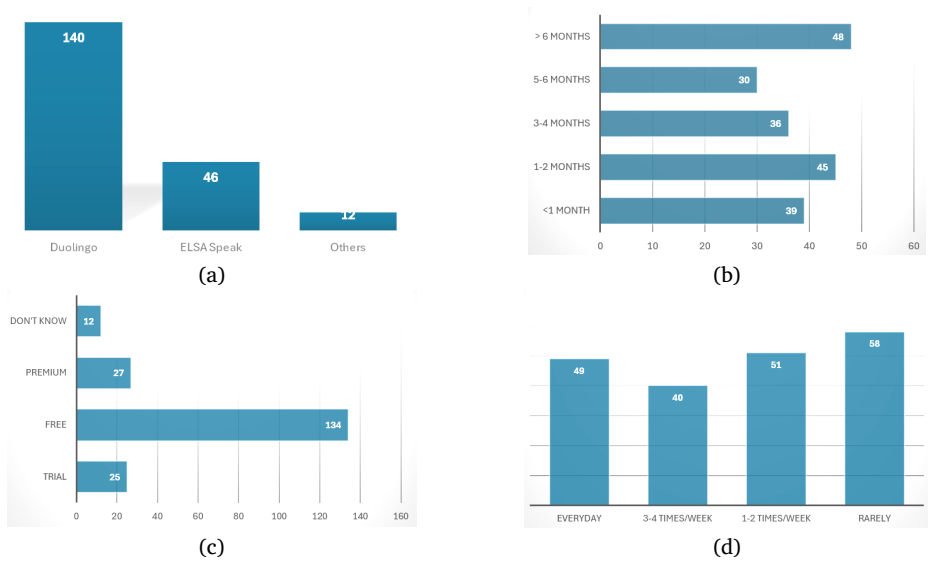


Figure 3. MLLAs General Questions Results: (a) App(s) used; (b) Duration; (c) Plan; (d) Intensity

With this data composition, the subsequent analysis will be able to better determine variations in patterns of use, retention, usability, and MLLA acceptance based on users' demographic and academic backgrounds.

4.1.2. Reliability and Validity

During pilot testing, two items (Q11 and Q16) were identified as underperforming items through overall reliability contribution and item-total correlation. These two items were excluded from the final analysis to improve the scale's validity and to offer more reliable results.

The result of full questionnaire's reliability ratings for all UX dimensions (Engagement, Retention, Usability, and Acceptance) was strong. Cronbach's Alpha varies between 0.858 and 0.868, implying that the measurement tools employed had high internal consistency on the whole.

| Section 1<br>(Engagement) |      |      | Section 2<br>(Retention) |      |      | Section 3<br>(Usability) |      |      | Section 4<br>(Acceptance) |      |      |
|---------------------------|------|------|--------------------------|------|------|--------------------------|------|------|---------------------------|------|------|
| CA: 0.868                 |      |      | CA: 0.859                |      |      | CA: 0.858                |      |      | CA: 0.868                 |      |      |
| Q                         | CI   | CA   | Q                        | CI   | CA   | Q                        | CI   | CA   | Q                         | CI   | CA   |
| Q1                        | 0.55 | 0.87 | Q6                       | 0.68 | 0.82 | Q11                      | 0.69 | 0.82 | Q15                       | 0.69 | 0.84 |
| Q2                        | 0.68 | 0.84 | Q7                       | 0.77 | 0.80 | Q12                      | 0.78 | 0.78 | Q16                       | 0.74 | 0.82 |
| Q3                        | 0.72 | 0.83 | Q8                       | 0.72 | 0.82 | Q13                      | 0.60 | 0.86 | Q17                       | 0.78 | 0.80 |
| Q4                        | 0.73 | 0.82 | Q9                       | 0.56 | 0.86 | Q14                      | 0.72 | 0.80 | Q18                       | 0.66 | 0.85 |
| Q5                        | 0.76 | 0.82 | Q10                      | 0.65 | 0.83 |                          |      |      |                           |      |      |

Table 1. Full Questions Reliability and Validity Tests Results

4.1.3. User Experience Patterns and Statistical Insights

*Descriptive Analysis*

The descriptive statistics among 198 MLLA users show that perceptions of engagement, retention, usability, and acceptance are on the moderate level as shown by the table below.

|       | Engagement | Retention  | Usability  | Acceptance |
|-------|------------|------------|------------|------------|
| count | 198.000000 | 198.000000 | 198.000000 | 198.000000 |
| mean  | 2.782828   | 2.779798   | 2.861111   | 2.925505   |
| std   | 1.000867   | 0.948842   | 0.937598   | 0.934977   |
| min   | 1.000000   | 1.000000   | 1.000000   | 1.000000   |
| 25%   | 2.000000   | 2.200000   | 2.250000   | 2.250000   |
| 50%   | 2.800000   | 2.800000   | 2.750000   | 3.000000   |
| 75%   | 3.400000   | 3.400000   | 3.250000   | 3.500000   |
| max   | 5.000000   | 5.000000   | 5.000000   | 5.000000   |

Table 2. Descriptive Analysis' Results

The mean scores for engagement, retention, usability, and acceptance ranged from 2.78 to 2.93 on a five-point Likert scale (1 = strongly disagree, 3 = neutral, 5 = strongly agree). These values are slightly below the neutral midpoint, indicating neither clear dissatisfaction nor strong endorsement of the applications. Accordingly, the results suggest a generally neutral to mildly positive user experience rather than a distinctly high level of engagement, retention, usability, or acceptance.

For data distribution, the measures of standard deviation range from 0.93 to 1.00, which depicts considerable variability in user ratings. This means that some users were satisfied, but others gave lower ratings. This is also evidenced in the lower quartile (25%) values that describe the 25% of the respondents who gave scores for engagement and retention below 2.0–2.2. On the contrary, the upper quartile (75%) signifies that 25% of the participants rated usability and acceptance over 3.25–3.5, meaning that more than half of the users rated the dimensions positively.

Although usability obtained comparatively higher central tendency values than the other UX dimensions, both the mean and median scores remained close to the neutral midpoint of the scale. Interquartile ranges further indicate moderate dispersion, suggesting variability in user evaluations. These statistics do not reflect strong endorsement of ease of use or readiness to adopt MLLAs, but rather indicate functional adequacy with

considerable room for improvement. Overall, these findings suggest a gap between “functional usability” and “motivational experience,” with users finding the apps easy to operate but less capable of sustaining long-term interest.

**UX Factor Prediction**

To build upon these correlations, a bootstrapped regression was run. The results reiterated the trend discovered in the correlations. Acceptance proved to be a strong predictor of usability ( $\beta = 0.755, p < 0.001$ ), and usability strongly impacted retention ( $\beta = 0.379, p < 0.001$ ).

The predictive models show strong relationships between engagement, retention, usability, and acceptance. All these elements have a strong positive effect on the others, resulting in a reinforcing loop. For example, users who can easily employ an app (usability) and comprehend its value (acceptance) will be in favor of using it for a long period (retention). This is indeed evident through the engagement model, where retention, usability, and acceptance have a positive effect on user engagement. Unexpectedly, users of ELSA were more engaged than for the majority of other apps, although in this instance, it did not necessarily transfer to long-term retention.

When comparing all models separately, retention was the best predictor overall, with particular influence on both engagement and acceptance. However, while ELSA had success in boosting early engagement, its negative correlation with retention suggests potential issues with sustained interest. Usability, though, appears to be less brand-specific and more broadly understood. Overall, the findings suggest that to build a successful MLLA, developers need to be worried not just about first-time appeal or technical usability, but also about sustaining user motivation and delivering perceived value throughout the learning process.

**Cluster Analysis**

The K-Means segmentation validated three UX pattern and engagement-intensive MLLA user groups with varying levels of retention, usability, acceptance, and UX patterns. Cluster 1 ( $n = 89$ ), the low-performing group, are users who are likely to drop off, and they would need ease of design, easier value propositions, and support mechanisms that are improved (all of which had low scores of  $\sim 2.0$ ). The medium cluster (Cluster 0,  $n = 24$ ) has moderate satisfaction and room to grow with focused educational and interactive capabilities (rated all aspects around 3.0). In contrast, the high-performance cluster (Cluster 2,  $n = 85$ ) is comprised of active, devoted users who perceive the app as valuable, intuitive, and suitable for their learning needs (rated all UX dimensions highly in  $\sim 4.5$ ). These results are portrayed in Figure 4 and Table 3 below.

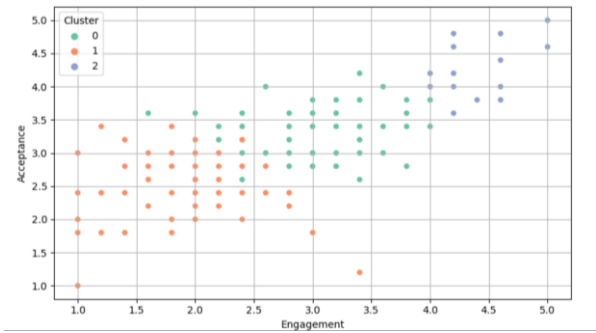


Figure 4. Users’ Cluster based on UX Profile

| Cluster | Engagement | Retention | Usability | Acceptance |
|---------|------------|-----------|-----------|------------|
| 0       | 3.01       | 3.00      | 3.2       | 3.25       |
| 1       | 1.90       | 1.96      | 2.5       | 2.52       |
| 2       | 4.56       | 4.38      | 4.5       | 4.43       |

Table 4. UX mean score per Cluster

These clusters offer actionable guidance for tailoring UX strategies, showing that different groups require different design emphases, from simplifying workflows for low-performing groups to sustaining challenge and value for advanced users.

Structure Equation Modelling (SEM)

The relationships among Engagement, Retention, Usability, and Acceptance were examined using a structural path model based on composite scale scores. Each construct was operationalized as the mean of its validated questionnaire items and treated as an observed variable in the model. Accordingly, the analysis should be interpreted as a path analysis of aggregated UX dimensions rather than as a latent-variable structural equation model. Given the cross-sectional nature of the data and the use of composite scores, the estimated paths indicate statistical associations among UX dimensions rather than causal effects.

|   | lval       | op | rval       | Estimate | Std. Err | z-value   | p-value      |
|---|------------|----|------------|----------|----------|-----------|--------------|
| 0 | Usability  | ~  | Engagement | 0.318846 | 0.055907 | 5.703165  | 1.176029e-08 |
| 1 | Usability  | ~  | Retention  | 0.379415 | 0.058972 | 6.433787  | 1.244636e-10 |
| 2 | Acceptance | ~  | Usability  | 0.754780 | 0.043290 | 17.435509 | 0.000000e+00 |
| 3 | Usability  | ~~ | Usability  | 0.196191 | 0.019718 | 9.949874  | 0.000000e+00 |
| 4 | Acceptance | ~~ | Acceptance | 0.228290 | 0.022944 | 9.949874  | 0.000000e+00 |

Table 4. Causal Relationships Among Key UX Factors in MLLA

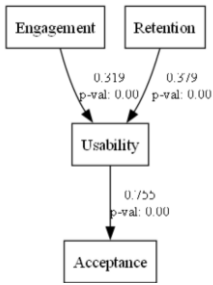


Figure 5. Key UX Factors in MLLA Causal Relationships

The results show that Engagement and Retention play a strong role in influencing perceptions of Usability. More specifically, Engagement showed a significant positive association with usability (estimate = 0.3188,  $p < 0.000001$ ) to show that more motivated and engaged users find the app easier to use. Similarly, Retention was positively associated with usability (estimate = 0.3794,  $p < 0.000001$ ), in that repeated usage of the app makes the user progressively more comfortable and familiar with its structure, further strengthening their sense of ease in using it. The strongest effect in the model was from Usability to Acceptance, highlighting that usability is the most basic element in determining whether users accept the MLLA as a learning tool. This finding supports the hypothesis that an effortless experience with simple navigation, understandable content arrangement, and easily accessible features can have a strong influence on user acceptance and future use.

In summary, the SEM analysis confirms that Usability actually serves as the central mediating variable in the relationship between users' engagement/retention and application acceptance. The feeling of the experience being intuitive and familiar not only provides support towards long-term use but also has a straight effect on increasing users' adoption willingness. Although there is a little bit of unexplained variance in the model, the overall low residual values are indicative of the fact that the model is well-suited to fitting the overall dynamics between these UX factors.

This result points towards designing MLLAs with a strong focus on usability in an attempt to maximize adoption rates. In order to validate the structural relationships, we correlated Spearman's Rho over the four pillars. All the cross-variable correlations were strong (0.67-0.78), validating the SEM-based relationships.

The model demonstrated acceptable fit based on incremental indices ( $CFI \approx 0.93$ ;  $TLI \approx 0.90$ ), while the RMSEA value was relatively high ( $\approx 0.21$ ). Because RMSEA is known to be unstable and prone to inflation in models with very low degrees of freedom, particularly when composite variables are used, overall model evaluation relied on multiple fit indices rather than RMSEA alone. A more detailed discussion of fit considerations is provided in Appendix E. Semopy does not compute SRMR for composite-based structural models, so SRMR is not reported. Overall, the CFI and TLI values support the acceptability of the model.

4.2. Qualitative Findings

Thematic analysis of interview responses yielded nine emergent themes, five of which were strongly dominant. The most frequently mentioned problem was repetitive and monotonous content emerged as a dominant concern among participants. Users described a learning experience that quickly became predictable, reducing cognitive challenge and diminishing motivation to continue. This perceived lack of progression helps explain why engagement and retention scores remained only moderate in the quantitative results, despite comparatively higher usability relative to other dimensions ratings. Participants did not disengage because the apps were difficult to use, but because repeated exposure failed to generate a sense of development or mastery. Despite progressing through various levels, users felt the apps failed to introduce advanced material or content relevant to real-life contexts, leaving their learning plateaued.

Another extremely close issue was the user interface design that was artful but superficial, especially for apps like Duolingo. Though 80% of the users loved the gamified interface and encouraging graphics, they didn't believe the approach dug deep enough and didn't support long-term or substantial learning. In addition, low retention and non-practical application were huge problems. Some 73% of those polled admitted to forgetting most of what they had learned and having trouble applying new grammar or vocabulary to real-life situations. Lastly, paid walls and advertisements were cited by 40% of users as being frustrated with the interruption caused by ads or limitations on free content.

| Theme                          | Description                                                | %   |
|--------------------------------|------------------------------------------------------------|-----|
| Repetitive Content             | Exercises felt predictable and lacked variety              | 87% |
| Stagnation at Basic Levels     | Lack of advanced or real-world content even after progress | 73% |
| Fun UI, Superficial Depth      | Visually engaging but did not support deep learning        | 80% |
| Poor Retention & Practical Use | Knowledge quickly forgotten and hard to apply              | 73% |
| Premium Barriers & Ads         | Ads and paywalls disrupted learning                        | 40% |

Table 6. Qualitative Results

Only 3 of 15 users perceived the apps as supporting long-term learning, indicating that while MLLAs may engage learners initially, they struggle to translate early engagement into sustainable progress. Users recommended improvements such as contextualized tasks, adaptive difficulty, and deeper speaking evaluation, reinforcing the quantitative findings that MLLAs need stronger engagement and retention mechanisms (See Appendix D for the full participant-by-theme matrix).

5. Discussions

This section synthesizes the study's findings by moving beyond statistical reporting to interpret the results in relation to theory and prior research. To avoid redundancy, detailed statistical results are presented in Section 4, while Section 5 focuses on interpretation and theoretical implications. The discussion is grounded in the four UX constructs empirically examined in the structural path model: Engagement, Retention, Usability, and Acceptance. The model reveals patterned associations among these dimensions rather than directional causal mechanisms: higher levels of engagement and retention were associated with more favorable usability perceptions, which in turn were strongly associated with acceptance. These relationships should be understood as co-occurring tendencies within users' reported experiences, not as evidence of causal progression. Concepts from Self-Determination Theory (SDT), FRAME, TPACK, and UTAUT-related literature are used solely as theoretical lenses to contextualize these associations and were not included as measured predictors in the quantitative model.

5.1. Quantitative Results

Quantitative findings showed moderate UX scores on four dimensions assessed (engagement, retention, usability, and acceptance). These values represent a middle ground: users neither strongly accepted nor rejected their experience with MLLAs. Still, the strength of the correlations among these variables, and particularly between engagement and retention ( $p = 0.78$ ) and usability and acceptance ( $p = 0.67$ ), implies a highly integrated UX system. Improving one facet, such as making an app more enjoyable, can indirectly heighten users' overall satisfaction and desire to continue using it.

Structural path modeling was used to examine patterned associations among the four UX dimensions. The results indicate that higher levels of engagement and retention were associated with more favorable

usability perceptions ( $\beta = 0.3794$  and  $\beta = 0.3188$  respectively,  $p < 0.001$ ). Usability, in turn, showed a strong association with acceptance ( $\beta = 0.7548$ ,  $p < 0.001$ ). Rather than implying a directional causal mechanism, these findings suggest that usability functions as a central relational link connecting users' affective experiences (engagement and retention) with their reported willingness to adopt or continue using an application. From this perspective, interface and experience design should be understood as embedded within broader patterns of user interaction, not evaluated in isolation. Notably, although usability scores were relatively higher, engagement and retention remained closer to neutral levels, pointing to a disconnect between ease of use and sustained learning impact. Consistent with prior research on MLLA user experience, this pattern indicates that high usability alone does not necessarily coincide with enduring engagement or deeper learning outcomes.

The qualitative interviews, however, introduced a more critical interpretive layer. Participants would praise the user-friendly interface and gamified design of popular applications like Duolingo but would commonly complain of a lack of meaningful progress in the material. While the SEM results suggest usability as a stimulant to acceptance, the qualitative data show that perceived usability does not always translate to actual learning value. For instance, most users reported content being repetitive or too easy, and that they felt stuck at a beginner level. Even the users who were clustered into the "Satisfied" cluster through K-Means analysis complained about the inability of the app to facilitate long-term retention or real-world application of language skills.

This discrepancy reveals a shortcoming in how users experience "usability" in learning environments. SEM posits that users who have fun using an app and find it easy to return to will find it more usable, but that this perceived usability can be an indicator of interface design instead of instructional effectiveness. That is, users may enjoy the app because it is comfortable and relatively accessible, not because it is necessarily making them more proficient in the language. Here is the key insight: acceptance driven by surface-level usability can lead to high adoption without corresponding learning outcomes.

Briefly, while the quantitative and SEM results affirm a well-functioning UX model wherein engagement, retention, usability, and acceptance predict each other in anticipated ways, the qualitative feedback suggests an underlying issue: a gap between surface-level fun and deep learning. This gap signals that UX satisfaction alone does not guarantee meaningful progress, especially when learning activities lack structure and cognitive challenge.

Furthermore, assessing MLLAs against theoretical frameworks like TPACK, FRAME, and Self-Determination Theory (SDT) helps to identify significant gaps that can account for the mismatch between usability ratings and more profound learning issues. Taking TPACK as a starting point, the apps show good technological knowledge (TK), something reflected in their smooth interfaces, responsive navigation, and gamification. Yet the same apps lack pedagogical knowledge (PK) and content knowledge (CK). The majority fail to scaffold learning or provide organized feedback, especially on speaking and productive language use. This imbalance creates a heavy T layout design, where interface elegance masks superficial instructional depth.

These patterns are reflected in the structural path model results. Usability showed a strong association with acceptance, indicating that users tended to value applications that were easy to navigate and comfortable to use. However, engagement and retention remained closer to neutral, suggesting that ease of use alone does not compensate for limitations in pedagogical depth or learning progression. This gap helps explain why technically polished applications may be readily accepted but still struggle to sustain long-term learning involvement.

The observed relationships among engagement, usability, and acceptance can be interpreted through the perspectives of TPACK and the FRAME framework. From a FRAME standpoint, the alignment between device usability and learner interaction appears central to positive acceptance outcomes. At the same time, FRAME also emphasizes social and contextual dimensions of mobile learning, which were largely absent in participants' accounts. Interview data revealed that learning activities were often decontextualized and repetitive, with limited opportunities for real-world use or peer interaction. This lack of contextual grounding may help explain why engagement and retention scores remained moderate despite relatively positive usability perceptions.

Similarly, TPACK highlights the importance of balancing technological efficiency with pedagogical and content integration. While participants generally reported low technical barriers and intuitive interfaces, qualitative responses suggested that instructional design and content sequencing were insufficient to support sustained learning development. These perspectives help contextualize why usability was strongly associated with acceptance in the present model without implying that pedagogical or content constructs were directly measured.

The association between engagement and retention can also be interpreted through concepts from Self-Determination Theory (SDT). Although psychological needs were not directly measured in the quantitative

model, qualitative findings suggest that when learners experienced a sense of progression or mastery, engagement was more likely to co-occur with continued use. Conversely, reliance on surface-level incentives attracted initial attention but did not sustain motivation in the absence of perceived competence, autonomy, or meaningful social connection. In this sense, SDT provides an explanatory framework for understanding why engagement and retention remained moderate despite favorable usability evaluations.

| Framework | Focus Area                                     | Findings                                            | Implication                                                      |
|-----------|------------------------------------------------|-----------------------------------------------------|------------------------------------------------------------------|
| TPACK     | Pedagogical & Content Fit                      | Weak feedback; little advanced/contextual content.  | Tech-driven, limited pedagogical depth.                          |
| FRAME     | Social & Contextual Learning                   | Few real-world or collaborative tasks.              | Learning is isolated and less authentic, reducing engagement.    |
|           | Device & Personal Usability                    | Good interface design; intuitive navigation.        | A strong interface, but limited instructional depth.             |
| SDT       | Motivation (Autonomy, Competence, Relatedness) | Motivation from streaks/peers; users feel isolated. | Low intrinsic motivation leads to poor retention and engagement. |

**Table 7.** Joint display of framework-based interpretation of quantitative and qualitative findings

These findings can be mapped across the three frameworks to show where current designs fall short (Table 7). MLLAs excel in interface quality and surface-level engagement but fail to offer the pedagogical integration, contextual richness, and psychological scaffolding that is required for deep, lasting learning. Quantitative associations from the structural path model are interpreted alongside qualitative themes to illustrate how theoretical frameworks contextualize user experience patterns. To advance beyond superficial engagement, future apps must integrate instructional design with contextual and motivational affordances from the outset.

In conclusion, SEM data reinforces what these models suggest: usability enhancement alone will not yield long-term learning or engagement results. The next wave of MLLAs must be instructionally smart, socially and contextually appropriate, and psychologically supportive if they are to realize their full educational potential.

**5.2. Qualitative Results Discussion**

Semi-structured interviews with 15 active MLLA users resulted in nine themes, five of which were strongly represented across participants. The themes describe perceptions of learning depth, task variety, feedback quality, and motivational experience with the current mobile language learning applications. On the whole, qualitative results point to the trend of initial enthusiasm followed by declining engagement due to repetitive content and lack of meaningful language improvement. Table 6 summarises the most salient themes.

**Repetitive and Predictable Content.** The issue most reported was related to task repetitiveness. Thirteen of the fifteen participants stated that activities became predictable after several days of use. Users described a feeling that the lessons recycle the same sentence structures or vocabulary; this makes progress feel "stagnant" and "no longer challenging." This was considered a central cause of disengagement and early drop-off.

**Stagnation at Basic Levels.** Eleven participants, or 73%, reported that even after progressing to higher levels within apps, advanced or real-world content was not introduced. Participants reported being "stuck in beginner mode" and having few opportunities to put learned vocabulary or grammar into useful applications. This perceived lack of progression contributed to frustration and ultimately reduced long-term retention.

**Visually Appealing yet Superficial UI.** Twelve of the participants, or 80%, addressed the aesthetically appealing, gamified aspects, and interface design of mainstream apps like Duolingo, but they followed up this positive framing by indicating the design "looked fun but didn't go deep". Gamified elements were motivating at the outset, but participants uniformly reported that interface attractiveness did not translate to deeper learning or improved competence.

**Weak retention and limited real-world application.** Eleven participants (73%) reported difficulty remembering what they had learned, especially when trying to apply it in conversations or academic tasks. Users noted that vocabulary learned through repeated drills was not reinforced within diverse contexts and thus could hardly be transferred outside the app. This problem has been stated to be most serious concerning speaking and productive language skills.

**Premium Paywalls and Advertisements.** Six participants (40%) reported frustration either with ads interrupting or limitations on free content. For example, some noted that such interruptions "break the learning flow" and "force a stop before I'm ready." Several participants said they stopped using the apps because essential features were locked behind paid subscriptions.

Only three of the fifteen interviewees felt that current MLLAs supported long-term learning or would be worth recommending to others. Most participants described a pattern in which initial motivation was high, especially because of visual appeal, streak systems, or ease of navigation, but this motivation declined sharply after one to two weeks of use. Several users mentioned that they continued using the apps only to maintain the streaks rather than actually feeling they were learning something new. Others reported that despite daily use, they were "unsure what progress actually meant" since the apps did not provide clear indicators of improvement beyond points or badges.

Participants also commented that the development of skills felt fragmented. For instance, while learners felt comfortable with recognition tasks, they were still unable to speak or construct sentences on their own. This mismatch between perceived progress in the app and actual language ability led users to question the value of continued use. In fact, several participants stated that they used additional materials such as YouTube lessons, classroom exercises, or conversation practice because the MLLAs alone did not help them make progress beyond beginning levels of proficiency.

Taken together, the interviews paint a picture of users who appreciate the convenience and accessibility of MLLAs yet believe that such tools cannot maintain meaningful development without supplementary support, richer content, or stronger feedback mechanisms. (See Appendix D for the full participant-by-theme matrix).

5.3. User Segmentation and Behavioural Patterns

To better understand the range of user experiences, a cluster analysis of interview and survey data was conducted. This identified patterns within how different users interact with mobile language learning apps (MLLAs), and more importantly, why some continue, and others drop off. The segmentation identifies not just preferences, but deeper behavioral trends that inform how MLLAs can better serve learners of variable motivation, autonomy, and digital readiness.

**Cluster 2: Highly Satisfied Users.** These learners had consistently good experiences. They found the apps to be relatively accessible and helpful, naming ease of use and perception of progress as reasons for continued use. Many were apparently self-regulated learners with high digital fluency and clear goals. Their usage suggests that current features in apps—like progress tracking, instant feedback, and conventional interface design—are effective for learners already proficient at controlling their own learning.

**Cluster 1: Frustrated or Passive Users.** This was the largest and most problematic cluster. While they generally gave neutral-to-positive first impressions, they also tended to lose interest early on. Interviews suggest this was due to shallow content, lack of feedback (especially for speaking exercises), and repetitious activities. These users describe a common experience: the application is usable at first, but the learning process becomes mechanical and unrewarding after a while. They expected more depth and real-world relevance in their practice but did not find it. This makes Cluster 1 a critical target group for redesign, as they mirror the majority of disengaged learners in real usage settings.

**Cluster 0: Neutral Users.** This group of users was in the middle. They did not despise the app, nor were they fully engaged. They used the app occasionally, questioned its long-term usefulness, or indicated that they were waiting for more personalized or meatier content. For this group, subtle design improvements, such as onboarding improvements, clearer learning goals, or smarter feedback, would go a long way in keeping them engaged. Their moderate stance suggests they are the easiest group to convert into active, long-term users if given the right nudges.

Those three clusters subsequently were labeled, key characteristics challenges and areas of improvement opportunities. The summary of those segmentations can be viewed on Table 8 below:

| Cluster   | Label                  | Key Characteristics                                                 | Challenges Identified                                                 | Opportunities for Improvement                                                       |
|-----------|------------------------|---------------------------------------------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Cluster 2 | Highly Satisfied Users | Consistent engagement, confident use, benefit from current features | May represent users with high self-regulation or extrinsic motivation | Maintain strong UX; consider adding advanced modules to retain long-term engagement |

| Cluster   | Label                       | Key Characteristics                                      | Challenges Identified                                          | Opportunities for Improvement                                                                   |
|-----------|-----------------------------|----------------------------------------------------------|----------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Cluster 1 | Frustrated or Passive Users | Initial interest fades; high early use but low retention | Content lacks depth; poor feedback on skills; repetitive tasks | Improve feedback quality, diversify content, connect learning to real-life use                  |
| Cluster 0 | Neutral Users               | Mixed engagement, unsure about app's long-term value     | Unclear purpose, weak onboarding, low personalization          | Add adaptive learning paths, set clearer goals, provide targeted nudges to increase involvement |

**Table 8.** User Segmentation and Behavioral Patterns

To sum up, usability alone is not enough to drive retention or real learning. While Cluster 2 represents a group that performs well within the current app structure, Clusters 1 and 0 represent clear design gaps. Users want more than just navigation, they want content that is meaty, feedback that is constructive, and learning that is compelling. Interview data supports this: most disengaged users drop off within a two-week window, not because of a buggy app, but one that does not develop with learners' demands. This early drop-off window highlights the urgency of improving depth, relevance, and feedback in the initial weeks of use.

**5.4. Towards Better MLLA Design: Analysis-Derived Recommendations**

Results of both statistical modeling and user interviews reveal several main gaps that current mobile language learning apps (MLLAs) should address. Firstly, content depth and thematic organization was a recurring problem. Repetitive drills and insufficient progression annoyed numerous users. Despite showing moderate to high usability scores, the quantitative data also revealed long-term stagnation of engagement. Addressing these issues should help make MLLAs more engaging and effective for users. That would mean that usability would not be synonymous with meaningful learning. More content that is structured and scenario-based, such as modules based on simulating real-life tasks like job interviews, campus life, or traveling, would be required to more actively support real language use and intellectual stimulation. This shift from drill-based content to scenario-driven learning is likely to enhance perceived usefulness and competence simultaneously.

The other central issue is the lack of substantive speaking feedback. While some apps offer pronunciation scoring, users consistently reported that such feedback appeared superficial and often inaccurate. What learners need is feedback that reflects genuine communicative competence, e.g., grammar integration, intonation cues, and markers of fluency. This would align with the SDT principle of competence, in which learners are more motivated as their progress is evaluated in terms that reflect actual real-world accomplishment, rather than merely technical performance.

From a motivational design standpoint, features such as streaks and badges may capture users' initial interest, but they are insufficient for sustaining motivation over time. Both interview themes and UTAUT2-related metrics suggest users need more than just "fun" to stay engaged. Features that allow users to set personal goals, reflect on their progress, and build mastery over time can help promote intrinsic motivation. Providing clear pathways with adaptive difficulty and personalized milestones can help support autonomy and relevance, two vital pillars of Self-Determination Theory. Integrating these features early in the learner journey may significantly reduce the two-week drop-off observed in Clusters 1 and 0.

Lastly, two practical but overlooked issues emerged: premium paywalls and the absence of social features. Many learners reported that their learning momentum was abruptly halted by ads or paywalls, leading to disengagement. While monetization is understandable, pushing premium products too aggressively undermines retention and user trust. On the other hand, none of the apps reviewed provided intrinsic possibilities for peer interaction or collaborative learning. Not only is this at odds with FRAME's emphasis on mobile learning's social-contextual dimension, but it also denies users the chance to use language in real communicative situations. This omission could be remedied while promoting learner relatedness by including peer review, discussion forums, or collaborative challenges. Adding meaningful social layers would align mobile learning with real communicative practice, a core part of language acquisition missing in most existing apps.

| Design Concern                      | Qualitative Evidence                             | Supported by SEM Results                                                                                    |
|-------------------------------------|--------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| Content Depth & Real-Life Scenarios | Repetitive drills, weak progression.             | <i>Perceived Usefulness</i> and <i>Content Quality</i> had significant positive effects on Engagement.      |
| Speaking Feedback Quality           | Inaccuracy, lack of real communicative value.    | <i>Perceived Competence</i> influences <i>Continuance Intention</i> (SDT path).                             |
| Shallow Gamification                | Initial attraction but short-term impact.        | <i>Hedonic Motivation</i> had weaker influence; <i>Intrinsic Motivation</i> was stronger in sustaining use. |
| Autonomy & Goal Setting             | Need for progress tracking and goal orientation. | <i>Autonomy</i> positively affected Engagement and Continuance Intention.                                   |
| Paywalls & Ads                      | Disruptive to learning flow.                     | <i>Effort Expectancy</i> and <i>Price Value</i> paths suggest negative impacts on intention.                |
| Lack of Social Interaction          | Absence of peer feedback or collaboration.       | FRAME's <i>Social</i> component missing; <i>Relatedness</i> in SDT unfulfilled.                             |

**Note:** Constructs listed under UTAUT2 and related frameworks are presented for theoretical interpretation and were not directly measured or included in the structural path model.

**Table 9.** Joint display aligning structural path model results with qualitative interpretations and theoretical lenses

Following user segmentation and statistical modeling (SEM) results, this section offers design guideline recommendations based on both quantitative results and qualitative observation. Usability and ease of access are crucial but are not sufficient to encourage ongoing use or compelling learning outcomes. SEM results indicated the central role of perceived usefulness, content quality, and motivational factors, all of which were confirmed through user interviews. The following recommendations address key areas of tension observed across user groups and modeled paths, with the goal to inform MLLA design with user-centered principles. Any references to additional theoretical constructs in this section are intended to support conceptual interpretation and do not reflect separate empirical models or estimated structural paths.

5.5. Final Reflection

This study confirms that most MLLAs are well-designed in the usability dimension, which aligns with both the technological aspect of TPACK and the device-focused aspects of the FRAME model. Users find such apps intuitive and visually appealing. However, both user interviews and SEM results reveal deeper concerns with pedagogy, context, and motivation, areas where many MLLAs are deficient.

These concerns are confirmed by the structural equation modeling results. While usability positively affects initial adoption, it has little effect on continued use. The observed relationships among engagement, retention, usability, and acceptance can be interpreted through constructs commonly discussed in UTAUT2, such as perceived usefulness and effort expectancy. However, these concepts are invoked here as theoretical lenses to contextualize the findings, rather than as empirically estimated predictors within the model.

Collectively, these findings point to the need for MLLAs to go beyond neat interfaces or extrinsic rewards. Effective mobile-supported language learning apps must include sound instructional design, facilitate authentic language use in daily life, and foster internal motivation. Adding features that allow learners to feel competent, make meaningful choices, and engage with others has the potential to transform MLLAs from tools of convenience to spaces of sustained learning and personal growth. Such integration would create a more durable motivational ecosystem, reducing dropout and enhancing long-term language development. Future research employing longitudinal or experimental designs would be required to examine causal sequencing among engagement, retention, usability, and acceptance.

Table 10 presents the mapping of SEM findings onto the theoretical frameworks, demonstrating how empirical evidence supports the conceptual underpinnings of good MLLA design.

| SEM Construct        | Key Finding                                                        | Related Framework | Theoretical Link                                                                                                                                      |
|----------------------|--------------------------------------------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Usability            | Positively affects initial acceptance but not long-term engagement | TPACK, FRAME      | Matches <i>Technological Knowledge</i> in TPACK and <i>Device Usability</i> in FRAME, showing ease of use alone is not enough for sustained learning. |
| Perceived Usefulness | Strong predictor of continuance intention                          | TPACK             | Reflects <i>Pedagogical-Technological</i> integration where users stay engaged when apps contribute meaningfully to their language development.       |

| SEM Construct                 | Key Finding                                                           | Related Framework | Theoretical Link                                                                                                                               |
|-------------------------------|-----------------------------------------------------------------------|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| Content Quality               | High-quality, structured content improves satisfaction and motivation | TPACK, SDT        | Tied to <i>Content Knowledge</i> in TPACK and <i>Competence</i> in SDT—users want rich, contextualized materials that mirror real-world usage. |
| Intrinsic Motivation          | Strongest predictor of sustained use and satisfaction                 | SDT               | Confirms the need for autonomy, competence, and relatedness to keep learners engaged beyond extrinsic incentives.                              |
| Social Interaction Absence    | Weak or missing in most apps; limits user engagement                  | FRAME, SDT        | Aligns with <i>Social Context</i> in FRAME and <i>Relatedness</i> in SDT—users benefit from peer interaction, collaboration, and feedback.     |
| External Barriers (e.g., Ads) | Negative influence on user retention and trust                        | FRAME             | Highlights how <i>Device</i> + <i>Context</i> interaction can disrupt flow—ads and paywalls reduce usability and perceived learning value.     |

Table 10. Linking SEM Results with Theoretical Frameworks

6. Conclusion and Recommendations

The study explored the user experience of Mobile Language Learning Applications (MLLAs) for engagement, retention, usability, and acceptance. While people accept MLLAs in general and find them relatively accessible, the majority of applications still fall short in maintaining long-term engagement and learning efficiently. Analysis via user feedback and statistics shows that the challenge lies mainly due to ineffective pedagogical approaches, too little personalization, and a lack of interactive or real-world aspects. These shortcomings suggest that current MLLAs prioritize interface design over deeper learning mechanisms needed for sustained progress.

The theoretical analysis using TPACK, FRAME, and Self-Determination Theory (SDT) highlights several key gaps. The majority of apps reflect technological design capability but are poor in pedagogical depth and learner-centered functionality. SDT, for instance, reports that the apps tend to neglect basic motivational drivers—autonomy, competence, and relatedness—necessary for long-term learning. When these needs are unmet, user engagement drops quickly despite initially positive impressions.

These theoretical results were confirmed by Structural Equation Modeling (SEM), which determined that usability and perceived ease of use are not enough to guarantee engagement and retention. SEM segregated the interaction among the four dimensions assessed, signaling that motivation and pedagogical alignment have more influential effects on retention than interface quality per se. This reinforces the idea that well-designed interfaces must be paired with meaningful learning pathways to achieve lasting impact.

Better, though, app developers require starting with better instructional design. In addition to the endless repetition of drills, features need to support scenario-based activities and simulations that reflect real communication. And students need more control—adaptive content, goal-setting tools, and interest-based modules can personalize the experience and create motivation. Embedding these elements early in the learner journey can help reduce the common two-week disengagement window identified in this study.

A good interface matters too. Clean navigation, timely feedback, and localization improve usability and user trust. For social interaction, functionality like discussion forums, peer activities, or live chat can build community. To address retention, developers need to look past badges and leaderboards, offering varied content progression and intrinsic rewards. Such enhancements would align MLLAs more closely with the social and contextual dimensions emphasized in FRAME and SDT.

Later MLLAs can also derive further benefits from context-aware or location-based features. Taking advantage of the use of GPS or geotagging, applications can link activities to geographical locations, compelling learners to exercise the language in natural settings. Such experiential learning mediates between computer-based learning and everyday communication. These location-aware interactions can transform passive practice into practical, meaningful engagement.

In brief, effective MLLAs must evolve beyond mere useful apps to seamless learning platforms. Supported by good pedagogy, for real-world application, and backed by motivational psychology, they have the capacity to foster deeper engagement and enhanced learning yield. By integrating insights from TPACK, FRAME, SDT, and SEM, future MLLAs can be designed in ways that meaningfully strengthen user motivation, contextual relevance, and long-term language development.

## References

- Al-Jamili, O., Aziz, M., Mohammed, F., Almogahed, A., & Alawadhi, A. (2024). Evaluating the efficacy of computer games-based learning intervention in enhancing English speaking proficiency. *Heliyon*, 10(16), e36440. <https://doi.org/10.1016/j.heliyon.2024.e36440>
- Alabi, O., & Bukola, T. (2023). Introduction to Descriptive statistics. *Recent Advances in Biostatistics*. <https://doi.org/10.5772/intechopen.1002475>
- Alshammare, H., Alshayeb, M., & Baslyman, M. (2025). Revealing the mobile UX horizon: Exploring user experience aspects, attributes, and measurement methods - A systematic mapping study. *Computer Standards and Interfaces*, 94(March), 103999. <https://doi.org/10.1016/j.csi.2025.103999>
- Andrade, C. (2021). The Inconvenient Truth About Convenience and Purposive Samples. *Indian Journal of Psychological Medicine*, 43(1), 86–88. <https://doi.org/10.1177/0253717620977000>
- Blanchin, M., Guilleux, A., Hardouin, J.-B., & Sébille, V. (2019). Comparison of structural equation modelling, item response theory and Rasch measurement theory-based methods for response shift detection at item level: A simulation study. In *Statistical Methods in Medical Research* (Vol. 29, Issue 4, pp. 1015–1029). SAGE Publications. <https://doi.org/10.1177/0962280219884574>
- Blatna, D. (2013). Bootstrapping In Regression Analysis Of Tertiary Education Attainment In European Countries. *The 7th International Days of Statistics and Economics*, 154–163.
- Bock, H.-H. (2007). *Clustering Methods: A History of k-Means Algorithms*. 1957, 161–172. [https://doi.org/10.1007/978-3-540-73560-1\\_15](https://doi.org/10.1007/978-3-540-73560-1_15)
- Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior*, 36(1), 3–15. <https://doi.org/10.1002/job.1960>
- Chen, M., & Hoe, T. W. (2025). K-Means Clustering: A Tool for English Language Teaching Innovations. *Forum for Linguistic Studies*, 7(2), 988–998. <https://doi.org/10.30564/fls.v7i2.8379>
- Cooper, R. (2010). Theoretical considerations in qualitative interviewing. *Qualitative Report*, 15(4), 1002–1005. <https://doi.org/10.46743/2160-3715/2010.1195>
- Creswell, J. W., & Clark, V. L. P. (2018). Designing and Conducting Mixed Methods Research - Third Edition. In *Encyclopedia of Research Design* (Third Edit). <https://doi.org/10.4135/9781412961288.n245>
- Dai, Y., & Wu, Z. (2022). Mobile-assisted peer feedback on EFL pronunciation: Outcome effects, interactional processes, and shaping factors. *System*, 111(July 2021), 102953. <https://doi.org/10.1016/j.system.2022.102953>
- DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, 40(4), 314–321. <https://doi.org/10.1111/j.1365-2929.2006.02418.x>
- Elhami, A. (2022). Conducting an Interview in Qualitative Research : *MEXTESOL Journal*, 46(1), 1–7.
- Elliott, R., & Timulak, L. (2015). A Handbook of Research Methods for Clinical and Health Psychology (DRAFT). *A Handbook of Research Methods for Clinical and Health Psychology (DRAFT)*, 147–160. <https://doi.org/10.1093/med:psych/9780198527565.001.0001>
- Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fangohr, H., Beg, M., Bergemann, M., Bondar, V., Brockhauser, S., Carinan, C., Costa, R., Antonia, F. D., Danilevski, C., Ehsan, W., Esenov, S. G., Fabbri, R., Fangohr, S., Flucke, G., Fortmann, C., Marsa, D. F., Giovanetti, G., Goeries, D., Hauf, S., ... Perrin, J. F. (2019). Data Exploration and analysis with Jupyter notebooks. *Proceedings of 17th International Conference on Accelerator and Large Experimental Physics Control Systems*, January. <https://doi.org/10.18429/JACoW-ICALEPCS2019-TUCPR02>
- Fisher, R. A. (2009). Descriptive Statistics Objective: Review the basic concepts of elementary statistics. *Mathematical Statistics with Applications*, 1–51. [www.stetson.edu/~efriedma/periodictable/jpg/Fisher.jpg](http://www.stetson.edu/~efriedma/periodictable/jpg/Fisher.jpg)
- Freedman, D. A. (2007). Bootstrapping Regression Models. *The Annals of Statistics*, 9(6), 1–14. <https://doi.org/10.1214/aos/1176345638>
- Guerrero, L. A., Ochoa, S., & Collazos, C. (2010). A mobile learning tool for improving grammar skills. *Procedia - Social and Behavioral Sciences*, 2(2), 1735–1739. <https://doi.org/10.1016/j.sbspro.2010.03.975>
- Hwang, G. J., Rahimi, M., & Fathi, J. (2024). Enhancing EFL learners' speaking skills, foreign language enjoyment, and language-specific grit utilising the affordances of a MALL app: A microgenetic perspective. *Computers and Education*, 214(February), 105015. <https://doi.org/10.1016/j.compedu.2024.105015>

ISO 9241-210. (2010). ISO 9241-210: Ergonomics of human–system interaction - Human-centred design for interactive systems. *International Organization for Standardization*, 2010, 32.

Janebi Enayat, M., Asadi Ghadim, N., & Arabmofrad, A. (2025). Effects of two mobile-assisted language learning apps on L2 receptive and productive vocabulary knowledge: A mixed-methods study. *System*, 133(April), 103763. <https://doi.org/10.1016/j.system.2025.103763>

Kelly, S., Kaye, S. A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, 77(December 2022), 101925. <https://doi.org/10.1016/j.tele.2022.101925>

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods and Research*, 44(3), 486–507. <https://doi.org/10.1177/0049124114543236>

Kessler, M., Loewen, S., & Gönülal, T. (2023). Mobile-assisted language learning with Babbel and Duolingo: comparing L2 learning gains and user experience. *Computer Assisted Language Learning*, May. <https://doi.org/10.1080/09588221.2023.2215294>

Koole, M. (2022). *The Framework for the Rational Analysis of Mobile Education (Frame) Model: An Evaluation of Mobile Devices for Distance Education*. February 2006.

Koole, M. L. (2009). Mobile Learning - A Model for Framing Mobile Learning. *Mobile Learning: Transforming the Delivery of Education and Training*, 1(2), 25–47.

Kukulska-Hulme, A. (2009). Will mobile learning change language learning? *ReCALL*, 21(2), 157–165. <https://doi.org/10.1017/S0958344009000202>

Kukulska-Hulme, A. (2021). Reflections on research questions in mobile assisted language learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 28–46. <https://doi.org/10.1515/jccall-2021-2002>

Kukulska-Hulme, A., & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology*, 49(2), 207–218. <https://doi.org/10.1111/bjet.12580>

Kunnan, A. J. (1998). An introduction to structural equation modelling for language assessment research. In *Language Testing* (Vol. 15, Issue 3, pp. 295–332). SAGE Publications. <https://doi.org/10.1177/026553229801500302>

Lin, C. J., Hwang, G. J., Fu, Q. K., & Cao, Y. H. (2020). Facilitating EFL students' English grammar learning performance and behaviors: A contextual gaming approach. *Computers and Education*, 152(November 2019), 103876. <https://doi.org/10.1016/j.compedu.2020.103876>

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293–311. <https://doi.org/10.1017/S0958344019000065>

Lubis, M., Sutoyo, E., Azuddin, M., & Handayani, D. (2019). User Experience in Mobile Application Design: Utility Defined Context of Use. *Journal of Physics: Conference Series*, 1361(1). <https://doi.org/10.1088/1742-6596/1361/1/012043>

Marshall, G., & Jonker, L. (2010). An introduction to descriptive statistics: A review and practical guide. *Radiography*, 16(4), e1–e7. <https://doi.org/10.1016/j.radi.2010.01.001>

McKinney, W. (2017). *Python for Data Analysis - 2nd edition*.

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1177/016146810610800610>

Okonkwo, C. (2024). Assessment of User Experience (UX) Design Trends in Mobile Applications. *Journal of Technology and Systems*, 6(5), 29–41. <https://doi.org/10.47941/jts.2147>

Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, 07(08), 64–69. <https://doi.org/10.31695/ijasre.2021.34050>

Priyatma, J. E., Sriwindono, H., Prima, P. H., & Polina, A. M. (2024). *The Application of K-Means Clustering Algorithm for Initial Analysis of Students Online Learning*. 6(07), 67–75. <https://doi.org/10.35629/5252-06076775>

Robinson, O. C. (2023). Probing in qualitative research interviews: Theory and practice. *Qualitative Research in Psychology*, 20(3), 382–397. <https://doi.org/10.1080/14780887.2023.2238625>

Royo, T., & Laborda, J. G. (2018). *Standardization of Design Interfaces Applied to Language Test on-line through Ubiquitous Devices*. <https://doi.org/10.3991/ijim.v12i4.9197>

Ryan, R. M., & Deci, E. L. (2017). Self-determination theory. Basic psychological needs in motivation, development and wellness. *Revue Québécoise de Psychologie*, 38(3), 231. <https://doi.org/10.7202/1041847ar>

Sakkir, G., & Syamsuddin, N. A. (2023). Students' Perceptions of Duolingo Mobile Assisted Language Learning (MALL) in Learning English Vocabulary. *EduLine: Journal of Education and Learning Innovation*, 3(3), 381–388. <https://doi.org/10.35877/454ri.eduline1970>

Sevara, M. (2024). The Impact of Language Learning Apps on English Proficiency. *European Journal of Emerging Technology and Discoveries*, 2(9), 45–58.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>

Xin, S., & Ren, B. (2016). Application of Regression Analysis for Small Samples Based on Bootstrap Method. *Proceedings of the 2016 International Conference on Education, Management, Computer and Society*, 37(Emcs), 816–819. <https://doi.org/10.2991/emcs-16.2016.200>

## APPENDICES

### Appendix A. Demographic and Application-Usage Questionnaire

This appendix presents the demographic items and preliminary usage questions administered prior to the main UX instrument.

#### A1. Demographic Questions

| No. | Question                                                   | Response Type                           |
|-----|------------------------------------------------------------|-----------------------------------------|
| 1   | Age                                                        | Dropdown: 18, 19, 20, 21, 22, $\geq 23$ |
| 2   | Gender                                                     | Multiple choice: Male, Female           |
| 3   | Study Program                                              | Short answer (program name)             |
| 4   | Have you ever used a mobile language learning application? | Multiple choice: Yes, No                |

#### A2. Application-Usage Questions

| No. | Question                                          | Response Type                                                      |
|-----|---------------------------------------------------|--------------------------------------------------------------------|
| 1   | Which application do you use?                     | Multiple choice: Duolingo, ELSA Speak, Other                       |
| 2   | How long have you used the application?           | Dropdown: <1 month, 1–2 months, 3–4 months, 5–6 months, > 6 months |
| 3   | What type of plan do you use?                     | Multiple choice: Trial, Free, Premium, Not sure                    |
| 4   | On average, how often do you use the application? | Multiple choice: Every day, 3–4 times/week, 1–2 times/week, Rarely |

### Appendix B. User Experience Questionnaire

This appendix contains the full set of UX items used in the study. The final instrument includes **18 validated items** measuring Engagement, Retention, Usability, and Acceptance on a **5-point Likert scale** (1 = *Strongly Disagree*, 5 = *Strongly Agree*). Items Q11 and Q16 were removed after the pilot due to low reliability.

#### B1. Engagement (5 items)

1. I feel motivated to keep using this application even when I am not required to.
2. The materials in this application feel boring because they repeat too much and rarely offer new challenges.
3. I rarely find opportunities to use the language in real-life contexts through this application.
4. I often feel like I am just clicking through the lessons without truly understanding the content.
5. The content in this application does not feel relevant to my daily life or local culture.

#### B2. Retention (5 items)

6. I still remember a good portion of the material I learned from this application over time.
7. I struggle to remember the material because the application emphasizes memorization rather than real usage.
8. The examples in this application are not based on situations I encounter in daily life.
9. Excessive repetition in the application makes the learning experience less memorable.

10. I still cannot apply what I learned in real conversations or practical tasks.

**B3. Usability (4 retained items)**

- 11. — *Removed after pilot testing (low item-total correlation)*
- 12. I find it difficult to locate features that help me practice the language in a practical and meaningful way.
- 13. The application does not support activities that resemble natural interaction in real-life situations.
- 14. Some lessons feel too repetitive, even when I have already mastered the material.
- 15. The content feels unfamiliar or disconnected from my everyday environment.

**B4. Acceptance (4 retained items)**

- 16. — *Removed after pilot testing (low reliability)*
- 17. I do not feel that I am actually learning how to use the language in real tasks or real-life situations.
- 18. The lessons focus too much on repetitive exercises and do not encourage deeper understanding.
- 19. I do not feel represented in the examples or cultural contexts shown in the application.
- 20. The repetitive format of the exercises reduces my motivation over time.

**Appendix C. Interview Protocol**

This appendix provides the complete interview guide used during qualitative data collection.

**C1. Warm-up and Demographics**

- 1. Could you tell me a bit about your major and year of study?
- 2. Have you used language-learning applications such as Duolingo or ELSA Speak? For how long?

**C2. Engagement (4 questions)**

*Linked with SDT and FRAME*

- 3. What usually makes you feel interested or excited to use this language-learning app?
- 4. Have there been moments when you felt bored or lost interest? What caused that?
- 5. Does the way the app presents lessons feel lively and enjoyable or more routine?
- 6. Are there features that make you feel more motivated to keep learning?

**C3. Retention (4 questions)**

*Linked with TPACK and SDT*

- 7. Do you still remember and use the material you learned from the app?
- 8. How often are you able to apply what you learned in real life?
- 9. Does the app focus more on memorization or deeper understanding?
- 10. Are there features that help strengthen your memory or language skills?

**C4. Usability (3 questions)**

*Linked with FRAME*

- 11. What was your first impression when using the app—easy or confusing?
- 12. Are there features that feel unintuitive or frustrating?
- 13. Have you experienced technical issues such as bugs, errors, or slow loading?

**C5. Acceptance (4 questions)**

*Linked with SDT, TPACK, FRAME*

- 14. Do you think this app is suitable for long-term use in language learning? Why?
- 15. Have you ever recommended the app to others? Why or why not?
- 16. What are your expectations for future improvements?
- 17. If you could change one thing to make the app better suited for your needs, what would it be?

**C6. Closing**

- 18. Is there anything else you would like to add?
- 19. Thank you for participating.

**Appendix D. Thematic Coding Matrix (15 interviewees)**

**Description:** This table lists each interview participant (anonymized ID), the app they used, and the themes/codes that emerged in their interview. Checkboxes indicate which themes were present for that participant.

| Theme/Code                      | Interviewee |   |   |   |   |   |   |   |   |    |    |    |    |    |    |
|---------------------------------|-------------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
|                                 | 1           | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Repetitive & monotonous content | ✓           | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Gamified UI Fun (superficial)   | ✓           | ✓ | ✓ | ✓ | ✓ | ✓ |   | ✓ | ✓ | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Real-world Use/Retention(No)    | ✓           | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Speaking Feature Limitations    |             |   | ✓ |   | ✓ |   |   | ✓ |   | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Premium Walls/Ads               |             |   |   |   |   |   | ✓ | ✓ | ✓ |    |    |    | ✓  |    | ✓  |
| Rarely Recommend                | ✓           | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Suggested Improvements          |             |   | ✓ |   | ✓ |   | ✓ | ✓ | ✓ | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |

Appendix E. SEM Model Results and Fit Indices

E1. Standardized Path Estimates

| Path                   | Estimate | p-value    | Interpretation                                                  |
|------------------------|----------|------------|-----------------------------------------------------------------|
| Engagement → Usability | 0.3188   | 1.18e−08   | Higher engagement is associated with higher perceived usability |
| Retention → Usability  | 0.3794   | 1.24e−10   | More consistent use predicts higher usability                   |
| Usability → Acceptance | 0.7548   | < 0.000001 | Strongest effect; usability strongly predicts acceptance        |

Residual variances: Usability = 0.196; Acceptance = 0.228

E2. Model Fit Indices

| Fit Index | Value                                 | Interpretation                                                          |
|-----------|---------------------------------------|-------------------------------------------------------------------------|
| CFI       | 0.9348                                | Acceptable model fit                                                    |
| TLI       | 0.8957                                | Marginal but acceptable for low-df structural models                    |
| RMSEA     | 0.2119                                | Inflated due to df = 5; RMSEA is unreliable with low degrees of freedom |
| SRMR      | Not computed in semopy for this model |                                                                         |

The relatively high RMSEA value should be interpreted with caution. RMSEA has been shown to be sensitive to models with very low degrees of freedom, where it may overestimate lack of fit even when other indices indicate acceptable model performance (Kenny et al., 2015). Given that the present structural path model uses composite variables and a parsimonious specification, the RMSEA value is likely influenced by these characteristics. Therefore, model adequacy was evaluated using a combination of absolute and incremental fit indices rather than relying solely on RMSEA. In low-df path models with observed composite variables, CFI and TLI are generally considered more informative indicators of comparative model fit.

Appendix F. Pilot Test: Validity & Reliability (Item and Scale Statistics)

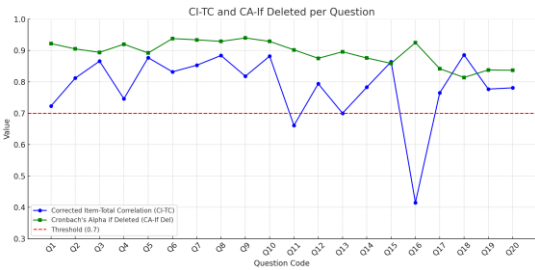
A pilot test was conducted with 20 participants to assess item clarity and internal consistency. Item-total correlations and factor diagnostics indicated that two items (Q11 and Q16) did not meet the pre-specified thresholds for retention (CI-TC Q11 = 0.661; Q16 = 0.415).

F1. Summary reliability (pilot → final)

| Construct         | Initial items (before pilot) | Final items (after removal) | Cronbach's α (final)   |
|-------------------|------------------------------|-----------------------------|------------------------|
| Engagement        | 5                            | 5                           | 0.868                  |
| Retention         | 5                            | 5                           | 0.859                  |
| Usability         | 5 → 4 (Q11 removed)          | 4                           | 0.858                  |
| Acceptance        | 5 → 4 (Q16 removed)          | 4                           | 0.868                  |
| Overall UX scales | 20 → 18                      | 18                          | α range: 0.858 – 0.868 |

F2. Item diagnostics table (Pilot)

| Section 1 (Engagement) Analysis |       |           | Section 2 (Retention) Analysis |       |           | Section 3 (Usability) Analysis |       |           | Section 4 (Acceptance) Analysis |       |           |
|---------------------------------|-------|-----------|--------------------------------|-------|-----------|--------------------------------|-------|-----------|---------------------------------|-------|-----------|
| Cronbach's Alpha: 0.924         |       |           | Cronbach's Alpha: 0.946        |       |           | Cronbach's Alpha: 0.903        |       |           | Cronbach's Alpha: 0.879         |       |           |
| Q Code                          | CI-TC | CA-If Del | Q Code                         | CI-TC | CA-If Del | Q Code                         | CI-TC | CA-If Del | Q Code                          | CI-TC | CA-If Del |
| Q1                              | 0.723 | 0.922     | Q6                             | 0.832 | 0.938     | Q11                            | 0.661 | 0.902     | Q16                             | 0.415 | 0.925     |
| Q2                              | 0.812 | 0.905     | Q7                             | 0.853 | 0.934     | Q12                            | 0.794 | 0.875     | Q17                             | 0.765 | 0.842     |
| Q3                              | 0.866 | 0.894     | Q8                             | 0.884 | 0.929     | Q13                            | 0.700 | 0.896     | Q18                             | 0.886 | 0.814     |
| Q4                              | 0.746 | 0.920     | Q9                             | 0.818 | 0.940     | Q14                            | 0.783 | 0.876     | Q19                             | 0.777 | 0.838     |
| Q5                              | 0.877 | 0.892     | Q10                            | 0.882 | 0.929     | Q15                            | 0.864 | 0.859     | Q20                             | 0.781 | 0.837     |





## Factors Affecting Lecturers' Intention to Adopt Artificial Intelligence Tools for Assessing Students' Performance in Malaysian Universities

Raghad Baker Sadiq<sup>1,2\*</sup>, Nur Fazidah Elias<sup>1</sup> and Hazura Mohamed<sup>1</sup>

<sup>1</sup>Faculty of Information Science and Technology, The National University of Malaysia, Selangor, Malaysia

<sup>2</sup>College of Business Informatics, University of Information Technology and Communications, Baghdad, Iraq

\*Correspondence: bakiermajid@yahoo.com

### PAPER INFO

#### *Paper history:*

Received 11 January 2026

Accepted 22 March 2026

#### *Citation:*

Sadiq, R. B., Elias, N. F. & Mohamed, H. (2026). Factors Affecting Lecturers' Intention to Adopt Artificial Intelligence Tools for Assessing Students' Performance in Malaysian Universities. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 129 - 155

#### *Copyright:*

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

### ABSTRACT

Empowering the student assessment process in higher education through artificial intelligence has become essential for delivering more adaptive, accurate, and equitable evaluations in a rapidly evolving educational landscape. As Malaysia seeks to integrate such technology into its assessment practices, lecturers' support and willingness are essential for successful implementation. Thus, this research investigates the factors influencing lecturers' intentions to adopt AI-based assessment tools in Malaysian universities by extending the meta-analysis based Modified Unified Theory of Acceptance and Use of Technology (meta-UTAUT). Data were collected from 414 lecturers in the Klang Valley and analyzed using Partial Least Squares Structural Equation Modelling (PLS-SEM). The results reveal that performance expectancy, effort expectancy, self-efficacy, trust, compatibility, and personal innovativeness influence lecturers' attitudes. For adoption intention, performance expectancy, self-efficacy, personal innovativeness, and attitude are key predictors. Moreover, lecturers' attitudes mediate the relationships between performance expectancy, effort expectancy, self-efficacy, trust, compatibility, and personal innovativeness and their intention. This study offers novel insights into AI-based assessment adoption and represents the first empirical application of the meta-UTAUT model in this context. From a practical perspective, the findings provide actionable guidance for policymakers, administrators, and developers to support the effective implementation of these tools in Malaysian universities.

**Keywords:** Assessment, Artificial intelligence, Higher education, Meta-UTAUT, Technology adoption, Intelligent teaching and learning, Student performance

## 1. Introduction

Artificial intelligence (AI) in higher education (HE) has reshaped teaching and learning (T/L) by enhancing effectiveness, adaptability, and overall quality (Chassab et al., 2021). It enables adaptive learning, provides immediate feedback, supports progress monitoring with early necessary interventions, and streamlines administrative tasks, thereby allowing lecturers to focus on other activities (Baker et al., 2019). These benefits encourage universities to adopt AI technologies to stay competitive and enhance the student learning experience (Naji et al., 2022). However, successful implementation heavily depends on lecturers' support, as they are the key stakeholders in HE (Wang et al., 2021). Thus, recent studies have examined factors influencing lecturers' acceptance and adoption of AI technologies using different models. For example, the

Unified Theory of Acceptance and Use of Technology (UTAUT) has been employed to examine learning analytics usage (Herodotou et al., 2023), the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) has been applied to investigate ChatGPT usage (Strzelecki et al., 2024), the Innovation Diffusion Theory has been used to study the adoption of intelligent tutoring systems (Wang et al., 2020), and the Technology Acceptance Model (TAM) has been adopted to explore the adoption of AI-based applications (Wang et al., 2021). Student assessment is a cornerstone of the educational process, and Malaysian universities are enhancing it through AI, enabling evaluations that are faster, more accurate, and more authentic than traditional methods (Ishaq et al., 2020). Thus, this study aims to investigate factors influencing lecturers' intentions to adopt AI assessment tools (AIAT) in Malaysian universities. The structure of this article is as follows: Section 2 reviews the current assessment process, AI in the assessment process and its adoption in Malaysia, and the development of the proposed model. Section 3 describes the methodology, and Section 4 presents the results. Section 5 provides the discussion, followed by the study's contributions and limitations in Sections 6, 7, 8, respectively. Finally, the conclusions are presented in Section 9.

## 2. Literature Review

### 2.1. Assessment Process

Student assessment plays a central role in HE by measuring learning outcomes and informing instructional practices (Akil & Matore, 2023). Effective assessment is closely linked to student satisfaction, motivation for academic achievement, and successful academic progression (Rigopoulos, 2022). Despite its importance, the literature consistently highlights several limitations associated with traditional assessment systems. Mainly, assessment practices in many universities remain largely summative and end-of-course oriented rather than being embedded throughout the learning process (Camacho-Miñano et al., 2020). Such an approach provides isolated snapshots of performance rather than a holistic view of learning, failing to capture the dynamic and continuous nature of learning (IIUM, 2015). In addition, concentrating assessments at the end of a course imposes a time constraint for both students and faculty. This limitation reduces students' ability to consolidate learning, engage deeply with the material, and manage stress and anxiety effectively (Braga et al., 2024). It also pressures faculty to prepare students for examinations within a limited timeframe, leading to "teaching to the exam," which hinders the development of a dynamic learning environment (Braga et al., 2024). Such assessment methods are also inadequate for accurately measuring students' understanding and capabilities (Braga et al., 2024).

Equally important, the rigidity of traditional assessment methods limits their capacity to adapt to rapid curricular changes, technological advancements, and evolving educational needs (Ali, 2024). Traditional assessments often possess a narrow scope and may struggle to ensure consistent, valid, and reliable measurement of learning outcomes across diverse contexts (Song, 2024). Moreover, approaches such as standardised tests frequently emphasise rote memorisation and basic cognitive abilities; failing to capture higher-order skills, such as creativity, critical thinking, collaboration, and problem-solving, that are essential for real-world applications (Ali, 2024). The emergence of generative AI tools, such as ChatGPT, further challenges conventional assessment practices, as students may exploit these technologies to submit non-original work, raising concerns regarding authentic assessment, academic integrity, and the credibility of higher education outcomes (Smolansky et al., 2023).

Another significant challenge in academic assessment is evaluation bias, which can distort student performance based on personal characteristics such as gender, ethnicity, or socioeconomic status (Ali, 2024). Bias may also emerge from subjective judgments when different educators evaluate the same assignment, leading to inconsistencies in grading. Uniform assessment approaches that fail to account for students' diverse prior knowledge, abilities, learning experiences, or cultural backgrounds can further disadvantage some students, limiting their capacity to demonstrate true skills and creating an unequal playing field (Smolansky et al., 2023).

Moreover, assessment is often perceived as burdensome, time-consuming, and complex for educators to design and implement effectively (Flodén, 2024). Limited human resources and teacher-centred approaches exacerbate this issue, with approximately 40% of educators' time devoted to grading and related tasks (Kamalov et al., 2023). Assessment quality may also be compromised by repetitive grading, fatigue, and subjective influences, including an educator's mood, particularly in large cohorts (Ye & Manoharan, 2018). Even widely used computer-based systems primarily support objective-type questions (e.g., multiple-choice) and are not designed to assess descriptive or higher-order responses such as essays (Pinto et al., 2023). Consequently, current assessment practices, manual or semi-automated, remain costly, error-prone, and increasingly inflexible (Machicao, 2019).

The digital era has further exposed the limitations of traditional methods, as AI tools facilitate cheating and misrepresentation of skills, raising concerns about the accuracy and integrity of student evaluation (AIBIG, 2024). As a result, these challenges underscore the urgent need for authentic assessment strategies and systematic reform to maintain the accuracy and integrity of student evaluation while keeping pace with technological advancements.

## 2.2. Adopting AI-based Assessment Tools in Malaysia

AI is transforming assessment and grading by addressing the limitations of traditional methods and enabling immediate, more precise, and inclusive evaluation practices suited to the evolving educational landscape (Bond et al., 2024). Unlike conventional grading, which is constrained by human limitations, intelligent assessment technologies analyze student responses in real time and provide instant feedback (Onesi-Ozigagun et al., 2024). In large-enrolment courses, these systems can reduce grading time by 30–60% (Alfaleh, 2026). AI technologies also facilitate the adaptation of traditional assessment methods to support the continuous evaluation of student performance across courses, offering deeper insights into learning progress (Sporrong et al., 2025). This enables students to identify their strengths and weaknesses more rapidly, thereby enhancing learning outcomes (Strielkowski et al., 2025). In addition, such capabilities allow instructors to detect learning gaps at an earlier stage and provide timely interventions, helping students remain on track and achieve their academic goals (Onesi-Ozigagun et al., 2024).

AI systems can further enhance the reliability and validity of assessments while reducing errors and biases associated with manual grading by applying standardized evaluation procedures through algorithmic models (Xia et al., 2024). These systems can efficiently evaluate complex student work, including written and verbal responses, particularly in classes with large numbers of students (Alfaleh, 2026). In the same vein, adaptive AI-based assessments dynamically adjust question difficulty according to students' abilities, providing tailored assessment experiences that improve accuracy and fairness while maintaining appropriate levels of challenge (Khlaif et al., 2024). This indicates that AIAT functions not only as tools for improving efficiency but also as a mechanism for delivering genuinely personalized assessments, thereby overcoming the limitations of traditional "one-size-fits-all" examinations.

Furthermore, the integration of AI content detection tools, such as Writer, Copyleaks, GPTZero, and CrossPlag, marks a shift in enforcing academic integrity. Rather than serving solely as punitive mechanisms, these tools support a broader move toward AI-aware pedagogy (Deep et al., 2025). By automating originality checks, institutions can address challenges posed by generative AI while reinforcing academic standards. This shifts the focus from merely detecting misconduct to promoting transparent and ethical AI use, helping maintain academic integrity in an increasingly technology-driven learning environment (Deep et al., 2025). Ultimately, by leveraging AIAT, HE institutions can move toward a continuous assessment model. This provides a more holistic view of student progress than traditional high-stakes exams, which only capture a single point in time (Onesi-Ozigagun et al., 2024). Examples of AIAT currently available include ChatGPT, Smodin's AI Grader, Moodie.ai, and Noodle Factor.

Malaysia's strategic landscape for AI integration is characterized by a robust, top-down push through initiatives, such as the Malaysia Education Blueprint 2015–2025 (Higher Education), MyDIGITAL, and the National Artificial Intelligence Roadmap 2021–2025 (MOSTI, 2021; National AI Office, 2021, 2025). Despite these policies, a significant gap exists between the national vision and its implementation in universities. AIAT adoption remains fragmented and reactive, constrained by differences in governance, faculty readiness, and technological infrastructure (Osman et al., 2024). This policy-practice gap, where strategic ambition exceeds institutional capacity, reflects a common challenge in emerging AI education contexts (Saman et al., 2024) and highlights that adoption involves complex human and organizational factors. At the core of this challenge are lecturers, who serve as the ultimate gatekeepers of pedagogical innovations. Their behavioral intention to adopt new technologies is the crucial factor determining the successful integration of AIAT into practice (Wang et al., 2021). While AIAT offers great opportunities to improve assessment, it also raises ethical and pedagogical concerns, including potential neglect of core educational values and reduced human involvement, or "human stunting" (Sporrong et al., 2025). Resistance may also arise from perceived complexity, unfamiliarity, or fundamental doubts about the pedagogical value of these tools (Zhai et al., 2021).

To better understand lecturers' intentions to adopt AIAT, different technology adoption models have been applied and extended in prior research. For example, Sánchez-Prieto et al. (2019) proposed an extended TAM model that incorporates factors such as trust, AI anxiety, and relative advantage. Shahid et al. (2024) adapted UTAUT and revealed that anxiety and adoption readiness significantly influence lecturers' attitudes toward AIAT adoption. Despite these valuable contributions, empirical research on AIAT adoption in HE remains limited. The specific challenges and enabling factors that shape lecturers' adoption decisions are still

not fully understood, particularly within specific contexts. Moreover, some studies have focused primarily on attitudes as the key outcome variable, overlooking behavioural intention as the immediate predictor of actual use. In contrast, research on AI adoption in other sectors suggests that behavioural intention is influenced by a range of domain-specific factors. In service contexts, behavioral intention is influenced by psychological traits, perceived empathy, risk, satisfaction, and interaction quality (Meidute-Kavaliauskiene et al., 2021; Pillai et al., 2020; Seo & Lee, 2021). In healthcare, perceived usefulness and subjective norms are key determinants (So et al., 2021), while in finance, trust and religiosity are significant predictors (Alkhowaiter, 2022). Similarly, studies in management contexts highlight the influence of performance expectancy, effort expectancy, facilitating conditions, satisfaction, and system quality (Chatterjee et al., 2021). Although these studies offer valuable insights into AI adoption, they often overlook the unique challenges of adopting AIAT in higher education, such as academic integrity, ethical considerations, and alignment with institutional policies. Addressing this gap is essential to translate policy objectives into effective, sustainable practices that enhance learning outcomes while maintaining academic integrity. Based on the discussion above, this study investigates the factors influencing lecturers' intentions to adopt AIAT in Malaysian universities. Specifically, the study addresses the following research questions:

**RQ1:** What are the key factors that influence lecturers' intentions to adopt AIAT in Malaysian universities?

**RQ2:** How can these factors be integrated into a theoretical model to explain lecturers' intentions to adopt AIAT?

**RQ3:** To what extent is the proposed model valid and reliable in explaining lecturers' intentions to adopt AIAT?

### 2.3. Research Model and Development of Hypotheses

To address RQ1, we critically evaluated the prominent technology adoption models based on logical consistency, scope, explanatory power, and parsimony (Bhattacharjee, 2012). This comparative analysis was essential given the scarcity of empirical research on AIAT. While foundational models like the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1975), Theory of Planned Behavior (TPB) (Ajzen, 1991), TAM, and Model of PC Utilization (MPCU) (Thompson et al., 1991) offer consistency, they prove insufficient for the multi-dimensional nature of AIAT. Specifically, TRA and TPB focus almost exclusively on psychological determinants, neglecting environmental and organizational influences, while TAM lacks social dimensions, and MPCU is restricted to personal computer use. The evaluation further indicates that the Technology Acceptance Model 2 (TAM2) (Venkatesh & Davis, 2000) and the Technology Acceptance Model 3 (TAM3) (Venkatesh & Bala, 2008) neglects the direct impact of external factors on intention, whereas UTAUT overlooks the relationship between facilitating conditions and behavioral intention. Notably, UTAUT and its extensions (i.e., UTAUT2 and UTAUT3) omit the mediating role of attitude, which is a factor identified as vital during early adoption stages (Upadhyay et al., 2022). Furthermore, Social Cognitive Theory (SCT) (Compeau et al., 1999) introduces interpretive ambiguity and fails to capture IT-specific determinants, such as facilitating conditions. Neglecting these factors and relationships would provide a reductionist view of technology adoption, overlooking the enablers and barriers that ultimately determine the successful integration of AIAT in Malaysia. Additionally, it would reduce explanatory power, as models such as TRA, TPB, TAM, MPCU, and SCT have frequently been shown to explain a low proportion of variance in complex environments (Bellet & Banet, 2023). Even UTAUT demonstrates relatively low predictive power when applied to academic staff, with explained variance ranging between 17% and 35% (Gunasinghe et al., 2020). From a parsimony perspective, while simpler models (e.g., TPB and TAM) are easy to apply, complex models, such as TAM3 or the full SCT, are difficult to operationalize (Daniel Mehari Alemayehu, 2023) or rely on moderators that lack universal relevance (i.e., UTAUT and UTAUT2) (Dwivedi et al., 2020).

The meta-UTAUT model (Dwivedi et al., 2019) addresses these limitations and is therefore adopted as the theoretical foundation for this study. It incorporates attitude as a mediating construct, providing a balanced structure that integrates individual, social, contextual, and technological dimensions relevant to technology adoption. This configuration offers the necessary theoretical depth for explaining technology adoption within educational contexts (Kannan et al., 2018). Furthermore, the inclusion of attitude enhances the model's explanatory power, while the removal of moderating variables simplifies its structure (Hakim et al., 2023), thereby improving its applicability for examining lecturers' intentions to adopt AIAT.

Dwivedi et al. (2020) recommended extending the meta-UTAUT model with context-specific factors to provide a more comprehensive representation of the phenomenon under investigation. In line with this recommendation, this study conducted a systematic review of 919 studies indexed in Scopus and Web of Science that cited the original model (Dwivedi et al., 2019). Studies were included if they used meta-UTAUT

as their theoretical foundation and incorporated external factors. As presented in Table 1, 12 studies met these criteria, collectively forming a pool of 21 potential external factors.

| Source                             | Purpose                                                                                                                     | External factors and their relationships                                                                                                                                                                                                                                                                                                                                                                     |
|------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Alkhowaiter (2022)                 | Developed and validated a model to investigate the factors affecting the use of mobile payments                             | Trust → intention                                                                                                                                                                                                                                                                                                                                                                                            |
| Upadhyay et al. (2022)             | Developed and validated a model examining factors influencing consumer adoption of mobile payment services during COVID-19  | Self-efficacy → effort expectancy<br>Perceived severity → attitude                                                                                                                                                                                                                                                                                                                                           |
| Balakrishnan et al. (2022)         | Developed and validated a model that investigates the factors influencing users to adopt chatbots in services               | Perceived intelligence → attitude and intention<br>Perceived anthropomorphism → attitude and intention<br>Social self-efficacy moderates the relationships between variables and intention                                                                                                                                                                                                                   |
| Hermanto et al. (2022)             | Developed and validated a model that examines the factors impacting taxpayers' adoption of online tax return reporting      | Trust → attitude<br>Grievance redressal moderates the relationship between facilitating conditions and intention                                                                                                                                                                                                                                                                                             |
| Chatterjee et al. (2021)           | Proposed a model to examine the factors influencing organizational users' adoption of AI-based CRM systems                  | Compatibility → attitude and intention<br>CRM satisfaction → attitude and intention<br>CRM quality → attitude and intention<br>Attitude as a mediator between all variables and intention                                                                                                                                                                                                                    |
| Roy et al. (2021)                  | Develop and validate a model to investigate the factors affecting customer engagement behaviour in smart retailing          | SRT novelty → performance expectancy, effort expectancy, social influence, and facilitating conditions<br>SRT effectiveness → performance expectancy, effort expectancy, social influence, and facilitating conditions<br>SRT interaction quality → performance expectancy, effort expectancy, social influence, and facilitating conditions<br>Attitude as a mediator between UTAU variables and engagement |
| Tamilmani et al. (2022)            | Developed and validated a model to investigate the factors influencing consumers' intentions to use Airbnb                  | Self-efficacy → intention<br>Trust → intention<br>Hedonic motivation → effort expectancy<br>Attitude as a mediator between UTAUT variables and intention                                                                                                                                                                                                                                                     |
| Sarker et al. (2020)               | Proposed a model to identify the factors influencing consumer adoption of social commerce                                   | Trust → intention<br>Perceived risk → intention                                                                                                                                                                                                                                                                                                                                                              |
| Patil et al. (2020)                | Developed and validated a model that examines the factors impacting the adoption of consumer mobile payments                | Personal innovativeness → attitude<br>Anxiety → attitude<br>Trust → attitude<br>Attitude as a mediator between all variables and intention                                                                                                                                                                                                                                                                   |
| Wiafe et al. (2020)                | Developed and validated a model to investigate the factors affecting a multi-carrier booking and shipping system usage      | Self-efficacy → intention<br>Anxiety → intention                                                                                                                                                                                                                                                                                                                                                             |
| Huseynov and Özkan Yıldırım (2019) | Developed and validated a model to examine the factors affecting consumers' shopping behaviours on B2C e-commerce platforms | Perceived enjoyment → attitude and intention<br>Compatibility → attitude, intention, and performance expectancy<br>Perceived information security → attitude<br>Perceived social pressure → attitude and intention                                                                                                                                                                                           |

|                       |                                                                                                        |                                                                         |
|-----------------------|--------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Hossain et al. (2019) | Developed and validated a model examining factors influencing physicians' adoption of e-health records | Personal innovativeness → intention<br>Resistance to change → intention |
|-----------------------|--------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|

**Table 1.** Summary of the final selected studies that adapted the meta-UTAUT model

To answer RQ2, the conceptual research model was developed by extending the meta-UTAUT model with selected external factors identified through the systematic review. At this stage, maintaining a balance between parsimony and comprehensiveness was critical (Sekaran & Bougie, 2016); therefore, the most frequently cited factors across the reviewed studies were prioritized. The prevalence of these factors suggests that they effectively address the limitations of meta-UTAUT, while their widespread application in AI-related research confirms their specific relevance to the present study. Ultimately, five external factors were integrated into the meta-UTAUT model, as illustrated in Figure 1.

As a critical step in the model's development, the initial conceptual model was evaluated by 10 experts through semi-structured interviews. This evaluation yielded strong consensus on the model's validity, confirming the inclusion of all proposed factors and relationships. Based on the recommendations of two experts, the model was refined by adding a link between self-efficacy and attitude and testing the mediating effect. The definitions of the factors with their corresponding hypotheses for the model are presented below.

**Performance expectancy:** is the degree to which a lecturer believes that using AIAT would facilitate and improve performance assessment, enhance learning outcomes, and improve teaching quality (Venkatesh et al., 2003). Performance expectancy positively influences faculty members' intention to adopt AI (Ghimire et al., 2024). Similarly, it positively shapes employees' attitudes towards AI-based CRM systems, with attitude serving as a mediating factor between performance expectancy and intention (Chatterjee et al., 2021). Thus, we proposed the following hypotheses:

**H1:** Performance expectancy has a positive effect on lecturers' attitudes towards adopting AIAT.

**H2:** Performance expectancy has a positive effect on lecturers' intention to adopt AIAT.

**H3:** Lecturers' attitudes mediate the relationship between performance expectancy and their intentions to adopt AIAT.

**Effort expectancy:** is the degree to which a lecturer believes that AIAT would be easy to use (Venkatesh et al., 2003). Effort expectancy positively influenced users' intention to accept intelligent systems (Wanner et al., 2022). Likewise, it positively influences users' intention to adopt AI in HE, with attitude mediating the relationship between effort expectancy and adoption intention (Chatterjee & Bhattacharjee, 2020). Accordingly, we proposed the following hypotheses:

**H4:** Effort expectancy has a positive effect on lecturers' attitudes towards adopting AIAT.

**H5:** Effort expectancy has a positive effect on lecturers' intention to adopt AIAT.

**H6:** Lecturers' attitudes mediate the relationship between effort expectancy and their intention to adopt AIAT.

**Social influence:** is the degree to which a lecturer perceives that important people, such as colleagues, supervisors, or the university, expect him/her to use AIAT (Venkatesh et al., 2003). Social influence positively affects users' intention to adopt learning analytics in universities (Bahari et al., 2023) and shapes students' attitudes towards MOOCs (Altalhi, 2021). Users' attitudes towards robo-advisors mediate the relationship between social influence and intention (Roh et al., 2023). Thus, we proposed the following hypotheses:

**H7:** Social influence has a positive effect on lecturers' attitudes towards adopting AIAT.

**H8:** Social influence has a positive effect on lecturers' intention to adopt AIAT.

**H9:** Lecturers' attitudes mediate the relationship between social influence and their intention to adopt AIAT.

**Facilitating conditions:** refer to the extent to which a lecturer perceives that organizational and technical support is available for using AIAT (Venkatesh et al., 2003). The adoption of AIAT depends on adequate resources (e.g., internet access, equipment), training, and IT support, which institutions must provide to foster positive attitudes and encourage adoption (Chatterjee & Bhattacharjee, 2020). Facilitating conditions positively influence users' intention to adopt AI in HE (Chatterjee & Bhattacharjee, 2020). Additionally, they shape positive attitudes towards robo-advisors, with attitude acting as a mediating factor between facilitating conditions and adoption intention (Roh et al., 2023). Accordingly, we proposed the following hypotheses:

**H10:** Facilitating conditions have a positive effect on lecturers' attitudes towards adopting AIAT.

**H11:** Facilitating conditions have a positive effect on lecturers' intention to adopt AIAT.

**H12:** Lecturers' attitudes mediate the relationship between facilitating conditions and their intention to adopt AIAT.

**Self-efficacy:** refers to the extent to which lecturers believe in their ability to use AIAT in the future (Venkatesh et al., 2003). It affects their engagement, effort, emotional responses, and persistence in overcoming challenges (Compeau & Higgins, 1995). Low self-efficacy can hinder engagement with a new system, limiting its success, while high self-efficacy has the opposite effect (Compeau & Higgins, 1995). Self-efficacy positively influences lecturers' intention to adopt AI in HE (Wang et al., 2021). Similarly, it shapes students' attitudes towards AI-based healthcare solutions, with attitude serving as a mediator (Kwak, Ahn, et al., 2022). In addition, it enhances perceptions of effort expectancy in adopting AI-based education systems (Altalhi, 2021). Therefore, the proposed hypotheses are:

**H13:** Self-efficacy has a positive effect on lecturers' attitudes towards adopting AIAT.

**H14:** Self-efficacy has a positive effect on lecturers' intention to adopt AIAT.

**H14a:** Self-efficacy has a positive effect on effort expectancy regarding AIAT.

**H15:** Lecturers' attitudes mediate the relationship between self-efficacy and their intention to adopt AIAT.

**Anxiety:** is defined as a lecturer's apprehension or fear of using AIAT to assess students' performance (Venkatesh et al., 2003). Such anxiety can reduce lecturers' engagement with AI tools, leaving them unmotivated, regardless of their potential benefits (Gupta & Bhaskar, 2020). Anxiety negatively influences users' intention to adopt AI-enabled tools (Kleine et al., 2023). In addition, it influences students' attitudes towards AI-based healthcare systems, with attitude serving as a mediating factor between anxiety and intention (Kwak, Seo, et al., 2022). Thus, the following hypotheses are proposed:

**H16:** Anxiety has a negative effect on lecturers' attitudes towards adopting AIAT.

**H17:** Anxiety has a negative effect on lecturers' intention to adopt AIAT.

**H18:** Lecturers' attitudes mediate the relationship between anxiety and their intention to adopt AIAT.

**Trust:** refers to the extent to which lecturers believe that AIAT will meet their expectations when assessing student performance (Patil et al., 2020). Trust plays a fundamental role in shaping behaviour (Kim & Prabhakar, 2004). Low trust may lead to non-use of technology, whereas high trust can increase the risk of misuse or abuse (Parasuraman & Riley, 1997). Trust positively influences both users' attitudes and intentions to adopt robo-advisors, with attitude as a mediator (Roh et al., 2023). Thus, the proposed hypotheses are:

**H19:** Trust has a positive effect on lecturers' attitudes towards adopting AIAT.

**H20:** Trust has a positive effect on lecturers' intention to adopt AIAT.

**H21:** Lecturers' attitudes mediate the relationship between trust and their intention to adopt AIAT.

**Compatibility:** refers to the degree to which lecturers believe that adopting and using AIAT aligns with their needs, existing knowledge, assessment methods, pedagogical practices, and work tasks (Moore & Benbasat, 1991). The perceived alignment between AIAT and lecturers' perceptions is critical for driving adoption (Gupta & Bhaskar, 2020). Compatibility positively influences users' intentions to adopt smart devices (Yang et al., 2022). Moreover, it affects employees' attitudes towards AI-based CRM systems, with attitude mediating the relationship between compatibility and intention (Chatterjee et al., 2021). Furthermore, compatibility strongly predicts performance expectancy by shaping users' perceptions of efficiency and the value of AI tools (Lee & Pan, 2023). Thus, the proposed hypotheses are:

**H22:** Compatibility has a positive effect on lecturers' attitudes towards adopting AIAT.

**H23:** Compatibility has a positive effect on lecturers' intention to adopt AIAT.

**H23a:** Compatibility has a positive effect on the performance expectancy of AIAT.

**H24:** Lecturers' attitudes mediate the relationship between compatibility and their intention to adopt AIAT.

**Personal innovativeness:** refers to the extent to which a lecturer is willing to try new technologies, such as AIAT (Agarwal & Prasad, 1998). Highly innovative lecturers drive technology adoption, as they are more receptive to change and confident in integrating AI-based tools to enhance T/L (Gupta & Bhaskar, 2020). Personal innovativeness has a positive effect on users' intention to adopt AI tools in HE (Bahari et al., 2023) and positively influenced users' attitudes towards using AI tools (Wang & Dai, 2020). Moreover, users' attitudes mediate the relationship between personal innovativeness and intention to adopt mobile payment (Patil et al., 2020). Hence, we proposed the following hypotheses:

**H25:** Personal innovativeness has a positive effect on lecturers' attitudes towards adopting AIAT.

**H26:** Personal innovativeness has a positive effect on lecturers' intention to adopt AIAT.

**H27:** Lecturers' attitudes mediate the relationship between personal innovativeness and their intention to adopt AIAT.

**Attitude towards using AIAT:** refers to a lecturer's positive or negative disposition towards using AIAT for student assessment (Venkatesh et al., 2003). It represents the lecturer's overall emotional response to technology and determines the willingness to adopt and engage with a new system (Dwivedi et al., 2019). Positive attitudes facilitate active engagement and effective utilization, whereas negative attitudes may impede adoption and limit the impact of the technology (Wang et al., 2021). Empirical evidence highlights

the critical role of attitude in shaping lecturers' intentions to adopt AI technologies (Chatterjee & Bhattacharjee, 2020). Based on the previous discussion, we derived the following hypothesis:  
**H28:** Lecturers' attitude positively affects their intention to adopt AIAT.

**Behavioural intention to adopt AIAT:** refers to a lecturer's intention to use AIAT for assessing students' performance (Venkatesh et al., 2003). It is widely regarded as the earliest indicator of technology adoption (Venkatesh et al., 2003).

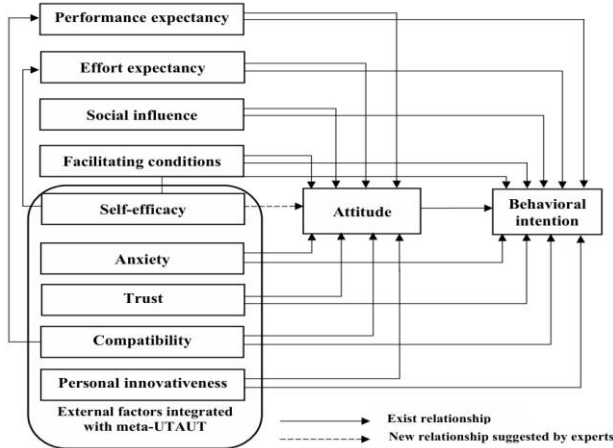


Figure 1. The final proposed model

3. Methodology

A quantitative survey method was used to test the hypotheses, as it effectively captures cause-and-effect relationships. The target population comprises lecturers from 41 universities in the Klang Valley, as they are renowned for their technological advancements in T/L, as reported by the Ministry of Education (2018). This region includes six public universities with 17,277 lecturers and 35 private universities with 7,280 lecturers, totalling 24,557 lecturers, according to the Ministry of Higher Education (2022). Using the online sample size calculator (Raosoft), the minimum required sample size was 379 participants. This sample size is adequate for providing valid results for Structural Equation Modelling (Kline, 2016). Stratified random sampling was used to ensure the accurate representation of the target population. This approach enhances the credibility of the sample by ensuring that all relevant subgroups are adequately represented, particularly in the context of heterogeneous populations with homogeneous subgroups (Taherdoost, 2016). Based on Kothari (2004), the following formula was used to calculate proportional stratified random sampling:

$$\text{The sample size for each stratum} = \frac{\text{Stratum population}}{\text{Total population}} \times \text{Identified sample}$$

The sample should include 267 respondents from public universities and 112 from private ones. Simple random sampling was employed to select participants from each stratum using Calculator.net's Random Number Generator.

The instrument was first reviewed for clarity by two IS academics and then validated by 11 experts to ensure content validity, following Lynn's guidance (1986). After two validation rounds, it was refined to 42 items. A pilot study was subsequently conducted to evaluate the instrument's reliability. As noted by Hertzog (2008), an appropriate sample size for pilot studies ranges from 10 to 40 participants per group. Accordingly, 39 responses were collected for the pilot test, comprising 28 participants from public universities and 11 from private universities. The results indicated that all constructs achieved Cronbach's alpha ( $\alpha$ ) values above the recommended threshold of 0.7, except for one item, which was subsequently removed (Hair et al., 2017). The final instrument and the corresponding internal consistency results are presented in Table 2.

| Code                    | Items                                                                                                                | Source                   | $\alpha$ |
|-------------------------|----------------------------------------------------------------------------------------------------------------------|--------------------------|----------|
| Performance expectancy  |                                                                                                                      |                          |          |
| PE1                     | Using the AIAT would enable me to accomplish assessment tasks more quickly.                                          | (Venkatesh et al., 2003) | 0.889    |
| PE2                     | Using the AIAT would enhance my productivity.                                                                        |                          |          |
| PE3                     | Using the AIAT would enhance the learning performance of students.                                                   |                          |          |
| PE4                     | Using an AIAT would improve the accuracy of students' assessments.                                                   |                          |          |
| PE5                     | Using the AIAT would help reduce my workload.                                                                        |                          |          |
| Effort expectancy       |                                                                                                                      |                          |          |
| EE1                     | It would be easy for me to become skilled at using the AIAT.                                                         | (Venkatesh et al., 2003) | 0.935    |
| EE2                     | The usage of the AIAT would be easy for me.                                                                          |                          |          |
| EE3                     | Learning to use the AIAT would be easy for me.                                                                       |                          |          |
| EE4                     | My interaction with the AIAT would be clear and understandable.                                                      |                          |          |
| Social influence        |                                                                                                                      |                          |          |
| SI1                     | My colleagues will support the use of the AIAT.                                                                      | (Venkatesh et al., 2003) | 0.972    |
| SI2                     | My superiors will support the use of the AIAT.                                                                       |                          |          |
| SI3                     | The university's top management will support the use of the AIAT.                                                    |                          |          |
| Facilitating conditions |                                                                                                                      |                          |          |
| FC1                     | I have the necessary resources to use the AIAT.                                                                      | (Venkatesh et al., 2003) | 0.789    |
| FC2                     | I have the necessary knowledge to use the AIAT.                                                                      |                          |          |
| FC3                     | The AIAT would be compatible with other technologies or tools I use.                                                 |                          |          |
| FC4                     | I can get assistance from the IT department when I have difficulties using AIAT.                                     |                          |          |
| Self-efficacy           |                                                                                                                      |                          |          |
| SE1                     | I could operate the AIAT correctly even if there is no one around to tell me what to do.                             | (Venkatesh et al., 2003) | 0.918    |
| SE2                     | I could use the AIAT correctly if I could get assistance when I am stuck.                                            |                          |          |
| SE3                     | I could learn to use the AIAT correctly if I have a lot of time.                                                     |                          |          |
| SE4                     | I could operate the AIAT correctly if the tool provides the appropriate user guidance and feedback.                  |                          |          |
| Anxiety                 |                                                                                                                      |                          |          |
| ANX1                    | It scares me to think that I could cause the AIAT to destroy a large amount of information by hitting the wrong key. | (Venkatesh et al., 2003) | 0.762    |
| ANX2                    | I hesitate to use the AIAT for fear of making mistakes I cannot correct.                                             |                          |          |
| ANX3                    | The AIAT is somewhat intimidating to me.                                                                             |                          |          |
| Trust                   |                                                                                                                      |                          |          |
| TR1                     | I trust an AIAT to be reliable.                                                                                      | (Patil et al., 2020)     | 0.931    |
| TR2                     | I trust an AIAT to be secure.                                                                                        |                          |          |
| TR3                     | I believe an AIAT is trustworthy.                                                                                    |                          |          |
| Compatibility           |                                                                                                                      |                          |          |
| COM1                    | Using the AIAT is compatible with all aspects of my work.                                                            | (Wang et al., 2020)      | 0.948    |
| COM2                    | Using the AIAT is completely compatible with my academic needs.                                                      |                          |          |
| COM3                    | I think that using the AIAT fits with the way I like to assess students' performance.                                |                          |          |
| COM4                    | Using the AIAT fits with my educational goals.                                                                       |                          |          |
| Personal innovativeness |                                                                                                                      |                          |          |
| PI1                     | If I heard about a new technology like AIAT, I would look for ways to experiment with it.                            | (Agarwal & Prasad, 1998) | 0.939    |
| PI2                     | Among my peers, I am usually the first to try out a new technology, like AIAT.                                       |                          |          |
| PI3                     | In general, I am willing to try out a new technology like AIAT.                                                      |                          |          |
| PI4                     | I like to experiment with new technologies like AIAT.                                                                |                          |          |
| Attitude                |                                                                                                                      |                          |          |
| ATT1                    | The AIAT is an appropriate tool for lecturers to use.                                                                | (Venkatesh et al., 2003) | 0.941    |
| ATT2                    | I like the idea of using the AIAT to assess students' performance.                                                   |                          |          |
| ATT3                    | I think using the AIAT will be advantageous for assessing students' performance.                                     |                          |          |
| ATT4                    | Overall, my attitude towards using the AIAT is positive.                                                             |                          |          |
| Behavioural intention   |                                                                                                                      |                          |          |

|     |                                                                            |                    |       |
|-----|----------------------------------------------------------------------------|--------------------|-------|
| BI1 | I would like to use the AIAT if I have an opportunity.                     | (Fan et al., 2018) | 0.985 |
| BI2 | I would like to use the AIAT as much as possible if I have an opportunity. |                    |       |
| BI3 | I make sure I use the AIAT if I have an opportunity.                       |                    |       |

**Table 2.** Construct measurement items and internal consistency reliability results

**3.1. Data Collection**

The questionnaire, developed using Google Forms, was distributed via email to 379 lecturers across 41 universities in the Klang Valley region. At the initial stage of data collection, the response rate was relatively low (approximately 27.7%), as many lecturers were on extended leave during the holiday season. Accordingly, the number of distributed questionnaires was increased to 758, comprising 534 for public universities and 224 for private universities. This adjustment was made to account for potential nonresponse and unusable data and to ensure that the final dataset was sufficiently robust for meaningful analysis (Pedersen & Nielsen, 2014). Ultimately, 414 of the 758 distributed questionnaires were returned, including 298 from public universities and 116 from private universities. This resulted in an overall response rate of 54.6%, with response rates of 55.9% for public universities and 51.6% for private universities, which are considered acceptable for generating reliable estimates (Sue & Ritter, 2012).

**3.2. Data Analysis**

Partial Least Squares Structural Equation modelling (PLS-SEM) was used to evaluate the hypotheses using SmartPLS 4. This approach is particularly appropriate when the primary objective is theory development rather than theory confirmation, especially in studies emphasizing prediction and the explanation of key target constructs (Hair et al., 2017).

Before conducting SEM analysis, data preparation and cleaning should be performed. Data cleaning was performed using SPSS 29. No missing data was found; however, 10 responses exhibiting suspicious response patterns (straight-lining) were identified and subsequently removed (Hair et al., 2017), resulting in 404 valid cases. Subsequently, two extreme univariate outliers with highly negative z-scores on performance expectancy (beyond the recommended  $-4$  to  $+4$  range) were detected and removed (Tabachnick & Fidell, 2013). In addition, 18 multivariate outliers were identified and removed using Mahalanobis distance at  $p < 0.001$ . In total, 20 cases were removed, yielding a final dataset of 384 cases for final analysis. Although PLS-SEM is robust to non-normal data, normality assessment was conducted to strengthen the statistical rigour of the study. The results showed that univariate skewness and kurtosis values for all variables were within acceptable thresholds (skewness:  $-2$  to  $+2$ ; kurtosis:  $-7$  to  $+7$ ) (Hair et al., 2017). Furthermore, the Mardia multivariate normality test revealed no significant deviations from normality (Cain et al., 2017), confirming that the dataset met the assumptions of normality. Finally, common method bias was assessed using Harman's single-factor test (1976). The initial component accounted for 39.706% of the total variance, which is below the 50% threshold, indicating that common method variance is unlikely to have significantly influenced the results (Podsakoff & Organ, 1986). SmartPLS 4 was employed to perform the PLS-SEM analysis, as it is widely used and freely available software application for implementing the PLS-SEM technique.

**4. Results**

**4.1. Descriptive Analyses**

Descriptive statistics were used to summarize the respondents' demographic characteristics and provide a clear overview of the sample composition. Table 3 explains the characteristics of the respondents.

**4.2. Measurement Model Results**

Before proceeding to the structural model, the measurement model was rigorously examined to confirm that all latent variables met the required reliability and validity standards. As depicted in Table 4, all factor loadings ranged from 0.515 to 0.956, exceeding the recommended threshold of 0.4; therefore, no items were removed (Hair et al., 2017). In addition, the average variance extracted (AVE) values ranged from 0.6 to 0.906, indicating that each variable accounted for at least 50% of the variance of its indicators and

demonstrating adequate convergent validity (Hair et al., 2017). Cronbach's alpha values ranged from 0.781 to 0.948, while composite reliability (CR) values ranged from 0.857 to 0.967, reflecting strong internal consistency and confirming the reliability of the measurement model.

| Demographic                                                          | Categories                          | N (384) | %    |
|----------------------------------------------------------------------|-------------------------------------|---------|------|
| Age                                                                  | 21-30                               | 13      | 3.4  |
|                                                                      | 31-40                               | 129     | 33.6 |
|                                                                      | 41-50                               | 141     | 36.7 |
|                                                                      | 51-60                               | 79      | 20.6 |
|                                                                      | More than 60                        | 22      | 5.7  |
| Gender                                                               | Male                                | 140     | 36.5 |
|                                                                      | Female                              | 244     | 63.5 |
| Academic qualification                                               | Bachelor's degree                   | 6       | 1.6  |
|                                                                      | Master's degree                     | 127     | 33.1 |
|                                                                      | PhD                                 | 251     | 65.3 |
| Type of university                                                   | Public university                   | 272     | 70.8 |
|                                                                      | Private university                  | 112     | 29.2 |
| Faculty                                                              | Computer science                    | 33      | 8.6  |
|                                                                      | Social science                      | 46      | 12.0 |
|                                                                      | Education                           | 39      | 10.2 |
|                                                                      | Business                            | 70      | 18.2 |
|                                                                      | Engineering                         | 36      | 9.4  |
|                                                                      | Medicine                            | 37      | 9.6  |
|                                                                      | Others                              | 123     | 32.0 |
| Academic position                                                    | Assistant Lecturer\Tutor\Instructor | 11      | 2.9  |
|                                                                      | Lecturer                            | 99      | 25.8 |
|                                                                      | Senior Lecturer\Assistant Professor | 164     | 42.7 |
|                                                                      | Associate Professor                 | 69      | 18.0 |
|                                                                      | Professor                           | 37      | 9.6  |
| Tenure of teaching (years)                                           | Others                              | 4       | 1.0  |
|                                                                      | Less than 3                         | 42      | 10.9 |
|                                                                      | 4-7                                 | 65      | 16.9 |
|                                                                      | 8-11                                | 63      | 16.4 |
|                                                                      | 12-15                               | 66      | 17.2 |
| General knowledge of AI                                              | More than 15                        | 148     | 38.5 |
|                                                                      | Terrible                            | 14      | 3.7  |
|                                                                      | Poor                                | 41      | 10.7 |
|                                                                      | Average                             | 187     | 49.0 |
|                                                                      | Good                                | 122     | 31.9 |
| AI-enabled services or tools have been used for educational purposes | Excellent                           | 18      | 4.7  |
|                                                                      | None                                | 9       | 2.4  |
|                                                                      | Turnitin                            | 341     | 90.7 |
|                                                                      | Grammarly                           | 261     | 69.4 |
|                                                                      | QuillBot                            | 153     | 40.7 |
|                                                                      | ChatGPT                             | 122     | 32.4 |
|                                                                      | Scholarly                           | 70      | 18.6 |
|                                                                      | Semantic Scholar                    | 57      | 15.2 |
|                                                                      | Trinka                              | 4       | 1.1  |
|                                                                      | Elicit                              | 3       | 0.8  |
|                                                                      | Connected Papers                    | 37      | 9.8  |
|                                                                      | Others                              | 18      | 4.8  |

**Table 3.** Demographic information of respondents

| Variable                | Items | Factor loadings | Average variance extracted | Cronbach's alpha ( $\alpha$ ) | Composite reliability |
|-------------------------|-------|-----------------|----------------------------|-------------------------------|-----------------------|
| Anxiety                 | ANX1  | 0.799           | 0.756                      | 0.852                         | 0.903                 |
|                         | ANX2  | 0.886           |                            |                               |                       |
|                         | ANX3  | 0.920           |                            |                               |                       |
| Attitude                | ATT1  | 0.883           | 0.833                      | 0.933                         | 0.952                 |
|                         | ATT2  | 0.935           |                            |                               |                       |
|                         | ATT3  | 0.925           |                            |                               |                       |
|                         | ATT4  | 0.907           |                            |                               |                       |
| Behavioural intention   | BI1   | 0.948           | 0.906                      | 0.948                         | 0.967                 |
|                         | BI2   | 0.956           |                            |                               |                       |
|                         | BI3   | 0.951           |                            |                               |                       |
| Compatibility           | COM1  | 0.869           | 0.805                      | 0.919                         | 0.943                 |
|                         | COM2  | 0.891           |                            |                               |                       |
|                         | COM3  | 0.924           |                            |                               |                       |
|                         | COM4  | 0.904           |                            |                               |                       |
| Effort expectancy       | EE1   | 0.878           | 0.805                      | 0.919                         | 0.943                 |
|                         | EE2   | 0.913           |                            |                               |                       |
|                         | EE3   | 0.916           |                            |                               |                       |
|                         | EE4   | 0.881           |                            |                               |                       |
| Facilitating conditions | FC1   | 0.852           | 0.625                      | 0.793                         | 0.866                 |
|                         | FC2   | 0.864           |                            |                               |                       |
|                         | FC3   | 0.875           |                            |                               |                       |
|                         | FC4   | 0.515           |                            |                               |                       |
| Performance expectancy  | PE1   | 0.856           | 0.620                      | 0.845                         | 0.890                 |
|                         | PE2   | 0.867           |                            |                               |                       |
|                         | PE3   | 0.729           |                            |                               |                       |
|                         | PE4   | 0.728           |                            |                               |                       |
|                         | PE5   | 0.747           |                            |                               |                       |
| Personal innovativeness | PI1   | 0.886           | 0.771                      | 0.901                         | 0.931                 |
|                         | PI2   | 0.807           |                            |                               |                       |
|                         | PI3   | 0.881           |                            |                               |                       |
|                         | PI4   | 0.934           |                            |                               |                       |
| Self-efficacy           | SE1   | 0.749           | 0.600                      | 0.781                         | 0.857                 |
|                         | SE2   | 0.818           |                            |                               |                       |
|                         | SE3   | 0.724           |                            |                               |                       |
|                         | SE4   | 0.803           |                            |                               |                       |
| Social influence        | SI1   | 0.889           | 0.787                      | 0.865                         | 0.917                 |
|                         | SI2   | 0.921           |                            |                               |                       |
|                         | SI3   | 0.850           |                            |                               |                       |
| Trust                   | TR1   | 0.909           | 0.811                      | 0.883                         | 0.928                 |
|                         | TR2   | 0.895           |                            |                               |                       |
|                         | TR3   | 0.897           |                            |                               |                       |

**Table 4.** Convergent validity and composite reliability of the variables

Discriminant validity was assessed using two widely recognized methods: Fornell and Larcker (1981) criterion and the Heterotrait-Monotrait (HTMT) ratio (Henseler et al., 2015), as presented in Table 5. According to Fornell and Larcker (1981), discriminant validity is confirmed when the square root of a variable's AVE exceeds its correlations with other variables. The results show that all variables in the measurement model satisfy this criterion, consistent with Chin (1998). Additionally, all HTMT values were below the 0.90 threshold, indicating clear distinctions among the variables and confirming adequate discriminant validity (Hair et al., 2017).

|                                    | Variable | ANX    | ATT   | BI    | COM   | EE    | FC    | PE    | PI    | SE    | SI    | TR    |
|------------------------------------|----------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Fornell–Larcker criterion          | ANX      | 0.873  |       |       |       |       |       |       |       |       |       |       |
|                                    | ATT      | -0.189 | 0.923 |       |       |       |       |       |       |       |       |       |
|                                    | BI       | -0.203 | 0.829 | 0.957 |       |       |       |       |       |       |       |       |
|                                    | COM      | -0.198 | 0.764 | 0.679 | 0.906 |       |       |       |       |       |       |       |
|                                    | EE       | -0.256 | 0.6   | 0.573 | 0.528 | 0.897 |       |       |       |       |       |       |
|                                    | FC       | -0.153 | 0.444 | 0.393 | 0.462 | 0.632 | 0.794 |       |       |       |       |       |
|                                    | PE       | -0.094 | 0.695 | 0.669 | 0.666 | 0.57  | 0.456 | 0.803 |       |       |       |       |
|                                    | PI       | -0.255 | 0.652 | 0.719 | 0.562 | 0.59  | 0.408 | 0.517 | 0.88  |       |       |       |
|                                    | SE       | -0.163 | 0.526 | 0.572 | 0.448 | 0.595 | 0.537 | 0.493 | 0.514 | 0.793 |       |       |
|                                    | SI       | 0.042  | 0.368 | 0.328 | 0.384 | 0.318 | 0.494 | 0.458 | 0.217 | 0.394 | 0.889 |       |
|                                    | TR       | -0.059 | 0.617 | 0.546 | 0.693 | 0.418 | 0.372 | 0.587 | 0.423 | 0.374 | 0.347 | 0.914 |
| Heterotrait-Monotrait (HTMT) ratio | ANX      |        |       |       |       |       |       |       |       |       |       |       |
|                                    | ATT      | 0.211  |       |       |       |       |       |       |       |       |       |       |
|                                    | BI       | 0.217  | 0.850 |       |       |       |       |       |       |       |       |       |
|                                    | COM      | 0.229  | 0.801 | 0.687 |       |       |       |       |       |       |       |       |
|                                    | EE       | 0.275  | 0.632 | 0.587 | 0.559 |       |       |       |       |       |       |       |
|                                    | FC       | 0.180  | 0.502 | 0.421 | 0.539 | 0.712 |       |       |       |       |       |       |
|                                    | PE       | 0.145  | 0.732 | 0.676 | 0.695 | 0.641 | 0.549 |       |       |       |       |       |
|                                    | PI       | 0.273  | 0.683 | 0.759 | 0.587 | 0.631 | 0.454 | 0.544 |       |       |       |       |
|                                    | SE       | 0.215  | 0.595 | 0.639 | 0.486 | 0.672 | 0.614 | 0.555 | 0.600 |       |       |       |
|                                    | SI       | 0.069  | 0.367 | 0.308 | 0.384 | 0.337 | 0.633 | 0.488 | 0.196 | 0.456 |       |       |
|                                    | TR       | 0.097  | 0.651 | 0.539 | 0.725 | 0.457 | 0.458 | 0.627 | 0.438 | 0.397 | 0.331 |       |

Notes: PE: Performance expectancy, EE: Effort expectancy, SI: Social influence, FC: Facilitating conditions, SE: Self-efficacy, ANX: Anxiety, TR: Trust, COM: Compatibility, PI: Personal innovativeness, ATT: Attitude towards AIAT, BI: Behavioural intention to adopt AIAT

Table 5. Discriminant validity assessment

#### 4.2.1. Structural Model Results

The analysis revealed that BI was positively influenced by PE, SE, PI, and ATT, with ATT exerting the strongest effect. Accordingly, H2, H14, H26, and H28 were supported, as summarized in Table 6. FC exhibited a statistically significant but weak negative effect on BI, thereby rejecting H11. The effects of EE, SI, ANX, TR, and COM on BI were non-significant, thereby rejecting H5, H8, H17, H20, and H23.

In contrast, ATT was positively influenced by PE, EE, SE, TR, COM, and PI, with COM having the strongest effect, supporting H1, H4, H13, H19, H22, and H25. Whereas SI, FC, and ANX had an insignificant impact, rejecting H7, H10, and H16. COM positively influenced PE, and SE positively influenced EE, thereby supporting H23a and H14a, respectively.

Additionally, the results demonstrate that ATT partially mediates the relationships between PE, SE, PI, and BI, supporting H3, H15, and H27. Moreover, ATT fully mediates the relationships between EE, TR, COM, and BI, supporting H6, H21, and H24. In contrast, ATT did not exert a significant mediating effect on the relationships between SI, FC, ANX, and BI, thereby rejecting H9, H12, and H18.

| Hypotheses |                                       | Direct effects |         |          | Indirect effects |         |          | Decision  |
|------------|---------------------------------------|----------------|---------|----------|------------------|---------|----------|-----------|
|            |                                       | $\beta$        | T-value | P-value  | $\beta$          | T-value | P-value  |           |
| H1         | PE $\rightarrow$ ATT                  | 0.177          | 3.859   | 0.000*** |                  |         |          | Supported |
| H2         | PE $\rightarrow$ BI                   | 0.107          | 2.327   | 0.020*   |                  |         |          | Supported |
| H3         | PE $\rightarrow$ ATT $\rightarrow$ BI |                |         |          | 0.087            | 3.518   | 0.000*** | Supported |
| H4         | EE $\rightarrow$ ATT                  | 0.097          | 2.237   | 0.025*   |                  |         |          | Supported |
| H5         | EE $\rightarrow$ BI                   | -0.022         | 0.555   | 0.579    |                  |         |          | Rejected  |
| H6         | EE $\rightarrow$ ATT $\rightarrow$ BI |                |         |          | 0.048            | 2.149   | 0.016*   | Supported |
| H7         | SI $\rightarrow$ ATT                  | 0.022          | 0.655   | 0.513    |                  |         |          | Rejected  |
| H8         | SI $\rightarrow$ BI                   | 0.005          | 0.159   | 0.874    |                  |         |          | Rejected  |
| H9         | SI $\rightarrow$ ATT $\rightarrow$ BI |                |         |          | 0.011            | 0.649   | 0.258    | Rejected  |
| H10        | FC $\rightarrow$ ATT                  | -0.057         | 1.423   | 0.155    |                  |         |          | Rejected  |

|      |                |        |        |          |        |       |          |           |
|------|----------------|--------|--------|----------|--------|-------|----------|-----------|
| H11  | FC → BI        | -0.086 | 2.262  | 0.024*   |        |       |          | Rejected  |
| H12  | FC → ATT → BI  |        |        |          | -0.028 | 1.407 | 0.080    | Rejected  |
| H13  | SE → ATT       | 0.087  | 2.097  | 0.036*   |        |       |          | Supported |
| H14  | SE → BI        | 0.137  | 3.852  | 0.000*** |        |       |          | Supported |
| H14a | SE → EE        | 0.600  | 17.090 | 0.000*** |        |       |          | Supported |
| H15  | SE → ATT → BI  |        |        |          | 0.043  | 2.089 | 0.018*   | Supported |
| H16  | ANX → ATT      | -0.008 | 0.251  | 0.802    |        |       |          | Rejected  |
| H17  | ANX → BI       | -0.009 | 0.346  | 0.729    |        |       |          | Rejected  |
| H18  | ANX → ATT → BI |        |        |          | -0.004 | 0.250 | 0.401    | Rejected  |
| H19  | TR → ATT       | 0.109  | 2.201  | 0.028*   |        |       |          | Supported |
| H20  | TR → BI        | -0.008 | 0.208  | 0.835    |        |       |          | Rejected  |
| H21  | TR → ATT → BI  |        |        |          | 0.054  | 2.214 | 0.013*   | Supported |
| H22  | COM → ATT      | 0.377  | 6.992  | 0.000*** |        |       |          | Supported |
| H23  | COM → BI       | 0.046  | 1.044  | 0.296    |        |       |          | Rejected  |
| H23a | COM → PE       | 0.612  | 15.637 | 0.000*** |        |       |          | Supported |
| H24  | COM → ATT → BI |        |        |          | 0.186  | 5.616 | 0.000*** | Supported |
| H25  | PI → ATT       | 0.218  | 4.650  | 0.000*** |        |       |          | Supported |
| H26  | PI → BI        | 0.295  | 6.273  | 0.000*** |        |       |          | Supported |
| H27  | PI → ATT → BI  |        |        |          | 0.108  | 3.923 | 0.000*** | Supported |
| H28  | ATT → BI       | 0.493  | 9.310  | 0.000*** |        |       |          | Supported |

Notes: \*\*\* p < 0.001, \* p < 0.05. PE: Performance expectancy, EE: Effort expectancy, SI: Social influence, FC: Facilitating conditions, SE: Self-efficacy, ANX: Anxiety, TR: Trust, COM: Compatibility, PI: Personal innovativeness, ATT: Attitude towards AIAT, BI: Behavioural intention to adopt AIAT

**Table 6.** Hypothesis testing results

Table 7 indicates that the model accounts for a significant proportion of variance in both ATT and BI ( $R^2 > 0.67$ ), affirming the statistical significance and theoretical robustness of the model (Cohen, 1988). The  $R^2$  values for EE and PE reflected moderate explanatory power ( $R^2 > 0.33$ ) (Hair et al., 2017).

| Variable                                            | $R^2$ | $Q^2$ predict |
|-----------------------------------------------------|-------|---------------|
| Attitude towards adopting AI-based assessment tools | 0.683 | 0.639         |
| Intention to adopt AI-based assessment tools        | 0.728 | 0.612         |
| Effort expectancy                                   | 0.360 | 0.354         |
| Performance expectancy                              | 0.375 | 0.371         |

**Table 7.** Coefficient of determination values ( $R^2$ ) and relevance predictive ( $Q^2$ ) for the model

The effect sizes range from 0.000 (negligible) to 0.600 (large), revealing substantial variation in the influence of the predictors, as shown in Table 8. COM is the strongest determinant of ATT, exhibiting a medium effect ( $f^2 \geq 0.15$ ) (Hair et al., 2017), underscoring its central role in shaping it. This is followed by PI and PE, each of which demonstrated a small effect ( $f^2 \geq 0.02$ ) (Hair et al., 2017). The remaining variables showed negligible influence. Regarding lecturers' intention, ATT exerted the strongest influence with a medium effect, followed by PI, which also exhibited a medium effect, revealing their central role in shaping BI. SE and PE showed small effects, whereas the others contributed negligibly to BI. Finally, the effects of COM on PE and SE on EE were both large ( $f^2 \geq 0.35$ ).

| Variable                | Endogenous variable |                       |                   |                        |
|-------------------------|---------------------|-----------------------|-------------------|------------------------|
|                         | Attitude            | Behavioural intention | Effort expectancy | Performance expectancy |
| Anxiety                 | 0.000               | 0.000                 |                   |                        |
| Attitude                |                     | 0.283                 |                   |                        |
| Compatibility           | 0.185               | 0.003                 |                   | 0.600                  |
| Effort expectancy       | 0.012               | 0.001                 |                   |                        |
| Facilitating conditions | 0.005               | 0.013                 |                   |                        |
| Performance expectancy  | 0.047               | 0.019                 |                   |                        |

|                         |       |       |       |  |
|-------------------------|-------|-------|-------|--|
| Personal innovativeness | 0.082 | 0.161 |       |  |
| Self-efficacy           | 0.013 | 0.037 | 0.562 |  |
| Social influence        | 0.001 | 0.000 |       |  |
| Trust                   | 0.020 | 0.000 |       |  |

**Table 8.** Effect size ( $f^2$ ) for the endogenous variables

The model's ability to predict new observations was evaluated and found to be strong. Table 7 revealed that ATT and BI have high predictive relevance ( $Q^2 \geq 0.50$ ), whereas EE and PE exhibited moderate predictive relevance ( $Q^2 \geq 0.25$ ) (Hair et al., 2017). To provide additional insights into the quality of PLS-SEM path model estimates, the  $q^2$  effect size was assessed. Table 9 indicate that most of the predictive relevance effect sizes were below the 0.02 threshold, implying no meaningful predictive effect (Hair et al., 2017). This is likely attributable to multiple predictors that did not exhibit significant effects, while for BI, only COM, TR, and PI yielded small effects, indicating modest predictive relevance overall.

| Exogenous variable      | Endogenous variable                                 | $Q^2$ included | $Q^2$ excluded | Effect size ( $q^2$ ) | Result    |
|-------------------------|-----------------------------------------------------|----------------|----------------|-----------------------|-----------|
| Anxiety                 | Attitude towards adopting AI-based assessment tools | 0.639          | 0.641          | -0.006                | No effect |
| Compatibility           |                                                     |                | 0.603          | 0.100                 | Small     |
| Effort expectancy       |                                                     |                | 0.641          | -0.006                | No effect |
| Facilitating conditions |                                                     |                | 0.642          | -0.008                | No effect |
| Performance expectancy  |                                                     |                | 0.64           | -0.003                | No effect |
| Personal innovativeness |                                                     |                | 0.597          | 0.116                 | Small     |
| Self-efficacy           |                                                     |                | 0.638          | 0.003                 | No effect |
| Social influence        |                                                     |                | 0.64           | -0.003                | No effect |
| Trust                   |                                                     |                | 0.629          | 0.028                 | Small     |
| Anxiety                 | Intention to adopt AI-based assessment tools        | 0.612          | 0.614          | -0.005                | No effect |
| Attitude                |                                                     |                | 0.612          | 0.000                 | No effect |
| Compatibility           |                                                     |                | 0.615          | -0.008                | No effect |
| Effort expectancy       |                                                     |                | 0.612          | 0.000                 | No effect |
| Facilitating conditions |                                                     |                | 0.611          | 0.003                 | No effect |
| Performance expectancy  |                                                     |                | 0.615          | -0.008                | No effect |
| Personal innovativeness |                                                     |                | 0.556          | 0.144                 | Small     |
| Self-efficacy           |                                                     |                | 0.603          | 0.023                 | Small     |
| Social influence        |                                                     |                | 0.613          | -0.003                | No effect |
| Trust                   |                                                     |                | 0.614          | -0.005                | No effect |
| Self-efficacy           | Effort expectancy                                   | 0.354          |                |                       | Large     |
| Compatibility           | Performance expectancy                              | 0.371          |                |                       | Large     |

**Table 9.** Effect size ( $q^2$ ) for the endogenous variables

## 5. Discussion

This research employed the extended meta-UTAUT model as the theoretical framework to investigate Malaysian lecturers' attitudes and intentions to adopt AIAT in universities. The results show that 18 hypotheses were supported, including 12 direct relationships and six indirect relationships mediated by ATT, while 12 hypotheses were not supported, comprising nine direct relationships and three mediated relationships. These findings indicate that the meta-UTAUT model was partially supported. ATT mediated the effects of certain independent factors on BI but did not mediate others, suggesting that the mediating role of ATT is context-dependent and construct-specific rather than universal. This highlights limitations in the meta-UTAUT assumption that ATT consistently bridges all antecedents and BI.

PE is a key predictor of both ATT and BI. Lecturers are more likely to develop favourable attitudes and stronger intentions to adopt AIAT when they perceive these technologies as capable of accelerating the assessment process, increasing their productivity, enhancing student learning outcomes, improving grading accuracy, and reducing workload, as reflected in the measurement items. They tend to value technologies that enhance teaching effectiveness and assessment quality (González-Calatayud et al., 2021), which in turn

motivates their adoption of AIAT rather than reliance on colleagues' opinions or institutional mandates (Nadzirah Muhammad et al., 2025). Previous studies in HE consistently identify PE as a critical determinant of technology adoption among lecturers, either by shaping ATT (Helmiatin et al., 2024), directly influencing BI (Strzelecki et al., 2024), or indirectly affecting intention through the mediating role of ATT (Prakash & Das, 2021). Consistent with these findings, lecturers' positive attitudes reinforce their intention to adopt AIAT, highlighting ATT as a key psychological mechanism through which performance-related beliefs translate into BI. These findings are theoretically supported by the Decomposed Theory of Planned Behaviour (DTPB) (Taylor & Todd, 1995), UTAUT and its extensions (UTAUT2, UTAUT3, meta-UTAUT), and TAM and its extensions (TAM2 and TAM3), which emphasise PE as a driver of ATT and/or BI. However, unlike UTAUT, PE was not the strongest direct predictor of BI in this study. This may be attributed to the novelty of AIAT, as lecturers remain cautious due to limited experience and uncertainty regarding reliability and integration. Similar patterns have been observed in studies of other emerging educational technologies (Elatrachi & Oukarfi, 2020).

EE significantly influenced lecturers' ATT toward AIAT, indicating that when lecturers perceive AIAT as easy to learn and use, requiring minimal effort (i.e., based on the items), they are more likely to develop positive attitudes toward them. This finding aligns with prior studies (Helmiatin et al., 2024) and with the meta-UTAUT, TAM, and DTPB models. However, the modest effect suggests that although lecturers value ease of use, it mainly reduces perceived barriers to adoption rather than acting as a strong motivational driver shaping attitudes. This relatively weak influence may be attributed to Malaysian lecturers' baseline digital competence and their prior exposure to learning management systems, online assessment platforms, and other educational technologies, which reduces sensitivity to usability concerns (Mansour & Mansour, 2025). Consequently, further reductions in perceived effort contribute only marginally to attitude formation. The non-significant direct effect of EE on BI further supports this interpretation. Lecturers may be willing to invest effort in using AIAT when clear benefits are expected (Strzelecki et al., 2024), reflecting a shift from effort avoidance to outcome maximization, where perceived usefulness outweighs usability concerns (Wang et al., 2021). As a result, the influence of EE diminishes when tangible performance gains are anticipated (Zaim et al., 2024). Although this finding contrasts with the direct-effect assumptions of UTAUT and TAM, the significant indirect effect through ATT supports the meta-UTAUT perspective. EE shapes positive attitudes toward AIAT, which in turn increases adoption intention, consistent with prior research (Wang et al., 2021). In this context, ATT functions as a mediating mechanism that translates EE perceptions into BI. However, the modest mediation effect suggests that simplifying AIAT alone is unlikely to substantially increase adoption without clear performance benefits (Wang et al., 2021).

Meanwhile, SI was not a significant predictor in this context, showing no effect on ATT or BI, either directly or indirectly through ATT. This suggests that lecturers' attitudes and intentions are largely unaffected by the views of colleagues, superiors, or top management (i.e., based on the items). These findings align with prior studies on AI adoption among academics (Baig & Yadegaridehkordi, 2025; Bervell et al., 2020; Enang & Christopoulou, 2025). However, they diverge from the meta-UTAUT model, indicating that the salience of this construct is highly context dependent. In Malaysian higher education institutions (MHEI), AIAT adoption remains at an early stage with limited uptake (AIBIG, 2024), reflecting a voluntary environment driven primarily by individual initiative rather than institutional mandates (Mohamad Mozie et al., 2025). Consequently, the effect of SI may be diminished in early-stage adoption contexts, where lecturers with limited experience encounter minimal peer or managerial pressure (Venkatesh et al., 2003).

FC did not significantly influence lecturers' attitudes toward AIAT. In MHEI, this can be attributed to structural and cultural factors. Malaysian universities have invested extensively in digital infrastructure, learning management systems, and institutional initiatives aligned with national digitalization agendas, such as the Malaysia Digital Economy Blueprint (Mohamad Mozie et al., 2025). Consequently, facilitating conditions, such as access to systems and technical support, are increasingly perceived as standard or taken-for-granted, rather than as factors shaping positive attitudes (Bahari et al., 2023). Similar findings have been reported in Malaysian universities' technology adoption studies (Bahari et al., 2023; Chun & Yunus, 2023). Interestingly, FC exhibited a significant but small negative direct effect on AIAT adoption intention, consistent with prior studies (Zaim et al., 2024). This contrasts with Perez (2024), who found that adoption is more likely when institutional support and resources are adequate, and users are confident and knowledgeable. In the present context, limited AIAT knowledge and experience among lecturers may heighten perceptions of complexity, risk, or misalignment with pedagogical practices, thereby reducing their willingness to adopt these tools (Adirinekso et al., 2020). The indirect effect of FC on BI through ATT was non-significant, confirming that ATT does not mediate this relationship. Lecturers tend to prioritise cognitive and performance-oriented evaluations over contextual enablers, weakening the effect of FC. This pattern is consistent with evidence from Asian higher education settings where FC fail to translate into intention through

ATT (Kwak, Seo, et al., 2022). These results challenge the meta-UTAUT assumption that FC universally enable technology adoption.

SE influenced lecturers' attitudes, with higher SE fostering more positive attitudes, consistent with Wang et al. (2021). In MHEI, lecturers often demonstrate high SE due to regular engagement with digital systems integral to teaching, such as learning management systems. This experience builds competence, confidence, and familiarity with digital tools, reinforcing their belief in their ability to learn and use new technologies effectively (Teo, 2018). However, the relatively small effect of SE on ATT suggests that other factors, such as PE and PI, play a more prominent role in shaping ATT. Similarly, the significant direct effect of SE on BI indicates that enhancing lecturers' competence and confidence is a key mechanism for promoting AIAT adoption, consistent with prior studies (Herodotou et al., 2023). Lecturers are more likely to engage with AIAT when they feel capable of using the tools independently, can access support as needed, have sufficient time to learn, and receive clear guidance and feedback (i.e., based on the items). The mediated effect through ATT further emphasizes that attitudinal mechanisms transmit the influence of SE to adoption intentions, as supported by Wang et al. (2021). This indicates that confidence not only directly drives intention but also reinforces positive perceptions, increasing the likelihood of engagement with AIAT. This pattern underscores the central role of ATT as a bridge between perceptions and BI, consistent with the meta-UTAUT model, which emphasizes that users' beliefs and perceptions are translated into intention primarily through their attitudinal response. Additionally, SE strongly shapes EE, aligning with TAM3 and prior research (Wang et al., 2021). Lecturers with higher confidence are more likely to perceive new tools as manageable, reducing anxiety and perceived complexity even before actual interaction (Teo, 2018).

ANX was not a barrier in this study, as it did not significantly predict ATT or BI, either directly or indirectly, consistent with previous research (Wang et al., 2021). This suggests that Malaysian lecturers' technology-related apprehension has largely diminished due to their accumulated experience with educational information systems, including AI-related tools, which reduces uncertainty and perceived risk associated with AIAT use (Henderson & Corry, 2021). High levels of SE further enable lecturers to perceive AIAT as controllable and manageable, shifting evaluations from emotional concerns to task- and performance-oriented considerations (Kai et al., 2026). Consequently, the absence of a significant anxiety effect reflects a context in which experience, confidence, and rational evaluation have largely supplanted anxiety as determinants of AIAT adoption. Within TAM and UTAUT frameworks, ANX is often considered a peripheral factor whose influence diminishes as users gain experience and familiarity, explaining its non-significant role in more digitally mature environments (Venkatesh et al., 2003).

TR significantly influenced ATT, consistent with prior findings (Xiong et al., 2023). When lecturers perceive AIAT as reliable, secure, and trustworthy (i.e., based on the items), they are more likely to develop a positive attitude toward integrating these tools into teaching and assessment. In MHEI, where assessment integrity and accountability are critical, TR reduces perceived uncertainty and professional risk, allowing lecturers to view AIAT as legitimate and pedagogically appropriate. Thus, TR primarily operates at the attitudinal level, shaping lecturers' cognitive and emotional evaluations. In contrast, the non-significant direct effect of TR on BI suggests that TR is insufficient to drive adoption decisions, consistent with prior research (Chuyen, 2023). Trust in AI depends on perceptions of system capability, reliability, security, and ethical use, which are critical considerations for all stakeholders in the educational ecosystem (Salloum, 2024). Lecturers often require clear evidence of AI systems' effectiveness before developing confidence in their use (Bhat et al., 2024), as AI-generated outputs may still be inaccurate or misaligned with factual information (U.S. Department of Education, 2023). Furthermore, the use of AI in T/L introduces potential safety and privacy risks, particularly regarding student data, which may be vulnerable to cyberattacks or data breaches, thereby undermining trust (Dhawan & Batra, 2020). Familiarity and prior experience are key antecedents of trust, as repeated interaction enables users to form expectations, reduce uncertainty, and evaluate system attributes such as reliability and competence (Teo, 2018). In the context of MHEI, lecturers have limited hands-on experience and knowledge with AIAT, making it more difficult to assess these tools and potentially amplifying perceived risks and uncertainty, thereby weakening trust and limiting its direct influence on adoption intentions (Kuen et al., 2023). Whereas TR indirectly affected BI through ATT, consistent with prior findings (Roh et al., 2023). Lecturers who trust AIAT are more likely to develop positive attitudes toward their use, which subsequently strengthens their adoption intentions. Thus, TR enhances intention primarily when it fosters a favourable evaluative response rather than acting as an independent driver. This is consistent with the meta-UTAUT model, where ATT functions as a critical intermediary that translates trust perceptions into BI. However, the small effect size suggests limited practical significance in this context (Hair et al., 2017).

COM had the strongest effect on ATT, supporting the DTPB model and prior findings (Nuha et al., 2018). Compatibility reduces disruption and makes technology use feel natural when it aligns with lecturers' work tasks, academic needs, assessment practices, and pedagogical goals (i.e., based on the items) rather than being

burdensome (Gupta & Bhaskar, 2020). This alignment lowers cognitive load and reduces resistance, explaining its strong positive effect on ATT (Gunnoo et al., 2023). However, COM did not directly predict BI, consistent with prior research (Yedilbayev et al., 2023). In the MHEI context, AIAT remains at an early stage, with lecturers largely overlooking these tools, resulting in limited knowledge and experience (AIBIG, 2024). Experience shapes how users perceive fit and task alignment; without familiarity or hands-on exposure, compatibility may have little effect on intention, as users struggle to evaluate how a technology integrates into their routines (Rahman et al., 2019). Consequently, lecturers may find it difficult to assess AIAT effectively, leaving many uncertain about adoption due to their limited exposure and incomplete understanding of the technology. COM exerted a significant indirect effect on BI through ATT, consistent with prior findings (Gunnoo et al., 2023). This indicates that COM primarily influences adoption via the attitudinal pathway rather than directly, particularly when users are unfamiliar with the technology. The indirect effect underscores ATT's critical role as a mediator in the adoption decision-making process, in line with the meta-UTAUT framework. Moreover, COM is a strong predictor of PE, in line with previous research (Oluyinka et al., 2021). This suggests that when lecturers perceive AIAT as aligned with their teaching and assessment practices, they are more likely to view the tools as effective for enhancing the assessment process.

PI was a key determinant of both ATT and BI, directly and indirectly through ATT, consistent with prior research (Patil et al., 2020; Strzelecki et al., 2024). In MHEIs, where AIAT remains unfamiliar and lecturers have limited knowledge and experience, PI becomes particularly important. Lecturers with high PI are more likely to experiment with AIAT, evaluate its relevance, and learn its features, thereby compensating for limited institutional guidance and fostering more positive attitudes and stronger adoption intentions (Farooq et al., 2017). Consequently, PI acts as a self-driven catalyst that enables early adopters to overcome uncertainty and barriers associated with emerging technologies (Ciftci et al., 2021). The significant indirect effect through ATT further highlights its role in shaping adoption decisions, with attitude functioning as a conduit between PI and BI, consistent with the meta-UTAUT model.

Finally, ATT emerged as the strongest predictor of lecturers' intention to adopt AIAT. This finding is consistent with TAM, TRA, TPB, DTPB, meta-UTAUT, as well as previous studies (Wang et al., 2021). Lecturers are more likely to adopt AIAT when they perceive these technologies as suitable tools for student assessment, recognize their advantages in evaluating student performance, and hold positive attitudes toward their use, as reflected in the measured items, which ultimately translate into strong adoption intentions.

## 6. Theoretical Contributions

This study extends the literature on technology adoption in HE by applying the meta-UTAUT model in the context of AIAT. The empirical analysis reveals partial support for the meta-UTAUT model and highlights the role of attitude in shaping behavioural intention. Additionally, this study develops and tests a conceptual model that incorporates several factors associated with lecturers' adoption of AIAT, some of which have not been examined in this context. In contrast to many prior studies that treat attitude as the final dependent variable, this study examines lecturers' behavioural intention as the outcome variable. Moreover, the independent variables are specified as antecedents of both attitude and behavioural intention, allowing the relationships among these variables to be examined simultaneously. The analysis further evaluates the mediating role of attitude within the meta-UTAUT framework, providing additional empirical evidence on the relationships among these variables. By examining these factors and their relationships, the study contributes to the limited literature on AIAT adoption and enhances understanding of the adoption of AI-based educational systems in HE.

## 7. Practical Contributions

To promote AIAT adoption in HE, policymakers and institutional leaders should implement a comprehensive support strategy. First, targeted workshops can enhance lecturers' understanding of AIAT features and functionalities by emphasizing practical benefits and alignment with pedagogical objectives, assessment practices, and existing workflows to strengthen perceived usefulness. Second, structured training programs, detailed user documentation, and dedicated technical support should be provided to boost lecturers' technical competence, confidence, and perceived ease of use. Third, trust-building measures are essential; these include initiatives to ensure system reliability, implementation of certified security frameworks, and establishment of clear policies for ethical and responsible AI use to mitigate associated risks. Fourth, early adopters should be identified through pilot implementations involving diverse lecturers and positioned as opinion leaders to facilitate broader acceptance.

Additionally, developers and designers must ensure that AIAT delivers accurate, pedagogically aligned assessments and maintains high usability standards; involving lecturers in the design and development process is critical to align the tools with instructional needs. Taken together, these coordinated efforts can foster positive attitudes towards AIAT and behavioural intention, supporting its effective integration into higher education practices.

## 8. Limitations and Future Research

This study has several limitations. First, the limited availability of empirical research on AIAT adoption in higher education constrains the opportunity to compare and contextualize the results. Second, although the sample of 414 participants was sufficient for statistical analysis, a larger sample would likely provide more nuanced insights, particularly if it included more lecturers from private universities. Third, data were collected at a single point in time, resulting in a cross-sectional rather than longitudinal design. Longitudinal design is generally more effective for examining causal relationships and capturing changes over time; however, the cross-sectional design effectively addressed the research objectives.

Future research could extend the model by incorporating additional AI-related factors, such as system likability, perceived intelligence, and anthropomorphism, to provide a more comprehensive understanding of adoption behavior. Moreover, expanding the sample to include lecturers from all Malaysian universities would enhance the generalizability of the findings. Targeted sampling across different institutional types could also support comparative analyses between public and private universities, revealing potential contextual differences in adoption patterns. Finally, this study examined behavioral intention rather than actual usage. Therefore, future studies should investigate actual usage behavior, as behavioral intention does not always translate into practice.

## 9. Conclusions

This research examines lecturers' attitudes and behavioural intentions toward adopting AIAT in Malaysian universities. A meta-UTAUT-based model incorporating external factors was proposed and empirically tested using survey data collected from 414 lecturers in the Klang Valley. The findings indicate that performance expectancy, effort expectancy, self-efficacy, trust, compatibility, and personal innovativeness significantly shape lecturers' attitudes towards AIAT. Moreover, performance expectancy, self-efficacy, personal innovativeness, and attitude are key determinants of adoption intention. The results further demonstrate that attitude mediates the relationship between performance expectancy, effort expectancy, self-efficacy, trust, compatibility, personal innovativeness, and the intention to adopt AIAT. These findings would provide practical guidance for fostering the effective adoption and integration of AI technologies in Malaysian higher education institutions.

## References

- Adirinekso, G. P., Purba, J. T., & Budiono, S. (2020). Measurement of performance, effort, social influence, facilitation, habit and hedonic motives toward pay later application intention: Indonesia evidence. In the Proceedings of the 2nd African International Conference on Industrial Engineering and Operations Management, Harare, Zimbabwe.
- Agarwal, R., & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research*, 9(2), 204-215.
- AIBIG. (2024). *AI spectrum bulletin: AI ethics: Balancing innovation and responsibility (3009-1950)*. (AI Spectrum Bulletin Series, Issue. U. M. Kelantan. <https://aibig.umk.edu.my/img/buletin/AI%20Spectrum%20Edition%202.pdf>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211. [https://doi.org/https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/https://doi.org/10.1016/0749-5978(91)90020-T)
- Akil, W. F. H. W. M., & Matore, M. E. E. M. (2023). The 21st century assessment strategies in medical laboratory education: sharing experiences between two higher institutions in Malaysia. *International Journal of Academic Research in Business and Social Sciences*, 13(7), 107-130. <https://doi.org/http://dx.doi.org/10.6007/IJARBSS/v13-i7/17196>

- Alfaleh, M. (2026). Sustainable AI-driven assessment in higher education: A systematic review of fairness, transparency, pedagogical innovation, and governance. *Sustainability*, 18(2), 785. <https://www.mdpi.com/2071-1050/18/2/785>
- Ali, Q. I. (2024). Towards more effective summative assessment in OBE: A new framework integrating direct measurements and technology. *Discover Education*, 3(1), 107. <https://doi.org/https://doi.org/10.1007/s44217-024-00208-5>
- Alkhowaiter, W. A. (2022). Use and behavioural intention of m-payment in GCC countries: Extending meta-UTAUT with trust and Islamic religiosity. *Journal of Innovation and Knowledge*, 7(4), 100240. <https://doi.org/https://doi.org/10.1016/j.jik.2022.100240>
- Altalhi, M. (2021). Toward a model for acceptance of MOOCs in higher education: The modified UTAUT model for Saudi Arabia. *Education and Information Technologies*, 26(2), 1589-1605. <https://doi.org/https://doi.org/10.1007/s10639-020-10317-x>
- Bahari, M., Arpacı, I., Azmi, N. F. M., & Shuib, L. (2023). Predicting the intention to use learning analytics for academic advising in higher education. *Sustainability*, 15(21), 15190. <https://doi.org/https://doi.org/10.3390/su152115190>
- Baig, M. I., & Yadegaridehkordi, E. (2025). Factors influencing academic staff satisfaction and continuous usage of generative artificial intelligence (GenAI) in higher education. *International Journal of Educational Technology in Higher Education*, 22(5). <https://doi.org/https://doi.org/10.1186/s41239-025-00506-4>
- Baker, T., Smith, L., & Anissa, N. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. Nesta. [https://media.nesta.org.uk/documents/Future\\_of\\_AI\\_and\\_education\\_v5\\_WEB.pdf](https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf)
- Balakrishnan, J., Abed, S. S., & Jones, P. (2022). The role of meta-UTAUT factors, perceived anthropomorphism, perceived intelligence, and social self-efficacy in chatbot-based services? *Technological Forecasting and Social Change*, 180, 121692. <https://doi.org/https://doi.org/10.1016/j.techfore.2022.121692>
- Bellet, T., & Banet, A. (2023). UTAUT4-AV: An extension of the UTAUT model to study intention to use automated shuttles and the societal acceptance of different types of automated vehicles. *Transportation research part F: traffic psychology and behaviour*, 99, 239-261. <https://doi.org/https://doi.org/10.1016/j.trf.2023.10.007>
- Bervell, B., Nyagorme, P., & Arkorful, V. (2020). LMS-enabled blended learning use intentions among distance education tutors: Examining the mediation role of attitude based on technology-related stimulus-response theoretical framework. *Contemporary Educational Technology*, 12(2), ep273. <https://doi.org/https://doi.org/10.30935/cedtech/8317>
- Bhat, M. A., Tiwari, C. K., Bhaskar, P., & Khan, S. T. (2024). Examining ChatGPT adoption among educators in higher educational institutions using extended UTAUT model. *Journal of Information, Communication and Ethics in Society*, 22(3), 331-353. <https://doi.org/https://doi.org/10.1108/JICES-03-2024-0033>
- Bhattacharjee, A. (2012). *Social Science Research: Principles, Methods, and Practices* (2nd ed.). University of South Florida. [https://digitalcommons.usf.edu/oa\\_textbooks/3](https://digitalcommons.usf.edu/oa_textbooks/3)
- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education* 21(4). <https://doi.org/https://doi.org/10.1186/s41239-023-00436-z>
- Braga, P. R. V., Granero, C. M. O., & Buck, E. (2024). Student and faculty perceptions of summative assessment methods in a block and blend mode of delivery. *Journal of University Teaching and Learning Practice*, 21(2). <https://doi.org/https://doi.org/10.53761/1.21.2.04>
- Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, 49(5), 1716-1735. <https://doi.org/https://doi.org/10.3758/s13428-016-0814-1>
- Camacho-Miñano, M.-d.-M., del Campo, C., Urquía-Grande, E., Pascual-Ezama, D., Akpınar, M., & Rivero, C. (2020). Solving the mystery about the factors conditioning higher education students' assessment: Finland versus Spain. *Education + Training*, 62(6), 617-630. <https://doi.org/https://doi.org/10.1108/ET-08-2019-0168>
- Chassab, R. H., Zakaria, L. Q., & Tiun, S. (2021). Automatic essay scoring: A review on the feature analysis techniques. *International Journal of Advanced Computer Science and Applications*, 12(10), 252-264. <https://doi.org/http://dx.doi.org/10.14569/IJACSA.2021.0121028>

- Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Education and Information Technologies*, 25(5), 3443-3463. <https://doi.org/https://doi.org/10.1007/s10639-020-10159-7>
- Chatterjee, S., Rana, N. P., Khorana, S., Mikalef, P., & Sharma, A. (2021). Assessing organizational users' intentions and behavior to AI integrated CRM systems: A Meta-UTAUT approach. *Information Systems Frontiers*, 25(4), 1299-1313. <https://doi.org/https://doi.org/10.1007/s10796-021-10181-1>
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-336). Psychology Press.
- Chun, T. W., & Yunus, M. M. (2023). Factors affecting Malaysian ESL teachers' behavioral intentions for technology use in the post-COVID-19 era [Brief Research Report]. *Frontiers in psychology, Volume 14 - 2023*. <https://doi.org/10.3389/fpsyg.2023.1127272>
- Chuyen, N. T. H. (2023). An empirical analysis of predictors of AI-powered design tool adoption. *TEM Journal*, 12(3), 1482-1489. <https://doi.org/https://doi.org/10.18421/TEM123-28>
- Ciftci, O., Berezina, K., & Kang, M. (2021). Effect of personal innovativeness on technology adoption in hospitality and tourism: Meta-analysis. In W. Wörndl, C. Koo, & J. L. Stienmetz, *Information and Communication Technologies in Tourism 2021* Cham.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/https://doi.org/10.4324/9780203771587>
- Compeau, D., Higgins, C. A., & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly*, 23(2), 145-158. <https://doi.org/10.2307/249749>
- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 19(2), 189-211. <https://doi.org/https://doi.org/10.2307/249688>
- Daniel Mehari Alemayehu, D. N. K. D. (2023). Evolution of technology adoption theories and models: A review of literature. *European Economic Letters (EEL)*, 13(5), 1472-1486. <https://www.eelet.org.uk/index.php/journal/article/view/932>
- Deep, P. D., Edgington, W. D., Ghosh, N., & Rahaman, M. S. (2025). Evaluating the effectiveness and ethical implications of AI detection tools in higher education. *Information*, 16(10), 905. <https://www.mdpi.com/2078-2489/16/10/905>
- Dhawan, S., & Batra, G. (2020). Artificial intelligence in higher education: Promises, perils, and perspective. *OJAS Expanding Knowledge Horizon*, 9(2), 11-22.
- Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers*, 21(3), 719-734. <https://doi.org/https://doi.org/10.1007/s10796-017-9774-y>
- Dwivedi, Y. K., Rana, N. P., Tamilmani, K., & Raman, R. (2020). A meta-analysis based modified unified theory of acceptance and use of technology (meta-UTAUT): A review of emerging literature. *Current Opinion in Psychology*, 36, 13-18. <https://doi.org/https://doi.org/10.1016/j.copsyc.2020.03.008>
- Elatrachi, M., & Oukarfi, S. (2020). Determinants of ICT integration by teachers in higher education in Morocco. In A. Kavoura, E. Kefallonitis, & P. Theodoridis, *Strategic Innovative Marketing and Tourism* Cham.
- Enang, E., & Christopoulou, D. (2025). Exploring academic's intentions to incorporate ChatGPT into their teaching practice. *Journal of University Teaching and Learning Practice*, 21(08). <https://doi.org/https://doi.org/10.53761/rn5y5614>
- Fan, W., Liu, J., Zhu, S., & Pardalos, P. M. (2018). Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research*, 294(1), 567-592. <https://doi.org/https://doi.org/10.1007/s10479-018-2818-y>
- Farooq, M. S., Salam, M., Jaafar, N., Fayolle, A., Ayupp, K., Radovic-Markovic, M., & Sajid, A. (2017). Acceptance and use of lecture capture system (LCS) in executive business studies. *Interactive Technology and Smart Education*, 14(4), 329-348. <https://doi.org/https://doi.org/10.1108/ITSE-06-2016-0015>
- Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley. <http://lib.ugent.be/catalog/rug01:000839739>
- Flodén, J. (2024). Grading exams using large language models: A comparison between human and AI grading of exams in higher education using ChatGPT. *British Educational Research Journal*, 51(1), 201-224. <https://doi.org/https://doi.org/10.1002/berj.4069>

- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.  
<https://doi.org/10.1177/002224378101800104>
- Ghimire, A., Imran, M. A. U., Biswas, B., Tiwari, A., & Saha, S. (2024). Behavioral intention to adopt artificial intelligence in educational institutions: A hybrid modeling approach. *Journal of Computer Science and Technology Studies*, 6(3), 56-64. <https://doi.org/10.32996/jcsts.2024.6.3.6>
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, 11(12).  
<https://doi.org/https://doi.org/10.3390/app11125467>
- Gunasinghe, A., Hamid, J. A., Khatibi, A., & Azam, S. M. F. (2020). The adequacy of UTAUT-3 in interpreting academicians' adoption to e-Learning in higher education environments. *Interactive Technology and Smart Education*, 17(1), 86-106. <https://doi.org/https://doi.org/10.1108/ITSE-05-2019-0020>
- Gunnoo, L., Subadar, U. A., & Fauzel, S. (2023). The mediating effect of attitude on adoption of mobile payment services in small island developing states. The case of Mauritius. *Social Sciences & Humanities Open*, 8(1), 100724. <https://doi.org/https://doi.org/10.1016/j.ssaho.2023.100724>
- Gupta, K. P., & Bhaskar, P. (2020). Inhibiting and motivating factors influencing teachers' adoption of AI-based teaching and learning solutions: Prioritization using analytic hierarchy process. *Journal of Information Technology Education: Research*, 19, 693-723. <https://doi.org/https://doi.org/10.28945/4640>
- Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd ed.). Sage Publications.  
[https://eli.johogo.com/Class/CCU/SEM/\\_A%20Primer%20on%20Partial%20Least%20Squares%20Structural%20Equation%20Modeling\\_Hair.pdf](https://eli.johogo.com/Class/CCU/SEM/_A%20Primer%20on%20Partial%20Least%20Squares%20Structural%20Equation%20Modeling_Hair.pdf)
- Hakim, M. M., Afifah, A. N., & Aryotejo, G. (2023). The analysis of factors affecting behavioral intention and behavior usage of e-wallet using Meta-UTAUT model. *International Journal on Advanced Science, Engineering and Information Technology*, 13(2), 786-793. <https://doi.org/10.18517/ijaseit.13.2.18297>
- Harman, H. H. (1976). *Modern Factor Analysis* (3rd ed.). The University of Chicago Press.
- Helmiatin, Hidayat, A., & Kahar, M. R. (2024). Investigating the adoption of AI in higher education: A study of public universities in Indonesia. *Cogent Education*, 11(1).  
<https://doi.org/https://doi.org/10.1080/2331186X.2024.2380175>
- Henderson, J., & Corry, M. (2021). Teacher anxiety and technology change: A review of the literature. *Technology, Pedagogy and Education*, 30(4), 573-587.  
<https://doi.org/https://doi.org/10.1080/1475939X.2021.1931426>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115-135.  
<https://doi.org/10.1007/s11747-014-0403-8>
- Hermanto, A. H., Windasari, N. A., & Purwanegara, M. S. (2022). Taxpayers' adoption of online tax return reporting: Extended meta-UTAUT model perspective. *Cogent Business & Management*, 9(1), 2110724.  
<https://doi.org/https://doi.org/10.1080/23311975.2022.2110724>
- Herodotou, C., Maguire, C., Hlosta, M., & Mulholland, P. (2023, Mar. 2023). Predictive learning analytics and university teachers: Usage and perceptions three years post implementation. LAK23: 13th International Learning Analytics and Knowledge Conference, Arlington, TX, USA.
- Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health*, 31(2), 180-191. <https://doi.org/10.1002/nur.20247>
- Hossain, A., Quaresma, R., & Rahman, H. (2019). Investigating factors influencing the physicians' adoption of electronic health record (EHR) in healthcare system of Bangladesh: An empirical study. *International Journal of Information Management*, 44, 76-87.  
<https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2018.09.016>
- Huseynov, F., & Özkan Yıldırım, S. (2019). Online consumer typologies and their shopping behaviors in B2C e-commerce platforms. *SAGE Open*, 9(2). <https://doi.org/https://doi.org/10.1177/2158244019854639>
- IIUM. (2015). *Student academic performance evaluation (undergraduate) regulations 2015*. I. I. U. Malaysia.  
<https://kulliyah.iium.edu.my/koed/wp-content/uploads/sites/23/2025/10/Student-Academic-Performance-2015-SAPER.pdf>

- Ishaq, K., Rana, A. M. K., & Zin, N. A. M. (2020). Exploring summative assessment and effects: Primary to higher education. *Bulletin of Educational Research*, 42(3), 23-50.
- JPT. (2022). *Higher education institutions*. Jabatan Pendidikan Tinggi, Kementerian Pendidikan Tinggi, Malaysia. Retrieved Sep. 19 from <https://jpt.mohe.gov.my/portal/index.php/en/hei>
- Kai, C., Ping, W., & Xiaomin, J. (2026). AI anxiety and adoption intention in higher education based on an extended TAM-UTAUT and PLS-SEM analysis. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-35823-9>
- Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, 15(16), 12451. <https://www.mdpi.com/2071-1050/15/16/12451>
- Kannan, R. J., Punithavathi, P., & Sambandam, N. (2018, Nov. 2018). Technology adoption models – adoption of ICT in educational institutions in India. In Proceedings of the 2018 World Engineering Education Forum-Global Engineering Deans Council (WEEF-GEDC), Albuquerque, NM, USA.
- Khlaif, Z., Odeh, A., & Bsharat, T. R. K. (2024). Generative AI-powered adaptive assessment. In M. Sanmugam, D. Lim, N. M. Mohd Barkhaya, W. A. J. Wan Yahaya, & Z. Khlaif (Eds.), *Power of Persuasive Educational Technologies in Enhancing Learning* (pp. 157-176). IGI Global. <https://doi.org/10.4018/979-8-3693-6397-3.ch007>
- Kim, K. K., & Prabhakar, B. (2004). Initial trust and the adoption of B2C e-commerce: the case of internet banking. *Data Base for Advances in Information Systems*, 35(2), 50-64. <https://doi.org/10.1145/1007965.1007970>
- Kleine, A.-K., Kokje, E., Lerner, E., & Gaube, S. (2023). Attitudes toward the adoption of 2 artificial intelligence-enabled mental health tools among prospective psychotherapists: Cross-sectional study. *JMIR Human Factors*, 12(10), e46859. <https://doi.org/10.2196/46859>
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (T. D. Little, Ed. 4th ed.). Guilford Press. <https://books.google.com.my/books?id=EkMVZUxZrgIC>
- Kothari, C. R. (2004). *Research Methodology: Methods and Techniques* (2nd ed.). New Age International.
- Kuen, L., Westmattmann, D., Bruckes, M., & Schewe, G. (2023). Who earns trust in online environments? A meta-analysis of trust in technology and trust in provider for technology acceptance. *Electronic Markets*, 33(1), 61. <https://doi.org/10.1007/s12525-023-00672-1>
- Kwak, Y., Ahn, J.-W., & Seo, Y. H. (2022). Influence of AI ethics awareness, attitude, anxiety, and self-efficacy on nursing students' behavioral intentions. *BMC Nurs.*, 21(1), 267. <https://doi.org/https://doi.org/10.1186/s12912-022-01048-0>
- Kwak, Y., Seo, Y. H., & Ahn, J.-W. (2022). Nursing students' intent to use AI-based healthcare technology: Path analysis using the unified theory of acceptance and use of technology. *Nurse Education Today*, 119, 105541. <https://doi.org/https://doi.org/10.1016/j.nedt.2022.105541>
- Lee, C. T., & Pan, L.-Y. (2023). Smile to pay: Predicting continuous usage intention toward contactless payment services in the post-covid-19 era. *International Journal of Bank Marketing*, 41(2), 312-332. <https://doi.org/https://doi.org/10.1108/IJBM-03-2022-0130>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-386. <https://doi.org/https://doi.org/10.1097/00006199-198611000-00017>
- Machicao, J. C. (2019, Jan. 2019). Higher education challenge characterization to implement automated essay scoring model for universities with a current traditional learning evaluation system. In Á. Rocha, C. Ferrás, & M. Paredes, *Information Technology and Systems* Cham.
- Mansour, T. M., & Mansour, H. T. (2025). Employing digital marketers' use of artificial intelligence in Jordanian construction companies: An analytical study within the utaut framework. *An-Najah University Journal for Research-B (Humanities)*, 40(3).
- Meidute-Kavaliauskiene, I., Čiğdem, Ş., Yıldız, B., & Davidavicius, S. (2021). The Effect of Perceptions on Service Robot Usage Intention: A Survey Study in the Service Sector. *Sustainability*, 13(17), 9655. <https://www.mdpi.com/2071-1050/13/17/9655>
- MOE. (2018). *Malaysian education blueprint 2013-2025* (Annual Reports, Issue. M. o. Education. <https://www.moe.gov.my/storage/files/shares/Dasar/PPPM/MEB%20Annual%20Report%202018.pdf>

- Mohamad Mozie, N., Ghazali, N., & Ali Husin, L. I. (2025). Adopting artificial intelligence in higher education: insights from the UTAUT framework on student intentions. *Advances in Business Research International Journal*, 11(1), 71-79.
- Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2(3), 192-222. <https://doi.org/https://doi.org/10.1287/isre.2.3.192>
- MOSTI. (2021). *Malaysia national artificial intelligence roadmap 2021-2025 (AI-RMAP 2021-2025)*. T. a. I. Ministry of Science. <https://mastic.mosti.gov.my/publication/artificial-intelligence-roadmap-2021-2025/>
- Nadzirah Muhammad, M., Sri Fatiany Abdul Kader, J., Syukrina Alini Mat, A., & Noor'ain Mohamad, Y. (2025). AI integration in TVET: Examining educators' behavioral intentions and the moderating role of experience. *Information Management and Business Review*, 17(2(I)S). [https://doi.org/10.22610/imbr.v17i2\(I\)S.4617](https://doi.org/10.22610/imbr.v17i2(I)S.4617)
- Naji, M. G., Elias, N. F., Lam, M. C., Abusamhadana, G. A. O., & Ahmad, N. A. (2022). Big data analytics adoption at higher education institutions in Iraq. *Journal of Theoretical and Applied Information Technology*, 100(9), 3157-3170.
- National AI Office. (2021). *MyDIGITAL Corporation*. National AI Office (NAIO). Retrieved January, 18 from [https://ai.gov.my/faq/talent-development-and-education?utm\\_source=chatgpt.com](https://ai.gov.my/faq/talent-development-and-education?utm_source=chatgpt.com)
- National AI Office. (2025). *Building an AI nation to gether, now*. National AI Office (NAIO). Retrieved January, 18 from <https://ai.gov.my/>
- Nuha, A., Gary, W., & Mike, W. (2018). An evaluation of the model of acceptance of e-assessment among academics in Saudi universities. *Education Journal*, 7(2), 23-36. <https://doi.org/https://doi.org/10.11648/j.edu.20180702.11>
- Oluyinka, S., Endozo, A. N., & Cusipag, M. N. (2021). Integrating trialability and compatibility with UTAUT to assess canvas usage during covid-19 quarantine period. *Asia-Pacific Social Science Review*, 21(2), 31-47. <https://doi.org/https://doi.org/10.59588/2350-8329.1366>
- Onesi-Ozigagun, O., Ololade, Y. J., Eyo-Udo, N. L., & Ogundipe, D. O. (2024). Revolutionizing education through AI: A comprehensive review of enhancing learning experiences. *International Journal of Applied Research in Social Sciences*, 6(4), 589-607. <https://doi.org/10.51594/ijarss.v6i4.1011>
- Osman, Z., Alwi, N. H., Jodi, K. H. M., Khan, B. N. A., Ismail, M. N., & Yusoff, Y. (2024). Optimizing artificial intelligence usage among academicians in higher education institutions. *International Journal of Academic Research in Accounting Finance and Management Sciences*, 14(2), 1-19. <https://doi.org/https://doi.org/10.6007/IJARAFMS/v14-i2/20935>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253. <https://doi.org/https://doi.org/10.1518/001872097778543886>
- Patil, P., Tamilmani, K., Rana, N. P., & Raghavan, V. (2020). Understanding consumer adoption of mobile payment in India: Extending Meta-UTAUT model with personal innovativeness, anxiety, trust, and grievance redressal. *International Journal of Information Management*, 54, 102144. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2020.102144>
- Pedersen, M. J., & Nielsen, C. V. (2014). Improving survey response rates in online panels: Effects of low-cost incentives and cost-free text appeal interventions. *Social Science Computer Review*, 34(2), 229-243. <https://doi.org/10.1177/0894439314563916>
- Perez, R. C. L. (2024). AI in higher education: Faculty perspective towards artificial intelligence through UTAUT approach. *Ho Chi Minh City Open University Journal of Science-Social Sciences*, 14(4), 32-50. <https://doi.org/https://doi.org/10.46223/HCMCOUJS.soci.en.14.4.2851.2024>
- Pillai, R., Sivathanu, B., & Dwivedi, Y. K. (2020). Shopping intention at AI-powered automated retail stores (AIPARS). *Journal of Retailing and Consumer Services*, 57, 102207. <https://doi.org/https://doi.org/10.1016/j.jretconser.2020.102207>
- Pinto, A. S., Abreu, A., Costa, E., & Paiva, J. (2023). How machine learning (ML) is transforming higher education: A systematic literature review. *Journal of Information Systems Engineering and Management*, 8(2), 21168. <https://doi.org/10.55267/iadt.07.13227>
- Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, 12(4), 531-544. <https://doi.org/https://doi.org/10.1177/014920638601200408>

- Prakash, A. V., & Das, S. (2021). Medical practitioner's adoption of intelligent clinical diagnostic decision support systems: A mixed-methods study. *Information & Management*, 58(7), 103524. <https://doi.org/https://doi.org/10.1016/j.im.2021.103524>
- Rahman, M. M., Deb, S., Strawderman, L., Burch, R., & Smith, B. (2019). How the older population perceives self-driving vehicles. *Transportation research part F: traffic psychology and behaviour*, 65, 242-257. <https://doi.org/https://doi.org/10.1016/j.trf.2019.08.002>
- Raosoft, I. (2004). *Raosoft sample size calculator*. Retrieved Jul. 10 from <http://www.raosoft.com/samplesize.html>
- Rigopoulos, G. (2022). Assessment and feedback as predictors for student satisfaction in UK higher education. *International Journal of Modern Education and Computer Science*, 14(5), 1-9. <https://doi.org/https://doi.org/10.5815/ijmecs.2022.05.01>
- Roh, T., Park, B. I., & Xiao, S. (2023). Adoption of AI-enabled robo-advisors in fintech: Simultaneous employment of UTAUT and the theory of reasoned action. *Journal of Electronic Commerce Research*, 24(1), 29-47.
- Roy, S. K., Singh, G., & Shabnam, S. (2021). Modelling customer engagement behaviour in smart retailing: Modelling customer engagement behaviour. *Australasian Journal of Information Systems*, 25. <https://doi.org/https://doi.org/10.3127/ajis.v25i0.2967>
- Salloum, S. A. (2024). Trustworthiness of the AI. In A. Al-Marzouqi, S. A. Salloum, M. Al-Saidat, A. Aburayya, & B. Gupta (Eds.), *Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom* (pp. 643-650). Springer. [https://doi.org/10.1007/978-3-031-52280-2\\_41](https://doi.org/10.1007/978-3-031-52280-2_41)
- Saman, H. M., Noor, S. M., Isa, C. M. M., Lian, O. C., & Narayanan, G. (2024, 2024-11-30T16:00:00.000Z). Embracing Artificial Intelligence as a catalyst for change in reshaping Malaysian higher education in the digital era: A literature review.
- Sánchez-Prieto, J. C., Cruz-Benito, J., Therón, R., & García-Peñalvo, F. J. (2019). How to measure teachers' acceptance of AI-driven assessment in eLearning: A TAM-based proposal. In the Proceedings of the 7th International Conference on Technological Ecosystems for Enhancing Multiculturality, León, Spain.
- Sarker, P., Hughes, D. L., & Dwivedi, Y. K. (2020, May. 2020). Extension of META-UTAUT for examining consumer adoption of social commerce: Towards a conceptual model. In F. J. Martínez-López & S. D'Alessandro, *Advances in Digital Marketing and eCommerce (DMEC 2020)*, Cham.
- Sekaran, U., & Bougie, R. (2016). *Research Methods For Business: A Skill Building Approach* (7th ed.). John Wiley & Sons.
- Seo, K. H., & Lee, J. H. (2021). The emergence of service robots at restaurants: Integrating trust, perceived risk, and satisfaction. *Sustainability*, 13(8), 4431. <https://www.mdpi.com/2071-1050/13/8/4431>
- Shahid, M. K., Zia, T., Bangfan, L., Iqbal, Z., & Ahmad, F. (2024). Exploring the relationship of psychological factors and adoption readiness in determining university teachers' attitude on AI-based assessment systems. *The International Journal of Management Education*, 22(2), 100967. <https://doi.org/https://doi.org/10.1016/j.ijme.2024.100967>
- Smolansky, A., Cram, A., Radulescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). *Educator and student perspectives on the impact of Generative AI on assessments in higher education* Proceedings of the Tenth ACM Conference on Learning, Copenhagen, Denmark. <https://doi.org/10.1145/3573051.3596191>
- So, S., Ismail, M. R., & Jaafar, S. (2021). Exploring acceptance of artificial intelligence amongst healthcare personnel: a case in a private medical centre. *Int. J. Adv. Eng. Manag*, 3, 56-65.
- Song, J. (2024). Integrating artificial intelligence in smart course design: Innovative teaching methods for talent cultivation in higher education. *Education Insights*, 1(1), 10-16. <https://doi.org/10.70088/63t0zj06>
- Sporrong, E., McGrath, C., & Cerratto Pargman, T. (2025). Situating AI in assessment—an exploration of university teachers' valuing practices. *AI and Ethics*, 5(3), 2381-2394. <https://doi.org/10.1007/s43681-024-00558-8>
- Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., & Vasileva, T. (2025). AI-driven adaptive learning for sustainable educational transformation. *Sustainable Development*, 33(2), 1921-1947. <https://doi.org/https://doi.org/10.1002/sd.3221>
- Strzelecki, A., Cicha, K., Rizun, M., & Rutecka, P. (2024). Acceptance and use of ChatGPT in the academic community. *Education and Information Technologies*, 29(17), 22943-22968. <https://doi.org/https://doi.org/10.1007/s10639-024-12765-1>

- Sue, V., & Ritter, L. (2012). *Conducting Online Surveys* (2nd ed.). Sage Publications.  
<https://doi.org/https://doi.org/10.4135/9781506335186>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Pearson Education.  
<https://books.google.com.my/books?id=ucj1ygAACAAJ>
- Taherdoost, H. (2016). Sampling methods in research methodology; How to choose a sampling technique for research. *International Journal of Academic Research in Management*, 5(2), 18-27.  
<https://doi.org/https://doi.org/10.2139/ssrn.3205035>
- Tamilmani, K., Rana, N. P., Nunkoo, R., Raghavan, V., & Dwivedi, Y. K. (2022). Indian travellers' adoption of Airbnb platform. *Information Systems Frontiers*, 24(1), 77-96.  
<https://doi.org/https://doi.org/10.1007/s10796-020-10060-1>
- Taylor, S., & Todd, P. (1995). Decomposition and crossover effects in the Theory of Planned Behavior: A study of consumer adoption intentions. *International Journal of Research in Marketing*, 12(2), 137-155.  
[https://doi.org/https://doi.org/10.1016/0167-8116\(94\)00019-K](https://doi.org/https://doi.org/10.1016/0167-8116(94)00019-K)
- Teo, T. (2018). Students and teachers' intention to use technology: Assessing their measurement equivalence and structural invariance. *Journal of Educational Computing Research*, 57(1), 201-225.  
<https://doi.org/https://doi.org/10.1177/0735633117749430>
- Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal Computing: Toward a Conceptual Model of Utilization. *MIS Quarterly*, 15(1), 125-143. <https://doi.org/10.2307/249443>
- U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and future of teaching and learning: Insights and recommendations*. <https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf>
- Upadhyay, N., Upadhyay, S., Abed, S. S., & Dwivedi, Y. K. (2022). Consumer adoption of mobile payment services during COVID-19: Extending meta-UTAUT with perceived severity and self-efficacy. *International Journal of Bank Marketing*, 40(5), 960-991. <https://doi.org/https://doi.org/10.1108/IJBM-06-2021-0262>
- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273-315. <https://doi.org/http://dx.doi.org/10.1111/j.1540-5915.2008.00192.x>
- Venkatesh, V., & Davis, F. A. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies *Management science*, 46(2), 186-204. <https://ssrn.com/abstract=4062393>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478. <https://doi.org/https://doi.org/10.2307/30036540>
- Wang, L., & Dai, X. (2020). Exploring factors affecting the adoption of mobile payment at physical stores. *International Journal of Mobile Communications*, 18(1), 67-82.  
<https://doi.org/https://doi.org/10.1504/IJMC.2020.104420>
- Wang, S., Yu, H., Hu, X., & Li, J. (2020). Participant or spectator? Comprehending the willingness of faculty to use intelligent tutoring systems in the artificial intelligence era. *British Journal of Educational Technology*, 51(5), 1657-1673. <https://doi.org/https://doi.org/10.1111/bjet.12998>
- Wang, Y., Liu, C., & Tu, Y.-F. (2021). Factors affecting the adoption of AI-based applications in higher education: An analysis of teachers perspectives using structural equation modeling. *Educational Technology & Society*, 24(3), 116-129.
- Wanner, J., Herm, L.-V., Heinrich, K., & Janiesch, C. (2022). The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study. *Electronic Markets*, 32(4), 2079-2102.  
<https://doi.org/https://doi.org/10.1007/s12525-022-00593-5>
- Wiafe, I., Koranteng, F. N., Tettey, T., Kastriku, F. A., & Abdulai, J.-D. (2020). Factors that affect acceptance and use of information systems within the maritime industry in developing countries: The case of Ghana. *Journal of Systems and Information Technology*, 22(1), 21-45. <https://doi.org/https://doi.org/10.1108/JSIT-06-2018-0091>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), 40. <https://doi.org/10.1186/s41239-024-00468-z>
- Xiong, Y., Shi, Y., Pu, Q., & Liu, N. (2023). More trust or more risk? User acceptance of artificial intelligence virtual assistant. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 34(3), 190-205.  
<https://doi.org/https://doi.org/10.1002/hfm.21020>

- Yang, Q., Al Mamun, A., Hayat, N., Jingzu, G., Hoque, M. E., & Salameh, A. A. (2022). Modeling the intention and adoption of wearable fitness devices: A study using SEM-PLS analysis. *Frontiers in Public Health*, 10, 918989. <https://doi.org/https://doi.org/10.3389/fpubh.2022.918989>
- Ye, X., & Manoharan, S. (2018). Machine learning techniques to automate scoring of constructed-response type assessments. In the Proceedings of the 28th EAEEIE Annual Conference, Hafnarfjörður, Iceland.
- Yedilbayev, Y., Sarybayeva, A., Zharylgapova, D., Shektibayev, N., Usembayeva, I., & Kurbanbekov, B. (2023). Factors influencing future physics teachers' acceptance of information and communicative competence technologies: A survey study. *Cogent Education*, 10(1), 2212119. <https://doi.org/https://doi.org/10.1080/2331186X.2023.2212119>
- Zaim, M., Arsyad, S., Waluyo, B., Ardi, H., Al Hafizh, M., Zakiyah, M., Syafitri, W., Nusi, A., & Hardiah, M. (2024). AI-powered EFL pedagogy: integrating Generative AI into university teaching preparation through UTAUT and activity theory. *Computers & Education: Artificial Intelligence*, 7, 100335. <https://doi.org/https://doi.org/10.1016/j.caeai.2024.100335>
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021(1), 8812542.



# Why Communication Dimensions Matter: Change Readiness Under Workplace Stressors In SMEs

Matea Cvjetković<sup>1\*</sup>

<sup>1</sup>Institute of Public Finance, Zagreb, Croatia

\*Correspondence: [matea.cvjetkovic@ijf.hr](mailto:matea.cvjetkovic@ijf.hr)

## PAPER INFO

### Paper history:

Received 09 February 2026

Accepted 02 April 2026

### Citation:

Cvjetković, M. (2026). Why Communication Dimensions Matter: Change Readiness Under Workplace Stressors In SMEs. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 157-175

### Copyright:

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

The purpose of this study was to examine the role of internal communication satisfaction dimensions in shaping employee readiness for change in small and medium-sized enterprises (SMEs) in Croatia. Although internal communication is important for employee readiness for change, there is still little empirical evidence showing how the flow of information within organizations relates to that readiness. To address this gap, the study observed internal communication satisfaction through its eight dimensions and tested their associations with readiness for change, while accounting for information overload and work demands as information processing constraints, and perceived quality of change communication as a potential mediator. The resulting model was tested using 32 hypotheses, based on survey data from 454 employees in Croatian SMEs. Direct and moderating effects were examined using regression-based analyses. Mediation effects were evaluated with structural equation modelling. The results showed that three internal communication dimensions were significantly related to employee readiness for change: satisfaction with information about the organization, communication in meetings and communication climate. Other dimensions did not show significant associations. Moderation roles of information overload and work demands were confirmed for several dimensions, while mediating effects were not confirmed. The findings offer theoretical and practical contributions.

**Keywords:** Internal communication dimensions, readiness for change, information processing constraints, SMEs

## 1. Introduction

Organizations now well recognize the importance of internal communication at all levels (Tkalac Verčič, 2025). It is a core organizational process that enables information flows, sensemaking, and coordinated action across all hierarchical levels and functional units (Yue et al., 2025). By using different modes of exchange of information, such as face-to-face communication for building trust, boosting collaboration, and resolving conflicts, or digital channels to transfer ideas fast to several recipients, internal communication enables co-worker interaction with the goal of achieving organizational objectives (Santoso et al., 2022). The growing significance of studying internal communication in theory and practice has been impacted by numerous elements, the main being the implementation of new information and communication technologies, which have shifted workplace dynamics into interconnected, open, and transparent environments (Yue et al., 2023). To keep track of all the company's processes, to transcend the vision, mission, and future plans, to unite all the employees and make them want to stay, to boost their motivation, engagement, and commitment, companies need to implement a strategic approach to internal communication (Cvjetković, 2024). This

particularly refers to small and medium-sized enterprises (SMEs) whose internal communication processes are informal and less structured, mostly due to the lack of human and financial resources (Piwowar-Sulej et al., 2024). This can impact their organizational information processing capacity, make them vulnerable to external pressures and affect the successful adoption of change initiatives. Since SMEs are important for their national economies because they create new jobs, contribute to innovation development and to the overall GDP (Kazakov et al., 2021), they have been intensively studied from different perspectives such as entrepreneurship, market orientation, and business performance (e.g., Rincón Díaz & Albors Garrigós, 2017; Maseda et al., 2019; Šlogar et al., 2023). Existing research consistently shows that internal market-oriented behavior, that is, organizational practices focused on satisfying employees' needs, is linked to stronger performance in SMEs. This highlights the importance of investing into organization-employee relationships to achieve long-term competitiveness (Kazakov et al., 2021).

This study builds on earlier findings showing that internal communication satisfaction (ICS), treated as a single construct, is positively related to employees' readiness for change, both directly and through employees' perceptions of change communication quality (Cvjetković, 2024; Sun et al., 2021). Results further showed moderating effects of information overload in a way that it weakened the positive effect of internal communication satisfaction on readiness for change, whereas work demands did not significantly alter this relationship. Even though these findings provided initial evidence of the relevance of internal communication in change processes, they cannot distinguish whether all communication dimensions contribute equally to readiness for change. Similarly, it cannot be differentiated whether the observed effects of information overload, work demands, or perception of change quality communication apply equally for all eight ICS dimensions. Moreover, studies that empirically tested the effect of internal communication on employee change readiness in SMEs are limited. Existing studies mostly focused on larger firms (Yazici, 2002; Ahmad & Huvila, 2019; Sun et al., 2021) or specific industries (Robson & Tourish, 2005; Stoltenkamp & Kasuto, 2011), revealing the gap in the current literature (Abrantes et al., 2024; Cvjetković, 2024).

Given the above, the present study examines the associations between eight dimensions of internal communication on employee readiness for change (RFC) in small and medium-sized enterprises in Croatia. Internal communication is observed through the construct of Internal Communication Satisfaction (ICS), defined by Tkalac Verčič et al. (2009) as an important outcome of internal communication routines. It represents collective thoughts and feelings regarding various organizational relationships (Clampitt & Girard, 1993) at the individual, team, or organizational level (Tsai, Chuang, & Hsieh, 2009). The eight dimensions observed are: satisfaction with feedback, satisfaction with communication with immediate superior, satisfaction with horizontal communication, satisfaction with informal communication, satisfaction with information about the organization, satisfaction with communication climate, satisfaction with communication in meetings, and satisfaction with quality of communication media. Conceptually, these dimensions represent complementary aspects of organizational information systems. Feedback, supervisor communication, and horizontal communication refer to vertical and lateral information flows. Information about the organization and communication climate relate to the quality and openness of organizational communication. Communication media reflect the effectiveness of available channels, while informal communication mirrors emergent information networks.

The research offers several theoretical contributions to corporate communication, organizational information systems, and behavioral economics literature. First, it advances current insights on the impact of internal communications on employee change readiness by (1) observing it through its eight dimensions that represent different organizational information processes and (2) by positioning the research in the neglected context of small and medium-sized enterprises. Second, by taking into account workplace stressors as moderators and quality of change communication as a mediator, the study positions itself in the contemporary workplace arena and links change readiness to information processing constraints and the role of change-specific communication. It also recognizes that communication activities do not operate in a vacuum, but instead, their effects depend on the different pressures employees face (O'Connor et al., 2018). By testing 32 proposed hypotheses, the results, thus, bring new knowledge to the organizational change and organizational communication theories. As Piwowar-Sulej et al. (2024) note, the post-pandemic, rapid digitalization of business processes raises new questions about whether current theories on internal communication and change implementation still hold. Recent studies show that employees' internal communication preferences have changed and shifted towards communication that is more digital, interactive, and adjusted to their needs (Tkalac Verčič & Špoljarić, 2023). This new trend is now being recognized as hyper-personalization, and it has already been proclaimed as the future of workplace communication (Aspland, 2025). This suggests that companies may need to consider adjusting communication patterns to be more targeted and employee-specific. But, in order to fully utilize the power of internal communication, key people in SMEs need to be aware of what dimensions of internal communication contribute the most to successful change readiness.

The paper has six sections. Following this introduction, a literature review gives more detail on the theoretical approach and develops the hypotheses. Methodology presents sampling, data collection and processing, while the results section reports the findings of hypothesis testing. A discussion follows, together with theoretical and practical contributions, followed by research limitations and future recommendations. In the end conclusion summarizes the main findings.

## 2. The Role of Internal Communication in Change Management

Internal communication can be defined as a cooperative process between the employees of one organization aimed at achieving a mutually beneficial outcome and cultivating good relationships (Yue et al., 2025). It also represents a structured system of information flows through which employees receive, interpret, and exchange task-related and other important information. It comprises formal and informal communications, in upward, downward, and horizontal directions, with a focus on building a holistic and ethical communication network that considers the needs of every employee. The positive perception of internal communication practices is defined as internal communication satisfaction (Tkalac Verčič, 2021). By actively participating in the organizational communication processes, employees satisfy their personal needs for involvement (Anderson & Martin, 1995). In that way, they build good relationships with their co-workers and increase their intrinsic motivation to work (Gray & Laidlaw, 2004).

When internal communication fails to meet psychological and relational needs, the consequences can be significant. The outcomes of poor organizational communications (expressed as low internal communication satisfaction) include lower employee commitment, absenteeism, larger turnovers, and lower productivity (Jiang & Probst, 2014). At the individual level, poor internal communication can result in greater insecurity, increased stress, and burnout (Gray & Laidlaw, 2004). These negative effects can also transfer to change management processes. The doubt or uncertainty in new initiatives that arise due to the lack of change-related information can cause anxiety among employees and diminish their readiness for change. These behaviors can be positioned within Information Processing Theory (Galbraith, 1974), which perceives employees as cognitive agents with limited capacity to absorb and utilize organizational information. From this perspective, different dimensions of internal communication represent distinct organizational information structures that either support or constrain employees' capacity to process change-related information. The challenge-hindrance framework (LePine et al., 2005) serves as a complementary lens, explaining how workplace stressors interact with these communication structures and affect employees' reactions. When changes occur, employees make sense of it through communication. They also evaluate the impact of change on their own job position and their future within the company (Men et al., 2020; Sun et al., 2021). Transparent and open communication builds trust and reduces uncertainty, which in turn encourages greater openness to change (Yue et al., 2020; Li et al., 2021). This is especially relevant for SMEs because their internal culture is characterized by closer interpersonal relationships, as well as informal and direct communication patterns (Farr-Wharton & Brunetto, 2007).

Previous studies observing the internal communication in SMEs are limited and have focused on examining it either as a single construct (Fritzenschaft, 2014), or through its few isolated dimensions (Farr-Wharton & Brunetto, 2007; Arnaout & Esposito, 2018). Other studies conducted on organizations of different sizes have also mainly examined separate elements of internal communication, such as supervisor communication, communication climate, corporate information, or informal information exchange, and how they relate to various change outcomes (Rogiest et al., 2015; Rafferty & Restubog, 2016; Li et al., 2021). Studies most often focused on satisfaction with communication with superiors, emphasizing its importance in developing trust and encouraging positive employee reactions to change (Farr-Wharton & Brunetto, 2007; Men et al., 2020; Yue et al., 2023). The communication climate has also been frequently analyzed, showing that open surroundings increase affective commitment and boost change-supportive outcomes (Neill et al., 2020).

Other ICS dimensions have received far less attention. Studies on horizontal communication show inconsistent results, ranging from weak positive effects on readiness to change to more considerable links between peer relationships and employee support for change (Cunningham et al., 2002; Madsen et al., 2005). Informal communication has mainly been discussed indirectly, mostly through rumors, which have been proven to increase anxiety or resistance during transition periods (Rafferty & Restubog, 2016; Jiang et al., 2020). In studies on transparent corporate information, positive effects on reducing uncertainty and building employee trust have been demonstrated, especially during turbulent times such as the COVID-19 pandemic (Li et al., 2021). Satisfaction with communication on meetings has been linked with improved collective problem-solving and active participation in change initiatives (Ellis et al., 2023). Finally, communication media has shown to be important in how employees perceive the credibility of information, where personal

and direct communication channels have shown stronger positive effects (Men et al., 2022). These discrepancies across previous findings further motivate a multi-dimensional approach that independently tests each ICS sub-dimension, especially in small and medium-sized enterprises as little is known empirically about their communication practices, particularly when managing change. Taking the indicated into account, the first set of hypotheses is defined as:

H1.1: Satisfaction with feedback is positively associated with readiness for change.

H1.2: Satisfaction with communication with one's immediate superior is positively associated with readiness for change.

H1.3: Satisfaction with horizontal communication is positively associated with readiness for change.

H1.4: Satisfaction with informal communication is positively associated with readiness for change.

H1.5: Satisfaction with information about the organization is positively associated with readiness for change.

H1.6: Satisfaction with the communication climate is positively associated with readiness for change.

H1.7: Satisfaction with the quality of communication media is positively associated with readiness for change.

H1.8: Satisfaction with communication in meetings is positively associated with readiness for change.

Even though the defined hypotheses presume the positive association of all eight ICS dimensions with RFC, it is expected that not all dimensions will contribute equally to this outcome. Information Processing Theory (Galbraith, 1974) explains why. Some dimensions carry information about the broader organizational context, including priorities, goals, expectations and overall business direction. This makes them more proximal to readiness for change. Other dimensions are more pointed to relational and task coordinating functions and are therefore less directly tied to organization-wide change processes. This distinction is especially relevant in SMEs, due to their flatter hierarchies and informal internal structures. On this basis, organizational-level ICS dimensions are expected to show stronger associations with RFC than interpersonal ones. This is supported by previous research using the same instrument, which showed that ICS dimensions did not contribute equally to different organizational outcomes. For example, Sinčić Ćorić et al. (2020) found that although all eight ICS dimensions correlated with life satisfaction, only two (informal communication and climate) emerged as significant predictors in the regression model. Similarly, Pološki Vokić et al. (2021) showed that communication climate and supervisor communication were the strongest predictors of organizational trust. This is why multidimensional approach to internal communication is important to realize which concrete dimensions contribute most to development of employees' readiness for change.

## 2.1. Mediation and Moderation effects

Studies so far have acknowledged that the association between internal communication and change readiness is not exclusively linear, but can integrate additional mediation and moderation processes. Mediating effects, such as trust in management, organizational support (Cao & Le, 2024), or self-efficacy (Zainab et al., 2022) can be significant transmitting variables, while the leadership roles (Saeed et al., 2024), or employee emotional exhaustion (Raza et al., 2023) can strongly moderate the strength or the direction of the observed relationship. The role of moderators stems from the actual business contexts. The processes of globalization and digitalization have accelerated the dynamics of change acceptance in today's organizations, making the work environment more uncertain and unpredictable (Christensen et al., 2020). Digital communication has changed the way jobs need to and can be performed in traditional offices, and has influenced the formation of new job arrangements. Hence, today's typical workplace consists not only of the physical office but also of online spheres; both of which require employees' full presence and can blur work-life boundaries. In that kind of situation, in which employees are already overwhelmed with too many responsibilities, new changes can present an additional burden to them (Chênevert et al., 2019). Previous studies have shown that these stressors can negatively impact individuals' physical, physiological, and cognitive aspects, and lead to increased tension and turnover intentions (Hakro et al., 2022), diminished affective commitments (Chênevert et al., 2013), and burnout (Morić Milovanović & Cvjetković, 2024).

Work demands accumulate over time and deplete a person's physical or mental resources, leading to overtiredness (Presti & Nonnis, 2014). As a result, employees lack the energy to cope with constant job requirements (Bakker et al., 2023a), particularly when accompanied by tight deadlines, heavy workloads, or a perceived inability to meet organizational expectations (O'Connor et al., 2018; Chênevert et al., 2019). A similar effect has been shown with information overload, a very frequent stressor in contemporary business environments. It is defined as a condition in which individual effectiveness is reduced due to longer exposure to excessive information (Hoq, 2016). Although new technologies were expected to simplify workflows, they

often accelerate the generation and circulation of information, making it more difficult for employees to find and retrieve the data they need. Research has shown that information overload at work can lead to stress, job dissatisfaction, or anxiety (Bawden et al., 1999) and can negatively affect overall well-being (Edmunds & Morris, 2000).

When examining the mediating effects on the ICS-RFC relationship, previous research has shown that high-quality change communication can further strengthen the relationship between internal communication and positive change outcomes, because it justifies the necessity for change, and in turn, reduces uncertainty (Petrout et al., 2018). Perception of quality of change communication (PQCC) refers to employee assessment of specific change-related information; whether it is relevant, accurate, concrete, truthful and delivered on time. So, while ICS dimensions relate to employees' general satisfaction with different organizational communication practices, PQCC specifically comprises their evaluations of the quality of communication about concrete change. Previous studies have shown that the quality of change communication mediates the link between leadership communication and employee reactions to change (Rafferty & Restubog, 2016; Meyer, 2021). Similarly, its mediating role in the relationship between internal communication, communication climate, and positive employee responses to change, including identification and affective commitment, has been confirmed (Rogiest et al., 2015; Sun et al., 2021). However, these findings mainly come from studies conducted in large organizations, while evidence from SMEs is still scarce. In light of this, an additional set of variables is proposed:

H2.1 – H2.8: The relationship between each dimension of internal communication satisfaction: *satisfaction with feedback, satisfaction with communication with immediate superior, satisfaction with horizontal communication, satisfaction with informal communication, satisfaction with information about the organization, satisfaction with communication climate, satisfaction with communication in meetings, and satisfaction with quality of communication media*, and readiness for change is mediated by the perceived quality of change communication.

H3.1–H3.8: Information overload moderates the relationship between each dimension of internal communication satisfaction: *satisfaction with feedback, satisfaction with communication with immediate superior, satisfaction with horizontal communication, satisfaction with informal communication, satisfaction with information about the organization, satisfaction with communication climate, satisfaction with communication in meetings, and satisfaction with quality of communication media*, and readiness for change.

H4.1–H4.8: Work demands moderate the relationship between each dimension of internal communication satisfaction: *satisfaction with feedback, satisfaction with communication with immediate superior, satisfaction with horizontal communication, satisfaction with informal communication, satisfaction with information about the organization, satisfaction with communication climate, satisfaction with communication in meetings, and satisfaction with quality of communication media*, and readiness for change.

Since the literature holds limited prior evidence on moderating roles of IO and WD at the individual ICS dimension level, hypotheses H3 and H4 are formulated as exploratory. No directional predictions are made regarding the sign or pattern of these interactions, because the effects may vary depending on the specific communication dimension and the level of the stressor.

### 3. Methodology

This is a quantitative study, based on the data collected via a self-report questionnaire created in Google Forms. The research targeted employees working in Croatian small and medium-sized enterprises, excluding micro enterprises due to their specific communication patterns. According to the Croatian Financial Agency (2024), in 2022, there were 452,970 employees in 15,855 small and medium-sized enterprises. To ensure representativeness for large populations, the recommended sample size is 384 valid answers (Krejcie & Morgan, 1970). The sampling frame was compiled using the Fininfo database (El koncept d.o.o., 2024). Respondents were selected using random sampling, more precisely, proportional stratified sampling which reflected the sectoral structure of the Croatian economy (Croatian Bureau of Statistics, 2023). Within each stratum, firms were randomly chosen according to the national activity classification (Singh & Mangat, 2013). In case of missing email contacts, they were manually added by using information available on the company websites.

In the first part of the questionnaire, respondents had to answer a different set of demographic questions, such as gender, age, education, role, department, firm size, and industry. The second part contained the study variables. Independent variables included eight dimensions of internal communication satisfaction. Perceived quality of change communication was a mediator, while work demands and information overload were moderators. Readiness for change served as the dependent variable. All constructs were measured using Likert-type scales adapted from validated instruments. *Internal Communication Satisfaction dimensions* were

assessed using 32 items (4 items per dimension), measured on a 7-point scale. Questions were based on the instrument developed by Tkalac Verčič et al. (2009). The item order was partially modified to reduce common method bias. Cronbach's Alpha values were as follows: satisfaction with feedback ( $\alpha = .89$ ), satisfaction with communication with immediate superior ( $\alpha = .89$ ), satisfaction with horizontal communication ( $\alpha = .87$ ), satisfaction with informal communication ( $\alpha = .78$ ), satisfaction with information about the organization ( $\alpha = .87$ ), satisfaction with communication climate ( $\alpha = .94$ ), satisfaction with quality of communication media ( $\alpha = .90$ ), and satisfaction with communication in meetings ( $\alpha = .91$ ). *Perceived Quality of Change Communication* was measured with nine statements on a 7-point Likert scale, adapted from Sun et al. (2021) scale. Cronbach's Alpha value was 0.96. *Work Demands* were operationalised through eleven items from the third version of the Copenhagen Psychosocial Questionnaire (Burr et al., 2019), measured on a 7-point scale. Cronbach's Alpha value was 0.89. *Information Overload* was observed through sixteen items measured on a 5-point scale (Misra & Stokols, 2012), with Cronbach's Alpha value of 0.91. *Readiness for Change* was measured with nine items (Bouckennooghe et al., 2009), using a 5-point Likert scale. Cronbach's Alpha value was 0.92.

Negatively phrased items were reworded into positive ones to avoid coding mistakes. The standard ethical standards and procedures were followed (Martin & Marker, 2007). The accompanying letter was sent together with the link to fill the questionnaire, in which the participants were informed of the aims of the study and the way their data would be used, ensuring their confidentiality and anonymity. Moreover, their rights regarding voluntary participation and withdrawal were clearly stated, notifying them additionally that their answers would be used in aggregate form.

### 3.1. Sample

The collection of data started in May 2024. The first wave lasted until mid-June, during which 4,100 emails were sent to the chosen sample, followed by one additional reminder. The survey closed at the beginning of July 2024. A total of 608 answers were received (14.8 % response rate). Next step included data cleaning: business owners and employees working in micro or large firms were deleted, as several dimensions of internal communication do not apply to their role (e.g., communication with supervisors). Boxplot analysis also revealed several outliers. After clearing up the data, the final sample consisted of 454 respondents, whose answers were used in the statistical analysis and hypothesis testing.

In terms of demographics, the sample consists mostly of women (64.8%). Majority of the respondents belong to the 31–50 age group (62.4%), have a higher education (75.5%), and perform either managerial (38.1%) or administrative (41.4%) positions within their organizations. Most respondents work in privately owned enterprises (90.7%). In total, 28.9% of the respondents work in medium-sized enterprises, while 71.1% are employed in small enterprises. Respondents are distributed across various departments and industries, which corresponds to the employment structure of SMEs based on the national classification of business activities (Croatian Bureau of Statistics, 2020).

## 4. Results

Before conducting the regression, mediation, and moderation analyses, sampling and nonsampling errors were examined. Non-response bias was estimated using the method of wave analysis (Phillips et al., 2016). The obtained difference of  $-0.11$  showed no meaningful bias. Additionally, in the Harman test, the largest factor captured 29% of the total variance. This implies that common method bias is not an issue, as it is below 50 % threshold (Fuller et al., 2016). Moreover, regression diagnostics were examined, including linearity, homoscedasticity, and multicollinearity (Hayes, 2022), all of which were within acceptable values. This allowed for further statistical analysis.

The first step included calculating descriptive statistics for all the observed variables (means, standard deviations, and internal consistency reliability coefficients through Cronbach's alpha). In the next step, Pearson correlation coefficients were obtained to explore the associations between the variables. Overall, the analysis showed moderate to strong positive associations among internal communication satisfaction dimensions. Readiness for change was positively related to all communication dimensions, with the strongest connection observed for communication climate, meeting communication, and satisfaction with organizational information. Perceived quality of change communication showed strong positive associations with communication satisfaction and a moderate positive relationship with readiness for change, while being negatively related to information overload and work demands. Both stressor variables exhibited predominantly weak negative relationships with most communication dimensions.

Following the correlation analysis, direct effects were first estimated using multiple regression analysis in SPSS (IBM Corp., 2019). In addition, moderation analyses were conducted to examine whether the two

stressor variables, information overload (IO) and work demands (WD), modify the strength of the relationships between each of the internal communication satisfaction dimension and readiness for change. For this purpose, all continuous predictors (SwHC, SwIO, SwCC, SwQM, SwIS, SwCM, SwIC, SwF, IO, WD) were mean-centered, and interaction terms were computed by multiplying each centered communication predictor with the centered moderator (Dawson, 2014). Moderation models were estimated using hierarchical regression. In the first step, the main effects were entered, while interaction terms were added in the second step to assess incremental variance explained ( $\Delta R^2$ ) (Dawson, 2014). To complement the SPSS results, significant interactions were visualised in R (R Core Team, 2023), where interaction plots were generated at low (–1 SD), mean, and high (+1 SD) values of the moderators.

A mediation analysis was conducted using structural equation modelling (SEM) in the R package lavaan (Rosseel, 2012) to test the mediating role of employees’ perception of change communication quality (PQCC). The mediation analysis was based on the same mean-centered variables used in the moderation analyses. The mediation model specified paths from each internal communication satisfaction dimension to the mediator PQCC (a-paths), from PQCC to RFC (b-path), and direct paths from each internal communication dimension to RFC (c’-paths). Indirect (a  $\times$  b) and total effects were defined using lavaan’s model constraint operators (Rosseel, 2012). Parameter estimation was performed using maximum likelihood (ML). Indirect effects were evaluated using non-parametric bootstrap procedures with 5,000 resamples, which is recommended in contemporary mediation research because indirect effect distributions are often asymmetric and bootstrapping provides more accurate confidence intervals (Hayes, 2022; Huh et al., 2022). Since the mediation model was just-identified (df = 0), global model fit indices were not interpreted. Instead, model evaluation focused on the interpretation of parameter estimates, indirect effects, and explained variance rather than overall fit statistics. Coefficients of determination ( $R^2$ ) were therefore reported for both PQCC and RFC to indicate the proportion of explained variance. Results of multiple regression analysis on direct effects are presented in Table 1, while moderation and mediation analyses are presented in Tables 2, 3, and 4.

| Predictor             | M     | S.D.  | B      | SE B  | $\beta$ | t      | p              | VIF  |
|-----------------------|-------|-------|--------|-------|---------|--------|----------------|------|
| Feedback (SwF)        | 5,060 | 1,289 | 0.014  | 0.039 | 0.027   | 0.366  | .714           | 2.87 |
| Supervisor (SwIS)     | 5,420 | 1,361 | –0.026 | 0.031 | –0.052  | –0.832 | .406           | 2.11 |
| Horizontal (SwHC)     | 5,502 | 1,047 | –0.001 | 0.039 | –0.001  | –0.018 | .985           | 1.93 |
| Informal (SwIC)       | 4,823 | 1,157 | –0.027 | 0.037 | –0.046  | –0.739 | .460           | 2.11 |
| Organizational (SwIO) | 5,197 | 1,368 | 0.076  | 0.034 | 0.153   | 2.277  | <b>.023 **</b> | 2.44 |
| Climate (SwCC)        | 5,330 | 1,342 | 0.093  | 0.039 | 0.183   | 2.381  | <b>.018 **</b> | 3.21 |
| Media (SwQM)          | 5,303 | 1,230 | 0.028  | 0.038 | 0.049   | 0.715  | .475           | 2.60 |
| Meetings (SwCM)       | 4,997 | 1,334 | 0.071  | 0.040 | 0.139   | 1.785  | <b>.075 *</b>  | 3.31 |

Note:  $R^2 = .18$ , adj.  $R^2 = .16$ ,  $F(8, 445) = 12.12$ ,  $p < .001$ . Durbin-Watson = 2.19, Max Cook’s Distance = 0.064;  $p < .05^{**}$ ;  $p < .10^*$ .

**Table 1.** Multiple regression results: direct effects

Table 1 presents the results of the multiple regression analysis of direct effects of internal communication satisfaction dimensions on readiness for change (RFC). The model is statistically significant and explained 18% of the variance in RFC ( $R^2 = .18$ , Adj.  $R^2 = .16$ ). Results show that three predictors display positive and statistically significant associations: satisfaction with information about the organization ( $\beta = .153$ ,  $p = .023 < 0.05$ ), with communication climate ( $\beta = .183$ ,  $p = .018 < 0.05$ ), and with communication in meetings ( $\beta = .139$ ,  $p = .075 < 0.10$ ). The other communication dimensions exhibited non-significant associations ( $p > .10$ ). Results of the IO moderation analysis are presented in Table 2.

Results show that the full model explained approximately 20% of the variance in RFC ( $R^2 = .198$ ). After the inclusion of the interaction terms, there was a small, but statistically insignificant increase in explained variance ( $\Delta R^2 = .023$ ). This pattern is not uncommon in moderation research, as the incremental variance of individual interaction terms is diluted when multiple predictors are tested simultaneously (Dawson, 2014; Hayes, 2022). Nevertheless, as many interaction terms were tested simultaneously, there is a higher chance of the occurrence of Type I error (false positive finding), which is why marginally significant interactions ( $p < .10$ ) should be interpreted with caution.

| Predictor                  | B      | SE B  | $\beta$ | t      | p      | VIF  |
|----------------------------|--------|-------|---------|--------|--------|------|
| Information Overload       | 0.115  | 0.035 | 0.145   | 3.248  | .001** | 1.10 |
| <b>Interaction effects</b> |        |       |         |        |        |      |
| SwF $\times$ IO            | 0.039  | 0.043 | 0.064   | 0.889  | .374   | 2.93 |
| SwIS $\times$ IO           | -0.085 | 0.038 | -0.147  | -2.263 | .024** | 2.38 |
| SwHC $\times$ IO           | -0.060 | 0.048 | -0.079  | -1.231 | .219   | 2.30 |
| SwIC $\times$ IO           | -0.045 | 0.046 | -0.071  | -0.972 | .332   | 2.97 |
| SwIO $\times$ IO           | -0.031 | 0.039 | -0.056  | -0.800 | .424   | 2.75 |
| SwCC $\times$ IO           | 0.004  | 0.047 | 0.007   | 0.091  | .928   | 3.57 |
| SwQM $\times$ IO           | 0.075  | 0.042 | 0.122   | 1.784  | .075*  | 2.63 |
| SwCM $\times$ IO           | 0.023  | 0.048 | 0.040   | 0.478  | .633   | 3.97 |

Note: Model 1:  $R^2 = .198$ , Adj.  $R^2 = .182$ ; Model 2:  $R^2 = .221$ , Adj.  $R^2 = .191$ ,  $\Delta R^2 = .023$ , F change (8, 436) = 1.615, Durbin-Watson = 2.251. Max Cook's distance = .075;  $p < .05^{**}$ ,  $p < .10^{*}$ ;

**Table 2.** Moderation Effects of Information Overload

The same predictors identified in the direct effects analysis remained significant predictors of RFC, including satisfaction with information about the organization (SwIO), communication climate (SwCC), and communication in meetings (SwCM). In addition, information overload itself showed a positive and statistically significant main association on RFC ( $\beta = .145$ ,  $p = .001 < .01$ ).

Two interaction terms reached statistical significance. The interaction between supervisor communication and information overload (SwIS  $\times$  IO) was significant ( $\beta = -.147$ ,  $p = .024 < .05$ ), indicating that the positive association between supervisor communication and readiness for change weakens as information overload increases. In addition, the interaction between satisfaction with the communication media quality and information overload (SwQM  $\times$  IO) showed a marginal association ( $\beta = .122$ ,  $p = .075$ ), indicating that the quality of the communication channel becomes more important for RFC when employees experience higher levels of information overload. Graphical representations in Figures 1 and 2 support these results.

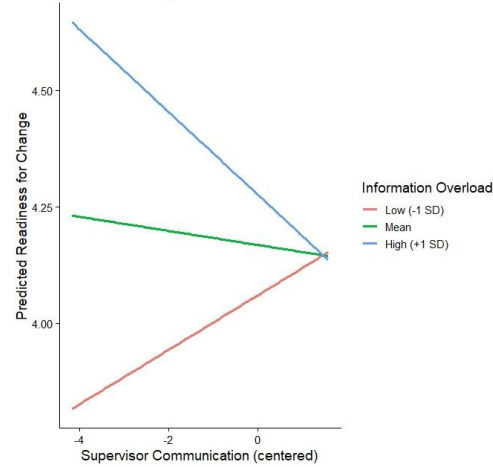


**Figure 1.** Interaction effect of IO 1

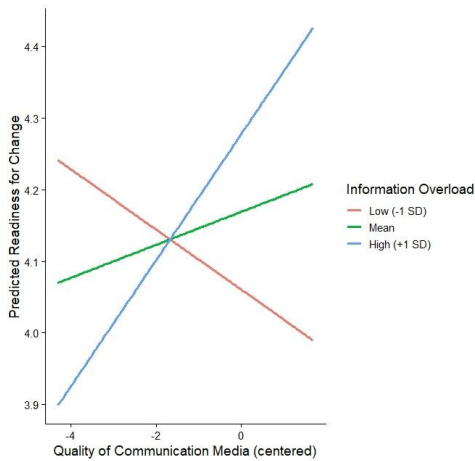


**Figure 2.** Interaction effect of IO 2

The interaction plot displayed in Figure 1 shows that satisfaction with supervisor communication is positively associated with readiness for change when information overload is low. But this relationship becomes weaker and turns negative at high levels of IO. Similarly, the interaction plot in Figure 2 shows that the relationship between the communication media quality and readiness for change also depends on the level of information overload. When information overload is high, a higher perception of the quality of communication media is connected to higher readiness for change. In contrast, under lower levels of information overload, this relationship is weaker and slightly negative. Next, the results of the WD moderation analysis are presented in Table 3.

| Predictor           | B      | SE B  | $\beta$ | t      | p      | VIF  |
|---------------------|--------|-------|---------|--------|--------|------|
| Work Demands        | 0.068  | 0.024 | 0.127   | 2.805  | .005** | 1.14 |
| Interaction effects |        |       |         |        |        |      |
| SwF $\times$ WD     | 0.014  | 0.029 | 0.035   | 0.486  | .627   | 2.82 |
| SwIS $\times$ WD    | -0.056 | 0.026 | -0.143  | -2.135 | .033** | 2.49 |
| SwHC $\times$ WD    | -0.073 | 0.034 | -0.140  | -2.130 | .034** | 2.39 |
| SwIC $\times$ WD    | 0.002  | 0.031 | 0.005   | 0.069  | .945   | 2.94 |
| SwIO $\times$ WD    | -0.038 | 0.028 | -0.103  | -1.393 | .164   | 3.01 |
| SwCC $\times$ WD    | 0.065  | 0.033 | 0.167   | 1.990  | .047** | 3.91 |
| SwQM $\times$ WD    | 0.010  | 0.031 | 0.026   | 0.336  | .737   | 3.23 |
| SwCM $\times$ WD    | 0.030  | 0.035 | 0.078   | 0.858  | .392   | 4.59 |

Model fit: Model 1:  $R^2 = .193$ , Adj.  $R^2 = .177$ ,  $F(9, 444) = 11.813$ ,  $p < .001$ ; Model 2:  $R^2 = .216$ , Adj.  $R^2 = .185$ ,  $\Delta R^2 = .022$ ,  $F \text{ change}(8, 436) = 1.552$ , Durbin-Watson = 2.218; Max Cook's Distance = .085;  $p < .05^{**}$ .

Table 3. Moderation Effects of Work Demands

As is visible in the table, Model 1, which tests the direct effects, explained 19.3% of the variance in readiness for change and was statistically significant. By adding the interaction terms, explained variance had a small, but non-significant increase ( $\Delta R^2 = .022$ ). This means that work demands contributed modestly to the model. The main, direct effects remained unchanged compared to the previous analysis. Work demands (WD) showed a significant positive main association ( $\beta = .127$ ,  $p = .005 < .01$ ), meaning higher perceived work demands were associated with higher readiness for change. Three interaction effects reached statistical significance. The interaction SwIS  $\times$  WD ( $\beta = -.143$ ,  $p = .033 < .05$ ) indicates that higher work demands weaken the positive association of supervisor communication on readiness for change. A similar pattern was observed for horizontal communication (SwHC  $\times$  WD;  $\beta = -.140$ ,  $p = .034 < .05$ ), suggesting that its association with readiness for change weakens as work demands increase. In contrast, the interaction between satisfaction with the communication climate and work demands is positive and significant (SwCC  $\times$  WD;  $\beta = .167$ ,  $p = .047 < .05$ ), suggesting that a supportive communication climate becomes increasingly important for promoting readiness for change under conditions of higher work demands. These interaction results are illustrated in Figures 3, 4, and 5.

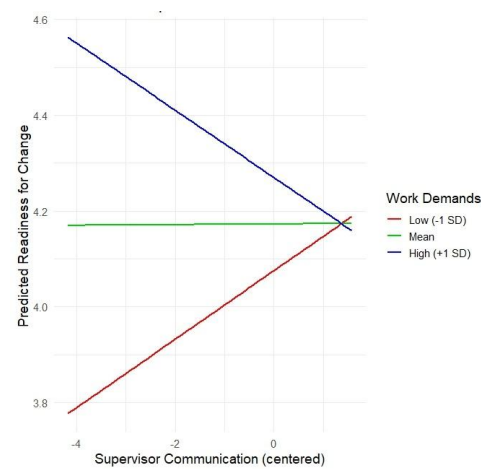


Figure 3. Interaction effect of WD 1

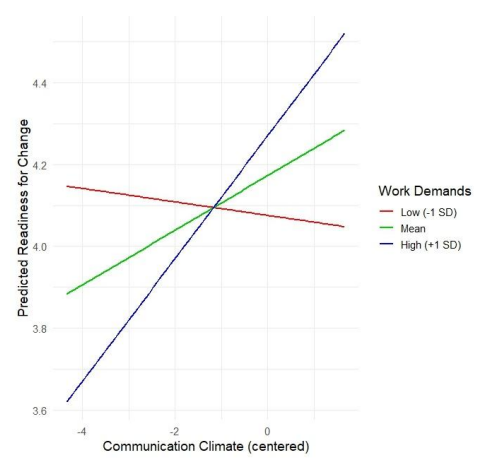


Figure 4. Interaction effect of WD 2

Figure 3 shows that when work demands are low, higher satisfaction with supervisor communication is associated with greater RFC, whereas under high work demands, this relationship becomes negative. At average levels of work demands, communication with the immediate superior is only weakly related to readiness for change, indicating that the association of communication with the immediate superior depends strongly on employees' workload. Figure 4 displays that the positive relationship between communication climate and RFC strengthens as work demands increase. Under high work demands, a supportive

communication climate is associated with a considerably higher level of readiness for change, whereas under low work demands, this relationship is weak.

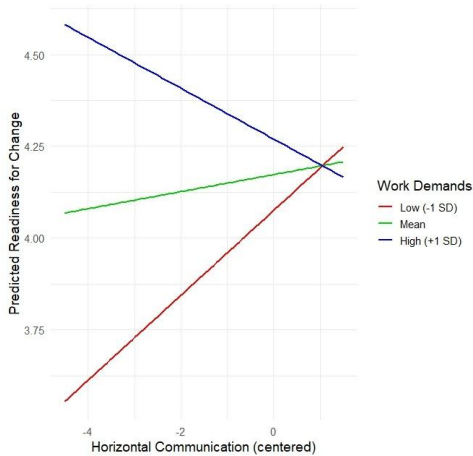


Figure 5. Interaction effect of WD 3

Figure 5 shows that the positive association between horizontal communication and readiness for change weakens as work demands increase. While better horizontal communication is linked to higher readiness for change under low work demands, this relationship becomes noticeably weaker and even turns negative when work demands are high.

To further examine the mechanisms linking internal communication satisfaction dimensions to readiness for change, the next analysis focuses on the mediating role of perceived quality of change communication, as presented in Table 4.

| Predictor             | ACME    | 95% CI (LL, UL) | P(ACME) | ADE     | P(ADE) | Total   | P (Total) |
|-----------------------|---------|-----------------|---------|---------|--------|---------|-----------|
| Feedback (SwF)        | − 0.001 | − 0.011, 0.009  | .865    | 0.000   | .998   | − 0.001 | .987      |
| Superior (SwIS)       | 0.010   | − 0.004, 0.024  | .187    | 0.067*  | .079   | 0.076** | .038      |
| Horizontal (SwHC)     | 0.009   | − 0.007, 0.025  | .262    | 0.084** | .043   | 0.093** | .024      |
| Informal (SwIC)       | 0.010   | − 0.006, 0.026  | .230    | 0.018   | .681   | 0.028   | .508      |
| Organizational (SwIO) | 0.014   | − 0.006, 0.034  | .173    | − 0.040 | .314   | − 0.026 | .514      |
| Climate (SwCC)        | 0.027   | − 0.006, 0.060  | .108    | 0.044   | .270   | 0.071*  | .067      |
| Media (SwQM)          | 0.005   | − 0.007, 0.017  | .321    | − 0.033 | .361   | − 0.027 | .446      |
| Meetings (SwCM)       | 0.002   | − 0.008, 0.012  | .743    | 0.013   | .769   | 0.014   | .742      |

Note. ACME = average causal mediation effect (indirect effect); ADE = average direct effect; Total = ACME + ADE. Estimated from 5,000 bootstrap resamples. \*\*p < .05, \*p < .10.

Table 4. Mediation effects

Table 4 presents the results of the mediation analysis. The indirect effect (ACME) captures the extent to which each communication dimension is associated with RFC through PQCC, while the average direct effect (ADE) reflects the remaining influence of the predictor on RFC when PQCC is included in the model. Total effects represent the combined direct and indirect associations (Hayes, 2022). Direct effects (ADE) are reported for completeness, although the primary focus of mediation analysis lies in the indirect pathways (ACME). The results show that none of the indirect effects (ACME) reached statistical significance, suggesting that mediation was not supported in this sample and model specification.

## 5. Discussion

This study examined the associations between eight dimensions of internal communication satisfaction and employee change readiness in SMEs. These dimensions represent distinct components of organizational information infrastructures, capturing how employees experience information flows, channel effectiveness, and information quality during change processes. Since these variables have not previously been thoroughly empirically tested, the results fill the identified research gap (Angwin et al., 2016). The research framework also included information overload and work demands as moderators, while also testing the mediating role of employees' perception of change communication quality. This resulted in setting up a total of 32 hypotheses. In terms of direct effects, multiple regression analysis revealed that only three internal communication dimensions were significantly associated with change readiness for SME employees: satisfaction with information about the organization ( $\beta = 0.153$ ,  $p = .023$ ), communication climate ( $\beta = 0.183$ ,  $p = .018$ ), and with communication in meetings ( $\beta = 0.139$ ,  $p = .075$ ), which confirms hypotheses H1.5, H1.6, and H1.8. These findings align with recent research showing that transparent communication reduces uncertainty and strengthens organizational relationships (Yue et al., 2023; Li et al., 2021), while a positive communication climate enhances affective commitment to change (Neill et al., 2020). The significance of communication in meetings supports earlier findings that group meetings reduce change-related uncertainty (Bordia et al., 2004) and ameliorate information transfer (Reinig & Shin, 2002) by allowing for direct response and reaction. All three significant dimensions relate directly to organizational-level information structures: satisfaction with information about the organization reflects perceived information quality and transparency, communication climate represents the broader informational environment, and communication in meetings functions as a rich channel that enables sensemaking and mutual interpretation of provided information.

On the other hand, other dimensions of ICS showed no significant influence on change readiness, meaning the remaining hypotheses could not be confirmed. These results may reflect the specific characteristics of SMEs, which have flatter organizational structures and informal communication practices (Psychogios et al., 2019; Blakčori & Aroles, 2021) that coexist with greater employee autonomy in decision-making (Lucas et al., 2008). When compared to other studies, these results are partially inconsistent with previous findings. For example, in terms of communication with the immediate supervisor, Endrejat et al. (2021) found that only a certain type of supervisor communication, identified as restrictive communication, reduced readiness for change, whereas collaborative communication had no significant effect. Similarly, Morić Milovanović et al. (2022) reported no significant effect of supervisor and peer relations on change readiness among Croatian employees in a medium-sized enterprise. Partially inconsistent results gained in this study should also be observed from the perspective of the post-pandemic context. Evidence suggests that employee perceptions of internal communication dimensions have shifted (Tkalac Verčič & Špoljarić, 2023), which may further clarify why these results do not completely comply with studies conducted before. For example, corporate information and communication climate gained more in importance while feedback, informal communication, and media quality became less notable (Tkalac Verčič & Špoljarić, 2023). This points to a possible contextual interpretation. During crises, employees tend to prefer formal communication channels to reduce uncertainty (Luo & Jiang, 2014), which may offer one possible explanation of why formal dimensions demonstrated stronger associations with RFC while more informal dimensions did not. However, this interpretation is based on prior literature and was not directly tested in the present study.

The analysis further showed that information overload and work demands have complex moderating roles. Both information overload ( $\beta = 0.145$ ,  $p = .001$ ) and work demands ( $\beta = 0.127$ ,  $p = .005$ ) had significant positive direct associations with change readiness. This finding can be interpreted through the 'challenge-hindrance framework' (Podsakoff et al., 2023), which advocates that reasonable levels of demands and information may function as motivators rather than barriers (LePine et al., 2005). In dynamic SME contexts, employees who are accustomed to multitasking and rapid information exchange may perceive organizational change as part of their everyday work environment (Cavanaugh et al., 2000; Majid et al., 2017). This is why these two stressor variables are not directly (when isolated) perceived as a threat to change readiness. But, when in interaction with other elements, both moderators shaped how specific internal communication dimensions were associated with change readiness. Information overload moderated the effects of supervisor communication ( $\beta = -0.147$ ,  $p = .024$ ) and communication media quality ( $\beta = 0.122$ ,  $p = .075$ ), supporting H3.2 and H3.7. From an information processing perspective, these findings indicate that work demands and information overload act as contextual constraints on employees' cognitive capacity to absorb and utilize organizational information. Higher information overload weakened the contribution of supervisor communication, which is consistent with the cognitive resource depletion perspective (Eppler & Mengis, 2004; Karr-Wisniewski & Lu, 2010; Arnold et al., 2023). In contrast, high-quality communication

media became more effective under information overload, as it facilitates more efficient information processing (Lee, 2018). Work demands showed a similar pattern. Negative moderation effects were found for supervisor communication ( $\beta = -0.143$ ,  $p = .033$ ) and horizontal communication ( $\beta = -0.140$ ,  $p = .034$ ), while a positive moderation effect emerged for satisfaction with communication climate ( $\beta = 0.167$ ,  $p = .047$ ), supporting H4.2, H4.3, and H4.6. High work demands tend to reduce employees' capacity to engage with supervisory and peer communication (Eppler & Mengis, 2004; Arnold et al., 2023), whereas a supportive communication climate becomes increasingly important in sustaining change readiness under demanding conditions (Johnston, 2005; Bakker et al., 2023b).

The study also examined whether PQCC mediates the relationship between ICS dimensions and RFC. The mediation analysis did not reveal any statistically significant indirect effects. This finding suggests that, in this sample and model specification, PQCC did not function as a mediator. However, this result should be interpreted carefully. The full measurement model was not tested in the present study, which means that the conceptual proximity between ICS dimensions and PQCC may have influenced the findings. This is why there remains a possibility that the absence of indirect effects reflects the model-specific constraints or the overlap of the measurements, rather than the actual absence of mediation. Nevertheless, this result does not diminish the relevance of the PQCC. Prior research, in which ICS was observed as a one-dimensional construct, showed that PQCC is a statistically significant mediator in the relationship between overall ICS and RFC (Cvjetković, 2024). So, when internal communication is aligned in a coherent way and through an integrated system, the way employees evaluate and perceive change communication becomes an important mechanism through which communication supports readiness for change.

### 5.1. Theoretical and practical implications

The results of this study contribute to internal communication, organizational information systems, change readiness, and management theories in several ways. First, by observing internal communication satisfaction as a multidimensional construct and testing its eight distinct dimensions simultaneously, the research provides empirical evidence that not all of its aspects equally contribute to change readiness in SMEs. The findings show that only three dimensions, satisfaction with information about the organization, satisfaction with communication climate, and satisfaction with communication in meetings, are significantly related to readiness for change. This indicates that general organizational information structures and everyday communication practices are more relevant for change readiness than interpersonal or channel-specific communication dimensions. Dimensions that are often emphasized in the change management literature, such as management communication, feedback, and peer communication, were not found to be significant predictors in the present study (Angwin et al., 2016). This specificity reveals that formal communication practices may be more significant for change readiness than interpersonal communication channels in SME contexts. From an information processing perspective, the results indicate that readiness for change in SMEs is shaped primarily through communication structures that reduce informational uncertainty, promote transparency, and support employees' cognitive processing of change-related information.

Second, this study incorporates information overload and work demands as workplace stressors and information processing constraints that shape the relationship between internal communication satisfaction and readiness for change. The moderation analysis showed that information overload weakens the effect of supervisor communication while strengthening the effect of communication media quality. Similarly, work demands reduce the effects of supervisor communication and horizontal communication while strengthening the importance of communication climate. These results indicate that communication practices that require greater cognitive effort are less effective when employees are under pressure from too much work or information. Additionally, the study contributes to existing change management theories by showing that mediation through PQCC was not supported in this sample and model specification when internal communication is examined in a multidimensional manner. Although several dimensions exhibit significant direct effects on readiness for change, these effects occurred independently of employees' evaluations of change-specific communication. This does not mean that the change communication quality does not play an important role in employee change readiness. On the contrary, previous findings have proven that it is a significant mediator when internal communication satisfaction is conceptualized as a unidimensional construct. In broader context, it means that coherent and integrated internal communication may shape employees' interpretations of change messages at an overall level, while, at the dimensional level, specific communication practices influence RFC more directly.

The results also carry important practical considerations for SME managers. The different effects of communication dimensions observed in this research suggest that a "blanket" approach to change communication is not efficient, especially when employees face different levels of information overload and

work demands. The significantly positive effects of satisfaction with communication about organization, communication climate, and communication in meetings suggest that these formal, organizational-level communication practices should be the core of change communication strategies in SMEs. This means prioritizing transparency in sharing information about financial performance, strategic plans, and regulatory changes, and investing in a supportive communication climate characterized by openness, trust, and psychological safety (Neill et al., 2020). This will not only directly promote change readiness but also serve as a protective mechanism against the negative effects of demanding work conditions, which is particularly valuable in contemporary work environments where hybrid and remote work arrangements are common. Structured meeting communication should be encouraged as a primary channel for change initiatives, providing opportunities for two-way dialogue, immediate feedback, and clarity about roles and expectations. Thus, these findings emphasize the need for more targeted and context-sensitive communication approaches.

## 5.2. Limitations and future recommendations

The results of the study should be interpreted taking into account the study's limitations. First, the research design is cross-sectional (based on perceptions and experiences at a single point in time) and relies on self-reported data, which restricts the possibility of drawing causal conclusions. This particularly applies to the positive direct associations of information overload and work demands with readiness for change. Although interpreted through the challenge-hindrance framework, there is a possibility that reverse causality is also present, meaning that employees with higher readiness for change may inherently perceive stressors as more manageable, rather than stressors driving readiness. Future research is encouraged to focus on longitudinal designs that observe the impact of ICS dimensions on RFC over a certain period of time. Second, the study focused exclusively on small and medium-sized enterprises, excluding micro-enterprises. Future studies are encouraged to include employees from micro-enterprises and complement quantitative findings with qualitative approaches. In addition, the proposed model could be further extended by adding additional organizational and individual-level variables, such as trust, employee involvement, and specific communication channel usage, while individual characteristics, such as demographics or personality traits, could provide a detailed understanding of other factors contributing to RFC.

Additionally, although the sampling frame was constructed using proportional stratified sampling to reflect the sectoral structure of the Croatian economy, the voluntary nature of individual participation and the achieved response rate of 14.8% introduce a self-selection element at the respondent level. As some employees could possibly be nested within the same organizations, their shared work contexts may have introduced clustering effects that were not accounted for in the present analyses. Future research should apply multilevel modelling to address this. Even though all survey instruments were applied as originally validated, the absence of confirmatory factor analysis, discriminant validity assessment, and measurement model fit evaluation in the present sample limits confidence in the distinctiveness of the constructs. This can be especially relevant due to the multidimensional nature of ICS and the conceptual similarity among several sub-dimensions. For this reason, future research should include full measurement model testing to provide stronger evidence of construct validity. Another limitation is that the study focused on RFC as an attitudinal construct, without examining its behavioral or performance-related outcomes, such as actual change implementation. Future research could investigate these mechanisms. Other organizational barriers could also be added as moderators, such as message competition or information filtering.

Finally, a methodological limitation relates to the use of different software and analytical procedures. Regression and moderation analyses were conducted in SPSS, while mediation was tested in R. Although this approach allowed the use of appropriate tools for each analysis, it may have reduced methodological consistency and made the integration of results more challenging. In particular, examining direct, moderating, and mediating effects in separate steps does not allow all relationships to be tested simultaneously within a single model, which may limit a full understanding of the interdependencies among the study variables. Future research could address this limitation by applying integrative analytical approaches that enable the simultaneous estimation of direct, indirect, and interaction effects. Despite these limitations, the study provides valuable empirical evidence on the complex and context-dependent role of internal communication in shaping employees' readiness for change, offering relevant theoretical insights and practical implications for managing change processes in small and medium-sized enterprises.

## 6. Conclusion

This research has shown that specific communication dimensions, particularly information about the organization, communication climate, and communication in meetings, play an important role in fostering employees' readiness for change, while this relationship is dependent on contextual factors such as information overload and work demands. Specifically, information overload weakened the association between supervisor communication and readiness for change, while strengthening the effect of communication media quality. Work demands reduced the effects of supervisor and horizontal communication, and strengthened the importance of communication climate. The findings indicate that in SMEs, where communication structures are less formalized, organizational-level communication practices have particular significance in preparing employees for change. This points to the importance of investing in structured, transparent, and climate-focused communication as a foundation for successful change management in small and medium-sized enterprises.

## References

- Abrantes, A. C. M., Bakenhus, M., & Ferreira, A. I. (2024). The support of internal communication during organizational change processes. *Journal of Organizational Change Management*, 37(5), 1030-1050. <https://doi.org/10.1108/JOCM-06-2023-0222>
- Ahmad, F., & Huvila, I. (2019). Organizational changes, trust and information sharing: An empirical study. *Aslib Journal of Information Management*, 71(5), 677–692. <https://doi.org/10.1108/AJIM-05-2018-0122>
- Anderson, C. M., & Martin, M. M. (1995). The effects of communication motives, interaction involvement, and loneliness on satisfaction: A model of small groups. *Small Group Research*, 26(1), 118–137.
- Angwin, D. N., Mellahi, K., Gomes, E., & Peter, E. (2016). How communication approaches impact mergers and acquisitions outcomes. *The International Journal of Human Resource Management*, 27(20), 2370–2397.
- Arnaut, B., & Esposito, M. (2018). The value of communication in turbulent environments: How SMEs manage change successfully in unstable surroundings. *International Journal of Entrepreneurship and Small Business*, 34(4), 500–517. <https://doi.org/10.1504/IJESB.2018.093576>
- Arnold, M., Goldschmitt, M., & Rigotti, T. (2023). Dealing with information overload: a comprehensive review. *Frontiers in psychology*, 14, 1122200. <https://doi.org/10.3389/fpsyg.2023.1122200>
- Aspland, W. (2025, February 10). *Hyper-personalisation for internal comms: The end of one-size-fits-all messaging*. LinkedIn. <https://www.linkedin.com/pulse/hyper-personalisation-internal-comms-end-messaging-wayne-aspland-5votc>
- Bakker, A. B., Xanthopoulou, D., & Demerouti, E. (2023a). How does chronic burnout affect dealing with weekly job demands? A test of central propositions in JD-R and COR-theories. *Applied Psychology*, 72(1), 389-410. <https://doi.org/10.1111/apps.12382>
- Bakker, A. B., Hetland, J., Olsen, O. K., & Espevik, R. (2023b). Daily transformational leadership: a source of inspiration for follower performance?. *European Management Journal*, 41(5), 700-708.
- Bawden, D., Holtham, C., & Courtney, N. (1999, October). Perspectives on information overload. In *Aslib proceedings* (Vol. 51, No. 8, pp. 249-255). MCB UP Ltd.
- Blakçori, F., & Aroles, J. (2021). Assessing the role of managerial feedback in changing routines in small and medium enterprises. *Journal of Organizational Change Management*, 34(3), 570–589. <https://doi.org/10.1108/JOCM-08-2020-0235>
- Bordia, P., Hunt, E., Paulsen, N., Tourish, D., & DiFonzo, N. (2004). Uncertainty during organizational change: Is it all about control? *European Journal of Work and Organizational Psychology*, 13(3), 345–366.
- Bouckennooghe, D., Devos, G., & Van den Broeck, H. (2009). Organizational change questionnaire – Climate of change, processes, and readiness: Development of a new instrument. *The Journal of Psychology*, 143(6), 559–599. <https://doi.org/10.1080/00223980903218216>
- Burr, H., Berthelsen, H., Moncada, S., Nübling, M., Dupret, E., Demiral, Y., & Pohrt, A. (2019). The third version of the Copenhagen Psychosocial Questionnaire. *Safety and Health at Work*, 10(4), 482–503. DOI: 10.1016/j.shaw.2019.10.002
- Cavanaugh, M. A., Boswell, W. R., Roehling, M. V., & Boudreau, J. W. (2000). An empirical examination of self-reported work stress among U.S. managers. *Journal of Applied Psychology*, 85(1), 65–74.

- Cao, T. T., & Le, P. B. (2024). Impacts of transformational leadership on organizational change capability: a two-path mediating role of trust in leadership. *European Journal of Management and Business Economics*, 33(2), 157-173. <https://doi.org/10.1108/EJMBE-06-2021-0180>
- Chênevert, D., Vandenberghe, C., Doucet, O., & Ayed, A. K. B. (2013). Passive leadership, role stressors, and affective organizational commitment: A time-lagged study among health care employees. *European Review of Applied Psychology*, 63(5), 277-286. DOI: 10.1016/j.erap.2013.07.002
- Chênevert, D., Kilroy, S., & Bosak, J. (2019). The role of change readiness and colleague support in the relationship between role stressors and withdrawal behaviours: the case of health care sector employees. *Journal of Organizational Change Management*, 32(2), 208-223.
- Christensen, J. O., Finne, L. B., Garde, A. H., Nielsen, M. B., Sørensen, K., & Vleeshouwers, J. (2020). *The influence of digitalization and new technologies on psychosocial work environment and employee health: a literature review*. Statens arbejdsmiljøinstitut, Det Nationale Forskningscenter for Arbejdsmiljø.
- Clampitt, P. G., & Girard, D. (1993). Communication satisfaction: a useful construct?. *Atlantic Journal of Communication*, 1(2), 84-102. <https://doi.org/10.1080/15456879309367255>
- Croatian Bureau of Statistics. (2020). Structural business indicators of enterprises, 2018** (First Release No. 15.1.2). [https://web.dzs.hr/Hrv\\_Eng/publication/2020/15-01-02\\_01\\_2020.htm](https://web.dzs.hr/Hrv_Eng/publication/2020/15-01-02_01_2020.htm)
- Croatian Bureau of Statistics. (2023, December 20). *Persons in paid employment, by activities, November 2023* (First Release, RAD-2023-2-1/11). <https://podaci.dzs.hr/2023/en/58006>
- Cunningham, C. E., Woodward, C. A., Shannon, H. S., MacIntosh, J., Lendrum, B., Rosenbloom, D., & Brown, J. (2002). Readiness for organizational change: A longitudinal study of workplace, psychological and behavioural correlates. *Journal of Occupational and Organizational psychology*, 75(4), 377-392.
- Cvjetković, M. (2024). *Utjecaj interne komunikacije na spremnost na promjene zaposlenika malih i srednjih poduzeća* (Doctoral Dissertation). Sveučilište Sjever.
- Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, 29(1), 1–19. <https://doi.org/10.1007/s10869-013-9308-7>
- Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. *International journal of information management*, 20(1), 17-28.
- El koncept d.o.o. (2024). Fininfo – Pretraga poduzeća.** <https://www.fininfo.hr>
- Ellis, L. A., Tran, Y., Pomare, C., Long, J. C., Churrua, K., Saba, M., & Braithwaite, J. (2023). Hospital organizational change: The importance of teamwork culture, communication, and change readiness. *Frontiers in public health*, 11, 1089252. <https://doi.org/10.3389/fpubh.2023.1089252>
- Endrejat, P. C., Klonek, F. E., Müller-Frommeyer, L. C., & Kauffeld, S. (2021). Turning change resistance into readiness: How change agents' communication shapes recipient reactions. *European Management Journal*, 39(5), 595-604. <https://doi.org/10.1016/j.emj.2020.11.004>
- Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The information society*, 20(5), 325-344. <https://doi.org/10.1080/01972240490507974>
- Farr-Wharton, R., & Brunetto, Y. (2007). Organisational relationship quality and service employee acceptance of change in SMEs: A social exchange perspective. *Journal of Management & Organization*, 13(2), 114-125.
- Financial Agency. (2024, July 1). *Rezultati poslovanja poduzetnika u 2023. godini – razvrstani po veličini*. <https://www.fina.hr/novosti/rezultati-poslovanja-poduzetnika-u-2023.-godini-razvrstani-po-velicini>
- Fritzenschaft, T. (2014). *Critical success factors of change management: An empirical research in German small and medium-sized enterprises*. Springer Gabler.
- Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common methods variance detection in business research. *Journal of business research*, 69(8), 3192-3198. <https://doi.org/10.1016/j.jbusres.2015.12.008>
- Galbraith, J. R. (1974). Organization design: An information processing view. *Interfaces*, 4(3), 28-36.
- Gray, J., & Laidlaw, H. (2004). Improving the measurement of communication satisfaction. *Management Communication Quarterly*, 17(3), 425–448. <https://doi.org/10.1177/0893318903257980>

- Hakro, T. H., Jhatial, A. A., & Chandio, J. A. (2022). Employee turnover intentions: Investigating the role of work overload, job satisfaction, employee engagement and job stress. *Research Journal of Social Sciences and Economics Review*, 3(2), 71-82. [https://doi.org/10.36902/rjsser-vol3-iss2-2022\(71-82\)](https://doi.org/10.36902/rjsser-vol3-iss2-2022(71-82))
- Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). New York, NY: The Guilford Press.
- Hoq, K. M. G. (2016). Information overload: Causes, consequences and remedies — a study. *Philosophy and Progress*, 55(1–2), 49–68. <https://doi.org/10.3329/pp.v55i1-2.26390>
- Huh, M., Li, X., Ialongo, N. S., Brown, C. H., & Wang, W. (2022). A structural equation modeling approach to meta-analytic mediation analysis. *Prevention Science*, 23, 147–159. <https://doi.org/10.1007/s11121-021-01318-4>
- IBM Corp. (2019). *IBM SPSS Statistics for Windows, Version 26.0*. Armonk, NY: IBM Corp.
- Jiang, X., Du, J., Zhou, J., & Cui, Y. (2020). The impact of negative informal information before a change on performance: A within-person approach. *International Journal of Environmental Research and Public Health*, 17(2), 670. <https://doi.org/10.3390/ijerph17020670>
- Jiang, L., & Probst, T. M. (2014). Organizational communication: A buffer in times of job insecurity?. *Economic and Industrial Democracy*, 35(3), 557-579. <https://doi.org/10.1177/0143831X13489356>
- Johnston, L. (2005). The relationship between organizational climate, occupational type and workaholism. *New Zealand Journal of Psychology*, 34(3), 181.
- Karr-Wisniewski, P., & Lu, Y. (2010). When more is too much: Operationalizing technology overload and exploring its impact on knowledge worker productivity. *Computers in Human Behavior*, 26(5), 1061–1072.
- Kazakov, S., Ruiz-Alba, J. L., & Muñoz, M. M. (2021). *The impact of information and communication technology and internal market orientation blending on organisational performance in small and medium enterprises*. *European Journal of Management and Business Economics*, 30(2), 129–149. <https://doi.org/10.1108/EJMBE-04-2020-0068>
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607–610. <https://doi.org/10.1177/001316447003000308>
- Lee, Y. (2018). Dynamics of symmetrical communication within organizations: The impacts of channel usage of CEO, managers, and peers. *International Journal of Business Communication*, 59(1), 3-21. <https://doi.org/10.1177/2329488418803661>
- LePine, J. A., Podsakoff, N. P., & LePine, M. A. (2005). A meta-analytic test of the challenge stressor–hindrance stressor framework: An explanation for inconsistent relationships among stressors and performance. *Academy of Management Journal*, 48(5), 764–775. <https://doi.org/10.5465/amj.2005.18803921>
- Li, J. Y., Sun, R., Tao, W., & Lee, Y. (2021). Employee coping with organizational change in the face of a pandemic: The role of transparent internal communication. *Public relations review*, 47(1), 101984.
- Lucas, G. J. M., Vermeulen, P. A. M., & Curseu, P. L. (2008). Entrepreneurial decision styles and cognition in SMEs. In *Entrepreneurial strategic decision making: A cognitive perspective* (pp. 105–122). Edward Elgar Publishing.
- Luo, Y., & Jiang, H. (2014). Effective public relations leadership in organizational change: A study of multinationals in mainland China. *Journal of Public Relations Research*, 26(2), 134-160.
- Madsen, S. R., Miller, D., & John, C. R. (2005). Readiness for organizational change: do organizational commitment and social relationships in the workplace make a difference?. *Human resource development quarterly*, 16(2), 213-234. <https://doi.org/10.1002/hrdq.1134>
- Majid, A., Yasir, M., & Yasir, M. (2017). Individual and work dynamics affecting the determinants of functional flexibility in SMEs: evidence from Pakistan. *Journal of Entrepreneurship in Emerging Economies*, 9(2), 144-160. [DOI:10.1108/jee-03-2016-0008](https://doi.org/10.1108/jee-03-2016-0008)
- Martin, J., & Marker, D. A. (2007). Informed consent: Interpretations and practice in social surveys. *Social Science & Medicine*, 65(11), 2260–2271. DOI: 10.1016/j.socscimed.2007.08.004
- Maseda, A., Iturralde, T., Aparicio, G., Boulkeroua, L., & Cooper, S. (2019). Family board ownership, generational involvement and performance in family SMEs: a test of the S-shaped hypothesis. *European Journal of Management and Business Economics*, 28(3), 285-300. <https://doi.org/10.1108/EJMBE-07-2018-0071>

- Men, L. R., Neill, M., & Yue, C. A. (2020). Examining the effects of symmetrical internal communication and employee engagement on organizational change outcomes. *Public Relations Journal*, 13(4), 1-19.
- Men, L. R., Neill, M. S., April Yue, C., & Verghese, A. K. (2022). The role of channel selection and communication transparency in enhancing employee commitment to change. *Journalism & Mass Communication Quarterly*, 10776990221100518. <https://doi.org/10.1177/10776990221100518>
- Meyer, H. (2021). *The role of transformational leadership style, work engagement, and quality of change communication on employee readiness for organizational change* (Master's thesis). MSH Medical School Hamburg, University of Applied Sciences and Medical University.
- Misra, S., & Stokols, D. (2012). Psychological and health outcomes of perceived information overload. *Environment and Behavior*, 44(6), 737-759. <https://doi.org/10.1177/0013916511404408>
- Morić Milovanović, B., Bubaš, Z., & Cvjetković, M. (2022). Employee readiness for organizational change in the SME internalization process: the case of a medium-sized construction company. *Social sciences*, 11(3), 131.
- Morić Milovanović, B., & Cvjetković, M. (2024). Analysis of antecedents and consequences of workplace-related burnout among Croatian employees in the post-COVID-19 era. *Business: Theory and Practice*, 25(1), 108-118. <https://doi.org/10.3846/btp.2024.19320>
- Neill, M. S., Men, L. R., & Yue, C. A. (2020). How communication climate and organizational identification impact change. *Corporate Communications: An International Journal*, 25(2), 281-298. <https://doi.org/10.1108/CCLJ-06-2019-0063>
- O'Connor, P. J., Jimmieson, N. L., & White, K. M. (2018). Too busy to change: High job demands reduce the beneficial effects of information and participation on employee support. *Journal of Business and Psychology*, 33(5), 629-643. <https://doi.org/10.1007/s10869-017-9515-8>
- Petrou, P., Demerouti, E., & Schaufeli, W. B. (2018). Crafting the change: The role of employee job crafting behaviors for successful organizational change. *Journal of management*, 44(5), 1766-1792. Phillips, A. W., Reddy, S., & Durning, S. J. (2016). Improving response rates and evaluating nonresponse bias in surveys: AMEE Guide No. 102. *Medical teacher*, 38(3), 217-228. DOI: 10.3109/0142159X.2015.1105945
- Piwowar-Sulej, K., Popowicz, E., & Sulich, A. (2024). *What is the link between internal communication, organizational culture and environmental strategy? The context of company size and employee perception*. *Central European Management Journal*, 32(2), 301-319. <https://doi.org/10.1108/CEMJ-06-2022-0073>
- Podsakoff, N. P., Freiburger, K. J., Podsakoff, P. M., & Rosen, C. C. (2023). Laying the foundation for the challenge-hindrance stressor framework 2.0. *Annual review of organizational psychology and organizational behavior*, 10(1), 165-199. <http://dx.doi.org/10.1146/annurev-orgpsych-080422-052147>
- Pološki Vokić, N., Rimac Bilušić, M., & Najjar, D. (2021). Building organizational trust through internal communication. *Corporate Communications: An International Journal*, 26(1), 70-85. <https://doi.org/10.1108/CCLJ-01-2020-0023>
- Presti, A. L., & Nonnis, M. (2014). Testing the Job Demands-Resources model: Evidence from a sample of Italian employees. *TPM-Test Psychom Methodol Appl Psychol*, 21, 89-101. <https://doi.org/10.4473/TPM21.1.6>
- Psychogios, A., Blakcori, F., Szamosi, L., & O'Regan, N. (2019). From feeding-back to feeding-forward: Managerial feedback as a trigger of change in SMEs. *Journal of Small Business and Enterprise Development*, 26(1), 18-42. <https://doi.org/10.1108/JSBED-01-2018-0034>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rafferty, A. E., & Restubog, S. L. (2016). Why do employees' perceptions of their organization's change history matter? The role of change appraisals. *Human Resource Management*, 56(3), 533-550. Raza, M. A., Imran, M., Rosak-Szyrocka, J., Vasa, L., & Hadi, N. U. (2023). Organizational change and workplace incivility: Mediated by stress, moderated by emotional exhaustion. *International Journal of Environmental Research and Public Health*, 20(3), 2008. <https://doi.org/10.3390/ijerph20032008>
- Reinig, B. A., & Shin, B. (2002). The dynamic effects of group support systems on group meetings. *Journal of Management Information Systems*, 19(2), 303-325. <https://doi.org/10.1080/07421222.2002.11045728>
- Rincón Díaz, C. A., & Albors Garrigós, J. (2017). Research and technology organizations' mobilizers of the regional environment: Competitive strategies. *European Journal of Management and Business Economics*, 26(2), 180-198. <https://doi.org/10.1108/EJMBE-07-2017-011>

- Robson, P. J., & Tourish, D. (2005). Managing internal communication: an organizational case study. *Corporate Communications: An International Journal*, 10(3), 213-222.
- Rogiest, S., Segers, J., & Van Witteloostuijn, A. (2015). Climate, communication, and participation impacting commitment to change. *Journal of Organizational Change Management*, 28(6), 1094–1106. <https://doi.org/10.1108/jocm-06-2015-0101>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. DOI:10.18637/jss.v048.i02
- Saeed, I., Khan, J., Zada, M., & Zada, S. (2024). Employee sensemaking in organizational change via knowledge management: leadership role as a moderator. *Current Psychology*, 43(7), 6657-6671.
- Santoso, N. R., Sulistyaningtyas, I. D., & Pratama, B. P. (2022). Transformational leadership during the COVID-19 pandemic: Strengthening employee engagement through internal communication. *Journal of Communication Inquiry*, 01968599221095182. doi: 10.1177/01968599221095182
- Sinčić Ćorić, D., Pološki Vokić, N., & Tkalac Verčić, A. (2020). Does good internal communication enhance life satisfaction? *Journal of Communication Management*, 24(4), 363–376. <https://doi.org/10.1108/JCOM-11-2019-0146>
- Singh, R., & Mangat, N. S. (2013). *Elements of survey sampling* (Vol. 15). Springer Science & Business Media.
- Stoltenkamp, J., & Kasuto, O. A. (2011). E-Learning change management and communication strategies within a HEI in a developing country: Institutional organisational cultural change at the University of the Western Cape. *Education and Information Technologies*, 16(1), 41-54. <https://doi.org/10.1007/s10639-009-9114-z>
- Sun, R., Li, J. Y. Q., Lee, Y., & Tao, W. (2021). The role of symmetrical internal communication in improving employee experiences and organizational identification during COVID-19 pandemic-induced organizational change. *International Journal of Business Communication*, 60(4), 1398-1426. <https://doi.org/10.1177/23294884211050628>
- Šlogar, H., Morić Milovanović, B., & Hrvatin, S. (2023). Does the relationship between entrepreneurial orientation and subjective financial firm performance have an inverted U-shape? Evidence from Southeast European SMEs. *Administrative sciences*, 13(2), 26. <https://doi.org/10.3390/admsci13020026>
- Tkalac Verčić, A. (2025). *Public relations, a home for internal communication?* *Journal of Public Relations Research*, 37(1–2), 1–3. <https://doi.org/10.1080/10627226X.2024.2434348>
- Tkalac Verčić, A.** (2021). The impact of employee engagement, organizational support, and employer branding on internal communication satisfaction. *Public Relations Review*, 47(1), Article 102009. <https://doi.org/10.1016/j.pubrev.2021.102009>
- Tkalac Verčić, A., Pološki Vokić, N., & Sinčić Ćorić, D. (2009). Razvoj mjernog instrumenta za procjenu zadovoljstva internom komunikacijom. *Društvena istraživanja: časopis za opća društvena pitanja*, 18(1-2 (99-100)), 175-202.
- Tkalac Verčić, A., & Špoljarić, A. (2023). Effects of an external crisis: How internal communication influences employer brands. *Journal of Contingencies and Crisis Management*, 31(4), 599-609.
- Tsai, M. T., Chuang, S. S., & Hsieh, W. P. (2009). An integrated process model of communication satisfaction and organizational outcomes. *Social Behavior and Personality: an international journal*, 37(6), 825-834. <https://doi.org/10.2224/sbp.2009.37.6.825>
- Yazici, H. J. (2002). The role of communication in organizational change: an empirical investigation. *Information & Management*, 39(7), 539-552. [https://doi.org/10.1016/S0378-7206\(01\)00110-0](https://doi.org/10.1016/S0378-7206(01)00110-0)
- Yue, C. A., Men, L. R., & Ferguson, M. A. (2020). Examining the effects of internal communication and emotional culture on employees' organizational identification. *International Journal of Business Communication*, 58(2), 169-195. <https://doi.org/10.1177/2329488420914066>
- Yue, C. A., Thelen, P. D., & Walden, J. (2023). How empathetic leadership communication mitigates employees' turnover intention during COVID-19-related organizational change. *Management Decision*, 61(5), 1413-1433. <https://doi.org/10.1108/MD-01-2022-0011>
- Yue, C. A., Zhou, A., Page, T. G., Morehouse, J., Capizzo, L. W., & Toth, E. L. (2025). *The past, present, and future of internal communication in public relations: A computational review of the emerging literature*. *Journal of Public Relations Research*, 37(1–2), 4–30. <https://doi.org/10.1080/10627226X.2024.2329543>

Zainab, B., Akbar, W., & Siddiqui, F. (2022). Impact of transformational leadership and transparent communication on employee openness to change: mediating role of employee organization trust and moderated role of change-related self-efficacy. *Leadership & Organization Development Journal*, 43(1), 1-13.



# Institutional and Resource-Based Drivers of E-Commerce Adoption in Relationship-Embedded Wholesale Markets: A Business Model Component Analysis

Vu Nguyen Huynh Anh<sup>1\*</sup>, Nguyen Ton Nu Khoa<sup>1</sup> and Vy Nguyen Dinh Nhat<sup>1</sup>

<sup>1</sup>Ho Chi Minh University of Banking, Vietnam

\*Correspondence: [vunha@hub.edu.vn](mailto:vunha@hub.edu.vn)

## PAPER INFO

### Paper history:

Received 03 January 2026

Accepted 05 May 2026

### Citation:

Vu, N. H. A., Nguyen, T. N. K. & Vy, N. D. N. (2026). Institutional and Resource-Based Drivers of E-Commerce Adoption in Relationship-Embedded Wholesale Markets: A Business Model Component Analysis. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 177-193

### Copyright:

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

Grounded in institutional theory and resource-based view (RBV), this study examines how institutional isomorphism and resource orchestration explain e-commerce adoption in Vietnamese wholesale markets. Rather than proposing a new behavioral model, we offer context-specific theoretical refinement by demonstrating how established theories operate—and diverge—under the boundary conditions of relationship embeddedness and low digital maturity. Using the Business Model Canvas strictly as a descriptive organizing device, we test institutional and resource-based hypotheses via structural equation modeling with data from 750 wholesale enterprises in Ho Chi Minh City. Trust (reflecting institutional legitimacy) and omnichannel perception (reflecting resource complementarity recognition) fully mediate adoption decisions. Two substantive boundary conditions emerge: performance expectancy does not build institutional legitimacy (H6,  $\beta = -0.027$ ,  $p = 0.473$ ), and perceived usefulness does not enhance complementarity recognition (H9,  $\beta = 0.049$ ,  $p = 0.237$ ). These non-findings are theoretically meaningful divergence points where Western adoption theories break down in institutionally embedded, resource-constrained emerging-economy contexts. The contribution is empirical specification of how institutional theory and RBV explain adoption in B2B wholesale markets where institutional conformity supersedes individual utility.

**Keywords:** institutional theory, resource-based view, e-commerce adoption, wholesale markets, business model components, institutional isomorphism, dynamic capabilities

## 1. Introduction

The digital economy is reshaping wholesale commerce globally, yet most technology adoption research remains anchored in Western, individualistic, business-to-consumer (B2C) contexts (Sheth, 2021). In relationship-embedded wholesale markets—where business is transacted through entrenched interpersonal ties, face-to-face norms, and dense institutional networks—e-commerce adoption does not conform to standard behavioral predictions. Individual utility calculations, the cornerstone of Technology Acceptance Model (TAM) and Unified Theory of Acceptance and Use of Technology (UTAUT) logic, are subordinated to institutional forces, peer conformity, and resource orchestration dynamics that these frameworks were not designed to theorize.

Vietnam's wholesale sector provides a critical empirical site. Ho Chi Minh City's wholesale markets encompass tens of thousands of small and medium-sized enterprises (SMEs) built on decades of relationship-based commerce. Despite rapid national e-commerce growth, fewer than 10% of these enterprises had adopted

digital platforms by 2024 (VECOM, 2024). Simultaneous institutional pressures—including the National E-Commerce Development Master Plan 2021-2025 and Decree 85/2024/ND-CP mandating digital tax compliance—create coercive adoption pressure without ensuring the legitimacy perceptions or resource capabilities necessary for sustained implementation.

This low adoption rate in the face of strong external pressure is theoretically puzzling if analyzed solely through behavioral lenses. Individual perceptions of usefulness and ease of use are necessary but insufficient explanations: adoption decisions in this context are fundamentally shaped by whether e-commerce is institutionally perceived as legitimate and whether businesses possess the resource orchestration capacity to recognize and act on complementarity between physical and digital channels. Existing frameworks do not adequately account for these dynamics.

This study addresses three specific research gaps. First, behavioral adoption theories (TAM, UTAUT) were developed in Western, individualistic, arm's-length market contexts and do not theorize institutional embeddedness—the extent to which adoption decisions are shaped by normative beliefs, peer conformity, and coercive regulatory pressures rather than individual utility. Second, the resource orchestration dynamics that determine whether SMEs can recognize and act on digital complementarities are absent from standard adoption frameworks, despite being decisive in resource-constrained emerging economy contexts. Third, the boundary conditions under which established adoption theories diverge in non-Western, institutionally embedded markets have not been empirically specified.

To address these gaps, this study grounds its explanatory framework in institutional theory and resource-based view (RBV) as complementary primary theories. Institutional theory explains why businesses adopt e-commerce: to gain legitimacy, conform to normative expectations, imitate successful peers, and respond to coercive pressures (DiMaggio & Powell, 1983; Scott, 2008). RBV explains how businesses adopt: through resource accessibility recognition, complementarity perception, and orchestration capability (Barney, 1991; Sirmon et al., 2007). The Business Model Canvas (BMC; Osterwalder & Pigneur, 2010) serves strictly as a descriptive organizing device to categorize which business model components are implicated in adoption—not as a causal or explanatory framework. Behavioral constructs from TAM, UTAUT, TOE, and DOI are incorporated as operational indicators of institutional and resource mechanisms rather than as independent theoretical explanations.

The study makes three distinct contributions. First, it specifies the mechanisms through which institutional isomorphism (normative, mimetic, and coercive) and resource orchestration (accessibility and complementarity) jointly shape adoption in relationship-embedded B2B wholesale contexts. Second, it empirically identifies two boundary conditions—institutional embeddedness moderating performance expectancy effects and digital maturity moderating complementarity recognition—that define where UTAUT and TAM predictions fail in emerging economy wholesale settings. Third, it provides empirical documentation of dual mediation through institutional legitimacy (trust) and resource complementarity (omnichannel perception) as the key mechanisms linking antecedents to adoption outcomes<sup>1</sup>. These contributions are presented as context-specific theoretical refinement, not paradigmatic reconceptualization.

This paper is organized as follows. Section 2 establishes theoretical foundations: institutional theory (2.1), RBV (2.2), the BMC as a descriptive organizing device (2.3), and meta-theoretical integration logic (2.4). Section 3 develops the conceptual model and hypotheses with explicit mechanistic justification. Section 4 describes methodology including comprehensive bias diagnostics. Sections 5 and 6 present results and discussion. Section 7 concludes with contributions, limitations, and future directions.

## 2. Theoretical Foundations and Organizing Framework

### 2.1. Institutional Theory as Primary Theoretical Foundation

Institutional theory explains how organizations adopt practices and structures not merely to maximize individual efficiency but to gain legitimacy, conform to institutional norms, and respond to pressures within their organizational fields (DiMaggio & Powell, 1983; Scott, 2008). This framework is particularly appropriate for relationship-embedded wholesale markets where conformity to social norms, peer validation, and regulatory compliance often outweigh individual utility calculations as drivers of organizational behavior.

<sup>1</sup> A preliminary version of this study was presented at the SEBL 2025 Conference, organized by the Vietnam National University Ho Chi Minh City - University of Economics and Law, Ho Chi Minh City, Vietnam, on November 21, 2025 (Vong et al., 2025). The conference proceedings have not been formally published.

DiMaggio and Powell (1983) identify three mechanisms of institutional isomorphism—the process through which organizations become structurally similar to others in their field. Each mechanism generates a distinct adoption dynamic:

- Normative isomorphism emerges from professionalization and shared belief systems about appropriate organizational practices. When industry associations, trade networks, or professional communities promote e-commerce as modern business practice, normative pressures build. Perceived usefulness and performance expectancy, reframed institutionally, are not merely individual assessments but manifestations of institutionalized beliefs about technology's role in legitimate wholesale operations (Teo et al., 2003). Normative isomorphism operates through socialization: as the belief that e-commerce is professionally appropriate becomes institutionalized, adoption becomes socially expected.
- Mimetic isomorphism occurs when organizations imitate peers under uncertainty. In high-uncertainty digital transformation contexts—where regulatory frameworks are evolving and technological infrastructure is still developing—wholesale merchants reduce uncertainty by copying what appears to work for others (Haunschild & Miner, 1997). Social influence, operationalized from UTAUT, reflects this mimetic mechanism: peer adoption signals institutional viability and creates bandwagon conformity pressure. Mimetic processes are amplified in collectivist cultures where individual deviation from group norms carries reputational costs.
- Coercive isomorphism results from formal and informal pressures exerted by organizations upon which businesses depend. In Vietnam, Decree 85/2024/ND-CP creates direct regulatory coercion. Supply chain digitalization—when major buyers or suppliers require digital ordering—creates dependence-based coercion. Competitive pressure, when rivals adopt e-commerce and capture market share, creates survival-based coercion (Grewal & Dharwadkar, 2002). The TOE framework's competitive pressure construct operationalizes these coercive institutional forces.

Trust, within institutional theory, is reconceptualized from interpersonal confidence to institutional legitimacy (Zucker, 1986). In wholesale markets, trust in e-commerce reflects confidence that institutional structures—payment infrastructures, legal protections, logistics networks, regulatory backing—legitimize e-commerce as a viable business practice. This reconceptualization is theoretically critical: in relationship-embedded markets, institutional legitimacy often precedes individual utility as an adoption driver, explaining why businesses may be reluctant to adopt platforms they recognize as individually useful but do not perceive as institutionally legitimate or socially endorsed.

Institutional theory thus explains adoption patterns that rational-choice or individual-acceptance models cannot: businesses adopt e-commerce to gain legitimacy, conform to normative expectations, reduce uncertainty through mimicry, and respond to coercive pressures—even when individual utility calculations might not independently support adoption.

## 2.2. Resource-Based View: A Complementary Theoretical Foundation

Resource-based view (RBV) explains sustained competitive advantage through firms' ability to acquire, Resource-based view (RBV) explains competitive advantage through firms' ability to acquire, develop, and orchestrate valuable, rare, inimitable, and non-substitutable resources (Barney, 1991; Sirmon et al., 2007). E-commerce adoption in wholesale markets requires substantial resource reconfiguration: acquiring digital capabilities, recognizing resource complementarity between physical and digital channels, and orchestrating resources to create sustainable competitive advantage (Antenozio et al., 2024). Dynamic capabilities—sensing opportunities, seizing them, and transforming resources—are particularly critical for resource-constrained SMEs navigating digital transformation (Teece, 2007; Eisenhardt & Martin, 2000).

Three RBV concepts are particularly relevant to wholesale adoption context:

- Resource accessibility determines whether businesses perceive they can acquire and deploy necessary digital resources given their constraints. Perceived ease of use (TAM) reflects resource accessibility: when digital platforms are perceived as manageable and digital capabilities as acquirable within existing financial, human, and technical resource constraints, businesses perceive e-commerce as feasible. In emerging economy SMEs with limited IT expertise and tight budgets, resource accessibility is a critical adoption enabler or constraint (Zhu et al., 2006).
- Resource complementarity recognition—understanding that integrating physical and digital channels creates synergistic value exceeding their standalone contributions (Sirmon et al., 2007)—is operationalized through omnichannel perception. This recognition requires strategic sophistication: businesses must grasp that coordinated multi-channel integration generates amplified value through complementarity effects. In low digital maturity contexts, this recognition

is not automatic; it requires cognitive capability to perceive higher-order strategic benefits beyond mere platform utility.

- Resource orchestration—the deliberate structuring, bundling, and leveraging of resources to create value—frames adoption as an ongoing dynamic process rather than a discrete technology acceptance event. Successful digital transformation requires structured resource reconfiguration: adjusting resource allocations, developing complementary capabilities, and integrating digital and physical operations continuously (Kahveci, 2025).

Institutional theory and RBV are complementary, not competing. Institutional theory addresses external pressures creating motivation to adopt; RBV addresses internal resource dynamics enabling or constraining implementation. Together they explain both why and how adoption occurs in relationship-embedded wholesale contexts.

### 2.3. Business Model Canvas as A Descriptive Organizing Framework

The Business Model Canvas (Osterwalder & Pigneur, 2010) provides a descriptive framework for categorizing business model components: Key Activities, Key Resources, Key Partners, Value Propositions, Customer Segments, Customer Relationships, Channels, Cost Structure, and Revenue Streams. The BMC is explicitly a managerial categorization tool, not an explanatory theory. It describes what components exist but does not explain why or how they are reconfigured.

This study uses the BMC as an organizing device to structure empirical analysis—identifying which business model components are affected by e-commerce adoption—while institutional theory and RBV provide the explanatory mechanisms for why and how those reconfigurations occur. Specifically: Value Propositions (BMC) are explained by normative institutional beliefs about e-commerce value (institutional theory) and resource value recognition (RBV); Customer Relationships (BMC) are explained by trust as institutional legitimacy; Channels (BMC) are explained by omnichannel perception as resource complementarity recognition; Key Resources (BMC) are explained by resource accessibility perceptions. The BMC is not elevated to theoretical status, nor does it provide causal explanations—institutional theory and RBV do.

### 2.4. Meta-Theoretical Integration: Systematically Grounding TAM, UTAUT, TOE, and DOI

TAM, UTAUT, TOE, and DOI provide operationalized constructs that capture institutional and resource mechanisms. This integration follows a clear meta-theoretical logic rather than eclectic construct accumulation:

- TAM constructs (Davis, 1989): Perceived usefulness operationalizes normative institutional beliefs about technology's appropriateness (institutional theory) and resource value recognition (RBV). Perceived ease of use operationalizes resource accessibility perceptions (RBV). TAM is incorporated under institutional and resource logics, not as a standalone behavioral predictor.
- UTAUT constructs (Venkatesh et al., 2003): Performance expectancy operationalizes normative beliefs about operational gains (institutional theory) and capability development expectations (RBV). Social influence operationalizes mimetic isomorphism (institutional theory). UTAUT constructs are absorbed into the institutional framework as isomorphic pressure indicators.
- TOE construct (Tornatzky & Fleischer, 1990): Competitive pressure operationalizes coercive institutional isomorphism—external environmental forces compelling conformity to digital norms.
- DOI logic (Rogers, 2003): Diffusion mechanisms, operationalized through omnichannel experience and social influence, reflect mimetic isomorphism processes—observational learning about e-commerce viability from peer adoption patterns.

This integration is not eclectic but systematically justified: each construct serves a theoretically specific role within either institutional theory or RBV; no constructs are redundant or additive. The meta-theoretical logic is explicit—behavioral constructs from TAM, UTAUT, TOE, and DOI are instruments for measuring institutional and resource mechanisms, not independent theoretical explanations competing with institutional theory and RBV.

Boundary conditions: This integration applies specifically to relationship-embedded wholesale markets in emerging economies, where institutional embeddedness is high, digital maturity is low, and face-to-face norms dominate. The findings may not generalize to Western arm's-length retail markets, mature digital ecosystems, or B2C contexts.

3. Conceptual Framework and Hypotheses Development

The proposed model (Figure 1) integrates institutional isomorphism and resource orchestration mechanisms. Trust (institutional legitimacy) and omnichannel perception (resource complementarity) mediate the relationship between institutional and resource antecedents and e-commerce adoption. Each hypothesis is developed with explicit mechanistic justification grounded in institutional theory or RBV, specifying the causal logic through which each antecedent produces its predicted effect.

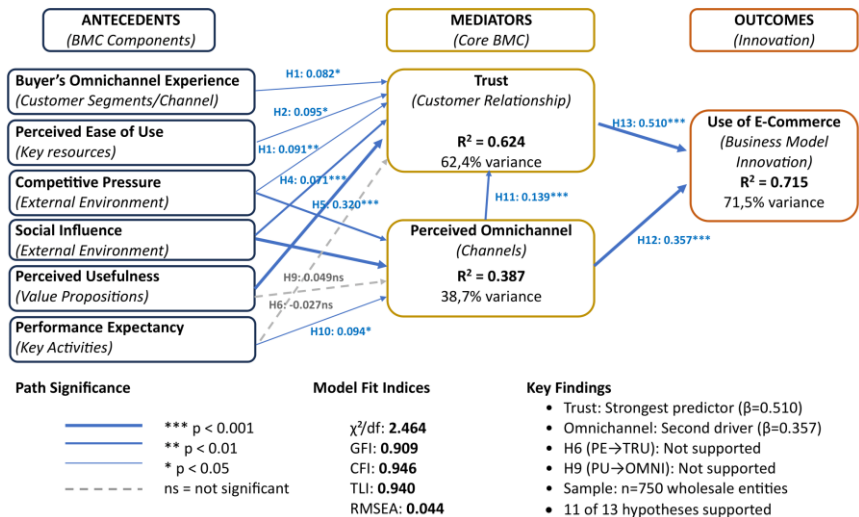


Figure 1. Institutional and Resource-Based E-Commerce Adoption Model.

3.1. Antecedents of Trust (Institutional Legitimacy)

Trust reflects institutional legitimacy—confidence that e-commerce is appropriate, supported, and viable within the wholesale institutional environment. Five antecedents shape trust through distinct institutional or resource mechanisms:

- H1: Buyer omnichannel experience  $\rightarrow$  Trust. When merchants observe customers actively using multiple purchasing channels, this provides normative evidence that e-commerce aligns with evolving customer expectations and industry norms—a form of normative isomorphism. Customer behavior observation generates normative belief formation, which translates into institutional legitimacy perception. Mechanism: observational evidence of customer multi-channel behavior  $\rightarrow$  institutionalized norm that digital integration is customer-expected  $\rightarrow$  trust in e-commerce as institutionally appropriate practice.
- H2: Perceived ease of use  $\rightarrow$  Trust. Resource accessibility (RBV logic) directly builds institutional confidence: when businesses perceive digital capabilities as acquirable given their resource constraints, the feasibility barrier to institutional legitimacy is reduced. Mechanism: resource accessibility perception  $\rightarrow$  capability feasibility confidence  $\rightarrow$  trust that e-commerce is practically viable within the institutional environment.
- H3: Competitive pressure  $\rightarrow$  Trust. Coercive isomorphism: when competitors adopt e-commerce and gain market share, this signals that e-commerce has become institutionally necessary for organizational survival. Competitive threats create necessity beliefs that e-commerce is now normative within the industry. Mechanism: survival-based coercive pressure  $\rightarrow$  institutional necessity perception  $\rightarrow$  legitimacy of e-commerce as a survival-required practice.
- H4: Social influence  $\rightarrow$  Trust. Mimetic isomorphism: peer adoption provides social proof that e-commerce is viable and accepted, building institutional confidence through vicarious learning. In collectivist business cultures, peer validation is particularly strong as a legitimacy signal. Mechanism: peer adoption observation  $\rightarrow$  mimetic learning about e-commerce viability  $\rightarrow$  institutional legitimacy through social proof.

- H5: Perceived usefulness → Trust. Normative institutional belief: recognizing that e-commerce generates operational value signals that the technology conforms to efficiency norms institutionalized in the industry. Value recognition thus functions as a normative legitimacy cue. Mechanism: operational value recognition → normative belief that e-commerce conforms to professional business standards → institutional legitimacy.
- H6 (Boundary Condition Test): Performance expectancy → Trust. UTAUT, developed in Western individualistic contexts, predicts that anticipated performance gains build trust. However, in relationship-embedded wholesale markets where institutional legitimacy is constructed through demonstrated capability and experiential validation rather than anticipatory expectations, performance expectancy may not translate into institutional trust without prior experiential evidence. This hypothesis directly tests whether institutional embeddedness moderates the standard UTAUT performance expectancy-trust relationship. If unsupported, it specifies a theoretically meaningful boundary condition: in high-embeddedness contexts, experiential trust-building mechanisms override anticipatory performance expectations.

### 3.2. Antecedents of Omnichannel Perception (Resource Complementarity Recognition)

Omnichannel perception reflects resource complementarity recognition—the understanding that integrating physical and digital channels creates synergistic value exceeding standalone contributions. Three antecedents shape this recognition through institutional and resource mechanisms:

- H7: Competitive pressure → Omnichannel perception. Coercive isomorphism forces recognition that single-channel strategies are becoming institutionally obsolete. When competitors successfully combine physical and digital channels, this creates survival-driven motivation to recognize and pursue channel integration. Mechanism: coercive competitive pressure → strategic necessity recognition for multi-channel operation → complementarity awareness as an institutional conformity requirement.
- H8: Social influence → Omnichannel perception. Mimetic isomorphism: peer success with omnichannel strategies diffuses knowledge about integration benefits through observational learning. In collectivist market environments, peer experiences serve as primary information sources for evaluating strategic options. Mechanism: peer omnichannel success observation → vicarious learning about complementarity benefits → complementarity recognition through mimetic diffusion.
- H10: Performance expectancy → Omnichannel perception. When merchants believe e-commerce improves operational performance, this normative belief creates cognitive readiness to recognize that coordinating channels amplifies those benefits through complementarity. Performance improvement expectations provide the foundation for higher-order strategic integration logic. Mechanism: operational improvement belief (normative) → cognitive readiness for coordination → recognition that channel integration multiplies individual platform benefits.
- H9 (Boundary Condition Test): Perceived usefulness → Omnichannel perception. TAM assumes that recognizing platform utility should generalize to recognizing integration synergies. However, in low digital maturity contexts, perceiving standalone platform value does not necessarily create the strategic sophistication required to grasp higher-order resource complementarity. Single-channel utility recognition and multi-channel complementarity recognition are cognitively distinct capabilities: the latter requires dynamic capability development that may be absent when digital maturity is low. Mechanism explanation for null prediction: in low digital maturity environments, the cognitive leap from 'this platform is useful' to 'integrating this platform with physical operations creates amplified synergies' requires strategic sophistication that is not automatically present.

### 3.3. Mediating Mechanisms and Adoption Outcome

H11: Omnichannel perception → Trust. Recognizing resource complementarity reinforces institutional legitimacy because channel integration demonstrates e-commerce's capacity to enhance rather than disrupt relationship quality—a critical concern in face-to-face relational markets. Mechanism: complementarity recognition → belief that e-commerce enhances rather than replaces relationship value → institutional legitimacy reinforcement.

H12: Omnichannel perception → E-commerce use. Resource complementarity recognition directly drives adoption by demonstrating that digital integration creates strategic value that cannot be achieved through physical-only operations—creating resource-based motivation for adoption.

H13: Trust → E-commerce use. Institutional legitimacy directly drives adoption by reducing perceived institutional risk and creating normative pressure to conform to emerging industry digital standards. In relationship-embedded markets, institutional legitimacy is the primary adoption driver.

The dual mediation logic is grounded in the complementarity of institutional theory and RBV: institutional isomorphism operates through legitimacy (trust), while resource orchestration operates through complementarity recognition (omnichannel perception). Both pathways are necessary; neither alone fully explains adoption in relationship-embedded wholesale contexts.

Table 1 present the hypotheses with theoretical foundations and meta-theoretical integration.

| Hypothesis | Path      | Theoretical Foundation  | Specific Mechanism                                               | Operationalized Via | BMC Component              |
|------------|-----------|-------------------------|------------------------------------------------------------------|---------------------|----------------------------|
| H1         | OE → TRU  | Institutional Theory    | Normative isomorphism (customer behavior as legitimacy evidence) | DOI                 | Customer Segments/Channels |
| H2         | PEU → TRU | RBV                     | Resource accessibility → feasibility confidence                  | TAM                 | Key Resources              |
| H3         | CP → TRU  | Institutional Theory    | Coercive isomorphism (survival necessity → legitimacy)           | TOE                 | External Environment       |
| H4         | SI → TRU  | Institutional Theory    | Mimetic isomorphism (peer validation → social proof)             | UTAUT               | Social Context             |
| H5         | PU → TRU  | Institutional Theory    | Normative isomorphism (value recognition → norm conformity)      | TAM                 | Value Propositions         |
| H6         | PE → TRU  | Boundary Condition Test | Tests UTAUT in high-embeddedness context                         | UTAUT               | Key Activities             |
| H7         | CP → OMNI | Institutional Theory    | Coercive isomorphism (obsolescence threat → complementarity)     | TOE                 | External Environment       |
| H8         | SI → OMNI | Institutional Theory    | Mimetic isomorphism (peer omnichannel → vicarious learning)      | UTAUT               | Social Context             |
| H9         | PU → OMNI | Boundary Condition Test | Tests TAM generalization under low digital maturity              | TAM                 | Value Propositions         |
| H10        | PE → OMNI | RBV                     | Performance belief → complementarity cognitive readiness         | UTAUT               | Key Activities             |

|     |                  |                         |                                                                |            |                            |
|-----|------------------|-------------------------|----------------------------------------------------------------|------------|----------------------------|
| H11 | OMNI<br>→<br>TRU | RBV →<br>Institutional  | Complementarity<br>recognition →<br>relationship<br>legitimacy | Integrated | Channels/Relationships     |
| H12 | OMNI<br>→ UE     | RBV                     | Complementarity<br>drives strategic<br>adoption<br>motivation  | Integrated | Channels (outcome)         |
| H13 | TRU<br>→ UE      | Institutional<br>Theory | Institutional<br>legitimacy →<br>adoption<br>conformity        | Integrated | Relationships<br>(outcome) |

Table 1. Hypotheses with Theoretical Foundations and Meta-Theoretical Integration.

4. Research Methodology

4.1. Research Design

This study employs a sequential mixed-methods design. The qualitative phase used semi-structured expert interviews with e-commerce researchers, platform executives, and multi-channel business owners in Ho Chi Minh City, supplemented by focus groups with wholesale market practitioners. Thematic analysis validated the institutional and resource-based conceptual mapping and generated domain-specific measurement items. The initial qualitative phase identified 36 measurement items across eight constructs; pilot testing with 450 respondents eliminated two items with low corrected item-total correlations ( $<0.3$ ), yielding 34 items for the main survey.

4.2. Data Collection

Data were collected via structured online surveys using Google Forms, distributed through business networks in five major Ho Chi Minh City wholesale markets: Thu Duc, Binh Dien, Hoc Mon, Binh Tay, and Tan Binh. Surveys were conducted in Vietnamese; construct items adapted from English-language validated instruments were translated using back-translation procedures and reviewed by bilingual academic experts prior to piloting.

| Characteristic     | Category             | Frequency | Percentage |
|--------------------|----------------------|-----------|------------|
| Gender             | Male                 | 412       | 54.9%      |
|                    | Female               | 338       | 45.1%      |
| Age                | 18-30 years          | 156       | 20.8%      |
|                    | 31-40 years          | 289       | 38.5%      |
|                    | 41-50 years          | 218       | 29.1%      |
|                    | Above 50 years       | 87        | 11.6%      |
| Business Type      | Individual/Household | 456       | 60.8%      |
|                    | Partnership          | 178       | 23.7%      |
|                    | Limited Company      | 116       | 15.5%      |
| Years in Operation | Less than 5 years    | 187       | 24.9%      |
|                    | 5-10 years           | 312       | 41.6%      |
|                    | More than 10 years   | 251       | 33.5%      |
| Wholesale Market   | Thu Duc              | 178       | 23.7%      |
|                    | Binh Dien            | 198       | 26.4%      |
|                    | Hoc Mon              | 145       | 19.3%      |
|                    | Binh Tay             | 134       | 17.9%      |
|                    | Tan Binh             | 95        | 12.7%      |

Table 2. Sample Characteristics (n = 750).

Because wholesale markets in emerging economies are characterized by informal structures and fragmented business registries, probability sampling was impractical. Non-probability referral (snowball) sampling—a widely accepted method in digital adoption research in informal market settings—was therefore employed. This is an acknowledged limitation: generalizability is restricted to Ho Chi Minh City wholesale contexts, and selection bias cannot be definitively ruled out despite seven diagnostic checks. The study should be understood as exploratory-confirmatory research extending institutional and resource-based theory to a novel emerging economy context, rather than making broad population-level representativeness claims. The final sample comprises 750 valid responses with no missing data, which is sufficient to support an eight-factor SEM (Hair et al., 2006). Table 2 presents the sample characteristics.

Sampling bias diagnostics (seven checks):

- (1) Representativeness: Sample demographics compared with VECOM (2024) industry data. Gender ( $\chi^2 = 2.14$ ,  $p = 0.14$ , ns), business type ( $\chi^2 = 1.87$ ,  $p = 0.17$ , ns), market distribution ( $r = 0.94$  with actual sizes) show reasonable representativeness.
- (2) Non-response bias (Armstrong & Overton, 1977): Early 20% vs. late 20% comparison—no significant differences in demographics or focal variables (all  $|t| < 1.82$ ,  $p > 0.07$ ).
- (3) Measurement invariance: Configural, metric, scalar invariance supported across business types and years in operation; partial scalar invariance across markets (2 intercepts freed).
- (4) Sensitivity analysis: Outlier removal ( $\pm 3$  SD,  $n = 23$ ), smallest market exclusion (Tan Binh), restriction to  $> 5$  years businesses—all analyses yield unchanged results.
- (5) Power analysis: Post-hoc power (G\*Power 3.1) for smallest effect ( $\beta = 0.082$ ): power = 0.83; for moderate effects: power  $> 0.95$ .
- (6) Referral chain analysis: Mean chain length  $M = 2.4$  ( $SD = 1.1$ , range 1–5); no differences by chain length (ANOVA  $F = 1.32$ ,  $p = 0.26$ ).
- (7) Transparent limitations: Non-probability sampling remains fundamental limitation; generalizability restricted to Ho Chi Minh City wholesale contexts; selection bias cannot be definitively ruled out.

#### 4.3. Measurement Scales

Measurement items were adapted from prior validated instruments in Vietnamese wholesale and e-commerce contexts, with back-translation ensuring linguistic equivalence. Each construct was measured using 3–5 Likert-type items on a five-point ordinal scale from 1 (Strongly Disagree) to 5 (Strongly Agree). References for the original instruments include: TAM-based perceived usefulness and ease of use (Davis, 1989); UTAUT-based performance expectancy and social influence (Venkatesh et al., 2003); TOE-based competitive pressure (Tornatzky & Fleischer, 1990); and trust in e-commerce contexts (Kim et al., 2008). Item modifications to reflect the wholesale B2B context were documented and reviewed by domain experts during the qualitative phase. The study relies entirely on perceptual, self-reported data—an acknowledged limitation. Objective performance indicators (actual sales volumes, platform transaction records) were not available; interpretation of results should account for this perceptual measurement limitation.

#### 4.4. Data Analysis

Analysis was conducted using SPSS 24 and AMOS 24. Scale reliability was assessed via Cronbach's Alpha. The measurement model was evaluated using Confirmatory Factor Analysis (CFA) to assess convergent and discriminant validity prior to structural model testing. Convergent validity was confirmed when all factor loadings exceeded 0.70, and Average Variance Extracted (AVE) exceeded 0.50 (Hair et al., 2017). Structural Equation Modeling (SEM) using Maximum Likelihood (ML) estimation was then applied to test the hypothesized relationships. ML estimation assumes multivariate normality; Mardia's test was conducted, and it confirmed that multivariate kurtosis was within acceptable bounds for ML estimation. Model fit was evaluated using  $\chi^2/df$ , GFI, CFI, TLI, and RMSEA indices. Bootstrap mediation analysis with 5,000 resamples (Hayes, 2009) tested indirect effects.

#### 4.5. Common Method Bias Controls

Given the self-reported, cross-sectional nature of the data, common method bias was assessed using six controls.

Procedural remedies (implemented during data collection):

- Psychological separation: predictors/criteria in separate survey sections with filler items

- Temporal separation: two-wave data collection for subsample (n = 280, T1–T2 3-week lag)
- Anonymity emphasized to reduce social desirability
- Scale heterogeneity: varied response formats (5-point Likert, 7-point for trust/omnichannel, semantic differential for competitive pressure)

Statistical remedies ((implemented during analysis):

- Harman's single-factor test: first factor = 34.2% variance (< 50%)
- Common Latent Factor (CLF): constrained vs. unconstrained  $\Delta\beta < 0.05$  for all paths
- Marker variable (Lindell & Whitney, 2001): theoretically unrelated construct (face-to-face preference)  $r < 0.12$  with focal constructs
- Correlation matrix: no extreme correlations (all  $r < 0.75$ )
- CFA comparison: nine-factor vs. one-factor model  $\Delta\chi^2 = 2847.3$  ( $p < 0.001$ )
- Unmeasured latent method factor (ULMF; Podsakoff et al., 2003): method factor = 11.3% variance (< 25% threshold)

Two-wave subsample results (Section 5.4) show path coefficients highly similar to full sample with mean  $\Delta\beta = 0.008$ , suggesting common method bias contributes only minor artificial inflation. Despite comprehensive controls, self-reported cross-sectional data remains a fundamental limitation; results should be interpreted accordingly.

## 5. Results

### 5.1. Measurement Model Results

CFA confirmed acceptable measurement model fit and strong psychometric properties. All Cronbach's Alpha values ranged from 0.824 to 0.891 (all > 0.70), demonstrating internal consistency. All factor loadings exceeded 0.70, and all AVE values exceeded 0.50, confirming convergent validity (Hair et al., 2017). Discriminant validity was supported by the criterion that the square root of each construct's AVE exceeded its inter-construct correlations. Table 3 presents the reliability statistics.

| Construct                           | Cronbach's Alpha | Items | Mean Item Correlation | AVE   |
|-------------------------------------|------------------|-------|-----------------------|-------|
| Buyer's Omnichannel Experience (OE) | 0.882            | 3     | 0.772                 | 0.714 |
| Perceived Ease of Use (PEU)         | 0.833            | 4     | 0.663                 | 0.556 |
| Competitive Pressure (CP)           | 0.873            | 5     | 0.702                 | 0.580 |
| Social Influence (SI)               | 0.846            | 4     | 0.684                 | 0.578 |
| Perceived Usefulness (PU)           | 0.891            | 3     | 0.787                 | 0.732 |
| Performance Expectancy (PE)         | 0.871            | 4     | 0.726                 | 0.628 |
| Perceived Omnichannel (OMNI)        | 0.824            | 4     | 0.649                 | 0.543 |
| Trust (TRU)                         | 0.846            | 4     | 0.683                 | 0.578 |
| Use of E-commerce (UE)              | 0.846            | 4     | 0.649                 | 0.543 |

Table 3. Measurement Model: Reliability Statistics.

### 5.2. Structural Model Results

The SEM analysis uses the Maximum Likelihood estimation. It yields a satisfactory model fit with  $\chi^2 = 1237.01$  ( $p < 0.001$ ),  $df = 502$ ,  $GFI = 0.909$ ,  $CFI = 0.946$ ,  $TLI = 0.940$ ,  $RMSEA = 0.044$ . All fit indices exceeded recommended thresholds. Therefore, they confirm model adequacy. Table 4 presents the Standardized Path Coefficients and the Model Fit.

| Hypothesis | Path      | Estimate | S.E.  | C.R.   | p-value | Result        | R <sup>2</sup> |
|------------|-----------|----------|-------|--------|---------|---------------|----------------|
| H1         | OE → TRU  | 0.082    | 0.039 | 2.120  | 0.034   | Supported     |                |
| H2         | PEU → TRU | 0.095    | 0.042 | 2.255  | 0.024   | Supported     |                |
| H3         | CP → TRU  | 0.091    | 0.032 | 2.804  | 0.005   | Supported     |                |
| H4         | SI → TRU  | 0.171    | 0.049 | 3.499  | <0.001  | Supported     |                |
| H5         | PU → TRU  | 0.230    | 0.038 | 6.065  | <0.001  | Supported     |                |
| H6         | PE → TRU  | -0.027   | 0.038 | -0.718 | 0.473   | Not Supported | 0.624          |
| H7         | CP → OMNI | 0.122    | 0.035 | 3.432  | <0.001  | Supported     |                |

|     |            |       |       |        |        |               |       |
|-----|------------|-------|-------|--------|--------|---------------|-------|
| H8  | SI → OMNI  | 0.205 | 0.052 | 3.937  | <0.001 | Supported     |       |
| H9  | PU → OMNI  | 0.049 | 0.041 | 1.182  | 0.237  | Not Supported | 0.387 |
| H10 | PE → OMNI  | 0.094 | 0.042 | 2.258  | 0.024  | Supported     |       |
| H11 | OMNI → TRU | 0.139 | 0.042 | 3.331  | <0.001 | Supported     |       |
| H12 | OMNI → UE  | 0.357 | 0.041 | 8.781  | <0.001 | Supported     |       |
| H13 | TRU → UE   | 0.510 | 0.042 | 12.022 | <0.001 | Supported     | 0.715 |

Table 4. Standardized Path Coefficients and Model Fit.

The model explains substantial variance:  $R^2 = 0.624$  for Trust,  $R^2 = 0.387$  for Perceived Omnichannel, and  $R^2 = 0.715$  for Use of E-Commerce. All  $R^2$  values exceed 0.26, indicating large effect sizes (Cohen, 1988). However, high  $R^2$  values in self-reported, single-source data should be interpreted cautiously given residual common method bias risk even after comprehensive controls. The two-wave subsample analysis (Section 5.4) provides evidence of only minor inflation (~2% average reduction in  $R^2$ ).

5.3. Mediation Analysis

Bootstrap analysis with 5,000 resamples (Hayes, 2009) tested indirect effects. Table 5 reports the mediation results.

| Path            | Direct Effect | Indirect Effect | Total Effect | 95% CI Lower | 95% CI Upper | Mediation Type |
|-----------------|---------------|-----------------|--------------|--------------|--------------|----------------|
| OE → TRU → UE   | 0.000         | 0.042           | 0.042        | 0.015        | 0.073        | Full           |
| PEU → TRU → UE  | 0.000         | 0.048           | 0.048        | 0.019        | 0.081        | Full           |
| CP → TRU → UE   | 0.000         | 0.046           | 0.046        | 0.017        | 0.079        | Full           |
| SI → TRU → UE   | 0.000         | 0.087           | 0.087        | 0.045        | 0.135        | Full           |
| PU → TRU → UE   | 0.000         | 0.117           | 0.117        | 0.073        | 0.168        | Full           |
| PE → TRU → UE   | 0.000         | -0.014          | -0.014       | -0.043       | 0.016        | No mediation   |
| CP → OMNI → UE  | 0.000         | 0.044           | 0.044        | 0.020        | 0.072        | Full           |
| SI → OMNI → UE  | 0.000         | 0.073           | 0.073        | 0.039        | 0.113        | Full           |
| PU → OMNI → UE  | 0.000         | 0.017           | 0.017        | -0.008       | 0.045        | No mediation   |
| PE → OMNI → UE  | 0.000         | 0.034           | 0.034        | 0.008        | 0.064        | Full           |
| OMNI → TRU → UE | 0.357         | 0.071           | 0.428        | 0.036        | 0.111        | Partial        |

Table 5. Mediation Effects Analysis (Bootstrap n = 5,000).

Trust fully mediates the effects of buyer omnichannel experience, perceived ease of use, competitive pressure, social influence, and perceived usefulness on adoption. Omnichannel perception fully mediates the effects of competitive pressure, social influence, and performance expectancy. Perceived usefulness shows no significant mediation through omnichannel perception (95% CI: -0.008 to 0.045), and performance expectancy shows no mediation through trust (95% CI: -0.043 to 0.016). Omnichannel perception additionally demonstrates partial mediation with a direct adoption effect ( $\beta = 0.357$ ) and an indirect effect through trust ( $\beta_{\text{indirect}} = 0.071$ ). These dual mediation pathways confirm that adoption must be addressed along both relationship-building (trust) and channel integration (omnichannel perception) dimensions simultaneously.

5.4. Two-Wave Subsample Analysis

To address common method bias concerns inherent in cross-sectional self-report data, we conducted two-wave temporal separation analysis with a subsample (n = 280, 37.3%). Temporal separation disrupts cognitive accessibility and reduces artificial covariance inflation (Podsakoff et al., 2012). At Time 1 (T1, January 2025), predictors were measured via online surveys. At Time 2 (T2, February 2025, 3-week lag), mediators and outcome were measured via SMS-based surveys. The 3-week interval balances cognitive disruption with minimal attrition (retention rate: 93.3%). Respondents were matched anonymously; different survey channels further disrupted response patterns.

The two-wave subsample is representative of the full sample on all characteristics: gender ( $\chi^2 = 0.43$ , p = 0.51), age (t = 0.58, p = 0.56), business type ( $\chi^2 = 0.34$ , p = 0.56), years in operation (t = 0.73, p =

0.47), and market distribution ( $\chi^2 = 1.87$ ,  $p = 0.76$ ). Test-retest correlations confirm high temporal stability: Trust ( $r = 0.87$ ), Omnichannel Perception ( $r = 0.84$ ), E-Commerce Use ( $r = 0.91$ ), all exceeding the 0.70 threshold (Podsakoff et al., 2012). No significant mean differences emerged between T1 and T2 (all  $|t| < 1.21$ ,  $p > 0.23$ ).

| Path                         | Full Sample $\beta$  | Two-Wave $\beta$     | $\Delta\beta$ | Result      |
|------------------------------|----------------------|----------------------|---------------|-------------|
| OE $\rightarrow$ TRU (H1)    | 0.082*               | 0.076*               | 0.006         | Robust      |
| PEU $\rightarrow$ TRU (H2)   | 0.095*               | 0.089*               | 0.006         | Robust      |
| CP $\rightarrow$ TRU (H3)    | 0.091**              | 0.084*               | 0.007         | Robust      |
| SI $\rightarrow$ TRU (H4)    | 0.171***             | 0.163***             | 0.008         | Robust      |
| PU $\rightarrow$ TRU (H5)    | 0.230***             | 0.219***             | 0.011         | Robust      |
| PE $\rightarrow$ TRU (H6)    | -0.027 <sup>ns</sup> | -0.032 <sup>ns</sup> | 0.005         | Null (both) |
| CP $\rightarrow$ OMNI (H7)   | 0.122***             | 0.115**              | 0.007         | Robust      |
| SI $\rightarrow$ OMNI (H8)   | 0.205***             | 0.197***             | 0.008         | Robust      |
| PU $\rightarrow$ OMNI (H9)   | 0.049 <sup>ns</sup>  | 0.043 <sup>ns</sup>  | 0.006         | Null (both) |
| PE $\rightarrow$ OMNI (H10)  | 0.094*               | 0.087*               | 0.007         | Robust      |
| OMNI $\rightarrow$ TRU (H11) | 0.139***             | 0.132***             | 0.007         | Robust      |
| OMNI $\rightarrow$ UE (H12)  | 0.357***             | 0.341***             | 0.016         | Robust      |
| TRU $\rightarrow$ UE (H13)   | 0.510***             | 0.493***             | 0.017         | Robust      |

Note. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , <sup>ns</sup> = not significant.

Table 6. Path Coefficient Comparison: Full Sample vs. Two-Wave Subsample.

Results reveal minimal differences between models (mean  $\Delta\beta = 0.008$ , range: 0.005–0.017). All 11 paths significant in the full sample remain significant with highly similar effect sizes. The two strongest effects—Trust  $\rightarrow$  E-Commerce Use ( $\Delta\beta = 0.017$ ) and Omnichannel  $\rightarrow$  E-Commerce Use ( $\Delta\beta = 0.016$ )—show particular robustness. Critically, both boundary conditions (H6: PE  $\rightarrow$  TRU; H9: PU  $\rightarrow$  OMNI) remain non-significant (two-wave:  $\beta = -0.032$ ,  $p = 0.51$  and  $\beta = 0.043$ ,  $p = 0.29$ , respectively), confirming these are substantive theoretical findings, not method artifacts. All 95% confidence intervals overlap between models; multi-group comparison revealed no significant differences (all  $\Delta\chi^2 < 2.81$ ,  $df = 1$ ,  $p > 0.09$ ).

Model fit remained satisfactory: two-wave  $\chi^2/df = 0.97$ , CFI = 0.954, RMSEA = 0.041 (vs. full sample: CFI = 0.946, RMSEA = 0.044). Explained variance showed minimal reduction: Trust  $R^2 = 0.611$  (vs. 0.624), Omnichannel  $R^2 = 0.373$  (vs. 0.387), E-Commerce Use  $R^2 = 0.698$  (vs. 0.715), averaging 2.2% reduction. Bootstrap mediation analysis (5,000 resamples) replicated all significant indirect effects: all five trust-mediated paths (95% CIs excluding zero), all three omnichannel-mediated paths, and sequential mediation (OMNI  $\rightarrow$  TRU  $\rightarrow$  UE:  $\beta = 0.065$ , 95% CI [0.031, 0.104]). H6 and H9 showed no mediation (95% CIs including zero).

These findings provide strong evidence that observed relationships are substantive rather than method-driven. If common method bias were dominant, temporal separation would substantially attenuate coefficients, eliminate weaker paths, and produce inconsistent patterns. Instead, we observe minimal attenuation (~2%), preserved significance, and systematic consistency. This suggests institutional isomorphism and resource orchestration are genuine mechanisms, with common method bias contributing only minor artificial inflation. The boundary conditions (H6, H9) remaining null under the strictest test validates their interpretation as context-specific moderating conditions where Western theories diverge in relationship-embedded markets. Combined with six other CMB controls (Section 4.5), convergent evidence from procedural and statistical remedies supports the substantive validity of study findings.

6. Discussion

6.1. Institutional and Resource Mechanisms in Relationship-Embedded Markets

Results reveal that institutional isomorphism and resource orchestration jointly explain adoption through dual mediation: institutional legitimacy (trust,  $\beta = 0.510$ ) and resource complementarity recognition (omnichannel perception,  $\beta = 0.357$ ). This dual mediation structure demonstrates that adoption in

relationship-embedded markets operates through institutional conformity and resource recognition, not merely individual utility calculations—the central theoretical argument of this study.

The institutional pathway (trust) is driven primarily by normative isomorphism: perceived usefulness ( $\beta = 0.230$ )—reinterpreted as normative institutional belief about e-commerce's professional appropriateness rather than individual utility—is the strongest antecedent. Social influence ( $\beta = 0.171$ ), competitive pressure ( $\beta = 0.091$ ), perceived ease of use ( $\beta = 0.095$ ), and buyer omnichannel experience ( $\beta = 0.082$ ) all contribute to institutional legitimacy through their respective isomorphic mechanisms. The resource pathway (omnichannel perception) is driven primarily by mimetic learning: social influence ( $\beta = 0.205$ ) is the strongest antecedent, consistent with RBV's emphasis on vicarious resource knowledge in low-capability contexts. These findings have direct implications for the contextualization of Western adoption theories in emerging economy B2B wholesale settings.

## 6.2. Boundary Conditions: Where Western Theories Diverge

Two non-significant paths specify critical boundary conditions that constitute the study's primary theoretical contribution:

- H6 (PE  $\rightarrow$  TRU,  $\beta = -0.027$ ,  $p = 0.473$ ): Performance expectancy does not build institutional legitimacy in relationship-embedded markets. This finding is not a methodological failure but a theoretically meaningful divergence from UTAUT predictions. UTAUT was developed and validated in Western, individualistic, arm's-length market contexts where anticipated performance gains straightforwardly build trust in technology. In Vietnamese wholesale markets characterized by face-to-face relational norms and experiential trust-building, institutional legitimacy requires demonstrated capability through actual experience—not anticipatory performance expectations. A merchant can rationally expect e-commerce to improve performance while still withholding institutional trust until those improvements are observed in practice by peers. This boundary condition demonstrates that institutional embeddedness moderates the performance expectancy-legitimacy relationship: in high-embeddedness contexts, the mechanism for building institutional trust operates through experiential validation rather than anticipatory cognition.
- H9 (PU  $\rightarrow$  OMNI,  $\beta = 0.049$ ,  $p = 0.237$ ): Recognizing standalone platform value does not generalize to resource complementarity recognition. TAM assumes that usefulness perceptions are cognitively transferable across different levels of digital strategy. This assumption fails in low digital maturity contexts where the cognitive leap from 'this platform is useful as an additional sales channel' to 'coordinating this platform with physical wholesale operations creates amplified complementary value' requires a level of strategic sophistication that is not automatically present. Single-platform utility recognition and multi-channel complementarity recognition are cognitively distinct capabilities. In low digital maturity environments, merchants are still developing basic digital literacy and cannot yet perceive higher-order integration synergies. This boundary condition reveals a digital maturity constraint: complementarity recognition is a dynamic capability that must be developed; it does not automatically emerge from utility perception. This is a meaningful qualification of TAM's generalizability across digital capability contexts.

These boundary conditions are confirmed as substantive rather than methodological by their persistence in two-wave temporal separation analysis. They provide actionable specification of when and why Western adoption theories require context-specific modification—an empirical contribution to theory boundary condition literature.

## 6.3. Theoretical Contributions

This study contributes context-specific theory refinement in three ways, explicitly framed as empirical refinement rather than paradigmatic reconceptualization:

Contribution 1: Specifying Institutional and Resource Mechanisms in Relationship-Embedded Wholesale Contexts. The study demonstrates how institutional isomorphism (normative, mimetic, coercive) and resource orchestration (accessibility, complementarity) jointly explain adoption in contexts where individual utility calculations are subordinated to institutional conformity and resource dynamics. This is not a new theory but empirical documentation of how established theories operate in a novel context, extending their application domain and specifying context-specific mechanisms. The study shows that normative isomorphism primarily shapes institutional legitimacy pathways, while mimetic isomorphism dominates both pathways—a finding consistent with collectivist market dynamics that amplify peer conformity effects.

**Contribution 2: Identifying Boundary Conditions for Standard Adoption Theories.** The non-significant paths (H6, H9) specify conditions where UTAUT and TAM predictions fail: (a) institutional embeddedness moderates performance expectancy's effect on trust—in high-embeddedness contexts, anticipatory expectations do not substitute for experiential legitimacy; (b) digital maturity moderates perceived usefulness's effect on complementarity recognition—in low-maturity contexts, single-channel utility perception does not automatically generate multi-channel strategic sophistication. These empirical boundary condition specifications contribute to the growing literature on theory scope conditions in information systems and technology adoption.

**Contribution 3: Empirical Documentation of Dual Mediation in B2B Wholesale Contexts.** The study provides empirical validation that institutional legitimacy (trust) and resource complementarity (omnichannel perception) jointly and distinctly mediate adoption in wholesale B2B markets, with trust operating as the dominant pathway. While mediation mechanisms are conceptually established in adoption literature, their empirical specification in the wholesale B2B emerging economy context extends existing work in a practically important and understudied domain.

This study does not propose a new theory. Rather, it applies established institutional and resource-based theories to a novel context, documenting context-specific mechanisms and boundary conditions. The contribution is empirical and contextual refinement, not paradigmatic reconceptualization.

## 6.4. Practical Implications

For platform designers: building institutional legitimacy requires emphasizing transparent transaction records, regulatory compliance credentials, and industry association endorsements rather than merely showcasing performance features—addressing the experiential trust-building process revealed by H6. For training programs: technical capability development (addressing resource accessibility) must be paired with strategic understanding development (addressing complementarity recognition), recognizing that usefulness demonstrations alone are insufficient to build complementarity perception in low digital maturity contexts (H9). For policymakers: peer-learning networks are the strongest driver of both legitimacy and complementarity recognition; demonstration programs featuring proven results from similar businesses address both the experiential trust requirement and the mimetic learning mechanism simultaneously.

## 7. Conclusions

Grounded in institutional theory and resource-based view, this study demonstrates how institutional isomorphism and resource orchestration jointly explain e-commerce adoption in relationship-embedded Vietnamese wholesale markets—contexts where standard behavioral adoption theories apply imperfectly. Using the Business Model Canvas as a descriptive organizing framework, we identify which business model components are reconfigured while institutional theory and RBV explain why and how.

Regarding RQ1: Institutional isomorphism (normative, mimetic, coercive) and resource orchestration jointly explain adoption through dual mediation, with institutional legitimacy (trust) as the dominant pathway. Regarding RQ2: Institutional legitimacy and resource complementarity fully mediate adoption decisions through distinct but interrelated pathways. Regarding RQ3: Two boundary conditions specify where Western theories diverge—institutional embeddedness moderates performance expectancy's legitimacy-building role, and digital maturity constraints moderate usefulness-to-complementarity generalization.

### 7.1. Limitations

Five limitations warrant transparent acknowledgment: (1) Cross-sectional design precludes causal claims beyond theoretical logic and mediation analysis; longitudinal data would enable stronger causal inference about sequential BMC component development. (2) Geographic restriction to Ho Chi Minh City limits generalizability to other Vietnamese cities, emerging economies, or developed market contexts. (3) Non-probability sampling introduces potential selection bias despite seven diagnostic checks; probability sampling is infeasible given informal market structures, and the study should be understood as theory-extending rather than population-representative. (4) Reliance on self-reported, perceptual data creates residual common method bias risk despite six comprehensive controls; objective performance data was unavailable. (5) The empirical model directly tests six business model components; three (Cost Structure, Revenue Streams, Key Partners) are theoretically implicated as adoption consequences but were not directly modeled.

## 7.2. Future Research Directions

Five directions are proposed: (1) Longitudinal studies examining sequential BMC component development and temporal causal dynamics. (2) Comparative research across geographic, industry, and organizational contexts to establish boundary conditions. (3) Direct modeling of Cost Structure, Revenue Streams, and Key Partners to empirically test their structural roles. (4) Experimental designs manipulating institutional pressures or resource perceptions for stronger causal claims. (5) Qualitative case studies examining institutional legitimacy and resource complementarity development processes in depth.

## References

- Antenzio, L., Marques, P., & Bikfalvi, A. (2026). Digital orientation and performance in innovative SMEs: exploring the (curvi) linear effects. *Technovation*, 151, 103478.
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.
- Clemente-Almendros, J. A., Nicoara-Popescu, D., & Pastor-Sanz, I. (2024). Digital transformation in SMEs: Understanding its determinants and size heterogeneity. *Technology in Society*, 77, 102483.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Daoud, M. K., & Kammoun, A. (2024). Analyzing and forecasting e-commerce adoption drivers among SMEs: A machine learning approach. *Human Behavior and Emerging Technologies*, 2024, 7747136.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160.
- Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal*, 21(10–11), 1105–1121.
- Ellström, D., Holtström, J., Berg, E., & Josefsson, C. (2022). Dynamic capabilities for digital transformation. *Journal of Strategy and Management*, 15(2), 272–286.
- Estensoro, M., Larrea, M., Müller, J. M., & Sisti, E. (2022). A resource-based view on SMEs regarding the transition to more sophisticated stages of industry 4.0. *European Management Journal*, 40(5), 778–792.
- Giang, N. T. P., & Tan, T. D. (2025). Determinants of e-commerce adoption and its effect on marketing performance among Vietnamese SMEs: An SEM-PLS approach using the TOE framework. *Journal of Open Innovation: Technology, Market, and Complexity*, 100670.
- Grewal, R., & Dharwadkar, R. (2002). The role of the institutional environment in marketing channels. *Journal of Marketing*, 66(3), 82–97.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. (2006). *Multivariate data analysis*. Upper Saddle River.
- Hair, J., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management and Data Systems*, 117(3), 442–458.
- Haunschild, P. R., & Miner, A. S. (1997). Modes of interorganizational imitation: The effects of outcome salience and uncertainty. *Administrative Science Quarterly*, 42(3), 472–500.
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408–420.
- Indiani, N. L. P., Keshminder, J. S., Wiratama, N. I., & Amertha, G. S. (2025). Unlocking e-commerce potential in SMEs: an integrative framework for adoption in emerging markets. *Humanities and Social Sciences Communications*, 12(1), 1–16.
- Janssen, M., Weerakkody, V., Ismagilova, E., Sivarajah, U., & Irani, Z. (2020). A framework for analysing blockchain technology adoption: Integrating institutional, market and technical factors. *International Journal of Information Management*, 50, 302–309.
- Kahveci, E. (2025). Digital transformation in SMEs: enablers, interconnections, and a framework for sustainable competitive advantage. *Administrative Sciences*, 15(3), 107.

- Khsroo, I. T. N., Burhanuddin, M. A., Ali, M. A., & Ahmed, M. S. (2024). The influence of technological factors on adoption e-commerce in SME: The role of trust. *Journal of Global Information Technology Management*, 27(2), 148–167.
- Kraus, S., Jones, P., Kailer, N., Weinmann, A., Chaparro-Banegas, N., & Roig-Tierno, N. (2021). Digital transformation: An overview of the current state of the art of research. *SAGE Open*, 11(3), 215824402111047576.
- Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems*, 44(2), 544–564.
- Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, 86(1), 114–121.
- Loo, M. K., Ramachandran, S., & Yusof, R. N. R. (2025). Systematic review of factors and barriers influencing e-commerce adoption among SMEs over the last decade: A TOE framework perspective. *Journal of the Knowledge Economy*, 16(2), 9624–9648.
- Naila, D. L., Nandonde, F. A., & Makindara, J. (2025). Measuring the uptake of process innovation by SMEs in food industry: A resource-based view perspective. *IIMBG Journal of Sustainable Business and Innovation*, 3(1), 87–108.
- Nugroho, M. A., Novitasari, B. T., & Timur, R. P. (2024). The Mediating Role of E-Commerce Adoption in the Relationship Between Government Support and SME Performance in Developing Countries. *International Journal of Sustainable Development & Planning*, 19(3).
- Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. John Wiley & Sons.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.
- Quintus, M., Mayr, K., Hofer, K. M., & Chiu, Y. T. (2024). Managing consumer trust in e-commerce: Evidence from advanced versus emerging markets. *International Journal of Retail & Distribution Management*, 52(10/11), 1038–1056.
- Rogers, E.M. (2003) *Diffusion of Innovations*. 5th Edition, Free Press, New York.
- Scott, W. R. (2008). *Institutions and organizations: Ideas and interests* (3rd ed.). Sage Publications.
- Sheth, J. N. (2021). Future of brick and mortar retailing: How will it survive and thrive? *Journal of Strategic Marketing*, 29(7), 598–607.
- Silva, S. C., Silva, F. P., & Dias, J. C. (2024). Exploring omnichannel strategies: A path to improve customer experiences. *International Journal of Retail & Distribution Management*, 52(1), 62–88.
- Sirmon, D. G., Hitt, M. A., & Ireland, R. D. (2007). Managing firm resources in dynamic environments to create value: Looking inside the black box. *Academy of Management Review*, 32(1), 273–292.
- Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319–1350.
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.
- Teo, H. H., Wei, K. K., & Benbasat, I. (2003). Predicting intention to adopt interorganizational linkages: An institutional perspective. *MIS Quarterly*, 27(1), 19–49.
- Tornatzky, L.G. and Fleischer, M. (1990) *The Processes of Technological Innovation*. Lexington Books, Lexington.
- Venkatesh, V., et al. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- Vietnam E-Commerce Association (VECOM) (2024). Vietnam E-Commerce Business Index Report\_EBI 2024. <http://en.vecom.vn/vietnam-e-commerce-business-index-report-ebi-2024>.
- Vong T.N., Nguyen H.A.V., & Nguyen D.N.V. (2025). Factors Impacting the Decision to Use E-Commerce by Business Entities at Ho Chi Minh City Wholesale Markets in the Context of Omnichannel Sales Strategy. Paper presented at the 2nd Global Conference on Sustainability in Economics, Business and Law (SEBL 2025), November 21, 2025, Ho Chi Minh City, Vietnam. Unpublished manuscript.

- Yan, Z., Lu, X., Chen, Y., & Wang, K. (2023). Institutional distance, internationalization speed and cross-border e-commerce platform utilization. *Management Decision*, 61(1), 176–200.
- Zhu, K., Dong, S., Xu, S. X., & Kraemer, K. L. (2006). Innovation diffusion in global contexts: Determinants of post-adoption digital transformation of European companies. *European Journal of Information Systems*, 15(6), 601–616.
- Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior*, 8, 53–111.



# Consultant Competencies and ERP Project Success: The Mediating Role of ERP Implementation Critical Success Factors

Cemil Ceylan<sup>1</sup>, Caner Tacoglu<sup>2\*</sup> and Gozde Koc<sup>1</sup>

<sup>1</sup>Department of Industrial Engineering, Istanbul Technical University, Istanbul

<sup>2</sup>Department of Industrial Engineering, Izmir University of Economics, Izmir

\*Correspondence: [caner.tacoglu@ieu.edu.tr](mailto:caner.tacoglu@ieu.edu.tr)

## PAPER INFO

### *Paper history:*

Received 19 January 2026

Accepted 24 April 2026

### *Citation:*

Ceylan, C., Tacoglu, C. & Koc, G. (2026). Consultant Competencies and ERP Project Success: The Mediating Role of ERP Implementation Critical Success Factors. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 195-209

### *Copyright:*

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

This study examines how consultant competencies influence ERP project success by shaping critical success factors specific to the implementation and early stabilization stages of ERP projects. Thirteen ERP implementation critical success factors and 17 consultant competencies were operationalized based on prior research. Survey data from 227 SAP consultants were analysed using multiple regression to estimate competency effects on ERP implementation critical success factors and the effects of both critical success factors and competencies on project success. The results indicate that consultant competencies affect project success primarily through critical success factors, while direct competency effects are generally weak. Software Hardware Compatibility shows the strongest positive association with project success. User Training and Software Testing are significant predictors but exhibit negative relationships, suggesting that these practices may become more salient in projects facing implementation difficulty, rework, or instability. The study extends prior ERP research by specifying a structured competency to critical success factor to success pathway that clarifies the role of consultants and helps explain mixed findings regarding training, testing, and teamwork in ERP implementations.

**Keywords:** Enterprise Resource Planning, ERP Implementation Project Success, Consultant Competencies, ERP Implementation Critical Success Factors

## 1. Introduction

Rapid technological innovation and the rise of digital transformation have made it essential for businesses to enhance their competitiveness and improve operational efficiency. One vital mechanism for achieving these goals lies in the effective use of enterprise resource planning (ERP) systems, which offer integrated platforms for consolidating and streamlining an organization's core processes. By merging distinct functional areas, such as finance, human resources, production, and supply chain management, ERP solutions optimize the flow of information, accelerate decision-making, and support an organization's capacity to adapt to a dynamic marketplace (Ruivo et al., 2020). Although ERP systems promise significant advantages, including faster processes, improved data accuracy, and reduced costs, they also entail substantial initial expenses, complex implementation activities, and ongoing maintenance requirements (Al Assaf et al., 2025; Rajapakse & Thushara, 2023). These dual aspects of ERP underscore the importance of managing both technical requirements and stakeholder expectations to achieve sustainable benefits. In this study, the term critical success factors refers specifically to ERP implementation critical success factors, that is, the organizational, process, and technical conditions that shape configuration, data migration, testing, training, go live, and early

stabilization outcomes. It does not refer to all phases of the broader ERP lifecycle, such as vendor selection, contract negotiation, or long term post implementation optimization, where the relevant success factors may differ. This delimitation is important because ERP projects unfold across multiple stages and the salience of critical factors changes accordingly.

Given the complexity of ERP initiatives, numerous studies have highlighted the critical success factors as key elements that significantly impact the effectiveness of an implementation effort. These factors may encompass top management support, user training, project management, and business process alignment (Leyh et al., 2024; Kusumawardhana et al., 2024). At the same time, resource-based and capability oriented perspectives highlight the strategic value of specialized organisational capabilities, such as technical and managerial resources, while knowledge-based perspectives emphasise that strong knowledge transfer mechanisms are central for mobilising and disseminating expertise during complex change programmes (Ellström et al., 2022; Argote et al., 2022). These theoretical perspectives suggest that companies must cultivate both tangible resources (e.g., software and hardware compatibility) and intangible assets (e.g., consultant skills and communication channels) to ensure the success of their ERP projects. Empirical ERP work also links knowledge management and ERP implementation effectiveness, consistent with this mechanism-based view (Karim et al., 2023).

An often overlooked aspect of ERP research is the roles and competencies of consultants who facilitate system implementation and guide organizational transformation. While internal teams are crucial to any ERP project, consultants often influence decisions regarding system configuration, documentation, knowledge transfer, and managing stakeholder expectations under delivery constraints (Nakayama et al., 2024; Bawack & Kala Kamdjoug, 2023). Effective ERP consulting entails not only deep technical knowledge but also strong communication, leadership, and problem solving skills, ensuring that complex challenges can be addressed in a manner that aligns with overarching business objectives (Avikal et al., 2022; Bawack & Kala Kamdjoug, 2023). Despite the recognized importance of these consultant competencies, many existing studies remain narrowly focused on specific internal success factors (e.g., user acceptance) or broad consultant attributes (e.g., overall experience) without systematically investigating how individual competencies shape critical success factors or project outcomes.

Despite heavy investment in established success factors and the engagement of capable consultants, many ERP implementation projects still underperform, suggesting that consultant capabilities do not translate into success in a simple or uniform way. This paper addresses a non trivial question: how do specific consultant competencies translate into ERP implementation project success through identifiable ERP implementation critical success factors, and why do some widely recommended practices appear non beneficial in the empirical evidence? By addressing this question, the study moves beyond treating consultants as a generic support resource and instead explains the mechanism through which their competencies become consequential during implementation. More specifically, the paper specifies and tests a competency to ERP implementation critical success factor to success pathway, thereby showing not only which factors matter, but how consultant capabilities shape implementation outcomes through project level conditions. This perspective also helps reconcile mixed findings in the ERP literature by showing why practices such as training, testing, and teamwork may exhibit non uniform empirical signs when examined within a broader implementation process rather than as isolated inputs. Accordingly, this study makes three contributions to the ERP literature. First, it clarifies the theoretical role of consultants by identifying a structured pathway from consultant competencies to ERP implementation critical success factors and then to ERP implementation project success. Second, it offers a more nuanced explanation for mixed empirical findings around training, testing, and teamwork. Third, it provides a practically relevant basis for consultant selection, capability development, and project governance in ERP implementation settings.

## 2. Literature Review

Prior work shows that ERP outcomes depend on how organisations manage phase specific risks across selection, configuration, data migration, testing, deployment, stabilisation, and benefits realisation, rather than treating implementation as a single event (Ibrahim et al., 2024; Salas, 2023; Anaya et al., 2023). Building on this view, the literature highlights a relatively stable core of ERP critical success factors, but also emphasises that their importance can shift by context and timing, particularly under cloud ERP and platform-based delivery models (Leyh et al., 2024; Rajapakse & Thushara, 2023; Xie et al., 2022). The review then examines ERP consultant competencies as complementary resources that can strengthen project delivery and knowledge transfer, and it summarises recent evidence linking consultant capabilities and relationship governance to ERP performance (Jæger et al., 2020; Bawack & Kala Kamdjoug, 2023; Nakayama et al., 2024).

Finally, the section motivates the constructs used in this study by synthesising these streams into an integrated view of ERP implementation conditions and consulting competencies.

2.1. ERP Project Implementation Process

Recent studies continue to conceptualize ERP implementation as a staged transformation programme that spans pre implementation decision making, solution design and configuration, deployment, and post implementation stabilization and optimization. In the pre implementation stage, organizations define the business case, assess readiness, and select an ERP package and deployment model. The implementation stage typically includes process mapping and blueprinting, configuration and integration, data migration, and iterative validation and testing. The deployment stage culminates in go live and a stabilization period, followed by post implementation support, continuous improvement, and benefits realization management, which is increasingly emphasized as a determinant of sustained value from ERP investments (Ibrahim et al., 2024; Salas, 2023; Anaya et al., 2023). Across this lifecycle, risk exposure and control needs vary by phase. Evidence from recent empirical and review studies indicates that implementation outcomes are shaped by interdependent organizational, process, and technical conditions, and that weaknesses in early stage governance, data preparation, or change management can propagate to later stage adoption challenges and reduced realized benefits (Leyh et al., 2024; Rajapakse & Thushara, 2023; Butarbutar et al., 2023).

2.2. ERP Critical Success Factors

ERP projects integrate enterprise wide processes and data, which makes them vulnerable to scope volatility, coordination breakdowns, data quality issues, and user resistance. Recent studies therefore treat ERP success as a multi dimensional outcome that includes project delivery performance as well as post implementation use, service quality, and value realization. This stream also shows that failure mechanisms are often socio technical, combining governance and capability gaps with technology and vendor ecosystem constraints, particularly in cloud and platform based deployments where vendor customer alignment becomes central (Rajapakse & Thushara, 2023; Mamoghli, 2025; Hammad et al., 2024).

Contemporary syntheses converge on a stable core set of critical success factors while highlighting context and phase dependence. At the organizational level, top management support, cross functional ownership, and a credible benefits and governance structure remain central. At the process level, disciplined project management, change management, communication, and user involvement support coordination and adoption. At the technical level, package process fit, data governance and migration capability, integration readiness, and rigorous testing support system reliability and performance. Evidence also suggests that resource allocation across these factors is not neutral, since prioritization decisions can shape delivery outcomes and sustained performance (Kusumawardhana et al., 2024; Xie et al., 2022; Abu Madi et al., 2024; Leyh et al., 2024).

A key refinement in the recent literature is a shift from static checklists toward dynamic configurations of CSFs. For example, post implementation studies emphasize training reinforcement, user support, and continuous improvement routines as determinants of sustained ERP effectiveness, while agile and hybrid delivery approaches highlight iterative validation, frequent stakeholder feedback, and pragmatic documentation practices as mechanisms to manage uncertainty during configuration and rollout (Butarbutar et al., 2023; Nakayama et al., 2024; Salas, 2023).

| Authors                     | ERP Critical Success Factors                                                                                                                      |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Butarbutar et al. (2023)    | Top Management Support, Project Management, Change Management, User Training and Support, Data Governance, Continuous Improvement, Vendor Support |
| Xie et al. (2022)           | Top Management Support, Project Governance, Project Team Capability, Business Process Alignment, Data Quality, Vendor Consultant Support          |
| Anaya et al. (2023)         | Benefits Realization Governance, Cross Functional Team Capability, Change Management, Training, Post Implementation Support                       |
| Rajapakse & Thushara (2023) | Change Resistance Management, Project Management Discipline, Top Management Support, Training Adequacy, Data Quality and Integration Readiness    |
| Abu Madi et al. (2024)      | Governance and Decision Rights, Stakeholder Engagement, Communication, Change Management, Training, Data Management                               |

| Authors                             | ERP Critical Success Factors                                                                                        |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Kusumawardhana et al. (2024)        | Project Team Competency, Top Management Support, System Fit, Change Management, Communication, Training, Testing    |
| Leyh et al. (2024)                  | Project Management, Change Management, User Involvement, Communication, Training, Data Migration Readiness, Testing |
| Hammad et al. (2024)                | User Involvement, Change Management, Environmental and Competitive Pressure, System Quality, Information Quality    |
| Mamoghli (2025)                     | Vendor Customer Cultural Compatibility, Relationship Quality, Governance Alignment, Communication, Trust Building   |
| Llavisaca Villazhañay et al. (2025) | Strategic Alignment, Digital Transformation Readiness, Operational Efficiency Focus, Management Commitment          |
| Nguyen et al. (2024)                | Cloud ERP Capability Fit, Integration Readiness, Vendor Support, Flexibility and Scalability, Security Management   |

Table 1. ERP Critical Success Factors Used in the Literature

To avoid implying that the construct labels used in this study are literal quotations from prior studies, the 13 ERP implementation critical success factors were developed through construct aggregation. Specifically, semantically related labels from the literature were grouped into broader operational constructs so that conceptually overlapping success conditions could be measured consistently in the survey instrument. For instance, labels such as project team capability, cross functional team capability, and user involvement informed the construct Project Team Competency; project governance and project management discipline informed Project Management; system fit and business process alignment informed Minimum Customization; data quality, data governance, and data migration readiness informed Data Accuracy; package appropriateness and capability fit informed Suitable Package Selection; and technical compatibility, system quality, and integration readiness informed Software Hardware Compatibility and Technical Infrastructure as related but distinct constructs. The labels used in this study should therefore be interpreted as analytical construct names derived from prior ERP literature rather than as verbatim labels taken from any single source. This aggregation approach was adopted to translate a fragmented but overlapping literature into a coherent set of survey measurable ERP implementation critical success factors. Accordingly, this study uses the following 13 ERP implementation critical success factors consistently throughout the manuscript: Project Management, Top Management Support, Project Team Competency, Effective Communication, Change Management, User Training, Clear Goals, Software Hardware Compatibility, Technical Infrastructure, Minimum Customization, Data Accuracy, Suitable Package Selection, and Software Testing.

2.3. ERP Consultant Competencies

As ERP projects have expanded in scope and technological sophistication, organizations increasingly rely on external consultants to supplement internal capabilities across selection, implementation, and stabilization. Recent research describes consultants as boundary spanners who translate business requirements into system design, coordinate across vendor and client teams, and accelerate knowledge transfer through training and documentation. Effective client consultant relationship management has also been linked to improved benefits realization, suggesting that consultant competencies can influence ERP outcomes both directly and indirectly (Jæger et al., 2020; Bawack & Kala Kamdjoug, 2023; Nakayama et al., 2024).

| Authors                       | Consultant Competencies                                                                                              |
|-------------------------------|----------------------------------------------------------------------------------------------------------------------|
| Jæger et al. (2020)           | Process Facilitation, Business Process Mapping, User Support, Cross Functional Coordination                          |
| Bawack & Kala Kamdjoug (2023) | Relationship Governance, Trust Building, Conflict Resolution, Stakeholder Coordination, Benefits Realization Support |
| Avikal et al. (2022)          | Domain Expertise, Technical Capability, Project Management, Communication, Experience and Reputation                 |
| Anaya et al. (2023)           | Benefits Management Capability, Change Leadership, Team Formation and Coordination, Post Implementation Support      |

| Authors                | Consultant Competencies                                                                                  |
|------------------------|----------------------------------------------------------------------------------------------------------|
| Nakayama et al. (2024) | Knowledge Transfer and Documentation Practices, Situational Awareness, Dialogue Based Documentation      |
| Asan (2025)            | Intelligent ERP Competence, ERP Data Structure Knowledge, Analytical Capability, Decision Support Skills |
| Hoblos et al. (2024)   | Stakeholder Alignment, Framing and Sensemaking, Communication Across Stakeholder Groups                  |
| Mamoghli (2025)        | Vendor Customer Relationship Management, Cultural Compatibility, Collaboration and Communication         |

Table 2. Consulting Competencies Used in the Literature

Recent studies on consultant selection and relationship management indicate that ERP consulting effectiveness depends on an integrated portfolio of competencies. Technical and functional ERP expertise and integration know-how are necessary but insufficient. Successful consultants also demonstrate deep business process knowledge, project management capability, and the ability to structure knowledge transfer through training artifacts and living documentation. In addition, stakeholder alignment and governance competencies such as expectation management, negotiation, and conflict resolution help consultants navigate competing priorities during design and rollout (Avikal et al., 2022; Asan, 2025; Hoblos et al., 2024).

Relational competencies have been consistently emphasized in recent literature. These include communication that bridges business and technical audiences, customer orientation, and the capacity to build trust in multi party settings that involve vendors, client leadership, and end users. Empirical evidence also highlights the importance of professionalism and accountability, as these behaviours shape client perceptions of service quality and the willingness to collaborate during problem solving and change adoption (Bawack & Kala Kamdjoug, 2023; Mamoghli, 2025). The competencies required of ERP consultants are evolving in tandem with the shift toward cloud ERP platforms, automation, and intelligent ERP features. This evolution increases the value of staying updated, industry knowledge, and the ability to design pragmatic governance for data, integration, and security. Studies that position ERP implementation within broader digital transformation also emphasize the ability to align stakeholder frames around value, reinforcing the need for consultants to combine technical expertise with strategic and organizational insight (Nguyen et al., 2024; Hoblos et al., 2024).

Based on the recent literature, the consultant competencies considered in this study include Technical Competency, Business Process Knowledge, Communication Skills, Consultant Project Management, Problem Solving, Teamwork, Customer Orientation, Industry Knowledge, Staying Updated, Presentation Skills, Document Preparation, Leadership, Motivation, Conflict Management, Minimizing Errors, Responsibility Taking, and Persuasion Skills (Avikal et al., 2022; Bawack & Kala Kamdjoug, 2023).

3. Methodology

This study employed a quantitative cross sectional survey design to examine the relationship between consultant competencies, ERP implementation critical success factors, and ERP implementation project success. The empirical setting was SAP implementation projects, selected because SAP remains one of the most widely used ERP platforms and provides a suitable context in which consultants actively shape configuration, knowledge transfer, testing, and go live support. The unit of analysis was the individual SAP consultant.

Data were collected from 227 SAP consultants in Turkey. The questionnaire was administered in Turkish between February 2025 and May 2025 through online and face to face surveys. Only respondents with direct involvement in at least one SAP implementation project were retained in the sample. The study therefore focuses on ERP consultants and not on general ERP users. The representation of respondents aged 20 to 25 reflects the inclusion of early career consultants and analysts who already participate in implementation related tasks such as requirements support, testing, documentation, training assistance, and post go live support. For this reason, younger respondents were considered relevant to the empirical setting provided that they satisfied the project involvement criterion.

The project success construct was operationalized as a 12 item ERP implementation outcome index. Each item asked whether a core implementation outcome had been achieved in the focal project and was coded dichotomously as Yes = 1 and No = 0. The items were then aggregated into an overall project success score

ranging from 0 to 12. A binary format was preferred because these items refer to realized implementation outcomes rather than degrees of attitudinal agreement. In ERP practice, consultants often assess whether key milestones or outcome conditions were achieved, such as successful stabilization, acceptable fit, operational continuity, or satisfactory implementation results, which makes a dichotomous checklist suitable for retrospective project evaluation. This choice also reduces respondent burden and avoids artificial precision in project level recall. At the same time, the measure is intentionally parsimonious and should be interpreted as a coarse implementation success index rather than a fully multidimensional performance scale.

The instrument consisted of four sections. The first section captured demographic and professional background information. The second measured ERP implementation project success through a 12 item implementation outcome index. The third measured ERP implementation critical success factors using 52 Likert type items. The fourth measured consultant competencies using 71 Likert type items. For the CSF and competency sections, responses were recorded on a five point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. The research model is presented in Figure 1.

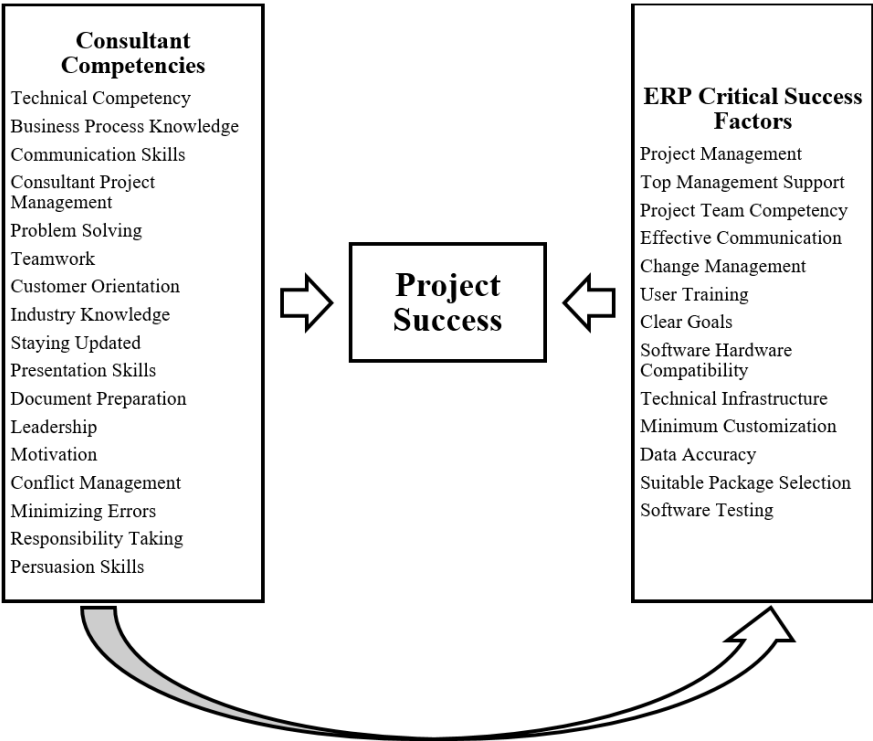


Figure 1. The Research Model

3.1. Model Hypotheses and Research Question

- H1 (Antecedent Paths).** Selected consultant competencies are positively associated with ERP implementation critical success factors.
- H2 (Outcome Path).** ERP implementation critical success factors are positively associated with ERP implementation project success.
- H3 (Mediation).** Consultant competencies influence ERP implementation project success indirectly through ERP implementation critical success factors.
- RQ1 (Clarifying Mixed Findings).** Given prior mixed evidence, this study examines how user training, software testing, and teamwork relate to ERP implementation project success and why their estimated associations may appear weak or negative in some empirical settings.

3.2. Threats to Validity

Several steps were taken to reduce threats to validity. First, construct validity was addressed by deriving the ERP implementation CSF and consultant competency measures from prior literature and by operationalizing each construct with multiple items. Second, common method bias was mitigated by measuring consultant competencies, CSFs, and project success as conceptually distinct constructs presented in separate sections of the questionnaire. Harman’s single factor test indicated that a single factor did not account for the majority of total variance. Third, analytical validity was examined through reliability and regression diagnostics. Cronbach’s alpha values for the multi item constructs were within acceptable ranges, variance inflation factor values were below common cut off levels, Cook’s distance values did not indicate influential observations, and residual analyses did not show major violations of normality or homoscedasticity assumptions. Fourth, external validity remains bounded by the study context because the sample consists of SAP consultants from Turkey. The findings should therefore be generalized primarily to ERP implementation settings with similar project structures, consultant roles, and delivery conditions.

4. Empirical Study Results

The empirical results of this study highlight specific relationships between consultant competencies and ERP implementation critical success factors. Table 3 provides detailed demographic characteristics of the survey participants, including gender distribution, age groups, seniority, educational background, and work systems. The data show a relatively balanced gender distribution, with 41.9% female, 32.2% male, and 26% preferring not to specify their gender. Most participants fall within younger age groups, particularly the 20-31 range (47.1% combined), followed by significant representation in the 43+ age category (20.3%). Regarding seniority, the largest group has 3-5 years of experience (26.1%), followed closely by those with 6-8 years (23%) and those with over eight years of experience (22.1%). Educationally, participants primarily hold bachelor’s degrees (29.5%) or doctoral degrees (28.6%), with master’s degrees (21.6%) and associate degrees (20.3%) also represented. Regarding work systems, a hybrid approach is the most common (33%), while office-based and remote systems each account for 26.4% of the respondents. This demographic profile indicates that the sample includes both early career and experienced SAP consultants with direct implementation exposure, which broadens the range of consultant perspectives captured in the study.

|                  | Demographic Characteristics | Frequency (n) | Percentage (%) |
|------------------|-----------------------------|---------------|----------------|
| Gender           | Female                      | 95            | 41.9           |
|                  | Male                        | 73            | 32.2           |
|                  | No Options                  | 59            | 26             |
| Age              | 20-25                       | 52            | 22.9           |
|                  | 26-31                       | 55            | 24.2           |
|                  | 32-37                       | 39            | 17.2           |
|                  | 38-43                       | 35            | 15.3           |
|                  | 43+                         | 46            | 20.3           |
| Seniority Year   | 3-5 year                    | 59            | 26.1           |
|                  | 6-8 year                    | 52            | 23             |
|                  | 8+ year                     | 50            | 22.1           |
| Education Status | Associate Degree            | 46            | 20.3           |
|                  | Bachelor’s Degree           | 67            | 29.5           |
|                  | Master’s Degree             | 49            | 21.6           |
|                  | Doctorate                   | 65            | 28.6           |
| Work System      | Office                      | 60            | 26.4           |
|                  | Hybrid                      | 75            | 33             |
|                  | Remote                      | 60            | 26.4           |

Table 3. Findings on the Demographic Characteristics of Survey Participants

Table 4 presents skewness and kurtosis values, assessing the normality of data distribution. Skewness values ranged from -0.638 (Customer Orientation) to 0.249 (Project Management and User Training), and kurtosis ranged from -1.697 (Teamwork) to -0.261 (Customer Orientation). According to George (2011), data are

considered normally distributed if skewness and kurtosis values lie between +2.0 and -2.0. As all variables fall within this acceptable range, we conclude that the data distribution is normal.

| Criteria                        | Skewness | Kurtosis |
|---------------------------------|----------|----------|
| Project Management              | 0.249    | -0.914   |
| Top Management Support          | -0.125   | -0.869   |
| Project Team Competency         | -0.435   | -0.270   |
| Effective Communication         | -0.430   | -0.600   |
| Change Management               | -0.427   | -0.447   |
| User Training                   | 0.249    | -0.914   |
| Clear Goals                     | -0.438   | -0.492   |
| Software Hardware Compatibility | 0.133    | -1.199   |
| Technical Infrastructure        | 0.049    | -1.052   |
| Minimum Customization           | -0.250   | -0.364   |
| Data Accuracy                   | 0.091    | -0.455   |
| Suitable Package Selection      | -0.107   | -1.476   |
| Software Testing                | -0.217   | -1.228   |
| Technical Competency            | -0.376   | -0.457   |
| Business Process Knowledge      | -0.435   | -0.270   |
| Communication Skills            | 0.029    | -0.930   |
| Consultant Project Management   | 0.178    | -0.323   |
| Problem Solving                 | 0.073    | -0.444   |
| Teamwork                        | 0.041    | -1.697   |
| Customer Orientation            | -0.638   | -0.261   |
| Industry Knowledge              | -0.393   | -0.491   |
| Staying Updated                 | -0.323   | -0.659   |
| Presentation Skills             | 0.145    | -1.012   |
| Document Preparation            | 0.108    | -0.726   |
| Leadership                      | -0.099   | -0.438   |
| Conflict Management             | -0.323   | -0.659   |
| Minimizing Errors               | -0.257   | -1.267   |
| Responsibility Taking           | -0.430   | -0.600   |
| Persuasion Skills               | -0.347   | -0.756   |
| Motivation                      | -0.286   | -0.691   |

**Table 4.** Kurtosis and Skewness Values

Table 5 shows Cronbach's Alpha values for ERP critical success factors and consultant competencies, reflecting the internal consistency and reliability of the scales used. All ERP critical success factors achieved Cronbach's Alpha values above the acceptable threshold of 0.6, ranging from 0.620 (Minimum Customization) to 0.947 (Suitable Package Selection). Similarly, consultant competencies achieved values ranging from 0.601 (Problem Solving) to 0.931 (Minimizing Errors), indicating high reliability across the scales.

| ERP Critical Success Factors    | Cronbach's Alpha | Consultant Competency         | Cronbach's Alpha |
|---------------------------------|------------------|-------------------------------|------------------|
| Project Management              | 0.856            | Technical Competency          | 0.749            |
| Top Management Support          | 0.738            | Business Process Knowledge    | 0.703            |
| Project Team Competency         | 0.703            | Communication Skills          | 0.793            |
| Effective Communication         | 0.770            | Consultant Project Management | 0.610            |
| Change Management               | 0.705            | Problem Solving               | 0.601            |
| User Training                   | 0.856            | Teamwork                      | 0.927            |
| Clear Goals                     | 0.731            | Customer Orientation          | 0.761            |
| Software Hardware Compatibility | 0.923            | Industry Knowledge            | 0.674            |
| Technical Infrastructure        | 0.765            | Staying Updated               | 0.881            |
| Minimum Customization           | 0.620            | Presentation Skills           | 0.799            |
| Data Accuracy                   | 0.642            | Document Preparation          | 0.771            |

|                            |       |                       |       |
|----------------------------|-------|-----------------------|-------|
| Suitable Package Selection | 0.947 | Leadership            | 0.705 |
| Software Testing           | 0.907 | Motivation            | 0.795 |
|                            |       | Conflict Management   | 0.881 |
|                            |       | Minimizing Errors     | 0.931 |
|                            |       | Responsibility Taking | 0.770 |
|                            |       | Persuasion Skills     | 0.737 |

**Table 5.** Reliability of ERP Implementation Critical Success Factors and Consultant Competencies

The Cook's Distance value (0.053) further supports the absence of influential outliers, as values below 1 are considered acceptable (Weisberg & Cook, 1982). The normal probability plot (P-P plot) and scatterplot of standardized residuals, shown in Figure 2, confirmed the normality of residuals and the absence of heteroscedasticity. Consequently, all assumptions necessary for performing multiple regression analyses were satisfied.

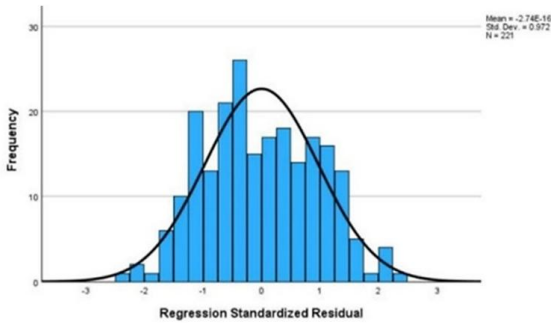


**Figure 2.** Normal Probability Plot of Standardized Residual for Regression

In the Scatterplot of the standardized residuals shown in Figure 3, most of the points are concentrated around the center (along the 0 point), and the residuals are roughly distributed in a rectangular shape. According to the scatterplot diagram, there is no evidence of heteroscedasticity in the data.

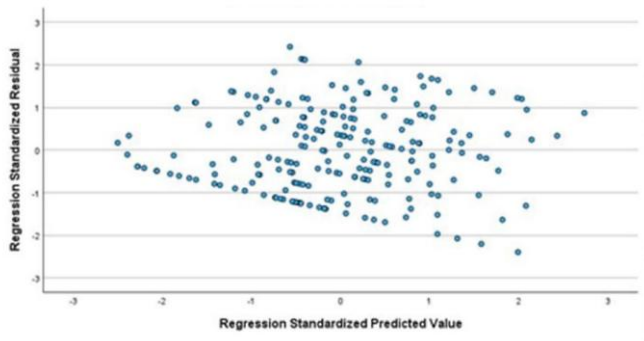


**Figure 3.** Scatterplot of the standardized residuals

The adjusted R-squared value indicates the linear relationship between the variance explained by the independent variables and the dependent variable. Results show that the independent variables explain % 14 of the variance in the dependent variable linearly (with an adjusted  $R^2$  value of 0.143). The regression analysis findings indicate that specific consultant competencies significantly influence several ERP implementation critical success factors. Leadership consistently and positively predicts Project Management, Top Management Support, User Training, Technical Infrastructure, and Minimum Customization. Problem Solving significantly influences Top Management Support and Minimum Customization. Communication Skills are important for User Training and Technical Infrastructure. Conflict Management and Teamwork positively affect Data

Accuracy, while Minimizing Errors and Teamwork enhance Suitable Package Selection. Persuasion Skills, Teamwork, and Minimizing Errors significantly improve Software Testing outcomes. However, no significant predictive relationships were identified between consultant competencies and Project Team Competency, Change Management, Clear Goals, or Software Hardware Compatibility. Table 6 presents selected regression results for the relationship between consultant competencies and ERP implementation critical success factors. Because reporting the complete set of regression outputs for all dependent variables would require a large number of tables and reduce readability, the manuscript reports the most substantively meaningful and representative results. Project Management and Software Testing are shown here because they capture important patterns in the broader set of estimated relationships.

| Variable                      | B      | Std. Error | Beta   | T      | p       |
|-------------------------------|--------|------------|--------|--------|---------|
| Constant (Project Management) | 2.153  | 0.826      |        | 2.607  | 0.010   |
| Communication Skills          | -0.231 | 0.117      | -0.182 | -1.972 | 0.050   |
| Industry Knowledge            | -0.064 | 0.106      | -0.053 | -0.610 | 0.543   |
| Presentation Skills           | -0.060 | 0.101      | -0.044 | -0.598 | 0.551   |
| Document Preparation          | -0.208 | 0.111      | -0.148 | -1.870 | 0.063   |
| Persuasion Skills             | 0.086  | 0.111      | 0.063  | 0.776  | 0.439   |
| Business Process Knowledge    | 0.099  | 0.111      | 0.070  | 0.896  | 0.371   |
| Problem Solving               | 0.088  | 0.126      | 0.067  | 0.701  | 0.484   |
| Teamwork                      | 0.110  | 0.161      | 0.108  | 0.686  | 0.494   |
| Customer Orientation          | -0.188 | 0.159      | -0.111 | -1.179 | 0.240   |
| Leadership                    | 0.387  | 0.117      | 0.323  | 3.314  | 0.001   |
| Conflict Management           | 0.088  | 0.100      | 0.078  | 0.874  | 0.383   |
| Minimizing Errors             | 0.060  | 0.166      | 0.046  | 0.362  | 0.718   |
| Constant (Software Testing)   | -0.088 | 0.650      |        | -0.135 | 0.893   |
| Communication Skills          | -0.158 | 0.092      | -0.122 | -1.718 | 0.087   |
| Industry Knowledge            | 0.009  | 0.083      | 0.007  | 0.103  | 0.918   |
| Presentation Skills           | 0.026  | 0.079      | 0.019  | 0.330  | 0.741   |
| Document Preparation          | -0.134 | 0.088      | -0.093 | -1.535 | 0.126   |
| Persuasion Skills             | 0.214  | 0.087      | 0.152  | 2.461  | 0.015   |
| Business Process Knowledge    | 0.024  | 0.087      | 0.017  | 0.280  | 0.780   |
| Problem Solving               | 0.040  | 0.099      | 0.029  | 0.399  | 0.691   |
| Teamwork                      | 0.368  | 0.126      | 0.351  | 2.918  | 0.004   |
| Customer Orientation          | -0.183 | 0.125      | -0.106 | -1.464 | 0.145   |
| Leadership                    | 0.033  | 0.092      | 0.027  | 0.364  | 0.716   |
| Conflict Management           | 0.108  | 0.079      | 0.094  | 1.373  | 0.171   |
| Minimizing Errors             | 0.598  | 0.130      | 0.448  | 4.594  | < 0.001 |

Note: Table 6 reports selected regression results for illustrative and interpretive purposes. Additional regressions were estimated for the full set of ERP implementation critical success factors but are not reproduced here in full in order to maintain readability.

**Table 6.** Consultant Competencies - ERP Critical Success Factors Regression Partial Results

Tables 7 and 8 present selected regression results examining the direct effects of ERP implementation critical success factors and consultant competencies on ERP implementation project success. Because the full set of outputs is extensive, the manuscript reports the most substantively meaningful results in order to preserve readability and interpretive focus. Table 7 shows that Software Hardware Compatibility positively and significantly predicts project success. In contrast, User Training, Suitable Package Selection, and Software Testing exhibit significant negative relationships with project success, suggesting that these factors may become more salient in projects facing implementation difficulty, rework, instability, or unmet expectations.

Table 8 shows that Customer Orientation among consultant competencies significantly and positively predicts ERP implementation project success, underscoring the importance of client focused behaviour in ERP implementations. By contrast, Teamwork shows a significant negative relationship with project success, indicating that higher coordination intensity may characterize more challenging implementation environments. The remaining reported consultant competencies do not exhibit statistically significant direct relationships with project success. Taken together, the results in Tables 7 and 8 further illustrate the complexity of consultant influence on ERP implementation outcomes.

| Variable                        | B      | Std. Error | Beta   | T      | p       |
|---------------------------------|--------|------------|--------|--------|---------|
| Constant (Project Success)      | 2.6222 | 0.151      |        | 17.344 | < 0.001 |
| Effective Communication         | -0.049 | 0.048      | -0.107 | -1.032 | 0.303   |
| Change Management               | -0.007 | 0.022      | -0.015 | -0.346 | 0.730   |
| User Training                   | -0.056 | 0.021      | -0.143 | -2.637 | 0.009   |
| Clear Goals                     | 0.019  | 0.050      | 0.041  | 0.386  | 0.700   |
| Software Hardware Compatibility | 0.087  | 0.027      | 0.192  | 3.290  | 0.001   |
| Minimum Customization           | 0.005  | 0.025      | 0.010  | 0.196  | 0.845   |
| Data Accuracy                   | 0.016  | 0.025      | 0.031  | 0.625  | 0.532   |
| Suitable Package Selection      | -0.166 | 0.036      | -0.408 | -4.578 | < 0.001 |
| Software Testing                | -0.210 | 0.027      | -0.559 | -7.811 | < 0.001 |
| Technical Infrastructure        | -0.040 | 0.022      | -0.082 | -1.810 | 0.072   |
| Top Management Support          | 0.000  | 0.024      | 0.000  | -0.005 | 0.996   |

**Table 7.** ERP Critical Success Factors - Project Success Regression Results

| Variable                   | B      | Std. Error | Beta   | T      | p       |
|----------------------------|--------|------------|--------|--------|---------|
| Constant (Project Success) | 1.527  | 0.238      |        | 6.405  | < 0.001 |
| Communication Skills       | 0.048  | 0.034      | 0.100  | 1.419  | 0.157   |
| Industry Knowledge         | -0.032 | 0.030      | -0.071 | -1.065 | 0.288   |
| Presentation Skills        | -0.019 | 0.029      | -0.036 | -0.636 | 0.525   |
| Document Preparation       | 0.052  | 0.032      | 0.097  | 1.616  | 0.108   |
| Persuasion Skills          | -0.039 | 0.032      | -0.074 | -1.213 | 0.227   |
| Business Process Knowledge | 0.003  | 0.032      | 0.005  | 0.087  | 0.930   |
| Problem Solving            | -0.048 | 0.036      | -0.096 | -1.320 | 0.188   |
| Teamwork                   | -0.343 | 0.046      | -0.881 | -7.401 | < 0.001 |
| Customer Orientation       | 0.328  | 0.046      | 0.510  | 7.128  | < 0.001 |
| Leadership                 | -0.002 | 0.034      | -0.004 | -0.049 | 0.961   |
| Conflict Management        | -0.020 | 0.029      | -0.047 | -0.690 | 0.491   |
| Minimizing Errors          | -0.041 | 0.048      | -0.083 | -0.856 | 0.393   |

**Table 8.** Consultant Competencies - Project Success Regression Results

## 5. Discussion

Our evidence provides substantial support for H1 and mixed support for H2, while the overall pattern is consistent with H3. Consultant competencies are clearly associated with several ERP implementation critical success factors, ERP implementation critical success factors are in turn associated with project success, and direct competency effects on success are generally weak, which is consistent with a primarily indirect pathway operating through critical success factors. This also helps clarify RQ1, as the non uniform signs for training, testing, and teamwork become more interpretable when viewed within the broader competency to critical success factor to success sequence. Our findings largely align with recent ERP research that positions consultants as central contributors to knowledge creation, skill development, and benefit realization across the implementation and post implementation phases (Bawack & Kala Kamdjoug, 2023). Leadership, problem solving, and teamwork appear especially important in shaping ERP outcomes, consistent with evidence that top management support and involvement, organizational fit, and a balanced cross functional project team remain among the most frequently cited ERP CSFs (Leyh et al., 2024). Our results reinforce this logic by showing that stronger consultant leadership and practical problem solving are associated with greater executive involvement, which in turn supports resource allocation to CSFs throughout the project lifecycle (Xie et al., 2022; Leyh et al., 2024). We also observed that teamwork skills significantly affect whether a suitable software and hardware compatibility package is selected and integrated, which is consistent with work showing that team effectiveness in enterprise systems implementation is an emergent property of interaction and coordination rather than a simple sum of individual capabilities (Bellini et al., 2020).

The negative coefficients for user training, software testing, and teamwork should not be interpreted as evidence that these practices are inherently detrimental to ERP implementation success. In complex ERP implementations, these activities are often intensified precisely when projects face misfit, user resistance, rework, schedule pressure, or post configuration instability. Under such conditions, greater training or more extensive testing may reflect a corrective response to implementation difficulty rather than a straightforward

positive input. Similarly, higher reported teamwork may indicate heavy coordination demands in projects with greater interdependence and unresolved issues. Accordingly, these negative signs are more plausibly interpreted as context dependent associations within multivariate models than as universal causal effects. Because the study is cross sectional and based on self reported project evaluations, modelling issues such as suppression effects, omitted project complexity variables, or reverse signaling cannot be fully ruled out. We therefore interpret these coefficients cautiously, as indicators of where managerial effort becomes concentrated under challenging implementation conditions.

Interestingly, we found mixed effects for communication skills, which positively influenced some areas (for example, documentation) but showed negative coefficients in others (for example, top management support and user training). Recent evidence helps interpret this pattern: “effective communication” in ERP implementations is often described as timely, accurate, and tailored to the recipient, but it also requires substantial attention and coordination effort across stakeholders. In parallel, contemporary ERP projects are increasingly characterized by fast moving, dialogue based, and transient documentation practices, which can improve situational awareness while also increasing the volume and velocity of messages that stakeholders must process (Nakayama et al., 2024). In this context, communication abundance can plausibly diminish perceived support or overwhelm end users when it is not carefully curated, aligning with broader evidence that many employees experience excessive volumes of organizational communication. Taken as a whole, our study reinforces the view that ERP success hinges on an integrated set of consultant skills (leadership, coordination, conflict handling, and communication discipline) combined with enabling organizational conditions such as executive involvement, fit focused configuration choices, and strong project governance around the CSFs that matter most (Leyh et al., 2024; Xie et al., 2022; Bawack & Kala Kamdjoug, 2023).

### 5.1. Managerial Implications

Our findings highlight that success in ERP initiatives often hinges on the skills of consultants, including leadership, teamwork, problem solving, and collaboration with key process enablers, such as top management support. Managers should act on the full competency to critical success factor to success pathway: invest in the consultant skills that most reliably strengthen executive support, project management discipline, and goal clarity, then ensure those CSFs are translated into execution. From a managerial standpoint, several actionable insights emerge. Leadership consistently stood out as a driver of multiple CSFs, including project management and user training. Managers hiring external consultants or assembling internal ERP teams should assess candidates’ technical credentials and probe their history of directing cross-functional teams, managing complex organizational dynamics, and influencing senior stakeholders.

The positive link between consultants’ problem solving ability and top management support indicates that skilled troubleshooting builds credibility at the highest levels of the organization. In consultant onboarding and development programs, providing real-life scenarios, simulation exercises, or “challenge sessions” can sharpen these skills. Consultants can maintain leadership commitment throughout the project lifecycle by demonstrating a reliable capacity to handle unexpected setbacks. Teamwork proved essential for software hardware compatibility and customizing ERP solutions. This suggests that managers would benefit from promoting work routines, such as agile stand-ups or cross-departmental workshops, where consultants and client teams jointly map requirements, resolve functional overlaps, and make collective decisions about configuration options. The result is a more unified, context-aware approach that reduces rework and fosters clarity.

Our results reveal a detailed effect of communication skills: well-targeted communication aids documentation and knowledge transfer, but excessive or poorly aligned messaging can lower top management support. Managers should train consultants to adapt their communication volume, frequency, and content to each stakeholder group, ensuring that executives, front-line staff, and IT specialists receive only the most relevant information. While baseline technical competency remains essential, it may not be the primary differentiator for achieving higher-level success factors. Consequently, managers should ensure that project teams feature a blend of technical specialists alongside consultants who excel at people-centric tasks such as conflict resolution, leadership, and creative problem solving. This combination helps harness the full spectrum of capabilities needed for a robust ERP rollout. Organizations can move beyond simple systems installation by adopting these measures and positioning their ERP projects for more profound, sustained benefits. Ultimately, managers who invest in well-rounded, leadership and teamwork-oriented consultants and align them with executive support and informed communication structures will be better positioned to achieve timely, efficient, and strategically meaningful ERP outcomes.

## 6. Conclusion

This study examined how specific consultant competencies influence ERP implementation project success through ERP implementation critical success factors in SAP implementation contexts. Using survey data from 227 SAP consultants, the study investigated whether consultant competencies shape ERP implementation critical success factors and whether these factors, together with consultant competencies, explain variation in ERP implementation project success. Based on the literature, three measurement instruments were used to capture consultant competencies, ERP implementation critical success factors, and project success. The consultant competency scale included 17 competencies, and the study operationalized 13 ERP implementation critical success factors derived from prior research.

Statistical analysis using multiple regression identified significant relationships between certain consultant skills and ERP success factors. Leadership skills have a positive influence on project management effectiveness, suggesting that consultants with strong leadership abilities manage ERP projects more effectively. Problem solving and leadership competencies significantly improved top management support, underscoring that skilled problem solvers and leaders facilitate better cooperation with senior management. Leadership and communication skills significantly enhanced the effectiveness of user training and the implementation of technical infrastructure. Consultants demonstrating these skills established stronger client relationships, thus improving training outcomes and infrastructure setup. Additionally, problem solving and leadership positively impacted minimal customization, indicating that these competencies help avoid unnecessary system complexity. Consultants' conflict management and teamwork significantly enhanced data accuracy. Similarly, minimizing error handling and effective teamwork had a positive influence on the selection of a suitable package. Furthermore, teamwork, persuasion skills, and error handling have notably improved the effectiveness of software testing. However, consultant competencies did not significantly impact ERP project team competency, change management, or the achievement of clear goals. Regarding overall project success, software hardware compatibility emerged as a positive predictor, aligning with existing literature emphasizing its importance. Conversely, compared to common literature findings, user training, suitable package selection, and software testing exhibited negative associations with project success. Among consultant skills, customer orientation has a positive influence on project outcomes, as supported by existing literature. Unexpectedly, teamwork had a negative impact on project success, indicating potential issues with team coordination.

In conclusion, this study provides valuable insights into how specific consultant competencies influence ERP implementation critical success factors and overall project success. While several findings align with existing literature, some unexpected negative relationships, such as the impact of teamwork on project success, reveal underlying complexities that warrant deeper investigation. Beyond listing "what matters," the paper specifies how consultant capabilities translate into success via critical success factors, strengthening the theoretical account. This pathway framing also reconciles mixed results for training, testing, and teamwork observed in prior studies and in our evidence. These insights offer practical guidance for organizations in selecting and developing consultants for ERP implementation projects.

Despite these contributions, several limitations should be acknowledged. First, the study focuses exclusively on SAP consultants in Turkey, which may limit generalizability to other ERP platforms, countries, or implementation environments. Second, the cross sectional survey design does not allow strong causal inference and cannot fully rule out reverse signaling, omitted contextual factors, suppression effects, or other modelling artifacts. Third, the project success construct was intentionally designed as a parsimonious binary implementation outcome index rather than a fully multidimensional success framework. Fourth, the analytical approach was based on multiple regression rather than a formal mediation model or a structural equation modelling framework. This choice was made to provide a transparent and interpretable assessment of the relationships between consultant competencies, ERP implementation critical success factors, and project success in an empirically direct manner. However, multiple regression does not estimate latent constructs, measurement error, simultaneous pathways, or indirect effects with the same level of integration as SEM based approaches. Therefore, the mediation logic proposed in this study should be interpreted as evidence of a regression based pathway pattern rather than as a full latent variable mediation test. Future research could extend the present model by using longitudinal designs, multi source data, mediation analysis, or covariance based or partial least squares SEM to examine the indirect competency CSF success mechanism more comprehensively. Future studies could also expand the analysis to other ERP vendors, compare different national or sectoral contexts, and investigate how project scale, organizational culture, and client consultant relationship quality condition the pathway from consultant competencies to ERP implementation project success.

## References

- Abu Madi, A., Ayoubi, R. M., & Alzbaidi, M. (2024). Spotting the critical success factors of enterprise resource planning implementation in the context of public higher education sector. *International Journal of Public Administration*, 47(2), 73-89.
- Al-Assaf, K., Alzahmi, W., Ahmed, V., & Bahroun, Z. (2025). Comprehensive Review of Enterprise Resource Planning (ERP) Systems And Performance Management Integration In Healthcare. *Management Systems in Production Engineering*.
- Anaya, M., Flak, L. S., & Abushakra, A. (2023). Realizing sustainable value from ERP systems implementation. *Sustainability*, 15(7), 5783.
- Argote, L., Guo, Z., Park, J., & Hahl, O. (2022). The mechanisms and components of knowledge transfer: The virtual special issue on knowledge transfer within organizations. *Organization Science*, 33(3), 1232–1249.
- Asan, H. (2025). Evaluation of Consultant Selection Criteria with Level Based Weight Assessment (LBWA) Method in Intelligent ERP (Enterprise Resource Planning) Perspective. *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 23(01), 29-52.
- Avikal, S., Nigam, M., & Ram, M. (2022). A hybrid multi criteria decision making approach for consultant selection problem in ERP project. *International Journal of System Assurance Engineering and Management*, 13(Suppl 2), 941-950.
- Butarbutar, Z. T., Handayani, P. W., Suryono, R. R., & Wibowo, W. S. (2023). Systematic literature review of Critical success factors on enterprise resource planning post implementation. *Cogent business & management*, 10(3), 2264001.
- Bawack, R. E., & Kala Kamdjoug, J. R. (2023). Managing client-consultant relationships to derive benefits from ERP projects. *Information Technology & People*, 36 (4), 1669–1702.
- Bellini, C., Pereira, R., & Becker, J. (2020). Emergent customer team performance and effectiveness: An ex post facto study on cognition and behavior in enterprise systems implementation. *Communications of the Association for Information Systems*, 47(1), 58.
- Cochran, W. G. (1977). *Sampling Techniques* (3rd edition). John Wiley & Sons, New York.
- Domagała, A. (2021). Post implementation ERP software development: Upgrade or reimplementation. *Applied Sciences*, 11(11), 4937.
- Ellström, D., Holtström, J., Berg, E., & Josefsson, C. (2022). Dynamic capabilities for digital transformation. *Journal of Strategy and Management*, 15(2), 272–286.
- Hammad, M. Z. M. B., Yahaya, J. B., & Mohamed, I. B. (2024). A model for enterprise resource planning implementation in the Saudi public sector organizations. *Heliyon*, 10(2).
- Hoblos, N., Sandeep, M. S., & Pan, S. L. (2024). Achieving stakeholder alignment in digital transformation: A frame transformation perspective. *Journal of Information Technology*, 39(4), 630-649.
- Hustad, E., & Stensholt, J. (2023). Customizing ERP systems: A framework to support the decision-making process. *Procedia Computer Science*, 219, 789-796.
- Ibrahim, M. F., Dharmawan, Y. S., Ngatini, Mustofa, M. I., & Utama, D. M. (2024). A literature review on ERP implementation: Methodologies, module, software, and policy. *AIP Conference Proceedings*, 2927, 050001.
- Jæger, B., Bruckenberg, S. A., & Mishra, A. (2020). Critical success factors for ERP consultancies, a case study. *Scandinavian Journal of Information Systems*, 32 (2), 7.
- Karim, M. R., Nordin, N., Yusof, M. F., Amin, M. B., Islam, M. A., & Hassan, M. S. (2023). Does ERP implementation mediate the relationship between knowledge management and the perceived organizational performance of the healthcare sector? Evidence from a developing country. *Cogent Business & Management*, 10(3), 2275869.
- Kumar Pandey, R., & Kumar, P. (2023). Challenges and opportunities of erp implementation in indian smes: A case study of Jharkhand. *Materials Today: Proceedings*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S2214785323006892>
- Kusumawardhana, R. H., Eitiveni, I., Yaziji, W., & Adriani, Z. A. (2024). Identifying critical success factors (CSF) in ERP implementation using AHP: A case study of a social insurance company in Indonesia. *Journal of Cases on Information Technology (JCIT)*, 26(1), 1-20.

- Leyh, C., Lorenz, A., Faruga, M. J., & Koller, L. (2024). Critical Success Factors for ERP Projects Revisited: An Update of Literature Reviews. In *2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS)* (pp. 131-140). IEEE.
- Llavisaca-Villazhañay, J., Flores-Siguenza, P., Guamán, R., Urdiales, C., & Gento-Municio, Á. M. (2025). Key Drivers of ERP Implementation in Digital Transformation: Evidence from Austro-Ecuadorian. *Administrative Sciences*, 15(6), 196.
- Mamoghli, S. (2025). Exploring the Impact of Customer Organizational Culture on Project Agility in ERP Implementation Projects. *Systems*, 14(1), 22.
- Naceur, R., Cimon, Y., & Pellerin, R. (2024). Information systems adoption and performance: Evidence from Canadian firms linking individual and organizational performance. *Information Systems Management*, 41 (1), 44–57.
- Nakayama, M., Hustad, E., Sutcliffe, N., & Beckfield, M. (2024). Organic transformation of ERP documentation practices: Moving from archival records to dialogue-based, agile throwaway documents. *International Journal of Information Management*, 74, 102717.
- Nguyen, N. D. K., Ali, I., Gupta, S., Chen, R., & Narescho, B. S. (2024). Bridging the nexus between cloud ERP and enterprise resilience. *Journal of Global Information Management (JGIM)*, 32(1), 1-22.
- Rajapakse, D. P. P. K., & Thushara, S. C. (2023). Critical failure factors in ERP implementation: A systematic literature review. *Journal of Business and Technology*, 7(1).
- Ruivo, P., Johansson, B., Sarker, S., & Oliveira, T. (2020). The relationship between ERP capabilities, use, and value. *Computers in Industry*, 117, 103209.
- Salas, W. H. (2023). Model to improve an ERP implementation based on agile best practice: A Delphi study. *Procedia Computer Science*, 219, 1785–1792.
- Weisberg, S., & Cook, R. D. (1982). Residuals and influence in regression.
- Xie, Y., Allen, C., & Ali, M. (2022). Critical success factor based resource allocation in ERP implementation: A nonlinear programming model. *Heliyon*, 8 (8).
- Yazıcıoğlu, Y., & Erdoğan, S. (2014). SPSS uygulamalı bilimsel araştırma yöntemleri (4. Baskı). *Ankara: Detay Yayıncılık*, 45.



## Smart Agriculture in Southeast Asia: Digitization of Smallholders through Organizational Commitment and Collective Ambition

Candraningrat<sup>1\*</sup>, Nimas A. Pratiwi<sup>1</sup> and Azmil Chusnaini<sup>2</sup>

<sup>1</sup>Faculty of Economics and Business, Universitas Dinamika, Surabaya, Indonesia

<sup>2</sup>Department of Management, Universitas Nahdlatul Ulama Surabaya, Surabaya, Indonesia

\*Correspondence: [candra@dinamika.ac.id](mailto:candra@dinamika.ac.id)

### PAPER INFO

#### *Paper history:*

Received 24 December 2025

Accepted 1 June 2026

#### *Citation:*

Candraningrat, Pratiwi, A. N. & Chusnaini, A. (2026). Smart Agriculture in Southeast Asia: Digitization of Smallholders through Organizational Commitment and Collective Ambition. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 211-232

#### *Copyright:*

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

### ABSTRACT

Smart agriculture aims to improve agricultural sustainability output and earnings, increase resilience to climate change, mitigate greenhouse gas emissions, and promote Sustainable Development Goal 2. Understanding the internal organizational factors influencing digital transformation and digital maturity among farmers is crucial to achieving smart agriculture. This study collected a sample of 287 farmers from Indonesia, Malaysia, Thailand, Vietnam, and Cambodia across 15 agricultural zones. The study examines collective ambition through the lens of the theory of acceptance model (TAM) and organizational commitment through the unified theory of acceptance and use of technology (UTAUT), using partial least square structural equation modelling (PLS-SEM). The findings show that collective ambition and organizational commitment have a significant and positive impact on digital transformation and digital maturity. The integration of TAM and UTAUT validates collective views of technology usability, social influence, and effort expectations as essential social mechanisms for coordinated action, sustained engagement, and enduring digital transitions.

**Keywords:** Collective ambition, digital maturity, digital transformation, organizational commitment, smallholders, and smart agriculture

## 1. Introduction

Agriculture is crucial for meeting food requirements and achieving sustainable development. Beyond food production, the sector contributes to ecosystem preservation, biodiversity conservation, and the sustainable use of natural resources, thereby supporting economic growth (Bistline et al., 2022). Southeast Asia (SEA) occupies a strategic position in global food production, serving both regional consumption and international food markets (Tan & Kuebbing, 2023). Despite its importance, the agricultural sector in SEA is highly vulnerable to climate change, which has become one of the most significant threats to food security in Asia. Rising temperatures, prolonged droughts, and extreme weather events have contributed to persistent declines in agricultural productivity across many regions (Lesk et al. 2016). These impacts are particularly severe for smallholder farmers, as they reduce crop yields, threaten household incomes, and undermine rural livelihoods, with broader consequences for global food security (IPCC, 2014). In response to these challenges, the *Sustainable Development Goals* (SDGs) emphasize the importance of digital transformation in agriculture (Damba et al., 2020). Smart agriculture integrates digital technologies such as the Internet of Things (IoT), cloud computing, big data analytics, and artificial intelligence (AI) to enable data-driven agricultural management across the value chain (Trendov et al., 2019). These technologies can enhance productivity,

reduce operational costs, and support more informed decision-making (Quy et al. (2022)). Therefore, smart agriculture is expected to improve the sustainability of agricultural production and income, strengthen adaptation to climate change, reduce greenhouse gas emissions, and contribute to the achievement of SDG 2 (*Zero Hunger*) and SDG 13 (*Climate Action*) (Azadi et al., 2021).

According to FAO (2023), smallholder farmers cultivating less than two hectares of land contribute more than one-third of global food production. Although five-sixths of the world's farms operate on less than two hectares and collectively occupy only 12% of agricultural land, they produce approximately 35% of the global food supply (Lowder et al., 2021), as smallholder farmers crucial role in ensuring food security. Nevertheless, global food systems continue to face significant challenges, as approximately 10.8% of the world's population remains undernourished, while billions of people suffer from micronutrient deficiencies or diet-related health problems (Veldhuizen et al., 2020). Addressing these challenges requires improving the productivity and resilience of smallholder farmers through access to technical support, knowledge, and innovative farming practices (Mujeyi et al., 2021). Smart agriculture offers a promising pathway by integrating digital technologies and data-driven management practices to improve agricultural efficiency and sustainability. Such approaches enable farmers to optimise resource use, reduce environmental impacts, and make more informed production decisions (Schnebelin et al., 2021). The successful implementation of smart agriculture requires technological infrastructure, analytical capabilities, experiential knowledge, and access to accurate information. These capabilities help farmers accelerate production processes, mitigate risks, monitor farm performance in real time, and optimise costs while maintaining product quality (Veldhuizen et al., 2020). Because smart agriculture combines biological, environmental, and digital systems, it represents a knowledge-intensive ecosystem that requires transdisciplinary insights to better understand the complex interactions between farming activities and their surrounding environments (Vishnoi & Goel, 2024).

While existing studies have extensively highlighted the technological benefits of smart agriculture (Bwambale et al., 2023; Javaid et al., 2022; Kernecker et al., 2020; Lova Raju & Vijayaraghavan, 2022; Michels et al., 2020), considerably less attention has been paid to the organizational mechanisms that enable farmers to adopt and effectively utilize digital technologies. This limitation suggests the need to examine how shared aspirations, organizational commitment, and technology acceptance processes collectively influence digital transformation in agriculture. According to Ready & Truelove, (2011), collective ambition can offer a clear and stable direction in environments characterized by continuous, large-scale, and rapid change. For the effective adoption of technology, collective ambition can be grounded in established frameworks within information science and business research. Previous empirical studies have demonstrated that the *technology acceptance model* (TAM) is a robust and widely utilized theory for explaining the adoption of new technologies. Davis et al. (1989) conceptualizes TAM as one of the most influential models for explaining technology adoption, with two main factors influencing an individual's intention to use technology: *perceived usefulness* (PU) and *perceived ease of use* (PEOU). Kets de Vries (2025) conceptualized collective ambition as strengthening an organization's dedication by harmonizing the shared aspirations and values of its members, which instills a sense of ownership, active involvement, and commitment to reaching shared goals. Meyer & Allen (2001) conceptualize organizational commitment as a multidimensional construct, encompassing normative, continuance, and affective components, each associated with an organization's willingness to invest in and institutionalize new technologies. Venkatesh et al. (2003, 2012) explained how individuals adopt and use technology within a more comprehensive framework, *the unified theory of acceptance and use of technology* (UTAUT), through the dimensions of *performance expectancy*, *effort expectancy*, *social influence*, and *facilitating conditions*. We argue that UTAUT can provide a relevant foundation for explaining collective ambition, as its constructs—particularly social influence and performance expectancy—capture the convergence of individual perceptions into a shared understanding to form collective goals and coordinated actions within an organization. Based on this argument, we formulate the following research questions

- RQ1: How do collective ambitions and organizational commitment influence digital transformation and digital maturity for smart agriculture among smallholder farmers?
- RQ2: How does the integration of TAM and UTAUT explain the organizational-level digital transformation beyond the perspective of individual adoption?

We employed *partial least square structural equation modelling* (PLS-SEM) to address the research questions, drawing on data from 287 smallholder farmers across Indonesia, Malaysia, Thailand, Vietnam, and Cambodia. Building on this framework, we make two key theoretical contributions. First, this study applies the TAM at the organizational level to uncover smallholders' intentions to use and adopt smart agriculture, considering two key factors, PU and PEOU. Previous research has applied TAM at the individual level within higher education institutions (Lai et al., 2023), digital marketing (Wistedt, 2024), and hybrid work modelling (Naqshbandi et al., 2023; Petani & Mengis, 2023). In recent years, UTAUT has been widely applied in agriculture, including Internet of Things (IoT) (Ronaghi & Forouharfar, 2020), communication technology (Beza et al., 2018; Narine et al., 2019), and mobile applications (Michels et al., 2020). This study repositions

UTAUT as a determinant influencing organizational commitment and shifts the analytical focus from individual behavioral intentions to the sustainability and resilience of organizations within smart agriculture implementation.

The remainder of this study is structured as follows. Section 2 reviews the relevant literature and establishes the theoretical background. Section 3 presents the development of the research hypotheses and the research design. Section 4 describes the data sources, and SEM methodology. Section 5 presents the empirical results. Section 6 discusses the findings. Finally, Section 7 presents the global implications, conclusion, managerial and theoretical implications, limitations, and future research directions.

## 2. Literature Review

### 2.1. Technology Acceptance Model

Introduced by Davis (1989), the *Technology Acceptance Model* (TAM) is an adaptation of the *Theory of Reasoned Action* (TRA) and the *Theory of Planned Behavior* (TPB), specifically designed to model user acceptance of information systems. The TAM is the predominant framework for explaining how users adopt technology (Krithika & Rajini, 2016). This theory posits that an individual's acceptance of technology is predicted by their behavioural intention, which is primarily influenced by perceived ease of use and the technology's usefulness (Taylor & Todd, 1995). *Perceived usefulness* (PU) represents an individual's perception of the extent to which the utilization of a specific technology enhances performance (Davis, 1989). *Perceived ease of use* (PEOU) is the degree to which a person believes that using a particular system is effort-free (Davis, 1989). Alqatan et al., (2017); Awa et al., (2015), and Kala Kamdjoug et al., (2021) shown that TAM is a robust and widely used theory for explaining the adoption of new technologies by users. Agrarian studies have traditionally emphasized transitions as critical junctures of structural change that provide opportunities to digital transformation processes (Azadi et al., 2021; Banluesapy et al., 2026; Vishnoi & Goel, 2024), and alters the social dynamics within agrarian communities (Castella et al., 2023). Fragomeli et al., (2026) reinforces TAM framework in the agriculture, demonstrating that integrating sustainability (environmental, social, and economic), significantly influence farmers' intention to adopt 4.0 technologies. Other research has shown that a potential user's perception of a technology's usefulness in improving farm performance does not necessarily lead to its adoption (Sayruamyat & Nadee, 2020). This mediating role of attitudes underscores the complexity of the adoption process, in which farmers' evaluations of technology benefits must be translated into positive attitudes before influencing intentions. This aligns with prior evidence that adoption is impeded during the early stages due to uncertainty and limited understanding of the technology's benefits (Marescotti et al., 2021). Consistent with previous research, farmers do evaluate the potential consequences of their actions towards Agriculture 4.0, and the PU of these tools significantly influences their intention to adopt them (Ulhaq et al., 2022).

Additionally, Alexander et al. (2017) and Beban & Gironde (2023) have underlined the transformation of agrarian societies from agriculture-dependent systems to more industrialized and market-oriented modes of production. However, in contemporary SEA, agrarian transition is influenced by complex interactions among various factors, including land dynamics, advancements in agricultural technology, climate change, and diverse forms of structural transformation (Tran & Touch, 2024). The increasing integration of digital technology into agricultural practices has introduced a novel dimension to agrarian change, necessitating a more comprehensive understanding that transcends traditional economic and structural explanations. The TAM framework for technology adoption predominantly focuses on perceptions and decision-making at the individual level, thus offering a limited understanding of the collective and organizational dynamics characteristic of smallholder-based agricultural systems. Therefore, in this study, TAM is positioned as a theoretical lens to understand the factors influencing smart agriculture adoption in smallholders in SEA. Thus, the constructs in the research are contextualized at the organizational level, in accordance with the characteristics of smallholder farming systems that are collective and community-based.

### 2.2. Unified Theory of Acceptance and Use of Technology

UTAUT was proposed by Venkatesh et al., (2003) to analyze research on the adoption of information technology. Reviewing eight classical adoption theories (*the task-technology fit model, diffusion of innovation theory, theory of rational action, theory of planned action, technology adoption model, theory of planned action model, motivational model, and social cognitive theory*), they integrated the arguments and extracted four core constructs that affect willingness to adopt new technology: (1) performance expectation, the degree to which individuals expect to be helped by the use of the technology; (2) effort expectation, the level of effort required

by individuals to use the technology; (3) social influence, the degree to which an individual's social environment affects their use of the technology; and (4) facilitating conditions, the degree to which the internal and external conditions of the individual's expected use of the technology promote their use of the technology. UTAUT is widely accepted in social research because of its explanatory power (up to 70%) (Venkatesh et al., 2003). In recent years, in the field of agriculture, UTAUT has been widely applied to IoT (Ronaghi & Forouharfar, 2020), communication technology (Beza et al., 2018; Narine et al., 2019), and mobile applications (Michels et al., 2020), among other areas of technology adoption intention research.

Despite the growing body of research on smart agriculture in SEA, existing studies provide limited insight into the collective dynamics and organizational structures that shape technology adoption among smallholder farmers (Parvathi & Waibel, 2016). This limitation is particularly important in ongoing agrarian transitions across the region. While agrarian transitions contribute to agricultural modernization and support adaptation to climate change and emerging development paradigms, their effects are often uneven across farming households (Tran & Touch, 2024). For instance, evidence from the Mekong Delta shows that even relatively better-off farmers, including landholders, face challenges in maintaining stable incomes due to volatility in rice markets (Betcherman et al., 2021). These conditions suggest that agricultural transformation, including digitalization, extends beyond the mere adoption of technology. Rather, it is closely associated with economic resilience, collective coordination, and the organizational capacity of farmers to respond to increasingly complex market and environmental challenges. Therefore, decisions regarding technology adoption are not solely individual choices but are shaped by social interactions, shared objectives, and coordinated actions within farming communities and organizations. To better understand these dynamics, this study employs the UTAUT as a theoretical framework for explaining technology adoption among smallholder farmers in SEA. Specifically, UTAUT provides a useful lens for examining how social and organizational factors influence the adoption of smart agriculture technologies within collective and community-based farming systems. The study seeks to explain how organizational commitment and collective ambition interact with technology acceptance processes to facilitate digital transformation and enhance digital maturity among smallholder farmers in Southeast Asia.

### 3. Hypotheses Development

#### 3.1. Collective Ambition and Digital Transformation

Collective ambition offers a clear direction in environments characterized by sustained, large-scale, and accelerating changes. A shared vision and common objectives generate social energy that facilitates the acceptance of transformation at both the individual and collective levels (Ready & Truelove, 2011). Within organizations, collective ambition enhances interdependence among members and aligns their efforts toward a common purpose. Digital transformation inherently necessitates cross-functional collaboration, openness to innovation, and a willingness to challenge established routines, which are processes fundamentally driven by a shared sense of purpose. When organizational goals are coherently aligned through core elements, such as purpose, vision, strategic targets, and leadership behaviors, digital transformation transcends a purely technological initiative and becomes a profound organizational shift. In smart agriculture, digital transformation extends beyond the deployment of digital tools to encompass aligning strategic directions, objectives, and perceived technology values. This alignment is particularly critical in smallholder communities, where farming activities are highly interconnected and dependent on shared resources and coordinated actions. Collective ambition enables a unified and coordinated response to technological change, reduces fragmentation, and fosters collective engagement in digital initiatives.

Building on the *technology acceptance model* (TAM), we posit that individual perceptions of *perceived usefulness* (PU) and *perceived ease of use* (PEOU) form the cognitive basis for technology adoption (Davis, 1989; Venkatesh et al., 2003). When organizational members hold positive perceptions of technological benefits and usability, these individual assessments coalesce into collective beliefs about the strategic importance of digital technology. Through this collective sensemaking process, individual evaluations are transformed into shared aspirations that align digital adoption with organizational goals, such as enhanced productivity, resource efficiency, and climate resilience. Accordingly, collective ambition facilitates coordinated action and sustained digital transformation rather than isolated or sporadic technology use. Therefore, based on TAM, it is anticipated that collective ambition will influence digital transformation through shared perceptions of its usefulness and ease of use. The following hypothesis is proposed:

H1. Collective ambition has a positive influence on digital transformation.

### 3.2. Organizational Commitment and Digital Transformation

Digital transformation in smart agriculture transcends the mere adoption of digital technologies, encompassing changes in work practices, inter-actor collaboration, and continuous learning processes. Organizational commitment functions as a binding mechanism that aligns individual interests with the broader transformation agenda, thus mitigating resistance to change and enhancing the consistency of digital implementation processes. Smallholders who cultivate strong affective and normative attachments to their organizations or communities are more inclined to support digital initiatives and actively engage in transformation efforts (Hanelt et al., 2021; Wessel et al., 2021). Organizational commitment has long been a central construct in organizational psychology and management research; however, its role in digital transformation has only recently garnered scholarly attention (Naima et al., 2021). A previous study demonstrated that technological developments reshape affective commitment in altering how individuals form emotional bonds with organizations in the digital era (Melović et al., 2020). Empirical evidence further highlights the significance of organizational commitment across diverse digital transformation, including brand marketing and e-businesses (Melović et al., 2020) and sustainability principles (Najm et al., 2023). Research on SMEs during the COVID-19 pandemic also emphasizes that sustained organizational commitment is critical for navigating digital transformation amid uncertainty (Winarsih et al., 2021). These studies indicate that organizational commitment is a decisive factor in effective implementation and long-term sustainability of digital transformation initiatives. From the UTAUT perspective, we argue organizational commitment supports continued technology use by strengthening performance expectancy, reinforcing social influence, and enabling facilitating conditions. Accordingly, we propose the following hypothesis:

H2. Organizational commitment positively influences digital transformation.

### 3.3. Collective Ambitions and Digital Maturity

Digital maturity pertains to the degree to which digital technologies are systematically integrated, sustainably utilized, and strategically leveraged to enhance organizational processes, decision-making, and long-term operational capabilities. In contrast to digital transformation, which focuses on the process of change, digital maturity signifies an organization's ability to consistently and purposefully exploit digital technologies over time, reflecting the stabilization and institutionalization of digital practices. From a socio-organizational perspective, this progression is not solely driven by technological adoption but also by shared cognitive and strategic alignment among organizational members (Blount and Leinwand, 2019). Collective ambition serves as a crucial social and cognitive foundation that facilitates the progression from digital transformation to advanced digital maturity. When organizational members, including smallholders, articulate explicit collective goals regarding the purpose and value of digital technologies, these shared beliefs function as a form of collective sensemaking that guides coordinated action and sustained engagement (Mian et al., 2025). As a result, digital practices are more likely to be institutionalized into daily routines rather than being perceived as temporary initiatives, promoting consistency in the development of digital capabilities, the use of data-driven decision-making, and collaboration among actors within the smart agriculture ecosystem. This perspective is further supported by prior studies emphasizing the importance of shared goals, collaboration, and organizational alignment in shaping digital maturity (Foguesatto et al., 2024; Mian et al., 2025). Inter-organizational collaboration and alignment around common objectives reduce fragmentation and enhance coordination, whereas knowledge-sharing routines and complementary resources contribute to sustained capability development, particularly in resource-constrained environments. Moreover, digital maturity is closely associated with dynamic capabilities, such as adaptability, continuous learning, and resource reconfiguration, all of which depend on a shared orientation toward long-term goals (Jie et al., 2025). In this regard, collective ambition acts as a unifying mechanism that aligns cognition, action, and organizational direction, enabling organizations to continuously assess, refine, and advance their digital capabilities while supporting the routinization of digital practices over time. Accordingly, this study proposes the following hypothesis:

H3. Collective ambition positively influences digital maturity.

### 3.4. Organizational Commitment and Digital Maturity

Organizational commitment, which denotes individuals' psychological attachment to and investment in organizations, is pivotal in decision-making processes and successful adaptation to digital change. Najm et al. (2023) contend that digital transformation serves as a mediating mechanism between sustainability principles and organizational commitment, highlighting that digitalization transcends mere technical processes by reshaping individual attitudes, identities, and organizational affiliations. Meanwhile, maturity

signifies optimal conditions, serving as an indicator of achievement and a framework for addressing or preventing issues (Nikkhou et al., 2016). The maturity model offers a comprehensive understanding of a company's status and the necessary steps to implement an Industry 4.0 strategy (Akdil et al., 2018). Maturity serves as a criterion for evaluation and beneficial for establishing the foundational phase, leading to a more advanced final phase (Proença and Borbinha, 2016). Digital maturity signifies an advanced stage of digital transformation in which technologies are deeply integrated into organizational structures, cultures, and routines, facilitating improved work organization, streamlined processes, and the application of data-driven decision-making and advanced analytical models (Rahman, 2025). These conditions enhance collaboration, strengthen alignment with shared organizational objectives, and reinforce long-term commitment to sustainability. Herscovitch and Meyer (2002) demonstrated that individuals with high levels of organizational commitment are more proactive in problem solving and contribute more positively to organizational performance. From a socio-technical perspective, organizational commitment facilitates the integration of human and technological elements by fostering loyalty, a sense of responsibility, and a willingness to learn and adapt continuously. This construct is also conceptually aligned with the UTAUT, which posits that performance expectancy, effort expectancy, social influence, and facilitating conditions support sustained technology use. When smallholders develop strong affective and normative attachments to their organizations, they are better positioned to leverage institutional support, maintain digital practices, and progressively strengthen their digital capabilities. Consequently, organizational commitment is conceptualized as a key driver of achieving higher and more sustainable levels of digital maturity. Based on the above argument, the following hypothesis is proposed:

H4. Organizational commitment positively influences digital maturity.

To establish a theoretical foundation, a conceptual framework was developed to capture the relationships between key technology acceptance constructs and digital capability outcomes. The proposed model, presented in Figure 2, delineates the hypothesized organizational factors of digital transformation and digital maturity.

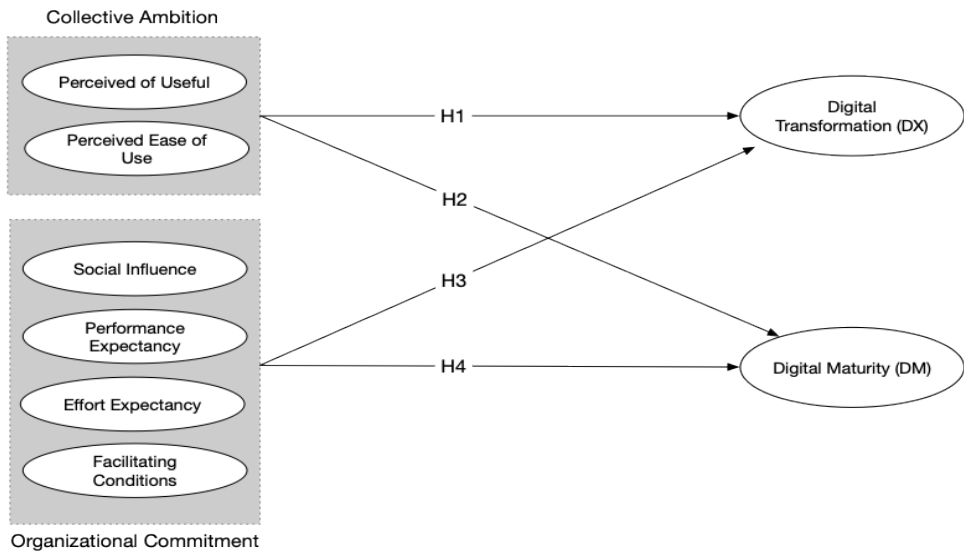


Figure 1. Proposed Conceptual Framework

4. Methodology

This study was conducted in the SEA region between August and September 2025 through field surveys in agricultural areas in Indonesia, Malaysia, Thailand, Vietnam, and Cambodia. The SEA region contributes to approximately 49.7% of Asia's total population (Population Reference Bureau, 2025). The region's food production systems are vital to regional and global food security. From 2011 to 2022, the region exported nearly 519 million metric tons of rice produced from its paddies, accounting for over a quarter of the world's production of rice (FAO, 2023). Mixed farming systems are the traditional form of rural livelihood in SEA and

are still practiced by a large number of the estimated 100 million smallholders in the region (Tan & Kuebbing, 2023). Attempting to cover the entire population in this study would be tedious owing to time and budget constraints; therefore, G-Power version 3.1 was utilized in the selection of an appropriate sample size. G-Power software recommended a sample size of 215 to test the proposed research model with an effect size of  $f^2 = 0.05$ ,  $\alpha = 0.05$ , power = 0.95, and two predictors. The PLS-SEM, on the other hand, required a minimum sample size of 100 (Reinartz et al., 2009). Participants were recruited using *purposive* and *snowball sampling*, covering 15 agricultural zones and involving approximately 35 respondents from each zone. A total of 525 questionnaires were distributed and collected, of which 287 were deemed valid, resulting in a response rate of 54.5%. The distribution of valid respondents is as follows: Indonesia (102), Malaysia (54), Thailand (51), Vietnam (47), and Cambodia (33), ensuring balanced representation across the region.

Survey questionnaires were employed to collect data from the study area. The distribution process was facilitated through farmers' cooperatives, agricultural extension officers, and local non-governmental organizations involved in digital agriculture initiatives. The questionnaire comprised two sections. The first section gathered information on respondents' demographic characteristics, including digital transformation readiness, age, education, annual income, cultivated land area, and land tenure status. The second section measured the study constructs, namely collective ambition, organizational commitment, digital transformation, and digital maturity. All measurement items were adapted from established scales reported in previous studies (see Appendix A). Responses were assessed using a five-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

#### 4.1. Data Analysis

A *structural equation model* (SEM) was employed to examine the relationships among collective ambition, organizational commitment, digital transformation, and digital maturity among smallholder farmers. Owing to the complexity of the developed model, which includes multiple latent constructs, the *partial least squares-based* method (PLS-SEM) was chosen instead of *covariance-based structural equation modelling* (CB-SEM). PLS-SEM consists of two main components. The first component is the measurement model, referred to as the outer model. The outer model estimates the contribution of each indicator in representing its associated latent variable and assesses how well the combined set of indicators measures a construct. The second component is the structural model, referred to as the inner model. The inner model measures the direct and indirect relationships between latent variables (Hair et al., 2014).

#### 4.2. Assessment of Measurement Model

Measurement model test revealed lower and higher factor loading values of 0.695 and 0.859, respectively (Table 1), indicating satisfactory internal consistency. As recommended by Gudergan et al. (2008), a factor loading greater than 0.70 is best for reliability assessment; however, if the AVE does not attain a satisfactory level, a value between 0.40 and 0.70 may be omitted. Hence, no item was dropped, as our AVE result attained the suggested cut off point of 0.50 (Hair et al., 2016).

| Construct                 |     | Items of the construct                                                    | Outer Loadings |
|---------------------------|-----|---------------------------------------------------------------------------|----------------|
| Collective ambitions      | CA1 | Clear understanding of organizational goals                               | 0,785          |
|                           | CA2 | Organizational vision provides clear direction for digital transformation | 0,766          |
|                           | CA3 | Realistic targets for sustainable digital transformation                  | 0,761          |
|                           | CA4 | Strategic priorities align with shared vision                             | 0,785          |
|                           | CA5 | Commitment to stakeholder promises                                        | 0,704          |
|                           | CA6 | Organizational values guide actions                                       | 0,783          |
|                           | CA7 | Leadership behavior aligns with vision and values                         | 0,718          |
| Organizational commitment | OC1 | Strong sense of belonging to organization                                 | 0,786          |
|                           | OC2 | Commitment to improving digital maturity                                  | 0,812          |
|                           | OC3 | Participation in digital transformation is meaningful                     | 0,815          |
|                           | OC4 | Responsibility toward digital transformation challenges                   | 0,774          |
| Digital transformation    | DX1 | Farming activities digitally connected with stakeholders                  | 0,859          |
|                           | DX2 | Digitalization across farming activities                                  | 0,825          |
|                           | DX3 | Digital integration from production to distribution                       | 0,785          |
|                           | DX4 | Use of digital tools for marketing                                        | 0,800          |
|                           | DX5 | Use of digital tools for transactions                                     | 0,803          |

|                  |     |                                                |       |
|------------------|-----|------------------------------------------------|-------|
| Digital maturity | DM1 | Digital systems are safe and accessible        | 0,701 |
|                  | DM2 | Ability to integrate and use agricultural data | 0,695 |
|                  | DM3 | Support through training and digital literacy  | 0,719 |
|                  | DM4 | Collaboration in digital farming practices     | 0,787 |
|                  | DM5 | Use of digital systems to improve efficiency   | 0,774 |

**Table 1.** Outer Loadings of the Measurement Model

The measurement model also reveals convergent and discriminant validity following Fornell–Larcker’s criterion and the *Heterotrait–Monotrait Ratio* (HTMT). Convergent validity is the degree to which indicators of a specific construct converge or share a high proportion of variance in common (Hair et al., 2016). The findings indicate that the *composite reliability* (CR) ranged between 0.773 and 0.882, while the AVE ranged between 0.515 and 0.663, both of which exceeded the cut-off point of 0.50 (Hair et al., 2010) and satisfied the convergent validity. To determine construct reliability and internal consistency, a cut-off point above 0.70 was considered for evaluating rhoA and *Cronbach’s alpha* (Dijkstra & Henseler, 2015). In our study, the rhoA threshold ranged from 0.773 to 0.882, indicating a satisfactory level of construct reliability. Cronbach’s alpha for this study exceeded the threshold of 0.70, indicating internal consistency of the data. Table 2 presents the summarized results of the construct reliability and validity.

| Constructs                | Cronbach’s Alpha | Rho_A | Composite Reliability | AVE   |
|---------------------------|------------------|-------|-----------------------|-------|
| Collective ambitions      | 0,877            | 0.882 | 0,882                 | 0,574 |
| Organizational commitment | 0,809            | 0.773 | 0,814                 | 0,635 |
| Digital transformation    | 0,873            | 0.878 | 0,878                 | 0,663 |
| Digital maturity          | 0,765            | 0.814 | 0,773                 | 0,515 |

**Table 2.** Construct Reliability and Validity

Discriminant validity was assessed by differentiating the square root of the AVE and factor correlation coefficients. In this study, three methods were used to verify discriminant validity. Table 3 presents the results of the Fornell–Larcker criterion for discriminant validity. Fornell & Larcker (1981) stated that the average variance shared between each construct and its measurements should exceed the variance shared between the construct and other constructs. The findings revealed that each variable conforms to a satisfactory level of discriminant validity. This is because there is no correlation (shown off-diagonal) exceeding the square root of the AVE (shown diagonal), subsequently indicating the validity of all constructs according to the discriminant validity test by Fornell and Larcker (1981).

|                           | Collective Ambition | Digital Maturity | Digital Transformation | Organizational Commitment |
|---------------------------|---------------------|------------------|------------------------|---------------------------|
| Collective Ambition       | <b>0,758</b>        |                  |                        |                           |
| Digital Maturity          | 0,402               | <b>0,717</b>     |                        |                           |
| Digital Transformation    | 0,409               | 0,476            | <b>0,815</b>           |                           |
| Organizational Commitment | 0,380               | 0,337            | 0,425                  | <b>0,797</b>              |

Note(s) : The square root of AVE is presented in diagonals and correlations, off-diagonals

**Table 3.** Discriminant Validity

Second, cross-factor loadings were examined to determine the discriminant validity of the measurement model. Table 4 presents the results of the cross-factor loadings. Hair et al. (2016) suggested that all indicators must have higher loadings for their respective variables. The results showed that all indicators had the highest values for their respective constructs, ensuring the attainment of discriminant validity in the existing study. Third, to accurately ascertain discriminant validity, the *heterotrait–monotrait* ratio (HTMT) was determined. An HTMT value lower than 0.85–0.90 is considered satisfactory discriminant validity (Gold et al., 2001; Henseler et al., 2015; Kline, 2023). The findings indicated that all HTMT values of correlations had values below 0.90, confirming satisfactory discriminant validity for all variables (Table 5).

| Item | Collective Ambition | Digital Maturity | Digital Transformation | Organizational Commitment |
|------|---------------------|------------------|------------------------|---------------------------|
| CA1  | 0.785               | 0.310            | 0.336                  | 0.355                     |
| CA2  | 0.766               | 0.372            | 0.313                  | 0.276                     |
| CA3  | 0.761               | 0.305            | 0.338                  | 0.303                     |
| CA4  | 0.785               | 0.258            | 0.314                  | 0.339                     |
| CA5  | 0.704               | 0.226            | 0.276                  | 0.160                     |
| CA6  | 0.783               | 0.364            | 0.333                  | 0.297                     |
| CA7  | 0.718               | 0.263            | 0.243                  | 0.267                     |
| DM1  | 0.244               | 0.701            | 0.264                  | 0.217                     |
| DM2  | 0.280               | 0.695            | 0.314                  | 0.194                     |
| DM3  | 0.309               | 0.719            | 0.321                  | 0.271                     |
| DM4  | 0.331               | 0.787            | 0.394                  | 0.300                     |
| DM5  | 0.265               | 0.680            | 0.408                  | 0.208                     |
| DX1  | 0.391               | 0.418            | 0.859                  | 0.382                     |
| DX2  | 0.373               | 0.395            | 0.825                  | 0.325                     |
| DX3  | 0.336               | 0.359            | 0.785                  | 0.336                     |
| DX4  | 0.285               | 0.393            | 0.800                  | 0.315                     |
| DX5  | 0.266               | 0.370            | 0.803                  | 0.371                     |
| OC1  | 0.314               | 0.276            | 0.316                  | 0.786                     |
| OC2  | 0.286               | 0.273            | 0.319                  | 0.812                     |
| OC3  | 0.359               | 0.296            | 0.392                  | 0.815                     |
| OC4  | 0.241               | 0.222            | 0.319                  | 0.774                     |

Table 4. Cross Loadings

| Variables                 | Collective Ambition | Digital Maturity | Digital Transformation | Organizational Commitment |
|---------------------------|---------------------|------------------|------------------------|---------------------------|
| Collective Ambition       | -                   |                  |                        |                           |
| Digital Maturity          | 0.479               | -                |                        |                           |
| Digital Transformation    | 0.460               | 0.579            | -                      |                           |
| Organizational Commitment | 0.442               | 0.419            | 0.501                  | -                         |

Table 5. Heteroit-Monotrait Ratio (HTMT)

5. Empirical Results

5.1. Respondents’ Socio-Demographic Profile

As shown in Table 6, 78.4% of smallholders indicated a willingness to engage in digital agricultural transformation. The mean age of the respondents was 46 years, with the predominant age group being 40–50 years (39.6%). Most smallholders had a lower secondary education background, with 8.2% having completed only primary school or not attending school at all and 52.7% having dropped out of junior high school. Nonetheless, 17.5% of the respondents had attained higher education. The average annual income of agricultural households was \$3,500, with 14.3% earning over \$6,000 per year. The average arable land area was 1.8 ha, and 51.2% of the smallholders had less than 1 ha. Land ownership was diverse, comprising 65.4% rice fields, 28.6% dry land, and 6.0% a combination of both.

| Characteristic                     | Percentage (%) | n   |
|------------------------------------|----------------|-----|
| Digital Transformation Readiness   |                |     |
| Willing to get involved            | 78.4           | 225 |
| Unprepared                         | 21.6           | 62  |
| Age                                |                |     |
| Average age (years)                | 46             | -   |
| 40–50 years                        | 39.6           | 114 |
| Education                          |                |     |
| Elementary school or not in school | 8.2            | 24  |
| Junior                             | 52.7           | 151 |

|                           |       |     |
|---------------------------|-------|-----|
| High School/Diploma       | 21.6  | 62  |
| College                   | 17.5  | 50  |
| Household Income per Year |       |     |
| Average (USD)             | 3,500 | -   |
| > 6,000 USD               | 14.3  | 41  |
| ≤ 6,000 USD               | 85.7  | 246 |
| Cultivated Land Area      |       |     |
| Average (ha)              | 1.8   | -   |
| < 1 ha                    | 51.2  | 147 |
| ≥ 1 ha                    | 48.8  | 140 |
| Land Type                 |       |     |
| Paddy                     | 65.4  | 188 |
| Dry land                  | 28.6  | 82  |
| Rice fields and dry land  | 6     | 17  |

Table 6. Respondents' Socio Demographic Information

5.2. Structural Model Evaluation

To assess potential collinearity issues within the structural model, the variance inflation factor (VIF) was examined. VIF values are commonly used to detect multicollinearity among predictor variables, which may distort model estimates and significance levels. Hair (2011) suggested that collinearity issues may arise when VIF values exceed 5.0, whereas Diamantopoulos & Siguaw (2006) recommended a more conservative threshold of 3.3. The results indicated that all VIF values ranged from 1.358 to 2.292, remaining well below the recommended thresholds and suggesting the absence of multicollinearity concerns (Hair et al., 2010). The predictive accuracy of the model was subsequently evaluated using the coefficient of determination ( $R^2$ ), which represents the proportion of variance explained by the exogenous constructs. The results showed that the  $R^2$  values for digital transformation and digital maturity were 0.252 and 0.201, respectively. These findings indicate that collective ambition and organizational commitment explain 25.2% of the variance in digital transformation and 20.1% of the variance in digital maturity (Wetzels et al., 2009).

The results indicate that collective ambitions have a significant positive effect on digital transformation ( $\beta = 0.289$ ,  $t = 5.827$ ), supporting H1. Similarly, organizational commitment significantly influences digital transformation ( $\beta = 0.315$ ,  $t = 6.203$ ), supporting H2. Furthermore, collective ambitions significantly affect digital maturity ( $\beta = 0.320$ ,  $t = 5.767$ ), and organizational commitment also has a significant effect on digital maturity ( $\beta = 0.215$ ,  $t = 3.854$ ), thus supporting H3 and H4 (Table 7). Figure 2 illustrates the results of the structural model assessment.

The  $F$ -square ( $f^2$ ) values were computed using the effect size with  $R$ -squared ( $R^2$ ). An  $F$ -square value of 0.35 reflects a high effect size, whereas 0.15 and 0.02 are considered medium and small effect sizes, respectively (Cohen, 1988). As shown in Table 7, the  $F$ -square value of the collective ambition for digital transformation (0.110) is considered small-medium, organizational commitment to digital transformation (0.096) is considered small, collective ambition for digital maturity (0.050) is considered small, and organizational commitment to digital maturity (0.114) is considered small-medium.

| Hypo | Path Relation | $\beta$ | St. Deviation | t-value | $f^2$ | Comment  |
|------|---------------|---------|---------------|---------|-------|----------|
| H1   | CA → DX       | 0.289   | 0.050         | 5.827   | 0.110 | Accepted |
| H2   | OC → DX       | 0.315   | 0.051         | 6.203   | 0.096 | Accepted |
| H3   | CA → DM       | 0.320   | 0.055         | 5.767   | 0.050 | Accepted |
| H4   | OC → DM       | 0.215   | 0.056         | 3.854   | 0.114 | Accepted |

Table 7. Results of Direct Effects among Constructs

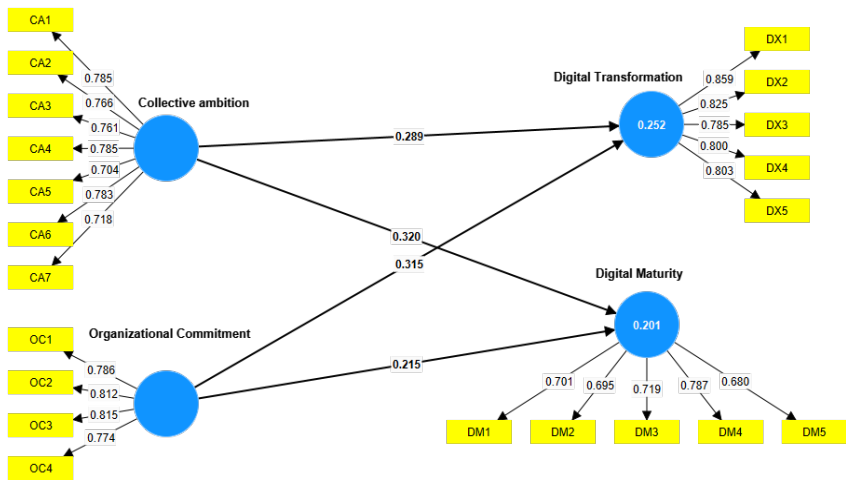


Figure 2. Structural Model Results from SEM-PLS Analysis

6. Discussion

This study represents one of the pioneering in the academic literature, a comprehensive evaluation of collective ambition in digital transformation within smart agriculture, specifically for smallholder farmers. Additionally, this research offers preliminary insights into the potential for achieving long-term objectives of digital technology at the community farming level. Beyond its findings, the study proposes a set of indicators designed to assess the integrity of the processes involved in forming and planning collective ambition for the adoption of digital technology.

Our initial findings indicate that collective ambition significantly impacts the digital transformation of smart agriculture among smallholders in SEA. There is a shared commitment to adopting digital technologies to enhance agricultural productivity, efficiency, and sustainability. The TAM framework elucidates technology acceptance through two primary constructs: PU and PEOU. For instance, if a farming community perceives that digital technology will improve crop yields (*perceived usefulness*), collective ambition will facilitate faster adoption. Furthermore, if the technology is deemed user-friendly, collective ambition encourages motivation to disseminate knowledge and provide mutual support (*perceived ease of use*). IoT technologies bring substantial benefits to farmers and agricultural businesses, including increased productivity, efficiency, cost reduction, and improved decision-making processes. Through IoT utilization, farmers can automate monitoring systems, employ sensors and actuators, and incorporate cloud-based technologies for data analysis and decision-making (Lova Raju & Vijayaraghavan, 2022). IoT acceptance in agriculture, industry influence refers to the impact and role that key stakeholders within the farming community and society have on individuals' decision-making processes (Yang et al., 2022). These influential figures include experienced agro-entrepreneurs, prominent members of the agricultural sector, and individuals who hold significance in a person's life and decision-making processes.

Our second finding indicates that organizational commitment positively influences digital transformation. *Effort expectancy* emerged as notable dimensions farmers' engagement with digitalization initiatives (OC3; mean = 4.20). *Social influence* significantly impacts the sense of belonging within the organization, thus accelerating adoption with group norms (OC1; mean = 4.15). *Facilitating conditions* represent the weakest dimension, that institutional and structural support mechanisms are insufficient to fully support long-term digital transformation (OC4; mean = 4.02). These findings are consistent with Ali et al. (2020) and Ulhaq et al., (2022) which emphasizes attitudes towards technology in influential farmers' adoption behavior. Sen et al. (2024) focus on behavioural factors, including subjective norms, beliefs, risk perceptions, and attitudes toward utility, which predominantly influence technology adoption in Cambodia, the Philippines, and Vietnam. Our findings confirm that individual behavioral characteristics and farmers' emotional attachment to their organizations are key factors in accelerating digital transformation. Organizational commitment is crucial for facilitating digital transformation in farmers' psychological readiness, strengthening their sense of behavioural control, and fostering trust in the long-term benefits of digital technology. When smallholder farmers possess a deep emotional attachment and identification with their organization, they exhibit greater confidence in the benefits of technology (*performance expectancy*), feel

more competent in its use (*effort expectancy*), are more responsive to collective influence (*social influence*), and optimize structural support (*facilitating conditions*).

The third finding highlights collective ambition in driving digital maturity, particularly through perceptions related to usefulness. Digital maturity is achieved when technology is viewed as a practical tool for realizing an organization's long-term objectives. Collective ambition serves as a strategic guide, aligning smallholder farmers with organizational management on digital transformation goals. This ensures consistency in the development of digital capabilities, readiness to adopt smart agriculture technologies, and drives innovation in agricultural production and management processes. According to Roeven et al., (2025), collective ambition serves as a shared vision and a platform for continuous learning. This facilitates the harmonization of understanding, the negotiation of differences, and the gradual enhancement of digital capacity among participants. Organisational managers had a role in improving digital literacy, expanding technological knowledge (e.g. soil sensors, crop monitoring software, and data management systems), and integrating technology into routine operational tasks. Because food communities must reconcile diverse perspectives and practical needs, smallholder farmers in SEA face similar dynamics when implementing collective digital ambitions. Without systematic learning mechanisms, collective visions risk becoming merely statements of intent rather than processes of continuous development.

The fourth finding indicates that organizational commitment exerts a positive influence on digital maturity, with relevant factors of effort expectancy and performance expectancy. A willingness to learn and confidence in one's ability to increase performance are essential for attaining digital maturity. Robust organizational commitment facilitates the effective integration of human and technological elements in SEA smart agriculture practices. From a socio-technical perspective, organizational commitment develops coherence, loyalty, and psychological readiness of smallholder farmers to adopt digital practices in agricultural activities (Paul et al., 2022). Smallholder farmers who possess an emotional attachment to and alignment with organizational goals are more likely to adopt new technologies, engage in training, and consistently utilize digital systems. Taghibaygi & Alibaygi (2024) investigated organizational ethical commitment toward digital integration and found that effectiveness positively influences ethical norms, which subsequently affect farmers' intentions to employ digital technology in agriculture in Iran. Farmers who perceive that digital agricultural technology enhances food security are more motivated to adopt it. Enhancing farmers' understanding of food security challenges and the significance of collaborative efforts can strengthen their ethical commitment to digital technology. Given that farmers rate training programs as effective, implementing educational initiatives, such as digital agriculture courses, demonstration farms, dialogues with landowners, and direct observation of the tangible outcomes of digital technology, could be a key strategy to bolster farmers' ethical commitment to the broader adoption of digital technology.

## 7. Conclusion

This study reveals collective ambition and organizational commitment to digital transformation and digital maturity among smallholder farmers in smart agriculture in SEA. Our initial finding indicates that collective ambition significantly impacts the digital transformation of smart agriculture among smallholders in SEA. The TAM framework elucidates technology acceptance through two primary constructs, PU and PEOU. Second, organizational commitment influences digital transformation. Our findings confirm that individual behavioral characteristics and farmers' emotional attachment to their organizations are key factors in accelerating digital transformation. Organizational commitment is crucial in facilitating digital transformation in farmers' psychological readiness, strengthening their sense of behavioral control, and fostering trust in the long-term benefits of digital technology. Third, collective ambition drives digital maturity, particularly through the dominant PU dimension. Collective ambition serves as a strategic guide, aligning smallholder farmers with organizational management on digital transformation goals. This ensures consistency in the development of digital capabilities, readiness to adopt smart agriculture technologies, and drives innovation in agricultural production and management processes. The fourth finding indicates that organizational commitment exerts a positive influence on digital maturity, with effort expectancy and performance expectancy serving as fundamental determinants. A willingness to learn and confidence in one's ability to increase performance are essential for attaining digital maturity. Robust organizational commitment facilitates the effective integration of human and technological elements in smart agriculture practices in SEA.

### 7.1. Managerial Implications

This study makes a significant contribution by introducing a novel conceptual framework for comprehending institutionally driven digital transformation in agriculture. Although the framework is rooted in SEA, it also applies to other developing regions that exhibit similar social, institutional, and structural characteristics. The

findings provide crucial managerial and policy-relevant insights for key stakeholders, including farmers, agricultural cooperatives, non-governmental organizations, and public institutions. Promoting a digitally oriented organizational commitment requires, foremost, articulating a clear digital vision and cultivating an organizational culture that values innovation and continuous learning. Leso et al. (2023) concluded that the results emphasize culture, organizational arrangements, and leadership in digital transformation, addressing earlier theoretical calls for the importance of strategic orientation in digitalization. This study advances the practice of smallholder groups to establish structured platforms for collective deliberation and coordination concerning digital transformation goals. Cooreman et al. (2021) argued that on-farm demonstrations can serve as effective learning spaces that support transformative learning when combined with facilitated dialogue and participatory approaches. The impact of such demonstrations is largely contingent upon their relevance to smallholders' real-world conditions and their capacity to integrate hands-on experience. Embedding knowledge exchange within field demonstrations enhances learning outcomes with stimulating critical reflection, constructive cognitive tension, and deeper knowledge sharing, compared to demonstrations that lack intentional dialogical facilitation. Furthermore, given that most smallholders remain at an intermediate stage of digital maturity, organizational leadership must proactively identify and address the key constraints that impede the complete adoption and effective utilization of digital technologies. An important, often under emphasized strategy is to integrate climate-related indigenous knowledge systems into the development and promotion of smart agriculture practices. Ewulo et al. (2025) suggested this integration can be achieved through participatory workshops and community-based consultations that recognize, value, and incorporate local knowledge and perspectives. Additionally, given smallholders' propensity to adopt smart farming practices in a bundled rather than isolated manner, agricultural extension services should advocate for integrated smart agriculture packages tailored to specific agroecological zones and farming systems. Such an approach can more effectively support the progression of digital maturity while enhancing the long-term sustainability and resilience of the agricultural sector.

## 7.2. Theoretical Contributions

The existing literature has extensively examined digital technology within the agricultural sector, focusing primarily on the following central themes: (1) the role and benefits of smart agriculture (Choruma et al., 2024; Hoang & Tran, 2023), (2) the factors influencing its adoption (Dibbern et al., 2024; Smidt & Jokonya, 2022), and (3) TAM and UTAUT adoption are conventionally employed to elucidate technology adoption at the individual level. However, studies have not fully elucidated the collective dynamics that define smallholder-based agricultural systems, particularly in SEA.

This study introduces the organizational dimension of collective ambition as a key antecedent of digital transformation and integrates it with the TAM to capture the social dynamics underlying technology adoption. Collective ambition is identified as a factor shaping the perception of usefulness and ease of use, as strong collective aspirations align the vision, goals, and beliefs of organizational members, thus increasing trust in ease of digital technology adoption. The TAM dimensions, PU and PEOU, elucidate the significance of collective ambition for digital transformation and digital maturity. The PU factor exerts a greater influence than PEOU, as smallholder farmers develop collective ambitions based on the belief that technology is advantageous for achieving the organization's long-term objectives. Meanwhile, PEOU serves as a supporting factor that facilitates technology adoption through ease of use and perceived value once its strategic advantages are collectively acknowledged. This study also contributes theoretically to the UTAUT framework through its integration with organizational commitment. Social influence and business expectations have emerged as the primary determinants of organizational commitment to digital transformation. Strong social influence reflects a robust social identity and collective norms within the farmer group, while business expectations suggest that digital transformation is perceived as a significant yet realistic endeavour. The key consistency across all these aspects lies in the cumulative ability to reduce uncertainty, anxiety, and resistance to change by affirming value, ease, social support, and organizational readiness to adopt technology. Ultimately, these variables integrate individual perspectives into collective commitment and behaviour that drive digital transformation and sustainable digital maturity. In accordance with Smidt & Jokonya (2022) to investigate the impact of digital technology adoption on the lives of rural communities, including aspects such as work patterns, social interactions, and access to information, to evaluate the genuine engagement of smallholder farmers.

Our research aims to emphasise the collective and organizational dimensions of digital transformation. To achieve this, the study integrates the TAM and UTAUT constructs with observational variables that reflect agricultural digitalization practices, such as the connectivity of farming businesses with stakeholders (DX1), integration of production processes through distribution (DX3), and utilization of digital technology for transactions and promotion of agricultural produce (DX4–DX5). Additionally, the level of digital maturity was

evaluated using indicators that assess system security, data integration capabilities, digital literacy support, external collaboration, and farm efficiency (DM1–DM5). The SEA region exhibits a significant gap in empirical and qualitative research on the adoption rates of smart farming services and technologies. Our study contributes empirical evidence from a psychological-organizational perspective, reinforcing the findings of Musa & Basir (2021), who suggested that smart farming possesses substantial potential to sustainably enhance food production while maintaining rigorous standards of food safety and quality. This approach offers a promising avenue towards achieving Sustainable Development Goal 2 (*Zero Hunger*) using innovative methods to construct food systems that are more profitable, resilient, and environmentally sustainable. For example, this entails aligning technological progress with best practices from countries that have further experience in smart farming, particularly concerning the adaptation of technology to each nation's specific climatic conditions. These elements are operationalized through the constructs of digital transformation (e.g., connectivity between farming enterprises and stakeholders, integration of processes from production to distribution, and the utilization of digital technology for transactions and the promotion of agricultural products) and digital maturity (e.g., system security, data integration capabilities, support for digital literacy, external collaboration, and the efficiency of farming businesses).

### 7.3. Global Practice Implications

#### 7.3.1. Youth Participation and STEM Learning

To establish a functional ecosystem for smart farming in SEA, more robust foundational strengths are imperative. At the grassroots level, particularly in less developed member states, skill deficiencies exist, notably in language and digital literacy (Hoang & Tran, 2023; Thi Hoa Sen et al., 2024). These barriers significantly impede rural farmers from adopting new technologies that can streamline and enhance their agricultural practices. Similar challenges have been observed in other countries, including developed regions, where most farmers are elderly. Therefore, training and capacity-building initiatives are essential to overcome entry barriers for young farmers, including access to markets, financing, and necessary inputs or equipment (Musa & Basir, 2021). Fundamentally, the government can support *science, technology, engineering, and mathematics* (STEM) education to cultivate interest among the younger generation in agriculture for technological disruption. A pertinent example is the green education system in the Netherlands (Kupper et al., 2012) and Australia (Gough et al., 2024), which has positively contributed to the development of high-quality human resources in the green sector (agriculture, nature, food). Similar initiatives could be considered globally, including in SEA, to develop a workforce capable of addressing the challenges of sustainable agriculture.

#### 7.3.2. Institutional Framework and Regional Cooperation

The government is crucial for establishing a framework for the adoption of new technologies in the agricultural sector. Close coordination between ministries is required to ensure that agricultural policies are relevant, effective, and inclusive. Given the accelerating pace of technological disruptions, policymakers must be adaptable and responsive to ever-changing technological interventions. In addition to labor and technology, significant investments are still required for research and development of products and smart farming solutions. The agri-finance subsector also needs to be encouraged without excessive regulation to create an ecosystem that supports digital agricultural innovation. Building a robust smart farming ecosystem is not limited to national efforts alone. The government, research institutions, and industry leaders must collaborate to facilitate the research and development of digital agricultural solutions. Disruptions in the food supply chain caused by the COVID-19 pandemic serve as a stark reminder of the risks that can arise if food systems and trade are disrupted. Given that none of the 10 ASEAN member states are self-sufficient in food production and that there is a high degree of intercountry dependence for food supplies (Lai, 2020), the adoption of smart farming technology must be undertaken collectively. In comparison, the European Union's *Common Agricultural Policy* (CAP) has long protected Europe's agricultural sector. This policy is one of the most integrated and absorbs approximately 40% of the EU budget to support the food production environment and farmers' incomes (Pe'er et al., 2019). Although ASEAN lacks a similar mechanism, it can still develop regional cooperation to intensify food security agendas. Macro-level strategies, such as the *ASEAN Food Security Reserve*, which sets aside and shares rice reserves during contingencies, as well as the *ASEAN Integrated Food Security (AIFS) Framework and the Strategic Plan of Action on ASEAN Food Security* (SPA-FS), have become important platforms for information exchange, the transfer of new technologies, and knowledge sharing.

#### 7.4. Limitations & Future Research Agenda

The primary limitation of this study is that it focused on smallholders in SEA, which may limit the applicability of the findings to agricultural studies in other regions. Although the study offers valuable insights into technology adoption, the results may not be generalizable to the broader agricultural sector because of variations in organizational environments, resource capabilities, and regulatory conditions. Additionally, the research achieved a response rate of 54.5%, which, although sufficient for SEM, may introduce potential non-response bias. Despite statistical tests indicating no significant bias, caution is advised when interpreting the generalizability of the findings. Future research could enhance external validity by incorporating a more extensive and diverse sample from various regions or smallholder stakeholder groups, such as agritech start-ups, the food processing industry, microfinance institutions, ministries of agriculture, and cross-regional bodies, and local communities and farmers' social networks. Moreover, the rapid advancement of technologies, such as artificial intelligence and blockchain, underscores the necessity for ongoing analysis of their adoption processes. Future research could assess the impact of these technologies on operational efficiency, strategic alignment, and competitive advantage, thereby providing actionable insights for organizations undergoing digital transformation. Although the proposed model explains important aspects of digital transformation and digital maturity among smallholders in SEA, additional organizational, technological, and contextual factors may further explain the variance in these constructs. Factors such as leadership, digital infrastructure, government support, socio-economic conditions, and technological readiness were not explicitly examined in this study and may provide valuable directions for future research. This approach will help ensure that the research remains pertinent to the practical and strategic needs of businesses in a rapidly evolving digital environment. As digital technology progresses, new studies must also investigate critical areas, such as cybersecurity, data privacy, and talent development. These aspects are crucial because they represent both the primary challenges and opportunities within the digital transformation landscape. Further research should focus on how organizations design strategies to manage these challenges while preparing their workforce to meet the demands of ever-evolving technology.

#### References

- Akdil, K. Y., Ustundag, A., & Cevikcan, E. (2018). Maturity and Readiness Model for Industry 4.0 Strategy. In A. Ustundag & E. Cevikcan, *Industry 4.0: Managing The Digital Transformation* (pp. 61–94). Springer International Publishing. [https://doi.org/10.1007/978-3-319-57870-5\\_4](https://doi.org/10.1007/978-3-319-57870-5_4)
- Alexander, K., Case, P., Jones, M., & Connell, J. (2017). Commercialising smallholder agricultural production in Lao People's Democratic Republic. *Development in Practice*, 27(7), 965–980. <https://doi.org/10.1080/09614524.2017.1353064>
- Ali, M., Man, N., Farrah, M. M., & Omar, S. Z. (2020). Factors influencing behavioral intention of farmers to use ICTs for agricultural risk management in Malaysia. *Pakistan Journal of Agricultural Research*, 33(2), 295.
- Alqatan, S., Noor, N. M. M., Man, M., & Mohamad, R. (2017). A theoretical discussion of factors affecting the acceptance of m-commerce among SMTEs by integrating TTF with TAM. *International Journal of Business Information Systems*, 26(1), 66–111. <https://doi.org/10.1504/IJBIS.2017.086057>
- Awa, H. O., Ojiabo, O. U., & Emecheta, B. C. (2015). Integrating TAM, TPB and TOE frameworks and expanding their characteristic constructs for e-commerce adoption by SMEs. *Journal of Science & Technology Policy Management*, 6(1), 76–94.
- Azadi, H., Moghaddam, S. M., Burkart, S., Mahmoudi, H., Van Passel, S., Kurban, A., & Lopez-Carr, D. (2021). Rethinking resilient agriculture: From climate-smart agriculture to vulnerable-smart agriculture. *Journal of Cleaner Production*, 319, 128602.
- Banluesapy, S., Ketcham, M., & Rattanasiriwongwut, M. (2026). Integrating Theories into a Model for the Acceptance of Internet of Things Technology for Developing Smart Farm Platforms: A Study of Readiness in the Agricultural Industry. *Global Journal of Flexible Systems Management*, 27(1), 85–98. <https://doi.org/10.1007/s40171-025-00479-3>
- Beban, A., & Gironde, C. (2023). Surviving cassava: Smallholder farmer strategies for coping with market volatility in Cambodia. *Journal of Land Use Science*, 18(1), 109–127. <https://doi.org/10.1080/1747423X.2023.2190744>
- Betcherman, G., Haque, I., & Marschke, M. (2021). Exploring livelihood transitions in the Mekong delta. *Singapore Journal of Tropical Geography*, 42(2), 222–240. <https://doi.org/10.1111/sjtg.12300>

- Beza, E., Reidsma, P., Poortvliet, P. M., Belay, M. M., Bijen, B. S., & Kooistra, L. (2018). Exploring farmers' intentions to adopt mobile Short Message Service (SMS) for citizen science in agriculture. *Computers and Electronics in Agriculture*, 151, 295–310. <https://doi.org/10.1016/j.compag.2018.06.015>
- Bistline, J. E., Blanford, G., Grant, J., Knipping, E., McCollum, D. L., Nopmongcol, U., Scarth, H., Shah, T., & Yarwood, G. (2022). Economy-wide evaluation of CO<sub>2</sub> and air quality impacts of electrification in the United States. *Nature Communications*, 13(1), 6693.
- Blount, S., & Leinwand, P. (2019). Why are we here?: If you want employees who are more engaged and productive, give them a purpose—One concretely tied to your customers and your strategy. *Harvard Business Review*, 2019(November-December), 1–9.
- Bwambale, E., Abagale, F. K., & Anornu, G. K. (2023). Data-driven model predictive control for precision irrigation management. *Smart Agricultural Technology*, 3, 100074.
- Castella, J.-C., Lu, J., Friis, C., Bruun, T. B., Cole, R., Junquera, V., Kenney-Lazar, M., Mahanty, S., Ornetsmüller, C., Pravalpruskul, P., & Vagneron, I. (2023). Beyond the boom-bust cycle: An interdisciplinary framework for analysing crop booms. *Global Environmental Change*, 80, 102651. <https://doi.org/10.1016/j.gloenvcha.2023.102651>
- Choruma, D. J., Dirwai, T. L., Mutenje, M. J., Mustafa, M., Chimonyo, V. G. P., Jacobs-Mata, I., & Mabhaudhi, T. (2024). Digitalisation in agriculture: A scoping review of technologies in practice, challenges, and opportunities for smallholder farmers in sub-saharan africa. *Journal of Agriculture and Food Research*, 18, 101286. <https://doi.org/10.1016/j.jafr.2024.101286>
- Cohen, J. (1988). Set Correlation and Contingency Tables. *Applied Psychological Measurement*, 12(4), 425–434. <https://doi.org/10.1177/014662168801200410>
- Cooreman, H., Debruyne, L., Vandenabeele, J., & Marchand, F. (2021). Power to the facilitated agricultural dialogue: An analysis of on-farm demonstrations as transformative learning spaces. *The Journal of Agricultural Education and Extension*, 27(5), 699–719. <https://doi.org/10.1080/1389224x.2021.1969958>
- Damba, O. T., Ansah, I. G. K., Donkoh, S. A., Alhassan, A., Mullins, G. R., Yussif, K., Taylor, M. S., Tetteh, B. K., & Appiah-Twumasi, M. (2020). Effects of technology dissemination approaches on agricultural technology uptake and utilization in Northern Ghana. *Technology in Society*, 62, 101294.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 982–1003. <https://doi.org/10.1287/mnsc.35.8.982>
- Deloitte. (2018). *Digital maturity pivot model*. <https://www2.deloitte.com/us/en/insights/focus/digital-maturity/digital-maturity-pivot-model.html>
- Diamantopoulos, A., & Siguaw, J. A. (2006). Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. *British Journal of Management*, 17(4), 263–282. <https://doi.org/10.1111/j.1467-8551.2006.00500.x>
- Dibbern, T., Romani, L. A. S., & Massruhá, S. M. F. S. (2024). Main drivers and barriers to the adoption of Digital Agriculture technologies. *Smart Agricultural Technology*, 8, 100459. <https://doi.org/10.1016/j.atech.2024.100459>
- Dijkstra, T. K., & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis*, 81, 10–23. <https://doi.org/10.1016/j.csda.2014.07.008>
- Ewulo, T. A., Akinseye, F. M., Teme, N., Agele, S. O., Yessoufou, N., & Kumar, S. (2025). Factors driving Climate-Smart Agriculture adoption: A study of smallholder farmers in Koumpentum, Senegal. *Frontiers in Agronomy*, 7, 1552720.
- FAO. (2023). *Crops FAOSTAT statistical database Rome: Updated December 23, 2023*. <http://www.fao.org/faostat/en/#data/QC>
- Foguesatto, C. R., Frare, A. B., Wegner, D., Mitregre, M., & Rodrigues, R. A. B. (2024). You will never stand alone: The role of inter-organizational collaboration and technological turbulence in shaping small business' digital maturity. *The Journal of High Technology Management Research*, 35(2), 100518. <https://doi.org/10.1016/j.hitech.2024.100518>

- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50.  
<https://doi.org/10.1177/002224378101800104>
- Fragomeli, R., Punzo, G., & Annunziata, A. (2026). Agriculture 4.0 technologies among Italian farmers: A C-TAM-TPB approach with sustainability and economic factors. *Journal of Cleaner Production*, 559, 148316.  
<https://doi.org/10.1016/j.jclepro.2026.148316>
- Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. *Journal of Management Information Systems*, 18(1), 185–214.  
<https://doi.org/10.1080/07421222.2001.11045669>
- Gough, A., Reid, A., & Stevenson, R. B. (2024). Environmental and Sustainability Education in Australia. In *World Review*. CRC Press.
- Gudergan, S. P., Ringle, C. M., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research, Formative Indicators*, 61(12), 1238–1249.  
<https://doi.org/10.1016/j.jbusres.2008.01.012>
- Hair, J. F. (2011). Multivariate data analysis: An overview. *International Encyclopedia of Statistical Science*, 904–907.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *SEM: An introduction. Multivariate data analysis: a global perspective* (pp. 629–686). Upper Saddle River, NJ: Pearson Education.
- Hair, J. F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, 26(2), 106–121.
- Hair Jr, J., Hair Jr, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2016). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. SAGE Publications.
- Hanelt, A., Firk, S., Hildebrandt, B., & Kolbe, L. M. (2021). Digital M&A, digital innovation, and firm performance: An empirical investigation. *European Journal of Information Systems*, 30(1), 3–26.  
<https://doi.org/10.1080/0960085X.2020.1747365>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.  
<https://doi.org/10.1007/s11747-014-0403-8>
- Herscovitch, L., & Meyer, J. P. (2002). Commitment to organizational change: Extension of a three-component model. *Journal of Applied Psychology*, 87(3), 474.
- Hoang, H. G., & Tran, H. D. (2023). Smallholder farmers' perception and adoption of digital agricultural technologies: An empirical evidence from Vietnam. *Outlook on Agriculture*, 52(4), 457–468.  
<https://doi.org/10.1177/00307270231197825>
- IPCC. (2014). *Climate Change 2014 Part A: Global and Sectoral Aspects. Climate Change 2014: Impacts, Adaptation, and Vulnerability*. [https://www.ipcc.ch/site/assets/uploads/2018/02/WGIIAR5-PartA\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WGIIAR5-PartA_FINAL.pdf)
- Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2022). Enhancing smart farming through the applications of Agriculture 4.0 technologies. *International Journal of Intelligent Networks*, 3, 150–164.
- Jie, H., Gooi, L. M., & Lou, Y. (2025). Digital maturity, dynamic capabilities and innovation performance in high-tech SMEs. *International Review of Economics & Finance*, 99, 103971.  
<https://doi.org/10.1016/j.iref.2025.103971>
- Kala Kamdjoug, J. R., Djuitchou Chengo, S. M., & Gueyie, J.-P. (2021). Factors affecting the adoption of information technologies by small woman-managed enterprises in Cameroon. *Journal of Small Business & Entrepreneurship*, 33(4), 433–451. <https://doi.org/10.1080/08276331.2020.1716134>
- Kernecker, M., Knierim, A., Wurbs, A., Kraus, T., & Borges, F. (2020). Experience versus expectation: Farmers' perceptions of smart farming technologies for cropping systems across Europe. *Precision Agriculture*, 21(1), 34–50. <https://doi.org/10.1007/s11119-019-09651-z>
- Kets de Vries, M. F. R. (2025). Navigating the Maze of Organizational Transformation. In M. F. R. Kets de Vries (Ed.), *The Path to Individual and Organizational Transformation: Confronting the Elephant in the Room* (pp. 99–113). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-88330-9\\_11](https://doi.org/10.1007/978-3-031-88330-9_11)
- Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford publications.

- Krithika, M., & Rajini, G. (2016). Comparative Review of Two Theoretical Frameworks in Technology Based Consumer Acceptance Behavior. *International Journal of Research in Management, Economics and Commerce*, 6(4), 16–26.
- Kupper, H., Laurentzen, R., & Mulder, M. (2012). Recent Policy Developments in Green Education in the Netherlands. *The Journal of Agricultural Education and Extension*, 18(2), 121–139. <https://doi.org/10.1080/1389224X.2012.655966>
- Lai, C. Y., Cheung, K. Y., & Chan, C. S. (2023). Exploring the role of intrinsic motivation in ChatGPT adoption to support active learning: An extension of the technology acceptance model. *Computers and Education: Artificial Intelligence*, 5, 100178. <https://doi.org/10.1016/j.caeai.2023.100178>
- Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, 529(7584), 84–87.
- Leso, B. H., Cortimiglia, M. N., & Ghezzi, A. (2023). The Contribution of Organizational Culture, Structure, and Leadership Factors in the Digital Transformation of SMEs: A Mixed-Methods Approach. *Cognition, Technology & Work*, 25(1), 151–179. <https://doi.org/10.1007/s10111-022-00714-2>
- Lova Raju, K., & Vijayaraghavan, V. (2022). A Self-Powered, Real-Time, NRF24L01 IoT-Based Cloud-Enabled Service for Smart Agriculture Decision-Making System. *Wireless Personal Communications*, 124(1), 207–236. <https://doi.org/10.1007/s11277-021-09462-4>
- Lowder, S. K., Sánchez, M. V., & Bertini, R. (2021). Which farms feed the world and has farmland become more concentrated? *World Development*, 142, 105455. <https://doi.org/10.1016/j.worlddev.2021.105455>
- Melović, B., Jocović, M., Dabić, M., Vulić, T. B., & Dudic, B. (2020). The impact of digital transformation and digital marketing on the brand promotion, positioning and electronic business in Montenegro. *Technology in Society*, 63, 101425.
- Meyer, J. P., & Allen, N. J. (2001). *Organizational Commitment. Personnel Psychology and Human Resource Management: A Reader for Students and Practitioners* (Vol. 289). Wiley Chichester.
- Mian, A. S., Vlahu-Gjorgievska, E., & Shen, J. (2025). Examining the Collaboration Framework for Achieving Government Performance: A Qualitative Case Study on Digital Transformation. *Digit. Gov.: Res. Pract.*, 6(4), 46:1-46:33. <https://doi.org/10.1145/3734696>
- Michels, M., Bonke, V., & Musshoff, O. (2020). Understanding the adoption of smartphone apps in crop protection. *Precision Agriculture*, 21(6), 1209–1226. <https://doi.org/10.1007/s11119-020-09715-5>
- Mujeyi, A., Mudhara, M., & Mutenje, M. (2021). The impact of climate smart agriculture on household welfare in smallholder integrated crop–livestock farming systems: Evidence from Zimbabwe. *Agriculture & Food Security*, 10(1), 4. <https://doi.org/10.1186/s40066-020-00277-3>
- Musa, S. F. P. D., & Basir, K. H. (2021). Smart farming: Towards a sustainable agri-food system. *British Food Journal*, 123(9), 3085–3099. <https://doi.org/10.1108/BFJ-03-2021-0325>
- Naima, B., Mustapha, M., & Sonia, K. (2021). Organizational commitment as a hard core of governments' digital transformation: Evidence from African social care services. *Dirassat Journal Economic Issue*, 12(1), 681–696.
- Najm, N. A., Yousif, A. S. H., Al-Ensour, J. A., & Alnidawy, A. A. B. (2023). The Impact of Sustainability Principles on Organizational Commitments and the Mediating Effect of Digital Transformation Process. In S. G. Yaseen (Ed.), *Cutting-Edge Business Technologies in the Big Data Era* (Vol. 136, pp. 273–288). Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-42455-7\\_25](https://doi.org/10.1007/978-3-031-42455-7_25)
- Naqshbandi, M. M., Kabir, I., Ishak, N. A., & Islam, Md. Z. (2023). The future of work: Work engagement and job performance in the hybrid workplace. *The Learning Organization: An International Journal*, 31(1), 5–26. <https://doi.org/10.1108/TLO-08-2022-0097>
- Narine, L. K., Harder, A., & Roberts, T. G. (2019). Farmers' intention to use text messaging for extension services in Trinidad. *The Journal of Agricultural Education and Extension*, 25(4), 293–306. <https://doi.org/10.1080/1389224X.2019.1629970>
- Nikkhou, S., Taghizadeh, K., & Hajiyakhchali, S. (2016). Designing a portfolio management maturity model (Elena). *Procedia-Social and Behavioral Sciences*, 226, 318–325.
- Parvathi, P., & Waibel, H. (2016). Adoption of Sustainable Palm Oil Practices by Indonesian Smallholder Farmers. *Journal of Southeast Asian Economies*, 291–316.
- Paul, F.-H., Brink, H., & Draxler-Weber, N. (2022). Characterizing Maturity of Digital Transformation in Organizations – A Socio-technical Framework. In Ė. Nazaruka, K. Sandkuhl, & U. Seigerroth (Eds),

- Perspectives in Business Informatics Research* (Vol. 462, pp. 189–204). Springer International Publishing. [https://doi.org/10.1007/978-3-031-16947-2\\_13](https://doi.org/10.1007/978-3-031-16947-2_13)
- Pe'er, G., Zinngrebe, Y., Moreira, F., Sirami, C., Schindler, S., Müller, R., Bontzorlos, V., Clough, D., Bezák, P., Bonn, A., Hansjürgens, B., Lomba, A., Möckel, S., Passoni, G., Schleyer, C., Schmidt, J., & Lakner, S. (2019). A greener path for the EU Common Agricultural Policy. *Science*, 365(6452), 449–451. <https://doi.org/10.1126/science.aax3146>
- Petani, F. J., & Mengis, J. (2023). Technology and the hybrid workplace: The affective living of IT-enabled space. *The International Journal of Human Resource Management*, 34(8), 1530–1553. <https://doi.org/10.1080/09585192.2021.1998186>
- Population Reference Bureau. (2025). *Population mid-2025*. <https://www.prb.org/international/indicator/population/snapshot>
- Proença, D., & Borbinha, J. (2016). Maturity models for information systems-a state of the art. *Procedia Computer Science*, 100, 1042–1049.
- Quy, V. K., Hau, N. V., Anh, D. V., Quy, N. M., Ban, N. T., Lanza, S., Randazzo, G., & Muzirafuti, A. (2022). IoT-enabled smart agriculture: Architecture, applications, and challenges. *Applied Sciences*, 12(7), 3396.
- Rahman, M. M. (2025). Data Analytics for Strategic Business Development: A Systematic Review Analyzing Its Role in Informing Decisions, Optimizing Processes, And Driving Growth. *Journal of Sustainable Development and Policy*, 1(01), 285–314.
- Ready, D. A., & Truelove, E. (2011). The power of collective ambition. *Harvard Business Review*, 89(12), 94–102.
- Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing*, 26(4), 332–344. <https://doi.org/10.1016/j.ijresmar.2009.08.001>
- Roeven, L., Wolf, S. A., Sengers, P., Liu, J., Rubambiza, G., Persaud, D., & Weatherspoon, H. (2025). Analyzing abstraction in critical agri-food studies and computer science: Toward interdisciplinary analysis of digital agriculture innovation. *Agriculture and Human Values*, 42(2), 1009–1026. <https://doi.org/10.1007/s10460-024-10655-3>
- Ronaghi, M. H., & Forouharfar, A. (2020). A contextualized study of the usage of the Internet of things (IoTs) in smart farming in a typical Middle Eastern country within the context of Unified Theory of Acceptance and Use of Technology model (UTAUT). *Technology in Society*, 63, 101415. <https://doi.org/10.1016/j.techsoc.2020.101415>
- Sayruamyat, S., & Nadee, W. (2020). Acceptance and Readiness of Thai Farmers Toward Digital Technology. In Y.-D. Zhang, J. K. Mandal, C. So-In, & N. V. Thakur (Eds), *Smart Trends in Computing and Communications* (pp. 75–82). Springer. [https://doi.org/10.1007/978-981-15-0077-0\\_8](https://doi.org/10.1007/978-981-15-0077-0_8)
- Schnebelin, É., Labarthe, P., & Touzard, J.-M. (2021). How digitalisation interacts with ecologisation? Perspectives from actors of the French Agricultural Innovation System. *Journal of Rural Studies*, 86, 599–610. <https://doi.org/10.1016/j.jrurstud.2021.07.023>
- Smidt, H. J., & Jokonya, O. (2022). Factors affecting digital technology adoption by small-scale farmers in agriculture value chains (AVCs) in South Africa. *Information Technology for Development*, 28(3), 558–584. <https://doi.org/10.1080/02681102.2021.1975256>
- Taghibaygi, M., & Alibaygi, A. (2024). The impact of ethical commitments on the intention to adopt digital agricultural technologies. *Scientific Reports*, 14(1), 23035.
- Tan, S. S. X., & Kuebbing, S. E. (2023). A synthesis of the effect of regenerative agriculture on soil carbon sequestration in Southeast Asian croplands. *Agriculture, Ecosystems & Environment*, 349, 108450. <https://doi.org/10.1016/j.agee.2023.108450>
- Taylor, S., & Todd, P. A. (1995). Understanding Information Technology Usage: A Test of Competing Models. *Information Systems Research*, 6(2), 144–176. <https://doi.org/10.1287/isre.6.2.144>
- Thi Hoa Sen, L., Chou, P., Dacuyan, F. B., Nyberg, Y., & Wetterlind, J. (2024). Barriers and enablers of digital extension services' adoption among smallholder farmers: The case of Cambodia, the Philippines and Vietnam. *International Journal of Agricultural Sustainability*, 22(1), 2368351. <https://doi.org/10.1080/14735903.2024.2368351>

- Tran, T. A., & Touch, V. (2024). How agricultural extension responds to amplified agrarian transitions in mainland Southeast Asia: Experts' reflections. *Agriculture and Human Values*, 41(4), 1773–1789. <https://doi.org/10.1007/s10460-024-10577-0>
- Trendov, N. M., Varas, S., & Zeng, M. (2019). *Digital technologies in agriculture and rural areas*. FAO; <https://openknowledge.fao.org/items/602e7346-9972-4a85-b881-fd2d05db8a32>
- Ulhaq, I., Pham, N. T. A., Le, V., Pham, H.-C., & Le, T. C. (2022). Factors influencing intention to adopt ICT among intensive shrimp farmers. *Aquaculture*, 547, 737407.
- Veldhuizen, L. J.L., Giller, K. E., Oosterveer, P., Brouwer, I. D., Janssen, S., van Zanten, H. H.E., & Slingerland, M. A. (2020). The Missing Middle: Connected action on agriculture and nutrition across global, national and local levels to achieve Sustainable Development Goal 2. *Global Food Security*, 24, 100336. <https://doi.org/10.1016/j.gfs.2019.100336>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.
- Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 157–178.
- Vishnoi, S., & Goel, R. K. (2024). Climate smart agriculture for sustainable productivity and healthy landscapes. *Environmental Science & Policy*, 151, 103600. <https://doi.org/10.1016/j.envsci.2023.103600>
- Wessel, L., Baiyere, A., Ologeanu-Taddei, R., Cha, J., & Jensen, T. B. (2021). Unpacking the Difference Between Digital Transformation and IT-Enabled Organizational Transformation. *Journal of the Association for Information Systems*, 22(1). <https://doi.org/10.17705/1jais.00655>
- Wetzels, M., Odekerken-Schröder, G., & van Oppen, C. (2009). Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration1. *MIS Quarterly*, 33(1), 177–195.
- Winarsih, Indriastuti, M., & Fuad, K. (2021). Impact of Covid-19 on Digital Transformation and Sustainability in Small and Medium Enterprises (SMEs): A Conceptual Framework. In L. Barolli, A. Poniszewska-Maranda, & T. Enokido (Eds), *Complex, Intelligent and Software Intensive Systems* (Vol. 1194, pp. 471–476). Springer International Publishing. [https://doi.org/10.1007/978-3-030-50454-0\\_48](https://doi.org/10.1007/978-3-030-50454-0_48)
- Wistedt, U. (2024). Consumer purchase intention toward POI-retailers in cross-border E-commerce: An integration of technology acceptance model and commitment-trust theory. *Journal of Retailing and Consumer Services*, 81, 104015. <https://doi.org/10.1016/j.jretconser.2024.104015>
- Yang, Q., Al Mamun, A., Hayat, N., Jingzu, G., Hoque, M. E., & Salameh, A. A. (2022). Modeling the Intention and Adoption of Wearable Fitness Devices: A Study Using SEM-PLS Analysis. *Frontiers in Public Health*, 10. <https://doi.org/10.3389/fpubh.2022.918989>
- Zahoor, N., Zopiatas, A., Adomako, S., & Lamprinakos, G. (2023). The micro-foundations of digitally transforming SMEs: How digital literacy and technology interact with managerial attributes. *Journal of Business Research*, 159, 113755. <https://doi.org/10.1016/j.jbusres.2023.113755>

**Appendix 1. Measurement Items**

| Latent Variable                                 | Theory Integration                                 |                         | Observational Variable                                                                                                                                                                          | Category Coding |
|-------------------------------------------------|----------------------------------------------------|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|
| Collective ambitions (Ready & Truelove, 2011)   | Technology Acceptance Model                        | Perceived Usefulness    | I clearly understand our organization's key goals and believe they are important for the future.                                                                                                | CA1             |
|                                                 |                                                    |                         | I believe our organization's vision provides a clear direction towards digital transformation.                                                                                                  | CA2             |
|                                                 |                                                    | Perceived Ease of Use   | Our organization has realistic targets and benchmarks to achieve sustainable digital transformation.                                                                                            | CA3             |
|                                                 |                                                    |                         | I see that our organization's strategic and operational priorities consistently support the achievement of a shared vision                                                                      | CA4             |
|                                                 |                                                    |                         | Our organization is committed to delivering on its promises to stakeholders (smallholders, communities, investors, and partners) regarding the experience and benefits they have been promised. | CA5             |
|                                                 |                                                    |                         | I feel that the core values of our organization are a real guide in facing challenges and successes.                                                                                            | CA6             |
|                                                 |                                                    |                         | Our organizational leaders demonstrate day-to-day behaviors that are consistent with the organization's vision, values, and commitments.                                                        | CA7             |
| Organizational commitment (Meyer & Allen, 2001) | Unified Theory of Acceptance and Use of Technology | Social Influence        | I have a strong sense of belonging to this smallholder organization, especially in our joint efforts to achieve digital transformation.                                                         | OC1             |
|                                                 |                                                    | Performance Expectancy  | I feel tied to my smallholder organization and am committed to supporting the increase of digital maturity.                                                                                     | OC2             |
|                                                 |                                                    | Effort Expectancy       | Participating in the agricultural digital transformation program at this organisation means a lot to me.                                                                                        | OC3             |
|                                                 |                                                    | Facilitating Conditions | I feel that the challenges that organizations face in the digital transformation process are also my responsibility.                                                                            | OC4             |
| Digital transformation (Zahoor et al., 2023)    |                                                    |                         | My farming business is increasingly digitally connected with stakeholders.                                                                                                                      | DX1             |
|                                                 |                                                    |                         | The digital transformation of agriculture is increasingly penetrating and connecting various aspects of my farming activities.                                                                  | DX2             |
|                                                 |                                                    |                         | My farming process, from production to distribution, is increasingly digitally integrated.                                                                                                      | DX3             |
|                                                 |                                                    |                         | I use digital technology to promote agricultural produce and reach buyers.                                                                                                                      | DX4             |
|                                                 |                                                    |                         | I use digital technology to support farming transactions.                                                                                                                                       | DX5             |

---

|                                      |                                                                                                                                          |     |
|--------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Digital maturity<br>(Deloitte, 2018) | I feel that the digital systems I use in agriculture are safe and accessible.                                                            | DM1 |
|                                      | I can integrate and leverage agricultural data to support my farming decisions.                                                          | DM2 |
|                                      | I received support through training, mentoring, or digital literacy programs that improved my ability to use digital farming technology. | DM3 |
|                                      | I collaborate with external stakeholders in developing digital farming practices.                                                        | DM4 |
|                                      | I use digital systems to improve efficiency in my farming business.                                                                      | DM5 |

---

# Embedding Retrieval-Augmented Generation into SME Customer Communication Work Practices

Rosalina<sup>1\*</sup>, Noor Lees Ismail<sup>2</sup>, Genta Sahuri<sup>3</sup> and Joseph Tedja Nugraha Wibawa<sup>4</sup>

<sup>1,3,4</sup>Study Program, Faculty of Computer Science, President University, Bekasi, Indonesia

<sup>2</sup>School of Information Technology, UNITAR International University, Selangor, Malaysia

\*Correspondence: [rosalina@president.ac.id](mailto:rosalina@president.ac.id)

## PAPER INFO

### *Paper history:*

Received 26 June 2025

Accepted 20 January 2026

### *Citation:*

Rosalina, Ismail, N. L., Sahuri, G., & Wibawa, J. T. N. (2026).

Embedding retrieval-augmented generation into SME customer communication work practices. In *Journal of Information and Organizational Sciences*, vol. 50, no. 1, pp. 233-246

### *Copyright:*

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## ABSTRACT

The increasing use of digital messaging platforms such as WhatsApp, Instagram, and Facebook has made real-time customer communication a key activity for Small and Medium Enterprises (SMEs). However, many SMEs struggle to access and apply business knowledge during live interactions, often relying on fragmented information and keyword-based retrieval that does not capture user intent. This study examines how a mobile keyboard-based Retrieval-Augmented Generation (RAG) system supports SME customer response work practices from an Information Systems perspective. The system organizes business knowledge into semantic chunks, represents them as vector embeddings, retrieves relevant information using similarity-based methods, and generates context-aware responses through a Large Language Model. It is implemented as a lightweight keyboard interface embedded directly within messaging applications. The system was evaluated using the RAGAS framework on 37 test queries and compared with a keyword-based baseline. The results show high faithfulness (0.997) and answer correctness (0.881), with an average response time of approximately 5 seconds. A preliminary User Acceptance Testing session with one SME stakeholder suggests that the system may help reduce response effort and support more consistent use of business knowledge in customer communication.

**Keywords:** Customer Response Systems, Indonesia, Knowledge Management, Large Language Models, Mobile Keyboard Interface, Retrieval-Augmented Generation, Small and Medium Enterprises (SMEs), Text Embedding, Vector Similarity

## 1. Introduction

Customer communication has become an essential part of how Small and Medium Enterprises (SMEs) operate on a daily basis. Many interactions now take place through mobile messaging platforms such as WhatsApp, Instagram, and Facebook Messenger, where customers expect quick and clear responses. This creates a recurring tension for SMEs: responses are expected in real time, yet the information needed to answer customer questions is often scattered across different sources. As the number and variety of inquiries increase, operators must search for product details, pricing rules, or service policies while continuing the conversation. This situation not only slows down response time but can also lead to inconsistent or incomplete answers.

Despite the widespread availability of digital tools, many SMEs still face difficulties in retrieving the right information quickly during ongoing conversations. Business knowledge is often fragmented across different sources, including product descriptions, pricing rules, delivery terms, and operational procedures.

As a result, responding to customer inquiries becomes time-consuming and cognitively demanding, especially when operators must manually search for relevant information while interacting with customers in real time.

Various traditional approaches have been used to support customer response activities, including rule-based chatbots, keyword-based search systems, and document repositories. However, these approaches exhibit significant limitations in practice. SME knowledge bases typically contain both structured and unstructured information, and keyword-based systems often fail to capture the underlying intent of customer queries. Instead, they rely on exact or partial word matching, which frequently leads to incomplete or irrelevant results. This limitation increases cognitive load for SME operators and may reduce the quality and consistency of responses.

Recent advances in Retrieval-Augmented Generation (RAG) have demonstrated the potential to address these limitations by grounding Large Language Models (LLMs) in external knowledge bases, thereby improving factual reliability and reducing hallucination (Church et al., 2024; Shan & Shan, 2024). RAG-based systems have been successfully applied across various domains, including e-commerce (Benita et al., 2024), technical support (Lee et al., 2024), university admissions (Nguyen et al., 2024), property management (Chen et al., 2024), and banking employee assistance (Sandakelum et al., 2025). More recent studies have extended RAG toward task-oriented dialogue, knowledge-graph-based retrieval, and personalized customer communication (Rao et al., 2025; Varga & Yamashita, 2025; Li et al., 2025; Singh et al., 2025). Despite these advances, existing research predominantly focuses on technical performance and system architecture, with limited attention to how RAG systems are embedded into everyday organizational practices, particularly in SME contexts. From an Information Systems perspective, customer response systems can be understood as socio-technical artifacts that shape how organizational knowledge is accessed, interpreted, and enacted in work practices. In SMEs, customer communication is often handled by a small number of individuals operating under time constraints, relying heavily on tacit knowledge and informal processes rather than formalized information systems.

To address this gap, this study investigates how a mobile keyboard-integrated Retrieval-Augmented Generation (RAG) system supports SME customer response activities as an information system embedded in everyday work practices. The proposed system integrates a RAG pipeline directly into a mobile keyboard interface, enabling users to retrieve and generate knowledge-grounded responses without switching between applications. Business knowledge is organized into semantic chunks, represented as vector embeddings, retrieved through similarity-based mechanisms, and used to generate context-aware responses via an LLM.

To examine both technical performance and organizational implications, the system was evaluated using the RAGAS framework on 37 test queries and compared with a conventional keyword-based approach. In addition, User Acceptance Testing (UAT) was conducted with an SME stakeholder to assess usability and practical applicability. Guided by a socio-technical and Information Systems perspective, this study addresses the following research questions:

- RQ1: How does integrating a RAG system into a mobile keyboard interface influence SME customer response work practices?
- RQ2: How does a keyboard-integrated RAG system support SME knowledge management during real-time customer communication?
- RQ3: How does human-AI collaboration manifest when SME operators use AI-generated, knowledge-grounded responses in customer interactions?

This paper makes three contributions to the Information Systems literature. First, it extends research on AI-enabled systems in SMEs by examining how such systems can be embedded into everyday work practices rather than used as standalone applications. Second, it contributes to knowledge management research by providing initial empirical insights on how semantic retrieval supports more consistent enactment of organizational knowledge in real-time communication. Third, it offers insights into human-AI collaboration by illustrating how AI can function as a cognitive support tool that augments, rather than replaces, human judgment in customer-facing activities.

The remainder of this paper is structured as follows. Section 2 reviews related work on knowledge management systems and RAG-based applications. Section 3 presents the system architecture and methodology. Section 4 reports the results and discusses the findings. Section 5 concludes the paper and outlines directions for future research.

## 2. Related Works

Recent research has expanded Retrieval-Augmented Generation (RAG) from simple text-based question answering toward more complex, domain-specific, and enterprise-oriented systems. Rather than treating RAG as a standalone retrieval mechanism, scholars increasingly position it as a socio-technical system that shapes

how organizations access knowledge, interact with users, and support decision-making processes. This shift is particularly relevant for SME contexts, where knowledge is often informal, distributed, and tightly embedded in daily operations.

A significant stream of research focuses on improving the technical performance of RAG systems through enhanced retrieval architectures and representation learning. For example, DeepSem has been proposed as a multilayer semantic retrieval framework that improves precision in industrial applications (Qi et al., 2025), while iterative retrieval feedback mechanisms have also been introduced to further enhance response quality (Zhu et al., 2025). These studies emphasize the importance of accurate and context-aware retrieval, which forms the technical foundation for supporting knowledge management in real-time communication environments.

Another line of work explores domain-specific applications of Retrieval-Augmented Generation (RAG) in organizational settings. Previous studies have demonstrated the effectiveness of RAG in specialized reference services, automotive information retrieval, manufacturing support, and healthcare communication, where responses must remain grounded in domain-specific knowledge to ensure reliability and trustworthiness (Chen & Chang, 2025; Yildirim & Samli, 2025; Wulf & Meierhofer, 2025; Coen et al., 2025). Collectively, these studies suggest that RAG can improve the quality and consistency of information-intensive tasks by providing contextually relevant and knowledge-grounded responses.

A growing body of research has also examined the application of RAG in business communication and customer service contexts. Prior work has shown that retrieval-enhanced approaches can improve customer feedback summarization, mitigate hallucination in multilingual customer service systems, and enhance performance in call center environments when combined with fine-tuning strategies (Praneeth et al., 2025; Patel et al., 2025; Sanjani et al., 2025). These findings support the use of retrieval-based methods for improving response quality, consistency, and reliability in customer-facing interactions.

In parallel, recent research has extended Retrieval-Augmented Generation (RAG) beyond text-based applications toward multimodal and hybrid systems that integrate diverse information sources. These approaches combine textual and visual retrieval capabilities to support tasks such as interpreting complex production documents, enhancing retail and advertising applications, and facilitating access to enterprise knowledge resources (Lystbæk & Lystbæk, 2025; Thiyagarajan, 2025; Choudhary et al., 2025). In organizational settings, RAG has also been applied to querying structured corporate documents, including financial reports and other enterprise records, through semantic retrieval mechanisms (Mehul et al., 2025). Collectively, these developments demonstrate the evolution of RAG from a standalone retrieval technique into an organizational intelligence layer that supports knowledge access, information integration, and decision-making across different business contexts.

From an Information Systems perspective, these studies can be grouped into three key research directions that align with the focus of this paper. First, prior work on retrieval performance and system architecture provides the foundation for examining how semantic retrieval supports knowledge management, which relates to RQ2. Second, research on enterprise integration and workflow support emphasizes the importance of embedding systems into operational contexts, forming the basis for RQ1, which investigates how keyboard-level integration influences SME work practices. Third, studies on human–AI collaboration highlight the role of AI as a cognitive support tool rather than a replacement for human judgment (Shneiderman, 2020; Raisch & Krakowski, 2021), directly informing RQ3.

Despite these advances, existing research predominantly focuses on system performance and architecture, with limited attention to how RAG systems are embedded into everyday work practices, particularly in SMEs. Most implementations are designed as standalone applications, requiring users to switch contexts when interacting with AI systems. This creates a gap between knowledge access and actual work activities, especially in real-time customer communication.

To address this gap, this study investigates a mobile keyboard–integrated RAG system as a lightweight, embedded information system that supports customer response activities directly within messaging environments. By examining both technical performance and organizational use, this research contributes to a more practice-oriented understanding of how AI-enabled systems can be effectively integrated into SME workflows.

### 3. Research Methodology

This study was conducted with Small and Medium Enterprises (SMEs) operating in Cikarang and Bekasi, Indonesia, where customer communication is predominantly carried out through mobile messaging platforms. These SMEs typically exhibit informal organizational structures, multitasking roles, and limited IT support, making them suitable for examining how AI-enabled systems can be embedded into everyday work practices.

This research adopts a design-oriented, mixed-methods approach, combining system development with empirical evaluation.

The study is classified as applied research, as it focuses on designing, implementing, and evaluating a practical solution for real-world SME customer communication. Quantitative evaluation is used to assess system performance, while qualitative feedback is used to understand usability and human–AI interaction. The research design is aligned with the research questions: RQ1 focuses on workflow integration, RQ2 on knowledge grounding and response quality, and RQ3 on human–AI collaboration. The proposed system is implemented as a distributed architecture, consisting of a mobile keyboard interface and a cloud-based backend. The mobile component enables users to interact with the system directly within messaging applications, while the backend handles embedding, retrieval, and response generation processes. The overall system architecture is illustrated in Figure 1, which shows how user input flows through retrieval and generation stages before producing a response.

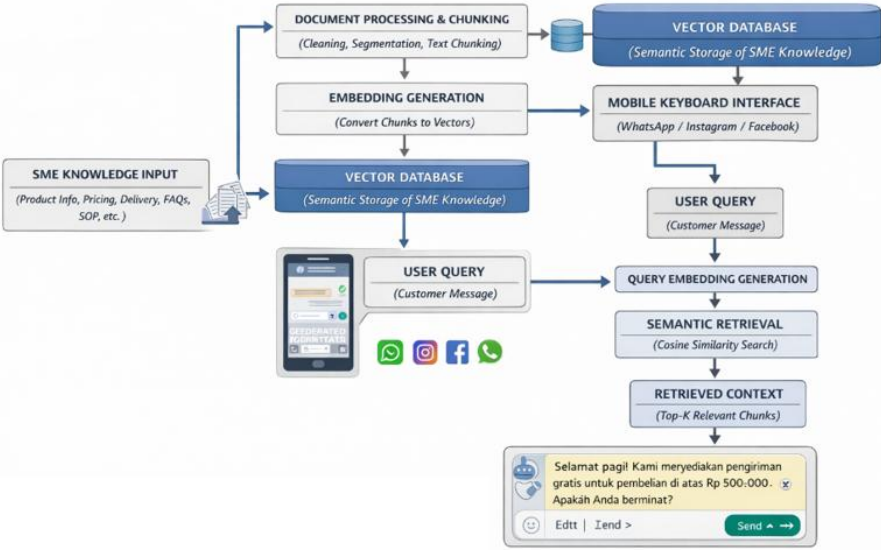


Figure 1. Architecture of the RAG system integration with the mobile application.

The system operates across multiple devices. The runtime environment includes a mobile device and a cloud server, whose specifications are presented in Table 1. In addition, the development environment is described in Table 2, which provides details of the hardware and software used during system implementation. These configurations reflect a lightweight and accessible setup suitable for SME contexts. The knowledge base used in this study was constructed from real-world SME business documents, including product descriptions, pricing policies, delivery terms, return procedures, Frequently Asked Questions (FAQs), and Standard Operating Procedures (SOPs). These documents were collected and converted into structured digital formats. Prior to processing, redundant and incomplete information was removed to ensure data consistency and quality.

| Specification | Mobile Device                     | Server              |
|---------------|-----------------------------------|---------------------|
| Model         | Oppo A78 2024                     | Superbase Free Tier |
| RAM, ROM      | 8GB, 256 GB                       | 500 MB, 8GB         |
| Display Size  | 6.43 In                           | -                   |
| CPU           | Qualcomm Snapdragon 680 Octa-core | Shared vCPU         |
| Battery       | 5000 mAh (TYP)                    | -                   |
| OS            | ColorOS 15.0                      | -                   |

Table 1. Running device specifications

| Specification         | Development Device                                                            |
|-----------------------|-------------------------------------------------------------------------------|
| Provide-Defined Model | MSI GF-63 Thin 11UD                                                           |
| RAM, ROM              | 16GB, 512 GB SSD NVME                                                         |
| Display Size          | 15.6' GHD (1920x1080), IPS-Level                                              |
| CPU                   | 11th Gen. Intel Core i7 Processor                                             |
| Battery               | 3-Cell, 52.4 Battery (Whr)                                                    |
| OS                    | Windows 11                                                                    |
| IDE                   | Android Studio (Version Koala - 2024.1.1)                                     |
| Additional Softwares  | Docker Desktop or Docker Engine<br>Node Package Manager (NPM), version 10.5.0 |

Table 2. Development device specifications

To support semantic retrieval, the knowledge base was preprocessed through a **chunking procedure**, where large documents were divided into smaller, semantically coherent segments based on markdown heading structures. The detailed procedure is formalized in **Algorithm 1**, which defines how document sections are identified and stored. This step improves retrieval precision by reducing contextual noise.

| Algorithm 1 Split Markdown by Headings |                                                                                       |
|----------------------------------------|---------------------------------------------------------------------------------------|
| 1:                                     | function SplitByHeading(markdownText)                                                 |
| 2:                                     | ▷ <b>Input:</b> A string 'markdownText'; containing Markdown content                  |
| 3:                                     | ▷ <b>Output:</b> A list of string 'sections', where each string is a section          |
| 4:                                     |                                                                                       |
| 5:                                     | <i>sections</i> ← new empty list                                                      |
| 6:                                     | <i>currentSection</i> ← new empty list                                                |
| 7:                                     | <i>lines</i> ← split <i>markdownText</i> by newline characters                        |
| 8:                                     |                                                                                       |
| 9:                                     | <b>for all</b> <i>line</i> in <i>lines</i> <b>do</b>                                  |
| 10:                                    | <b>If</b> <i>line</i> starts with a heading pattern ('#' through '#####') <b>then</b> |
| 11:                                    | <b>If</b> <i>currentSection</i> Is not empty <b>then</b>                              |
| 12:                                    | <i>joinedSection</i> ← join elements of <i>currentSection</i> with newlines           |
| 13:                                    | Add <i>joinedSection</i> to <i>sections</i>                                           |
| 14:                                    | <b>else</b>                                                                           |
| 15:                                    | Add <i>line</i> to <i>currentSection</i>                                              |
| 16:                                    | ▷ Add the last remaining section if it exists                                         |
| 17:                                    | <b>If</b> <i>currentSection</i> is not empty <b>then</b>                              |
| 18:                                    | <i>joinedSection</i> ← join elements of <i>currentSection</i> with newlines           |
| 19:                                    | Add <i>joinedSection</i> to <i>sections</i>                                           |
| 20:                                    |                                                                                       |
| 21:                                    | <b>return</b> <i>sections</i>                                                         |

After chunking, each text segment is transformed into a vector representation using the Voyage-3 embedding model. This model is based on Transformer architectures (Vaswani et al., 2017) and produces high-dimensional embeddings that capture semantic meaning. The embedding process is conceptually illustrated in Figure 2, which shows how textual input is converted into vector representations for similarity comparison.

During query processing, user input is encoded into the same vector space and compared with stored embeddings using a similarity-based retrieval mechanism. In this implementation, similarity is calculated using a negative inner product, as formalized in **Algorithm 2**. The system retrieves the top five most relevant chunks that exceed a predefined similarity threshold. These retrieved segments are then combined with the user query and system instructions to construct a structured prompt, as described in **Algorithm 3**, which is passed to a Large Language Model to generate a context-aware response.

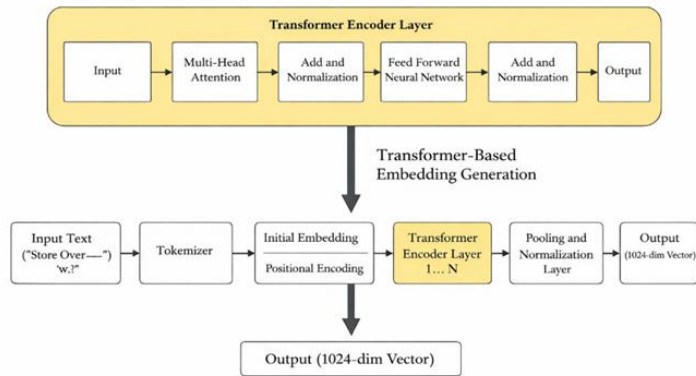


**Figure 2.** Transformer-based text embedding pipeline for the proposed RAG system.

---

**Algorithm 2** Match Chunks by Vector Similarity

---

```

1: function MatchChunks (queryEmbedding, matchThreshold, knowledgeBaseID)
2:   ▷ Input: A query vector 'queryEmbedding', a float 'matchThreshold', and an integer
   'knowledgeBaseID'
3:   ▷ Output: A list of 'matchingChunks' sorted by similarity
4:   results ← new empty list
5:
6:   for all chunk in the Chunks table do
7:     If chunk.knowledge_base_id = knowledgeBaseID then
8:       similarity ← NegativeInnerProduct (chunk.embedding, queryEmbedding)
9:       If similarity < -matchThreshold then
10:        Add (chunk, similarity) to results
11:   Sort results in ascending order based on similarity
12:
13:   return results

```

---



---

**Algorithm 3** Construct LLM Prompt

---

**Input:**

*userMessage*: The query from the user

*knowledgeDocs*: A collection of relevant document chunks

*agentSystemCommand*: An optional system instruction

*AgentTemp*: The temperature setting for the model

**Output:** A formatted prompt ready for the LLM

```

1: function Construct Prompt (userMessage, knowledgeDocs, agentSystemCommand, agentTemp)
2:   message ← new empty list
3:   ▷ Construct the System Message
4:   If agentSystemCommand is not empty then
5:     systemContent ← agentSystemCommand + "\n\nDocument:\n" + knowledgeDocs
6:   else
7:     systemContent ← "Documents:\n" + knowledgeDocs
8:   Add (role: "system", content: systemContent) to messages
9:   ▷ Construct the User Message
10:  Add (role: "user", content: userMessage) to messages
11:  ▷ Prepare for LLM call
12:  model ← "accounts/fireworks/model/deepseek-v3"
13:  temperature ← agentTemp or 0.7
14:
15:  return CallLLM (model, messages, temperature)

```

---

The system’s performance was evaluated using the RAGAS framework, focusing on three metrics: faithfulness, answer correctness, and response time. Faithfulness measures the extent to which generated responses are grounded in retrieved context as in equation 1, while answer correctness evaluates alignment with predefined reference answers using automated scoring within the RAGAS framework as in equation 2.

*Faithfulness* = (number of supported claims)/(Total claims in generated answer) (1)

*Correctness* = (Semantically correct answer components)/(Total expected answer components) (2)

Equation (1) defines faithfulness as the proportion of generated claims supported by the retrieved context, while Equation (2) defines answer correctness as the semantic similarity between the generated answer and the reference answer.

Response time is measured as the duration between user input and system output. The evaluation was conducted on a dataset of 37 test queries, consisting of 35 relevant and 2 irrelevant questions across multiple SME domains. In addition to quantitative evaluation, a User Acceptance Testing (UAT) session was conducted with one industry stakeholder experienced in SME digital operations. The testing followed a task-based approach, where the participant used the system to perform typical customer response activities, including retrieving information, generating responses, and editing outputs before sending messages. Data from the UAT session were collected through direct observation and structured feedback, focusing on usability, response quality, and workflow integration. While the qualitative evaluation is limited in scale, it provides initial insights into how the system supports human–AI collaboration and practical usability in SME contexts.

4. Result

The results are organized according to the three research questions, combining quantitative performance evaluation with qualitative insights from User Acceptance Testing (UAT). Quantitative metrics are used to assess knowledge grounding and response quality (RQ2), while workflow-related indicators and qualitative observations are used to explore system usage and human–AI interaction (RQ1 and RQ3). The mapping between research questions, data sources, and findings is summarized in Table 3.

| Research Question | Data Source                           | Empirical Evidence                                      | Key Observations                                                                                                               |
|-------------------|---------------------------------------|---------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| RQ1               | System logs, response time, usability | Response time reduced; user reported less app switching | Results suggest that keyboard-level integration may influence workflow by reducing interruptions during customer communication |
| RQ2               | RAGAS, baseline comparison            | Faithfulness 0.997; correctness 0.881                   | Findings indicate that the system can support more consistent use of business knowledge compared to keyword-based approaches   |
| RQ3               | UAT stakeholder) (1                   | User reviewed and edited responses                      | Observations suggest that the system supports a human-in-the-loop interaction pattern                                          |

Table 3. Mapping Research Questions to Empirical Evidence and Key Observations

4.1. Response Generation Evaluation

The performance of the proposed RAG system was evaluated using the RAGAS framework (Es et al., 2024) on a dataset of 37 queries, consisting of 35 relevant and 2 irrelevant questions. The evaluation focuses on three metrics: faithfulness, answer correctness, and response time. Faithfulness measures the extent to which generated responses are grounded in retrieved context, while answer correctness evaluates the alignment between generated responses and predefined reference answers using automated scoring within the RAGAS framework. Response time is measured as the duration between user input and generated output. As shown in Table 4, the system achieves a high average faithfulness score of 0.997 and an answer correctness score of 0.881, indicating that responses are generally well-grounded and aligned with expected answers. The average response time is approximately 5.02 seconds, suggesting that the system is capable of supporting near real-time interaction. However, it should be noted that the small dataset size and limited number of irrelevant queries may influence the stability of these results.

| Metric             | Average | Std. Deviation | 95% Confidence Interval |
|--------------------|---------|----------------|-------------------------|
| Faithfulness       | 0.997   | 0.0018         | [0.991, 1.000]          |
| Answer Correctness | 0.881   | 0.173          | [0.823, 0.938]          |
| Response Time (s)  | 5.02    | 0.80           | [4.735, 5.265]          |

Table 4. RAG System Performance Metrics (37 Trials)

To provide additional illustration of how these metrics are derived, **Figure 3** presents example outputs from the RAGAS evaluation framework, showing answer relevancy and correctness scores for selected queries. The figure **displays** how the system evaluates the alignment between generated responses and reference answers, as well as the degree to which responses are grounded in retrieved context. However, this visualization represents a limited set of instances and should be interpreted as an illustrative example supporting the aggregated results in Table 4, rather than as standalone evidence of performance.

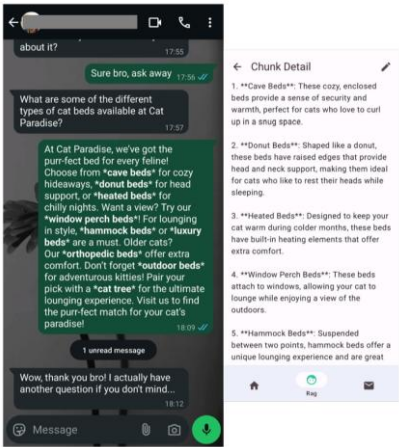


Figure 3. Answer Relevancy and Correctness

To provide a comparative perspective, the proposed system was evaluated against a keyword-based baseline, as presented in **Table 5**. The results indicate that the RAG-based approach achieves higher scores in both faithfulness and answer correctness, while also reducing response time. These findings suggest that semantic retrieval can improve response quality compared to keyword-based matching. However, the comparison should be interpreted with caution due to differences in retrieval mechanisms and the limited evaluation scale.

| Metric             | RAG System | Keyword System |
|--------------------|------------|----------------|
| Faithfulness       | 0.997      | 0.777          |
| Answer Correctness | 0.881      | 0.671          |
| Response Time(s)   | 5.02       | 10.15          |

Table 5. Average Performance Comparison: RAG vs. Keyword-based System

4.2. User Acceptance Testing

To assess practical usability and real-world applicability, User Acceptance Testing (UAT) was conducted with one industry stakeholder experienced in SME customer communication. The evaluation followed a task-based approach, where the participant interacted directly with the system. During the session, the participant performed typical customer response activities, including retrieving business information, generating responses through the keyboard interface, and editing outputs before sending messages. Given the exploratory nature of this study, the UAT was designed as an initial qualitative assessment rather than a large-scale user evaluation.

Feedback from the UAT session is summarized in Table 6. The participant reported that the keyboard-integrated interface was easy to use and reduced the need to switch between applications, which helped support smoother workflow execution. In terms of knowledge management, the participant indicated that adding and updating business information was relatively straightforward. The generated responses were generally perceived as relevant and aligned with expected business information, although minor edits were sometimes needed before sending.

| Component                 | Observation                              | User Feedback                                                                                                                                  |
|---------------------------|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| Keyboard Interface        | System integrated into mobile keyboard   | The participant found the interface easy to use and reported that it reduced the need to switch between applications                           |
| Knowledge Base Management | Uses markdown-based input                | The participant indicated that adding and updating knowledge was relatively straightforward                                                    |
| Response Quality          | AI-generated responses from RAG pipeline | The participant noted that responses were generally relevant and aligned with business information, although minor edits were sometimes needed |
| Workflow Integration      | Real-time usage during messaging         | The participant reported that the system supported response activities within the messaging workflow                                           |

Table 6. User Acceptance Testing Feedback Summary

Figure 4 illustrates the system interface, including the keyboard-based query interaction and knowledge base management components. These visualizations provide context for how the system is used in practice and support the interpretation of the qualitative findings. However, as the evaluation involved a single participant, these findings should be considered preliminary and not representative of broader SME populations.

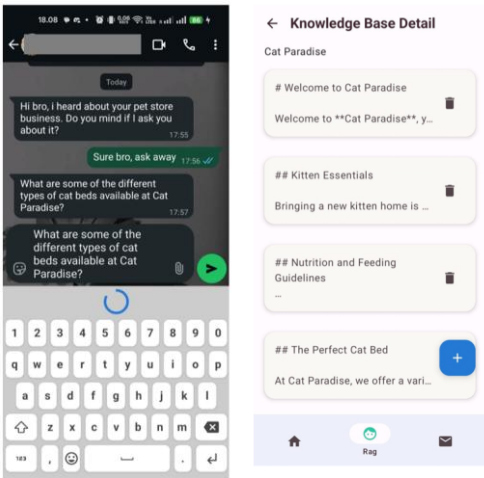


Figure 4. Keyboard Interface for RAG Query (left side), Knowledge Base Management Interface (right side)

4.3. Retrieval Evaluation

To further evaluate the effectiveness of the retrieval mechanism, the proposed RAG-based approach was compared with a conventional keyword-based method in terms of context precision as in equation 3, recall as in equation 4 and accuracy as in equation 5.

$Precision = TP / (TP + FP)$  (3)

$Recall = TP / (TP + FN)$  (4)

$Accuracy = (TP + TN) / (TP + TN + FP + FN)$  (5)

For retrieval evaluation, each of the 37 queries was manually annotated with relevant document chunks. Precision and recall were computed by comparing retrieved chunks against the ground-truth relevant chunks. Retrieval thresholds were varied from 0.1 to 1.0 in increments of 0.1 to generate the precision–recall curves. The results are presented in Figure 5, which shows the precision–recall relationship across different retrieval thresholds.

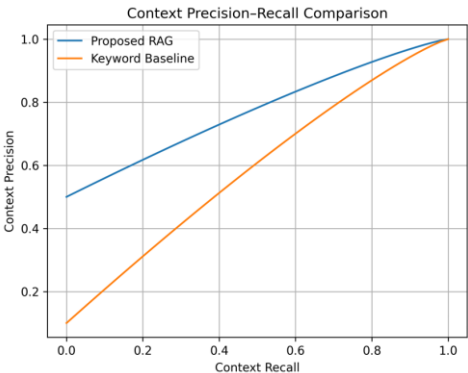


**Figure 5.** Comparison of Context Precision–Recall curves between the proposed mobile keyboard–based RAG system and the keyword-based baseline

The overall accuracy comparison between the keyword-based baseline and the proposed RAG-based system is shown in Figure 6. The results indicate that the RAG-based approach achieves higher accuracy in retrieving relevant knowledge and generating appropriate responses. This improvement can be attributed to the use of semantic embeddings, which allow the system to identify conceptually related information even when queries are expressed using different wording. Consequently, the system is better able to align generated responses with SME-specific knowledge, including pricing policies, operational rules, and service procedures.

From a practical perspective, this suggests that the proposed system has the potential to reduce response errors and improve consistency in customer communication. However, further evaluation with larger datasets and more diverse users is required to validate these findings.

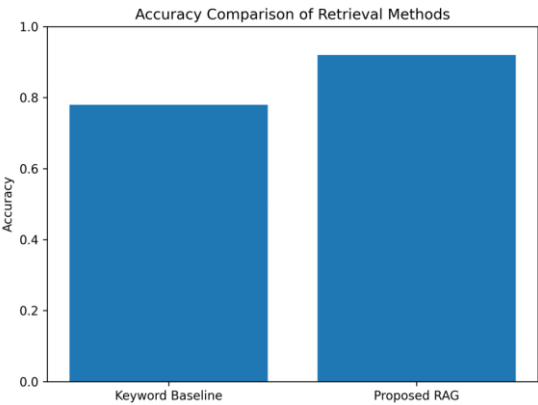


**Figure 6.** Accuracy comparison between the keyword-based baseline and the proposed RAG-based retrieval mechanism.

5. Discussion

The results of this study point to a consistent pattern: the proposed mobile keyboard–based RAG system can support both technical performance and the way SMEs handle customer communication in practice. When

the quantitative results are considered together with the User Acceptance Testing (UAT), the system does not only perform well in isolation but also fits reasonably into existing SME workflows.

Looking more closely at the operational side, the system reduced the average response time from 10.15 seconds in the keyword-based approach to 5.02 seconds in this study. This reduction was observed under controlled conditions, so it should not be interpreted as a guaranteed outcome in all settings. Still, it gives a practical sense of what might happen when retrieval and response generation are embedded directly into the keyboard interface. SME operators often respond to customers while handling multiple tasks at once, and reducing the need to switch between applications seems to make the interaction smoother. Rather than introducing entirely new ways of working, the system appears to support and streamline what users are already doing.

There are also some interesting implications for how SMEs handle knowledge during customer interactions. The high faithfulness and correctness scores suggest that the system is generally able to produce responses that stay aligned with the underlying business information. Compared to keyword-based search, which often requires users to interpret and rewrite retrieved content, the RAG-based approach produces responses that are closer to being ready to send. That said, the dataset used here is relatively small, so these results should be read as indicative rather than conclusive.

A similar pattern can be seen in the comparison with the keyword-based baseline. The proposed system tends to perform better in retrieving relevant information and generating usable responses, but this difference should be viewed with some caution. The evaluation was carried out in a controlled environment, and real customer queries are usually more varied and less predictable. In practice, the performance gap may not always be as clear as what is observed here.

One aspect that stands out is how the system is actually used. During the UAT session, the stakeholder did not simply accept the generated responses but reviewed and adjusted them before sending. This suggests that the system works more as a support tool than as a replacement for human input. The interaction feels closer to collaboration: the AI drafts a response, and the user refines it. In SME settings, where communication style and relationships with customers matter, this balance seems important.

From an Information Systems perspective, this points to the importance of how AI is embedded into everyday work. Much of the existing literature focuses on standalone tools or chatbot interfaces. Here, the system is integrated at the keyboard level, placing it directly within the flow of communication. This reduces the gap between accessing information and responding to customers, which may explain why the system feels usable even in a relatively simple setup. For SMEs, where resources are often limited, this kind of lightweight integration may be more practical than adopting more complex systems. The study suggests that a keyboard-integrated RAG system can support more efficient and consistent customer communication in SMEs, at least within the context examined here. At the same time, the scope of the evaluation remains limited. The dataset consists of a relatively small number of queries, and the qualitative insights are based on a single stakeholder. For that reason, the findings should be treated as exploratory. Further work involving more participants, broader datasets, and longer periods of use would help clarify how the system performs in real-world SME environments.

## 6. Conclusion

This study explores how a mobile keyboard-based Retrieval-Augmented Generation system can support customer communication in Small and Medium Enterprises from an Information Systems perspective. Instead of focusing only on technical performance, the study looks at how such a system fits into everyday work, particularly how SME operators access and use business knowledge while interacting with customers. The results suggest that embedding AI directly into a keyboard interface can make a noticeable difference in how customer responses are handled. Rather than acting as a separate tool, the system becomes part of the communication process itself. This seems to help reduce the effort required to retrieve information and draft responses, especially in situations where users need to respond quickly while managing other tasks. Another point that emerges from the study is how knowledge is used in practice. By combining semantic retrieval with generative models, the system is able to produce responses that are generally aligned with existing business information. This can be particularly useful in SME contexts, where knowledge is often scattered across documents and not formally structured. At the same time, the system does not replace human input. Users still review and adjust responses, which highlights a collaborative pattern where AI supports the task rather than fully automating it. These observations contribute to Information Systems research by showing that the value of AI does not only depend on model performance, but also on how and where the system is integrated. In this case, placing the functionality at the keyboard level reduces the gap between accessing knowledge and acting on it. For SMEs, this kind of lightweight integration may be more practical than adopting more complex

systems that require significant changes to existing workflows. There are, however, several limitations that need to be acknowledged. The evaluation is based on a relatively small number of queries and involves SMEs located in Cikarang and Bekasi, Indonesia. In addition, the qualitative insights come from a single stakeholder and reflect short-term system use. These factors mean that the findings should be interpreted with caution and cannot be generalized without further validation. Future research could explore how similar systems perform across different industries and regions, as well as over longer periods of use. It would also be useful to examine how such tools affect broader outcomes, such as customer satisfaction, response consistency, or sales performance. Looking at how these systems integrate with other organizational processes, such as training or onboarding, may also provide a deeper understanding of their role within SME information systems.

## References

- Aurora, C., & Tuga Mauritsius. (2025). Creation of RAG Chatbot in Answering Queries Related to Banking Terms Using Microsoft Azure. *International Journal of Cyber and IT Service Management*, 5(2), 144–155. <https://doi.org/10.34306/ijcitsm.v5i2.209>
- Benita, J., Tej, K. V. C., Kumar, E. V., Subbarao, G. V., & Venkatesh, CH. (2024). Implementation of Retrieval-Augmented Generation (RAG) in Chatbot Systems for Enhanced Real-Time Customer Support in E-Commerce. *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, pp.1381–1388. <https://doi.org/10.1109/icacrs62842.2024.10841586>
- Bostrom, R. P., & Heinen, J. S. (1977). MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. *MIS Quarterly*, 1(3), 17–32. <https://doi.org/10.2307/248710>
- Campi, R., Giudici, M., Pincirolì Vago, N. O., Brambilla, M., & Fraternali, P. (2025). Enhancing Human-AI Collaboration through a Conversational Agent for Energy Efficiency. *Proceedings of the AAAI Symposium Series*, vol. 5, no. 1, pp. 52–55. <https://doi.org/10.1609/aaais.v5i1.35554>
- Chen, C.-C., & Chang, C.-C. (2025). AI-enhanced reference services in special libraries: a case study of the Hakka Literary Museum. *The Electronic Library*, vol. 43, no. 5, pp. 715–732. <https://doi.org/10.1108/el-05-2025-0168>
- Chien, C.-Y., Wang, H.-C., Wang, S.-M., & Wu, K.-C. (2025). A Novel Intermediator of Food Experience: Applying Retrieval Augmented Generation Technology for Emotional Menu Design. *Proceeding of IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, pp. 87–88. <https://doi.org/10.1109/icce-taiwan66881.2025.11208075>
- Choudhary, N., Goyal, P., Devashish Siwatch, Atharva Chandak, Mahajan, H., Khurana, V., & Kumar, Y. (2025). AdQuestA: Knowledge-Guided Visual Question Answer Framework for Advertisements. *Proceeding of IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5812–5821. <https://doi.org/10.1109/wacv61041.2025.00567>
- Church, K. W., Sun, J., Yue, R., Vickers, P., Saba, W., & Chandrasekar, R. (2024). Emerging trends: a gentle introduction to RAG. *Natural Language Engineering*, vol. 30, no. 4, pp. 870–881. <https://doi.org/10.1017/s1351324924000044>
- Coen, E., Del Fioli, G., Kaphingst, K. A., Borsato, E., Shannon, J., Smith, H., Masino, A., & Allen, C. G. (2025). Chatbot for the Return of Positive Genetic Screening Results for Hereditary Cancer Syndromes: Prompt Engineering Project. *JMIR Cancer*, vol. 11, e65848. <https://doi.org/10.2196/65848>
- Cook, S. D. N., & Brown, J. S. (1999). Bridging Epistemologies: The Generative Dance Between Organizational Knowledge and Organizational Knowing. *Organization Science*, vol. 10, no. 4, pp. 381–400. <https://doi.org/10.1287/orsc.10.4.381>
- Decker, B., Grzemski, S., & Hengstebeck, I. (2025). Implementation of an AI Support Chatbot Based on Microsoft Azure OpenAI with Special Consideration of Quality. *Proceedings of EUNIS 2025 Annual Congress in Belfast*, vol. 107, pp. 237–226. <https://doi.org/10.29007/16ss>
- Es, S., James, J., Anke, L. E., & Schockaert, S. (2024). RAGAs: Automated Evaluation of Retrieval Augmented Generation. *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, 150–158. <https://doi.org/10.18653/v1/2024.eacl-demo.16>
- Mehul, Kanagavalli, V. R., K R, S., P N, G., M P, S., U, S., R, G., S, G., & R, N. (2025). Gen AI Driven FAQ Chatbot Using Advanced RAG Architecture for Querying Annual Reports. *2025 International Conference on Computing and Communication Technologies (ICCT)*, 1–6. <https://doi.org/10.1109/icct63501.2025.11020025>

- Mumford, E. (2006). The story of socio-technical design: reflections on its successes, failures and potential. *Information Systems Journal*, vol. 16, no. 4, pp. 317–342. <https://doi.org/10.1111/j.1365-2575.2006.00221.x>
- Lee, H.-C., Hung, K., Man, G. M.-T., Ho, R., & Leung, M. (2024). Development of an RAG-Based LLM Chatbot for Enhancing Technical Support Service. *TENCON 2024 - 2024 IEEE Region 10 Conference (TENCON)*, pp. 1080–1083. <https://doi.org/10.1109/tencon61640.2024.10902801>
- Li, Y., Zhao, H., Lin, J., Zhou, Z., & Zhang, Y. (2025). Research on Personalized Insurance Intelligent Customer Service Dialogue Generation Based on Large Language Models. *Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security*, pp. 178–183. <https://doi.org/10.1145/3728725.3728753>
- Lukman Arif Sanjani, Sarno, R., Sungkono, K. R., Agus Tri Haryono, Abdullah Faqih Septiyanto, & Dwi Sunaryono. (2025). Performance Analysis of LLM Models with RAG and Fine-Tuning T5 for Chatbot Optimization in Call Centers. *Proceeding of International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI)*, pp. 152–157. <https://doi.org/10.1109/icocseti63724.2025.11018908>
- Lystbæk, M. S., & Ulrik Sahl Lystbæk. (2025). A Norm-Chatbot: Local LLM Vision with Vision-Based RAG for Complex Production Documents and Task-Specific Responses. *Communications in Computer and Information Science*, pp. 17–30. [https://doi.org/10.1007/978-3-031-96199-1\\_2](https://doi.org/10.1007/978-3-031-96199-1_2)
- Nguyen, H. T., NGO, Q.-D., & DANG, Q.-D. (2024). An Intent-Filtered Retrieval-Augmented Generation Chatbot for University Admissions in Vietnam. *Proceedings of International Conference on Advanced Technologies for Communications (ATC)*, pp. 928–933. <https://doi.org/10.1109/atc63255.2024.10908188>
- Orlikowski, W. J., & Yates, J. (2002). It's About Time: Temporal Structuring in Organizations. *Organization Science*, vol. 13, no. 6, pp. 684–700. <https://doi.org/10.1287/orsc.13.6.684.501>
- Packowski, S., Halilovic, I., Schlotfeldt, J., & Smith, T. (2024). Optimizing and Evaluating Enterprise Retrieval-Augmented Generation (RAG): A Content Design Perspective. *Proceedings of the 2024 8th International Conference on Advances in Artificial Intelligence*, pp. 162–167. <https://doi.org/10.1145/3704137.3704181>
- Patel, N., Mouratidis, H., & Zhi, K. N. K. (2025). LLM-Based Automated Hallucination Detection in Multilingual Customer Service RAG Applications. *IFIP Advances in Information and Communication Technology*, pp. 360–373. [https://doi.org/10.1007/978-3-031-96235-6\\_26](https://doi.org/10.1007/978-3-031-96235-6_26)
- Praneeth, B., Mohana, Nattam, E. C., Jetti, K., Kavyashree, B. K., Rakshitha, D., Ramakanth Kumar, P., & Sreelakshmi, K. (2025). Optimization of Customer Feedback Summarization Using Large Language Models (LLM) and Advanced Retrieval-Augmented Generation. *IEEE Access*, vol. 13, pp. 124319–124332. <https://doi.org/10.1109/access.2025.3588337>
- Qi, Y., Gu, S., Ma, W., Wang, N., & Wang, Q. (2025). DeepSem: Multilayer Semantic Image Retrieval for High-Accuracy Multimodal RAG in Industrial Applications. *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1388–1393. <https://doi.org/10.1109/compsac65507.2025.00174>
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *The Academy of Management Review*, vol 46, no. 1, pp. 192–210. <https://doi.org/10.5465/amr.2018.0072>
- Rao, D., Zhuang, J., & Jiang, Z. (2025). Leveraging dynamic few-shot prompting and ensemble method for task-oriented dialogue with subjective knowledge. *Information Processing & Management*, vol. 63, no. 2, pp. 104317–104317. <https://doi.org/10.1016/j.ipm.2025.104317>
- Sandakelum, D., Rajapakse, C., & Jayalath, N. (2025). Real-Time Knowledge Retrieval for Banking Chatbots: A RAG-Based Approach to Employee Assistance. *Proceeding of International Research Conference on Smart Computing and Systems Engineering (SCSE)*, pp. 1–6. <https://doi.org/10.1109/scse65633.2025.11030999>
- Shan, R., & Shan, T. (2024). Retrieval-Augmented Generation Architecture Framework: Harnessing the Power of RAG. *Lecture Notes in Computer Science*, pp. 88–104. [https://doi.org/10.1007/978-3-031-77954-1\\_6](https://doi.org/10.1007/978-3-031-77954-1_6)
- Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction*, vo. 12, no. 3, pp. 109–124. <https://doi.org/10.17705/1thci.00131>
- Singh, R. K., Singh, K., Kunde, S., Mishra, M., Singhal, R., & Nambiar, M. (2025). RAGs to Riches: Cost-efficient Complex Query Orchestration. *Proceeding of ACM/SPEC International Conference on Performance Engineering*, pp. 114–120. <https://doi.org/10.1145/3680256.3721327>

- Thiyagarajan, K. (2025). Multimodal RAG for Enhanced Information Retrieval and Generation in Retail. *2025 International Conference on Visual Analytics and Data Visualization (ICVADV)*, pp. 102–106. <https://doi.org/10.1109/icvadv63329.2025.10961713>
- Varga, I., & Yamashita, Y. (2025). Inquiry Assistant Using LLM-Generated Knowledge Graphs. *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 4319–4323. <https://doi.org/10.1145/3726302.3731956>
- Vaswani, A., et al. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- Wang, J., & Shalaby, A. (2025). TransitTalk: Large language model-based digital assistants for enhancing transit customer experience and staff performance. *Journal of Intelligent Transportation Systems*, pp 1–18. <https://doi.org/10.1080/15472450.2025.2576883>
- Wulf, J., & Meierhofer, J. (2025). The Impact of Large Language Models on Task Automation in Manufacturing Services. *Procedia CIRP*, 134, pp. 1089–1094. <https://doi.org/10.1016/j.procir.2025.03.071>
- Yang, Y.-T., Jiang, J.-Y., Lin, Y.-T., & Chang, C.-Y. (2025). Enhancing Retrieval-Augmented Generation with Knowledge Graph-Based Soft-Labeling and Triplet Similarity SBERT. *Proceeding of IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, pp. 487–488. <https://doi.org/10.1109/icce-taiwan66881.2025.11208012>
- Yildirim, S., & Ruya Samli. (2025). A RAG-Based Automotive Sector AI Assistant for Enhanced Information Retrieval. *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, vol. 33, pp. 20–27. <https://doi.org/10.55549/epstem.1729714>

## An Advanced Stacked Ensemble Framework for Enhanced Software Bug Predictions

Sabrina Nasrin Saima<sup>1</sup>, Abdus Salam<sup>2\*</sup>, Mir Md. Kawsur<sup>2</sup> and K. M. Imtiaz-Ud-Din<sup>2</sup>

<sup>1</sup>Department of Computer Science & Engineering, Stamford University Bangladesh, Dhaka, Bangladesh

<sup>2</sup>Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh

\*Correspondence: [abdus.salam@aiub.edu](mailto:abdus.salam@aiub.edu)

### PAPER INFO

#### Paper history:

Received 03 March 2026

Accepted 29 June 2026

#### Citation:

Saima, S. N., Salam, A., Kawsur, M. M., & Imtiaz-Ud-Din, K. M. (2026). An advanced stacked ensemble framework for enhanced software bug predictions. In *Journal of Information and Organizational Sciences*, vol. 49, no. 1, pp. 247-262

#### Copyright:

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

### ABSTRACT

Software Bug Prediction (SBP) is a critical activity in software engineering that seeks to identify defect-prone modules before release, enabling focused testing and reducing post-release failures. With the greater scale and complexity of software, data-driven prediction techniques are necessary to maintain reliability and limit maintenance costs. Machine learning methods have led to recent advances in SBP by inducing defect patterns from historical data. However, existing models struggle to address class imbalance, feature redundancy, and limited generalization across projects. Furthermore, most previous studies focus on single classifiers, which are unable to capture complex data patterns effectively. These limitations reduce the utility of the real world and make it difficult to maintain consistent cross-project performance. To address these issues, we propose a new stacked ensemble framework that combines three base learners (Random Forest, XGBoost, CatBoost) and a Multi-Layer Perceptron model as a meta-learner. We performed a systematic ablation study on four preprocessing pipelines (SMOTE + RFE, SMOTE + Boruta, ADASYN + RFE, ADASYN + Boruta) to evaluate the state-of-the-art model. Empirical evaluation across ten benchmark datasets from the NASA and PROMISE repositories shows consistent gains over individual classifiers and competitive literature baselines, with the propose framework achieving a maximum accuracy of 96.52% while providing notable improvements in bug-prediction robustness.

**Keywords:** Software Bug Prediction, Stacked Ensemble, Neural Network, Meta-Model, Data Balancing, NASA Datasets, PROMISE Datasets

## 1. Introduction

The size and complexity of a software system has grown rapidly, which has greatly raised the possibility that flaws may be introduced during the development phase. Massive functional failures, safety risks, financial losses, and reputational harm can all result from undetected bugs in software. The quality assurance of software is therefore now a major imperative for organizations striving to develop reliable, flexible, and cost-effective products. One of the best means to determine software quality is Software Bug Prediction (SBP), the process of identifying defect-prone modules prior to deployment to employ testing resources in a more efficient manner (Al-Fraihat et al., 2024). So, anticipating software defects is one of the most important steps in the software development phase (Siva et al., 2023).

During the last two decades, a large number of machine learning (ML) methods have been utilized for SBP, ranging from classic classifiers such as Decision Trees and Naive Bayes to more recent algorithms such as Support Vector Machines, Random Forests etc. (Hammouri et al., 2018). In addition to classification, Ensemble Modelling (EM) has become a viable method to improve overall performance by integrating predictions from several machine learning models (Ali et al., 2024). These models perform well; however, their performance is often constrained by several lingering challenges. The most important challenge among them is class imbalance, where the vast majority of non-defective modules are so large in number compared to defective ones that the models are biased in favor of the majority class (Iqbal et al., 2019; Siva et al., 2023). Software datasets also have irrelevant or redundant features, such as noisy or overlapping data, which decrease the accuracy of predictions and increase computational cost (Das et al., 2024). One of the primary reasons why software bug prediction performs poorly is an imbalanced dataset (Tong et al., 2022). Another concern is the lack of generalizability, as models trained on one dataset do not perform well on other projects due to variations in software characteristics, metrics, and defect distributions (Gray et al., 2013). These issues point to the need for more powerful and robust solutions for software defect prediction.

To address such problems, this study proposes a stacked ensemble framework for software bug prediction. The method combines three best-performing base classifiers with a meta-classifier. The framework also has some advanced pre-processing methods, including feature selection and class balancing, to enhance predictive performance and reducing overfitting. To enhance the strength and generalizability of software bug prediction, this work tries to answer three key questions. First, can a stacked ensemble model outperform individual classifiers in software bug prediction? Second, do combined feature selection and class balance techniques achieve measurable improvements over raw datasets? Lastly, can the proposed framework outperform current methods in various project contexts and generalize successfully across different datasets? These are the questions that motivate our study, guiding a systematic evaluation of ten NASA and PROMISE datasets to discover how algorithmic and preprocessing advances interact to impact predictive accuracy and stability in software defect detection. The key contributions of this study are summarized as follows:

- We introduce a new stacked ensemble framework that combines three machine learning models, namely Random Forest, XGBoost, and CatBoost, with Multi-Layer Perceptron (MLP) as a neural network-based meta-classifier for enhanced software bug prediction.
- We perform a first-of-its-kind ablation study to analyze four combinations of pre-processing techniques: SMOTE + RFE, SMOTE + Boruta, ADASYN + RFE, and ADASYN + Boruta, in an attempt to determine which combination of class balancing and feature selection methods provides the optimal performance when combined with the proposed framework.
- We validate the effectiveness of the proposed framework through extensive experiments on ten benchmark datasets (NASA and PROMISE) using multiple evaluation metrics, demonstrating clear improvements over raw datasets.
- Our proposed framework is compared with ten baseline classifiers and the state-of-the-art method, demonstrating high accuracy and good generalization in different software defect prediction scenarios.

The remainder of this paper is structured as follows. Section 2 presents the related work in ensemble learning and software bug prediction. Section 3 describes the proposed methodology and experimental setup. Section 4 provides the results and analysis. Finally, Section 5 concludes the paper and suggests future work.

## 2. Related Work

Software Bug Prediction has been an area of active research during the last few decades, motivated by increasing software system complexity and increasing demands for dependable, maintainable, and secure software products. Early prediction of bugs is crucial for reducing post-release failures, reducing maintenance costs, and improving customer satisfaction. Rules-based and manual inspection methods have been helpful but are tedious, error-prone, and not scalable. Software bug prediction has been transformed by machine learning, which identifies defect-prone modules by learning patterns from historical metrics. Over the years, researchers have explored a wide range of techniques, from traditional classifiers to advanced deep learning models and hybrid optimization approaches, each of which aims to improve the accuracy and reliability of predictions. This section presents constant challenges, highlights gaps that drive ongoing research, and summarizes significant achievements.

2.1. Traditional Machine Learning Techniques

Machine Learning based bug/defect prediction models utilized algorithms such as Naive Bayes, Decision Tree, Artificial Neural Networks (ANN), Support Vector Machine (SVM), J48, Random Forest, Logistic Regression, AdaboostM1, XGBoost (Ali et al., 2024; Hammouri et al., 2018; Iqbal et al., 2019; Meenakshi & Singh, 2019; Naidu et al., 2022) etc. PROMISE datasets were used to conduct experiments, which confirmed the effectiveness of these models, which were even up to 90% accurate (Hammouri et al., 2018). Similarly, research (Iqbal et al., 2019) examined numerous algorithms using NASA datasets, including MLP, Radial Basis Function (RBF), SVM, and K-Nearest Neighbors (KNN). It found that RBF performed best in a number of instances, closely followed by MLP and SVM. Despite these successes, limitations remain, where models tuned for one dataset often degrade significantly when applied to others that cause performance inconsistency. Also, effect of class imbalance problem where datasets usually have significantly fewer faulty modules compared to non-faulty ones, with models leaning towards the majority class.

2.2. Data Preparation and Balancing Techniques

The quality and quantity of software metrics play a crucial role in determining the accuracy of defect prediction models. Over the years, researchers have used a variety of datasets for software bug prediction. To provide a clearer overview of dataset usage in prior studies, a summary is presented in Table 1. Among the many techniques that have been widely applied in dealing with class imbalance is the Synthetic Minority Oversampling Technique (SMOTE) (Tong et al., 2022) and adaptive variants of it. Adaptive SMOTE (ASMO) (Das et al., 2024), improved the performance of the classifier in modelling minority classes, as indicated in (Rizwan et al., 2017), where the maximum accuracy of 89.4% was achieved with a bagging ensemble on multiple datasets. XGBoost with SMOTE was tuned in another paper (Gupta et al., 2020) to enhance minority detection rates, while author (Ferenc et al., 2020) an introduced a new BugHunter dataset and recommended it, offering temporally precise bug labels, but fewer studies. While pre-processing has been demonstrated to work, no consensus exists on the optimal mixture of cleaning, balancing, and feature selection techniques to yield consistent, cross-dataset performance.

| Studies                                                                                                                                                                                                                                          | Dataset(s) Used  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
| (Al-Fraihat et al., 2024; Siva et al., 2023; Iqbal et al., 2019; Das et al., 2024; Gray et al., 2013; Dada et al., 2021; Amit et al., 2024; Wang et al., 2021; Gupta et al., 2020; Rizwan et al., 2017; Siva et al., 2024; Balogun et al., 2021) | NASA             |
| (Siva et al., 2023; Hammouri et al., 2018; Ali et al., 2024; Tong et al., 2022; Wang et al., 2021; Iqbal & Aftab, 2020; Siva et al., 2024)                                                                                                       | PROMISE          |
| (Puranik et al., 2016; Khleel & Nehez, 2023; Borandag, 2023)                                                                                                                                                                                     | Eclipse JDT Core |
| (Khleel & Nehez, 2023)                                                                                                                                                                                                                           | Bugcatchers Bug  |
| (Khleel & Nehéz, 2023)                                                                                                                                                                                                                           | GitHub Bug       |
| (Mahfoodh & Obediat, 2020)                                                                                                                                                                                                                       | Mozilla Core     |
| (Daza, 2025)                                                                                                                                                                                                                                     | Software Defect  |

Table 1. Summary of Datasets Used in Previous Studies.

2.3. Bug Prediction using Ensemble Learning

Ensemble learning has proven to be a promising method for improving prediction accuracy and generalizability (Wang et al., 2021; Bala et al., 2022). Bagging, boosting, and stacking techniques combine multiple classifiers into one to tap their strength in complementary directions (Sharma et al., 2023). Several ensemble models, including Vote, Bagging, Random Forest, and AdaBoostM1, were optimized using hyper parameter tuning in a study by (Al-Fraihat et al., 2024), which achieved an accuracy of over 81% on the JM1 dataset. Similarly to this, a different study (Dada et al., 2021) used a stacked architecture with Random Forest acting as the meta-learner and KNN, GLMNet, and Linear Discriminant Analysis (LDA) as the base learner, which achieved an average accuracy of 88% on a few chosen datasets. However, most of the ensemble methods (Al-Fraihat et al., 2024; Dada et al., 2021; Borandag, 2023) focus solely on accuracy, passing metrics such as ROC-AUC, precision, recall, and effort-aware metrics, particularly useful in real-world bug prediction contexts. Moreover, most of them address neither all meaningful challenges such as class imbalance, redundant features, nor overfitting in a single framework.

## 2.4. Hybrid Optimization and Deep Learning Techniques

Recent times have seen the integration of deep learning into SBP. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models have shown strong performance, particularly in managing large datasets (Borandag, 2023). Hybrid frameworks further enhance predictability by linking optimization algorithms with a deep learning architecture. Examples include Adaptive Artificial Jelly Optimization with LSTM (Siva et al., 2023) and Adaptive Golden Eagle Optimizer with LSTM (Siva et al., 2024), both of which utilize feature selection to improve learning efficacy and precision. Additional hybrid methods combine dimensionality reduction with ensemble methods, for instance, study (Amit et al., 2024) employed Principal Component Analysis (PCA) with AdaBoost and Random Forest to achieve 91.17% accuracy for different datasets without changing the fold count. Feature selection approaches based on bio-inspired algorithms, such as Spider Wasp Optimization (Das et al., 2024), have also been shown to improve prediction accuracy, although often at the cost of substantial computational overhead. In addition, current studies have shown that deep learning architectures such as CNN and LSTM networks can learn to learn discriminative features from software measures automatically and outperform most traditional machine learning models in defect prediction tasks (Shin et al., 2021). Similarly, a study (Alsghaier & Akour, 2020) integrated the Genetic Algorithm (GA) with a Support Vector Machine (SVM) classifier, showing that hybrid combinations can significantly enhance software fault prediction accuracy across large and small-scale datasets. Despite these advancements, many hybrid and deep learning methods still demand significant computational resources and are prone to overfitting when applied to small datasets, which limits their practical applicability.

To bridge these gaps, our work suggests a stacked ensemble approach that utilizes three very potent base learners, such as XGBoost, CatBoost, and Random Forest, within a Multi-Layer Perceptron (MLP) meta-classifier. The distinguishing feature of this approach is its focus on pre-processing as an end-to-end, carefully considered pipeline. Along with that, we explore how multiple pre-processing strategies can cooperate to increase the precision of bug prediction. By examining these attributes in an ablation study, our findings can provide actionable guidance to software engineers and researchers looking to build more resilient, generalizable bug prediction models. By testing this configuration on ten diverse sets of datasets from the NASA and PROMISE repositories, we aim to build a solution that not only performs well, but also repeatedly, on a variety of software project types

## 3. Methodology

This section presents our proposed stacked ensemble framework for software defect prediction. We begin by describing the NASA and PROMISE datasets used in this study, followed by the data pre-processing and stratified dataset splitting procedures. The ensemble learning approach is then detailed, including the selection of base classifiers and the meta-model, along with the rationale for their inclusion. We also introduce the ablation study design, which evaluates the effect of different combinations of class balancing and feature selection techniques. Finally, the evaluation metrics used to assess the performance of the proposed study are outlined. An overview of the complete methodology is illustrated in Figure 1.

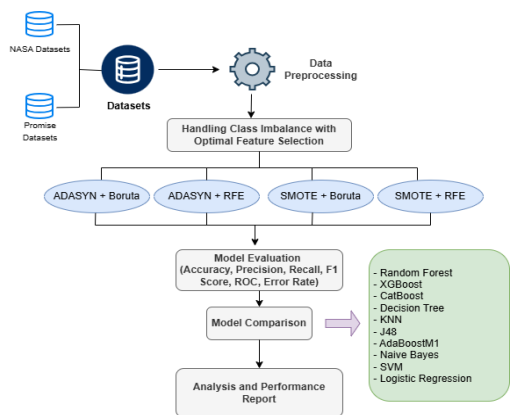


Figure 1. Overall Workflow of the proposed study.

### 3.1. Dataset

In this study, we used ten publicly available datasets that are widely recognized benchmarks in software defect prediction. Five datasets from the NASA Metrics Data Program (MDP), namely JM1, PC4, MW1, KC3, and MC2 and five from the PROMISE repository, including Camel-1.6, Ant-1.7, Prop-6, Synapse-1.2 and Xalan-2.4, were chosen because they are among the most frequently cited and extensively studied in the software defect prediction literature. The datasets were obtained from well-established public sources. The PROMISE datasets were obtained from the PROMISE'19 Replication Package hosted on Figshare, a curated and DOI-assigned release. The NASA datasets were gathered from the NASA Defect Dataset GitHub repository by Chakkrit Tantithamthavorn, which contains original MDP datasets. These repositories are publicly available.<sup>1</sup> Both of the sources are well-cited in previous research, guaranteeing reproducibility and authenticity of our experiments. Each dataset is composed of software modules described by multiple static attributes, such as size measures, complexity metrics, and coupling indicators. Along with these independent variables, every module is also labeled (Yes or No) with a defect indicator that is the prediction target. A summary of the datasets used in this study, including the number of modules, features, and class distributions before pre-processing, is presented in Table 2.

| Source        | NASA  |      |     |     |     | PROMISE   |         |           |        |             |
|---------------|-------|------|-----|-----|-----|-----------|---------|-----------|--------|-------------|
| Project Name  | JM1   | PC4  | KC3 | MW1 | MC2 | Camel-1.6 | Ant-1.7 | Xalan-2.4 | Prop-6 | Synapse-1.2 |
| Instances     | 10878 | 1458 | 458 | 403 | 161 | 965       | 745     | 723       | 644    | 234         |
| Defective     | 2102  | 178  | 43  | 31  | 52  | 188       | 166     | 110       | 61     | 64          |
| Non-Defective | 8776  | 1280 | 415 | 372 | 109 | 777       | 579     | 613       | 583    | 170         |
| Features      | 22    | 41   | 41  | 41  | 41  | 20        | 20      | 20        | 20     | 20          |

**Table 2.** Summary of raw datasets before pre-processing.

### 3.2. Data Pre-processing

To ensure the input data are clean, balanced, and appropriate for model training, effective data pre-processing is a crucial step in the suggested model for software bug prediction. The pre-processing steps have been designed to systematically handle data inconsistencies, label harmonization, balanced class distributions, select useful features, and ablation studies utilizing four combinations of feature selection and class balancing.

#### 3.2.1. Data Cleaning

The first was that NASA's initial datasets were accessed in the Weka Attribute-Relation File Format (ARFF) format. Therefore, all NASA datasets were converted to the CSV format. Data cleansing is the process of discovering and handling duplicates, null values, and missing entries. Incomplete or missing reading modules were not included as they can contribute bias and noise to the training. To ensure that each software module appears only once in the analysis, duplicate records were also removed, which can be found in publicly available datasets that were also removed. This was done to provide a clean base for further pre-processing.

#### 3.2.2. Label Harmonization

Since NASA and PROMISE datasets both retained defect data in different formats, label harmonization was necessary. The defective labels in the NASA datasets were categorical (Y for defective and N for non-defective). These were converted to numeric form (1 for defective, 0 for non-defective). The defect data in the PROMISE datasets were provided as bug counts. To normalize these to NASA format, we binarized the counts: modules with one or more defects were labeled defective (1), and no-defect modules were labeled as non-defective (0). The conversion gave us a consistent target variable for all datasets.

<sup>1</sup> PROMISE Dataset: [https://figshare.com/articles/dataset/Replication\\_Package\\_-\\_PROMISE\\_19/8852084?file=16224071](https://figshare.com/articles/dataset/Replication_Package_-_PROMISE_19/8852084?file=16224071);  
 NASA Dataset: <https://github.com/klainfo/NASADefectDataset>

### 3.2.3. Class Balancing

In Class imbalance, defective modules are significantly outnumbered by non-defective modules, which is a typical issue in software defect prediction. This imbalance relies on machine learning models to predict the majority class. To counteract this, we employ two resampling techniques: the Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN). SMOTE generates synthetic samples in the feature space of the minority examples, whereas ADASYN tunes its sampling to target hard-to-learn regions, creating more representative synthetic instances. Both methods are employed only on the training dataset, without any inclusion of information from the test dataset.

### 3.2.4. Feature Selection

To reduce redundancy and improve model interpretability, we employed two feature selection methods. RFE (Recursive Feature Elimination) and Boruta. RFE is a wrapper-based method that removes the least contributory features step by step based on their contribution to model performance. In this study, a Random Forest classifier was used as the estimator. The iteration continues until the top 15 important features are selected. In contrast, Boruta is a random feature selection algorithm whereby all the important features are selected through comparison between the importance of original features and the random forest importance score using shadow features. Features that score higher in significance when compared to their corresponding shadow features are classified as important.

### 3.2.5. Ablation Study

To contrast the effect of different pre-processing methods, we conducted an ablation study combining class balancing and feature selection in four configurations: SMOTE + RFE, SMOTE + Boruta, ADASYN + RFE, and ADASYN + Boruta. Each configuration was executed on the datasets and validated using the proposed stacked ensemble model. The controlled test allowed us to identify the optimal pre-processing pipeline and provided a better understanding of how balancing and feature selection in combination play a role in model performance. To demonstrate the effect of class balancing and feature selection, Figure 2 illustrates the class distribution after applying four preprocessing combinations (SMOTE + RFE, SMOTE + Boruta, ADASYN + RFE, and ADASYN + Boruta). It can be observed that the applied preprocessing techniques significantly reduce class imbalance, resulting in a more balanced distribution between defective and non-defective instances, where SMOTE-based techniques produce a perfectly balanced distribution, whereas ADASYN-based techniques result in a slight imbalance due to adaptive sampling. For brevity, only JM1 is presented, while the same preprocessing steps were consistently applied to all NASA and PROMISE datasets.

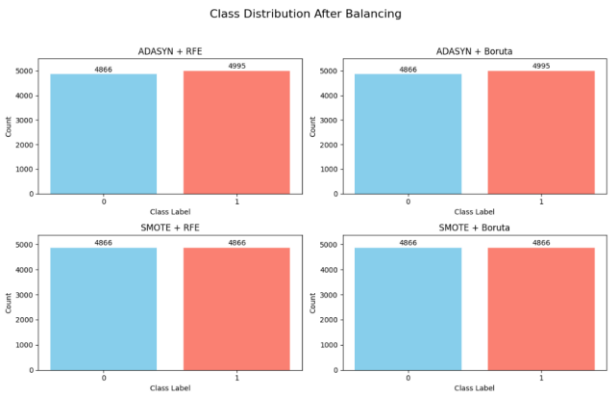


**Figure 2.** Class Distribution of the Training Dataset after Applying Preprocessing Techniques.

### 3.2.6. Data Partitioning

The dataset was partitioned into two distinct subsets: training and testing. The allocation of the dataset followed a 70:30 split, with 70% of the data reserved for training and another 30% reserved for testing.

### 3.3. Experimental Setup

The experimental setting for this study used Python 3.10.2 as the primary programming language within the Google Colab (cloud-based Jupyter Notebook) environment. For model development, we used TensorFlow 2.1.0 and Keras 2.3.1. All experiments were performed on a Linux-based runtime equipped with a cloud-based CPU (~ 2.20 GHz), 12 GB RAM, and the default Colab hardware configuration. Data pre-processing, machine learning pipeline construction, and visualization were performed using Scikit-learn, Pandas, NumPy, Matplotlib, and Seaborn libraries.

### 3.4. Evaluation Metrics

To comprehensively evaluate the performance of the proposed stacked ensemble model, some of the most widely accepted metrics were used, including Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Error Rate. Taken collectively, the metrics provide an equitable idea about the model's ability to identify defective modules while minimizing incorrect predictions. They also help to make the evaluation objective and consistent, even when dealing with datasets with imbalances of classes, giving a better representation of the reliability of the model and overall robustness.

### 3.5. Proposed Framework

In this section, we introduce a new stacked ensemble framework to enhance the reliability and predictive power of software defect prediction models. As shown in Figure 3, our approach operates in two principal layers: training and testing. The training layer is further subdivided into three sequential stages: data preprocessing, base classification, and meta-learning, and the testing layer (prediction stage) predicts the software as defective or non-defective.

#### 3.5.1. Base Classification Step

In the base classification stage, three well-established machine learning models, Random Forest, XGBoost, and CatBoost, are individually trained on the pre-processed training sets. The models were chosen because of their complementary strengths; Random Forest ensures strength against overfitting by means of bootstrap aggregation and random feature selection. XGBoost efficiently fits complex non-linear connections and performs gradient boosting. CatBoost is effective in feature interaction management and is highly efficient even with minimal parameter tuning. Each of the three base learners, generate probabilistic predictions (class probabilities) to being in each of the classes, which were then used as informative inputs to the meta-learner.

#### 3.5.2. Meta-Learner Step

The meta-learning phase uses a Multi-Layer Perceptron (MLP) model as the meta-classifier. To create the meta-features required by the stacking model, 5-fold stratified cross-validation was performed on the training dataset. In each cross-validation iteration, the base models (Random Forest, XGBoost, and CatBoost) were trained on the data subset and used to predict the unseen folds. These out-of-fold (OOF) predictions were collected from the base models and combined to create the meta-features. This approach helps reduce the risk of data leakage and enables the meta-model to learn from more reliable, unbiased predictions. The meta-learner architecture contains one input layer, two hidden layers, and one output layer. The input layer obtains probability predictions from the three input features of Random Forest, XGBoost, and CatBoost. There are two hidden layers that are fully connected (16 and 8 nodes) with ReLU functions, which is an activation function, and a dropout layer was used to prevent overfitting. The output layer consists of a single node that uses a sigmoid activation function to determine the highest possible probability that a software module is defective. During training, the meta-learner MLP was minimized by the Adam optimizer using a binary cross-entropy loss function because it is well suited for binary classification problem. The model was trained up to 15 epochs with a batch size of 32 to achieve stable convergence. To prevent overfitting and enhance generalization, an early stopping feature was employed by monitoring the validation loss with a patience of 5 epochs, restoring the best-performing model weights when stopped with L2 regularization factor of 0.001 and a dropout of 0.5 were employed. The dataset was split with 70:30 ratio where this splitting ratio ensured consistency with previous work in the area while providing a well-defined setup for comparing the various preprocessing techniques. A stratified 10-fold cross-validation was also done to test for the reliability of the model, the

holdout splitting technique remained the main approach for evaluation throughout the experiment. All of these hyper-parameters contributed to stable training, effective optimization, and improved generalization performance of the stacked ensemble model.

3.5.3. Prediction Stage

In the prediction or testing phase, the new instances are first passed to the trained base classifiers to obtain their class probabilities can be retrieved. These are then combined and fed into the MLP meta-model, which generates the final binary classification, that is, whether each module is defective or non-defective. Thus, the stacked ensemble strategy proposed in this study has certain advantages. As the ensemble effectively employ the best attributes of three different base learners most efficiently, the ensemble is not only more predictive, but also more robust and stable than any single classifier.

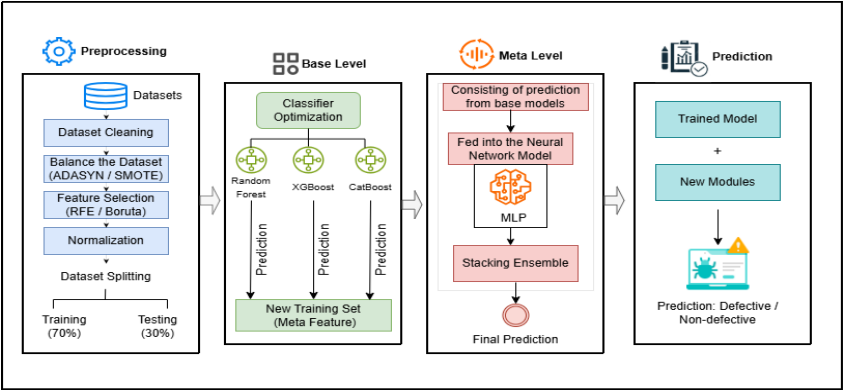


Figure 3. Proposed Stacked Ensemble Model Architecture.

4. Experimental Result Analysis

This section presents the experimental results of the proposed stacked ensemble model and compares its performance with ten baseline machine learning classifiers on ten datasets (JM1, KC3, MC2, MW1, PC4 and Ant-1.7, Camel-1.6, Synapse-1.2, Prop-6, Xalan-2.4). Performance is measured using a range of different metrics like accuracy, precision, recall, f1-score, ROC-AUC, and error rate to provide a highlight overview of the model’s effectiveness. The results are organized into three sections that are presented below.

4.1. Performance Analysis of Proposed Bug Prediction Model

Firstly, we show the results of an ablation study in which we compare the impact of four different combinations of pre-processing on the performance of the model. Although accuracy has been employed as an initial metric because the proposed framework is designed for software bug prediction, where overall prediction correctness is an important consideration, and accuracy is commonly used in comparing various software defect prediction that is widely used in prior studies, it is not entirely indicative of performance in imbalanced datasets. Therefore, additional evaluation metrics, including precision, recall, F1-measure, and ROC-AUC, were also considered to ensure the quality of the models. In several cases, despite comparatively lower accuracy, classifiers achieved higher recall or ROC-AUC values, which signifies better ability in detecting defects in modules.

| Technique Combination | Accuracy (%) | F1 Score | Recall | Precision | ROC-AUC | Error Rate (%) | Dataset |
|-----------------------|--------------|----------|--------|-----------|---------|----------------|---------|
| ADASYN + RFE          | 72.59        | 0.41     | 0.43   | 0.38      | 0.64    | 27.41          | JM1     |
| ADASYN + Boruta       | 72.31        | 0.40     | 0.42   | 0.37      | 0.64    | 27.69          |         |
| SMOTE + RFE           | 72.51        | 0.41     | 0.44   | 0.38      | 0.64    | 27.49          |         |
| SMOTE + Boruta        | 72.81        | 0.41     | 0.43   | 0.38      | 0.65    | 27.19          |         |
| ADASYN + RFE          | 77.37        | 0.41     | 0.42   | 0.42      | 0.66    | 22.63          | KC3     |

|                 |              |             |             |             |             |              |     |
|-----------------|--------------|-------------|-------------|-------------|-------------|--------------|-----|
| ADASYN + Boruta | 76.79        | 0.38        | <b>0.44</b> | 0.39        | <b>0.68</b> | 23.21        | MC2 |
| SMOTE + RFE     | 74.74        | 0.31        | 0.33        | 0.31        | 0.60        | 25.26        |     |
| SMOTE + Boruta  | 76.79        | 0.32        | 0.35        | 0.36        | 0.65        | 23.21        |     |
| ADASYN + RFE    | 55.19        | <b>0.51</b> | <b>0.68</b> | 0.38        | 0.64        | 44.81        |     |
| ADASYN + Boruta | 54.68        | 0.36        | 0.39        | <b>0.50</b> | 0.52        | 45.32        |     |
| SMOTE + RFE     | <b>58.85</b> | 0.44        | 0.52        | 0.44        | <b>0.66</b> | <b>41.15</b> |     |
| SMOTE + Boruta  | 53.65        | 0.45        | 0.57        | 0.42        | 0.55        | 46.35        | MW1 |
| ADASYN + RFE    | 85.78        | 0.36        | 0.40        | 0.33        | 0.36        | 14.22        |     |
| ADASYN + Boruta | <b>87.34</b> | 0.39        | 0.38        | 0.46        | 0.39        | <b>12.66</b> |     |
| SMOTE + RFE     | 86.97        | 0.39        | <b>0.40</b> | 0.40        | 0.63        | 13.03        |     |
| SMOTE + Boruta  | 85.78        | <b>0.39</b> | 0.38        | <b>0.47</b> | <b>0.69</b> | 14.22        | PC4 |
| ADASYN + RFE    | 89.51        | 0.63        | 0.67        | 0.62        | 0.88        | 10.49        |     |
| ADASYN + Boruta | <b>90.75</b> | 0.66        | <b>0.71</b> | <b>0.65</b> | <b>0.89</b> | <b>9.25</b>  |     |
| SMOTE + RFE     | 89.75        | 0.63        | 0.66        | 0.62        | 0.88        | 10.25        |     |
| SMOTE + Boruta  | 90.52        | <b>0.67</b> | 0.68        | 0.65        | 0.86        | 9.48         |     |

Table 3. Ablation study results for NASA datasets using four combinations techniques (10-fold CV).

| Technique Combination | Accuracy (%) | F1 Score    | Recall      | Precision   | ROC-AUC     | Error Rate (%) | Dataset     |
|-----------------------|--------------|-------------|-------------|-------------|-------------|----------------|-------------|
| ADASYN + RFE          | 84.52        | 0.28        | 0.22        | 0.39        | 0.64        | 15.48          | Ant-1.7     |
| ADASYN + Boruta       | <b>85.11</b> | 0.27        | 0.23        | <b>0.40</b> | <b>0.65</b> | <b>14.89</b>   |             |
| SMOTE + RFE           | 84.23        | 0.27        | 0.23        | 0.39        | 0.63        | 15.77          |             |
| SMOTE + Boruta        | 84.22        | <b>0.29</b> | <b>0.25</b> | 0.38        | 0.61        | 15.78          |             |
| ADASYN + RFE          | 85.53        | 0.19        | 0.15        | 0.28        | 0.53        | 14.47          | Camel-1.6   |
| ADASYN + Boruta       | 86.45        | 0.24        | 0.19        | <b>0.39</b> | 0.57        | 13.55          |             |
| SMOTE + RFE           | <b>86.79</b> | <b>0.25</b> | <b>0.19</b> | 0.37        | <b>0.60</b> | <b>13.21</b>   |             |
| SMOTE + Boruta        | 86.33        | 0.22        | 0.17        | 0.36        | 0.59        | 13.67          |             |
| ADASYN + RFE          | 72.67        | 0.51        | 0.57        | 0.48        | 0.74        | 27.33          | Synapse-1.2 |
| ADASYN + Boruta       | <b>73.93</b> | 0.59        | 0.65        | 0.55        | 0.76        | <b>26.07</b>   |             |
| SMOTE + RFE           | 72.81        | <b>0.60</b> | 0.65        | <b>0.61</b> | <b>0.77</b> | 27.19          |             |
| SMOTE + Boruta        | 72.79        | 0.59        | <b>0.66</b> | 0.60        | 0.75        | 27.21          |             |
| ADASYN + RFE          | 76.93        | <b>0.24</b> | <b>0.30</b> | <b>0.22</b> | 0.62        | 23.07          | Prop-6      |
| ADASYN + Boruta       | 75.73        | 0.20        | 0.26        | 0.18        | 0.59        | 24.27          |             |
| SMOTE + RFE           | <b>77.65</b> | 0.21        | 0.25        | 0.20        | <b>0.64</b> | <b>22.35</b>   |             |
| SMOTE + Boruta        | 74.29        | 0.20        | 0.29        | 0.17        | 0.57        | 25.71          |             |
| ADASYN + RFE          | 84.64        | 0.20        | <b>0.19</b> | 0.22        | 0.59        | 15.36          | Xalan-2.4   |
| ADASYN + Boruta       | 83.24        | 0.17        | 0.16        | 0.19        | 0.61        | 16.76          |             |
| SMOTE + RFE           | 85.11        | 0.18        | 0.15        | 0.22        | 0.61        | 14.89          |             |
| SMOTE + Boruta        | <b>85.69</b> | <b>0.21</b> | 0.16        | <b>0.27</b> | <b>0.63</b> | <b>14.31</b>   |             |

Table 4. Ablation study results for PROMISE datasets using four combinations techniques (10-fold CV).

| Technique Combination | Accuracy (%) | F1 Score    | Recall      | Precision   | ROC-AUC     | Error Rate (%) | Dataset |
|-----------------------|--------------|-------------|-------------|-------------|-------------|----------------|---------|
| ADASYN + RFE          | <b>83.58</b> | <b>0.85</b> | <b>0.89</b> | 0.81        | <b>0.87</b> | <b>16.42</b>   | JM1     |
| ADASYN + Boruta       | 81.68        | 0.83        | 0.87        | 0.79        | 0.86        | 18.32          |         |
| SMOTE + RFE           | 83.49        | 0.84        | 0.87        | <b>0.82</b> | 0.86        | 16.51          |         |
| SMOTE + Boruta        | 80.96        | 0.82        | 0.84        | 0.79        | 0.85        | 19.04          |         |
| ADASYN + RFE          | 83.82        | 0.83        | 0.76        | 0.90        | 0.88        | 16.18          | KC3     |
| ADASYN + Boruta       | 80.88        | 0.82        | 0.85        | 0.78        | 0.85        | 19.12          |         |
| SMOTE + RFE           | 83.82        | 0.85        | 0.91        | 0.79        | 0.91        | 16.18          |         |
| SMOTE + Boruta        | <b>92.65</b> | <b>0.93</b> | <b>0.91</b> | <b>0.94</b> | <b>0.93</b> | <b>7.35</b>    |         |
| ADASYN + RFE          | 80.00        | 0.81        | 0.88        | 0.75        | 0.80        | 20.00          | MC2     |
| ADASYN + Boruta       | 85.71        | 0.86        | 0.88        | 0.83        | 0.84        | 14.29          |         |

|                 |              |             |             |             |             |             |     |
|-----------------|--------------|-------------|-------------|-------------|-------------|-------------|-----|
| SMOTE + RFE     | <b>91.43</b> | <b>0.91</b> | <b>0.94</b> | <b>0.89</b> | <b>0.90</b> | <b>8.57</b> | MW1 |
| SMOTE + Boruta  | 80.00        | 0.80        | 0.82        | 0.78        | 0.86        | 20.00       |     |
| ADASYN + RFE    | 89.47        | 0.89        | 0.89        | 0.89        | 0.91        | 10.53       |     |
| ADASYN + Boruta | 91.58        | 0.91        | 0.89        | 0.93        | 0.94        | 8.42        |     |
| SMOTE + RFE     | 91.67        | 0.92        | <b>0.96</b> | 0.88        | 0.93        | 8.33        |     |
| SMOTE + Boruta  | <b>93.75</b> | <b>0.94</b> | 0.94        | <b>0.94</b> | <b>0.96</b> | <b>6.25</b> | PC4 |
| ADASYN + RFE    | 95.83        | 0.96        | 0.98        | 0.94        | 0.97        | 4.17        |     |
| ADASYN + Boruta | 96.25        | 0.96        | <b>0.99</b> | 0.94        | <b>0.98</b> | 3.75        |     |
| SMOTE + RFE     | 94.68        | 0.95        | 0.98        | 0.92        | 0.97        | 5.32        |     |
| SMOTE + Boruta  | <b>96.52</b> | <b>0.97</b> | 0.98        | <b>0.95</b> | 0.97        | <b>3.48</b> |     |

**Table 5.** Ablation study results for NASA datasets using four combinations techniques (70:30 split).

| Technique Combination | Accuracy (%) | F1 Score    | Recall      | Precision   | ROC-AUC     | Error Rate (%) | Dataset     |
|-----------------------|--------------|-------------|-------------|-------------|-------------|----------------|-------------|
| ADASYN + RFE          | 93.98        | 0.94        | 0.95        | 0.91        | 0.96        | 6.02           | Ant-1.7     |
| ADASYN + Boruta       | <b>94.58</b> | <b>0.95</b> | <b>0.95</b> | <b>0.94</b> | 0.96        | <b>5.42</b>    |             |
| SMOTE + RFE           | 93.29        | 0.93        | 0.95        | 0.92        | 0.95        | 6.71           |             |
| SMOTE + Boruta        | 94.51        | 0.95        | 0.95        | 0.94        | <b>0.97</b> | 5.49           |             |
| ADASYN + RFE          | <b>95.87</b> | <b>0.94</b> | <b>0.94</b> | <b>0.97</b> | <b>0.97</b> | <b>4.13</b>    | Camel-1.6   |
| ADASYN + Boruta       | 93.58        | 0.93        | 0.90        | 0.97        | 0.97        | 6.42           |             |
| SMOTE + RFE           | 93.15        | 0.93        | 0.92        | 0.94        | 0.95        | 6.85           |             |
| SMOTE + Boruta        | 93.61        | 0.94        | 0.93        | 0.94        | 0.96        | 6.39           |             |
| ADASYN + RFE          | 79.55        | 0.78        | 0.73        | 0.84        | 0.82        | 20.45          | Synapse-1.2 |
| ADASYN + Boruta       | <b>81.82</b> | <b>0.82</b> | <b>0.82</b> | 0.82        | <b>0.85</b> | <b>18.18</b>   |             |
| SMOTE + RFE           | 75.00        | 0.72        | 0.64        | 0.82        | 0.79        | 25.00          |             |
| SMOTE + Boruta        | <b>81.82</b> | 0.81        | 0.77        | <b>0.85</b> | 0.78        | 18.18          |             |
| ADASYN + RFE          | 83.33        | 0.82        | 0.76        | 0.89        | 0.90        | 16.67          | Prop-6      |
| ADASYN + Boruta       | 82.35        | 0.84        | 0.94        | 0.76        | 0.92        | 17.65          |             |
| SMOTE + RFE           | <b>91.18</b> | <b>0.91</b> | 0.90        | <b>0.92</b> | 0.93        | <b>8.82</b>    |             |
| SMOTE + Boruta        | 88.24        | 0.89        | <b>0.98</b> | 0.82        | <b>0.97</b> | 11.76          |             |
| ADASYN + RFE          | 89.94        | <b>0.91</b> | <b>0.98</b> | 0.85        | 0.93        | 10.06          | Xalan-2.4   |
| ADASYN + Boruta       | 89.94        | 0.90        | 0.94        | 0.87        | 0.93        | 10.06          |             |
| SMOTE + RFE           | <b>90.00</b> | 0.90        | 0.94        | <b>0.87</b> | 0.92        | <b>10.00</b>   |             |
| SMOTE + Boruta        | 89.41        | 0.90        | 0.94        | 0.86        | <b>0.93</b> | 10.59          |             |

**Table 6.** Ablation study results for PROMISE datasets using four combinations techniques (70:30 split).

Tables 5 and 6, present the results of the proposed stacked ensemble model on each of the ten datasets from the NASA and PROMISE repositories respectively, using 70:30 splits. For each dataset, we performed an ablation study by evaluating four pre-processing combinations: ADASYN + RFE, ADASYN + Boruta, SMOTE + RFE, and SMOTE + Boruta. These tables present the performance of different preprocessing combinations, allowing comparative evaluation across multiple metrics for each dataset. SMOTE-based techniques, SMOTE + Boruta and SMOTE + RFE, were superior to most datasets, and PC4 achieved the highest accuracy of 96.52% using SMOTE + Boruta. The SMOTE + RFE model performed better when calculating the accuracy, while the recall rate using SMOTE + Boruta reached 98%. This observation is also seen for MW1, where the SMOTE + RFE model produced a recall score of 96%, indicating the robustness of the framework. ADASYN-combination pairs also performed well, especially giving the higher results in JM1 (ADASYN + RFE) and Camel-1.6 (ADASYN + RFE with 95.87% accuracy). It is clear that SMOTE tends to perform better through balancing the synthetic samples generated, where ADASYN also works well on certain datasets. The ablation study shows that preprocessing choices are data-dependent, with SMOTE + Boruta outperformed overall. The 10-fold cross-validation technique was also used, and the results are illustrated in Tables 3 and 4. Compared to the 70:30 splits, the performance obtained using 10-fold CV was generally lower for several datasets, particularly the smaller ones. This difference can be due to the variability in data partitioning and the smaller training dataset size within each fold. While the 70:30 split produced comparatively higher results in the present study, the 10-fold cross-validation may provide a more balanced evaluation of model performance.

To provide an intuitive overview of the ablation study, Figure 4 illustrates the highest accuracy achieved for each dataset, selected from the four combinations of feature selection and class balance given in Tables 5 and 6 for each dataset where the SMOTE+Boruta combination technique worked best for PC4 (NASA datasets) and ADASYN+RFE for Camel-1.6 (PROMISE datasets). By showing which pre-processing methods worked best for each dataset, this combined overview allows an efficient evaluation of the effectiveness of different pre-processing methods. The final result for the particular dataset was determined taking the maximum accuracy achieved in each instance; this provided the basis for subsequent comparisons with the proposed framework. The highest performance was shown for PC4 (96.52%) with SMOTE + Boruta, followed by Camel-1.6 (95.87%) with ADASYN+RFE, implying that the SMOTE-based methods perform better compared to ADASYN for NASA datasets, while ADASYN performs similarly on selected PROMISE datasets. The horizontal bar chart represents the best performing datasets, with PC4 and Camel-1.6 shown in orange to emphasize the highest accuracies, while the remaining datasets are displayed in sky blue.

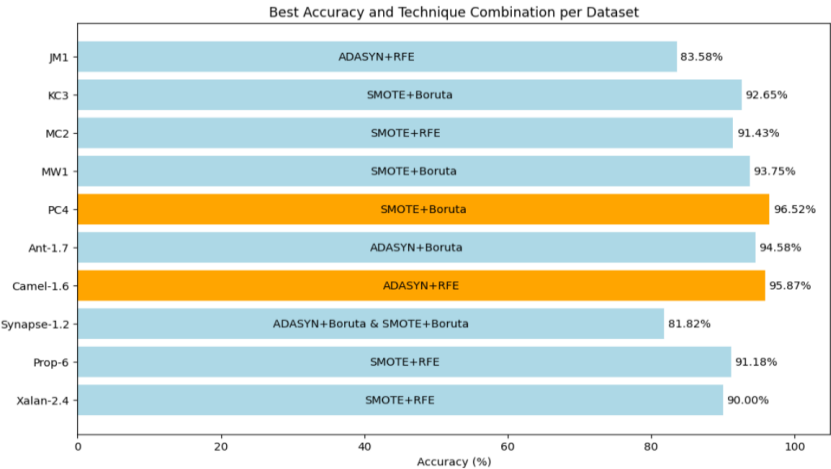


Figure 4. Comparative best Accuracy of the framework across Different Technique Combination.

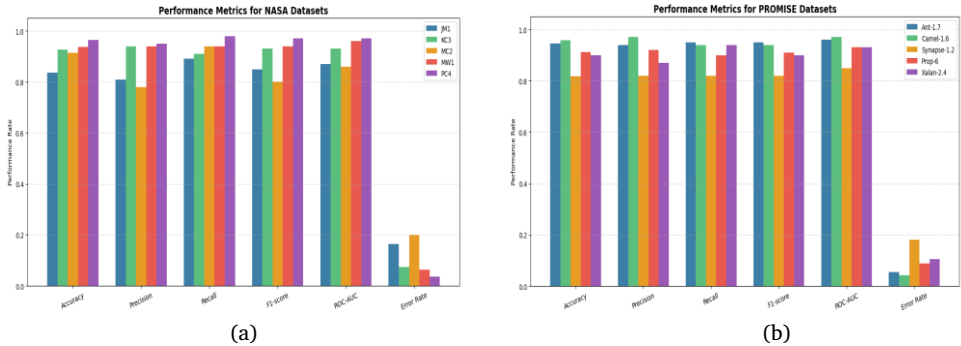


Figure 5. Performance Metrics of the Proposed Framework. (a) NASA Datasets (b) PROMISE Datasets.

To provide a comprehensive comparison of the proposed stacked ensemble model with both datasets, Figure 5 demonstrates the bar chart plots of relevant performance metrics, including Accuracy, Precision, Recall, F1-score, ROC-AUC, and Error Rate. From the figure, it is clear that the proposed stacked ensemble framework has proven its consistency by performing very well on NASA and PROMISE datasets, especially with respect to ROC-AUC and F1-score. Although some minor variations have been observed in precision and error rates, almost all other measures exhibit equally good performance on all datasets. For each dataset, the different combinations of preprocessing and feature selection techniques were evaluated, and the higher-performing combination was selected with respect to multi-metric evaluation. These top-performing results are displayed in Figure 5 for each dataset.

#### 4.2. Performance Analysis of Proposed Framework with Base Models

Next, we compare the performance of the proposed stacked ensemble model with ten baseline classifier models. Random Forest, XGBoost, CatBoost, Decision Tree, KNN, J48, AdaBoostM1, Naive Bayes, SVM, and Logistic Regression. The comparison indicates the effectiveness of the proposed method in predictive strength and generalization across different datasets. Additionally, we presented the results before and after pre-processing for the three base classifiers employed in the framework.

| Datasets    | Proposed Framework | Random Forest | XGBoost | CatBoost     | Decision Tree | KNN   | J48   | AdaBoostM1 | Naive Bayes | SVM   | Logistic Regression |
|-------------|--------------------|---------------|---------|--------------|---------------|-------|-------|------------|-------------|-------|---------------------|
| JM1         | <b>83.58</b>       | 82.77         | 80.89   | 79.93        | 72.88         | 71.47 | 73.08 | 73.08      | 55.62       | 57.16 | 62.84               |
| KC3         | <b>92.65</b>       | 88.24         | 89.71   | 89.71        | 79.41         | 61.76 | 86.76 | 86.76      | 60.29       | 54.41 | 80.88               |
| MC2         | <b>91.43</b>       | 82.86         | 74.29   | 82.86        | 68.57         | 57.14 | 71.43 | 71.43      | 65.71       | 66.67 | 65.71               |
| MW1         | <b>93.75</b>       | 92.19         | 90.53   | 91.58        | 86.32         | 69.47 | 88.42 | 88.42      | 82.11       | 77.89 | 93.68               |
| PC4         | <b>96.52</b>       | 94.79         | 95.21   | 95.63        | 90.83         | 68.13 | 90.62 | 90.62      | 75.83       | 59.17 | 86.88               |
| Ant-1.7     | <b>94.58</b>       | 94.58         | 93.29   | 94.18        | 87.20         | 79.27 | 89.02 | 89.02      | 70.73       | 70.73 | 75.61               |
| Camel-1.6   | <b>95.87</b>       | 93.61         | 93.15   | 94.06        | 89.95         | 72.60 | 89.04 | 89.04      | 63.01       | 54.34 | 74.89               |
| Synapse-1.2 | <b>81.82</b>       | 79.55         | 79.55   | 81.82        | 65.91         | 65.91 | 68.18 | 68.18      | 75.00       | 75.00 | 75.00               |
| Prop-6      | <b>91.18</b>       | 89.22         | 91.16   | 90.20        | 83.33         | 77.45 | 84.31 | 84.31      | 61.76       | 50.98 | 63.73               |
| Xalan-2.4   | 90.00              | 92.30         | 91.72   | <b>92.31</b> | 77.51         | 81.66 | 79.88 | 79.88      | 65.68       | 71.60 | 75.15               |

**Table 7.** Accuracy (%) comparison of the proposed framework and base classifiers for different datasets.

Table 8 presents a comparative analysis between the base classifiers of the performance of the proposed framework (Random Forest, XGBoost and CatBoost) in raw datasets and their performance after applying the proposed preprocessing pipeline, alongside the results of the proposed stacked ensemble model. The proposed stacked ensemble model is evaluated only after preprocessing, since the meta-learning stage was explicitly designed to work on balanced (ADASYN/SMOTE) and feature-refined (RFE/Boruta) model were used determined through ablation studies. The improvements are clear across almost all datasets. Significant increases in accuracy of 5.59% and 9.60% for JM1 and KC3 demonstrate that addressing class imbalance and removing irrelevant features result in better classification models. The most significant result is from the MC2 dataset, with 19.64% improvement, highlighting the role of preprocessing for strongly imbalanced datasets, one of the core contributions of this study. The only small drop is seen for Prop-6 (-0.22%), which is likely due to its already balanced class distribution, where the selection of features may have removed slightly useful predictors. This comparison is informative as it demonstrates why pre-processing should be done before stacking. It shows that the suggested preprocessing steps continually improve the performance of the given classifiers, justifying the strength of our preprocessing procedures and enabling the final stacked ensemble to produce more accurate and generalizable predictions on benchmark datasets.

| Datasets | After Preprocess   |               |         |          | Before Preprocess |         |          | Improvement |
|----------|--------------------|---------------|---------|----------|-------------------|---------|----------|-------------|
|          | Proposed Framework | Random Forest | XGBoost | CatBoost | Random Forest     | XGBoost | CatBoost |             |
| JM1      | <b>83.58</b>       | 82.77         | 80.89   | 79.93    | 77.25             | 76.98   | 77.99    | + 5.59%     |
| KC3      | <b>92.65</b>       | 88.24         | 89.71   | 89.71    | 74.58             | 83.05   | 81.36    | + 9.60%     |
| MC2      | <b>91.43</b>       | 82.86         | 74.29   | 82.86    | 65.79             | 71.79   | 68.42    | + 19.64%    |
| MW1      | <b>93.75</b>       | 92.19         | 90.53   | 91.58    | 91.74             | 90.08   | 91.74    | + 2.01%     |

|             |              |       |       |              |       |       |       |         |
|-------------|--------------|-------|-------|--------------|-------|-------|-------|---------|
| PC4         | <b>96.52</b> | 94.79 | 95.21 | 95.63        | 91.09 | 90.41 | 90.84 | + 5.43% |
| Ant-1.7     | <b>94.58</b> | 94.58 | 93.29 | 94.18        | 86.67 | 87.41 | 86.67 | + 7.17% |
| Camel-1.6   | <b>95.87</b> | 93.61 | 93.15 | 94.06        | 85.23 | 85.80 | 87.50 | + 8.37% |
| Synapse-1.2 | <b>81.82</b> | 79.55 | 79.55 | 81.82        | 78.72 | 76.60 | 80.85 | + 0.97% |
| Prop-6      | <b>91.18</b> | 89.22 | 91.16 | 90.20        | 89.92 | 91.40 | 90.10 | -0.22%  |
| Xalan-2.4   | 90.00        | 92.30 | 91.72 | <b>92.31</b> | 89.21 | 88.49 | 87.77 | + 0.79% |

**Table 8.** Accuracy (%) comparison of the stacked ensemble framework and base classifiers before and after preprocessing for different datasets.

Overall, the results are determinate that the use of feature selection and class balancing not only improves the performance of the basic classifiers but also strengthens the suggested stacked ensemble, which, after preprocessing, has the highest accuracy on nearly all of the datasets. This validates that the best accurate software defect prediction technique includes an adequate pre-processing step.

### 4.3. Comparative SOTA (state-of-the-art) Study with Proposed Framework

Finally, we perform a benchmarking against state-of-the-art (SOTA) methods reported in the literature on the same datasets. This demonstrates that our method not only performs competitively compared to individual machine learning models but also achieves promising results under the present experimental setup, relative to recently published approaches, indicating its potential for practical software defect prediction applications.

Table 9 compares the accuracy of our proposed stacked ensemble model with some existing work on the NASA and PROMISE datasets. Although the proposed study demonstrates competitive performance across all evaluated datasets, direct comparison with previous studies must be interpreted with caution due to the differences in methodology. Differences in dataset preprocessing, feature engineering, sampling strategies, and evaluation settings, etc. could affect the obtained results. Therefore, the comparative analysis intended to give a general indication of the performance level achieved. For the JM1 dataset, the existing works have reported accuracies ranging from 62.78% (Iqbal & Aftab, 2020) to 81.95% (Al-Fraihat et al., 2024), while the proposed framework achieves 83.58% under the present experimental setup. In the same way, on KC3, previous studies reported a highest accuracy of 86.84% (Dada et al., 2021), whereas our model achieves 92.65% in the current evaluation protocol. In MC2 and MW1, our proposed framework attains accuracies of 91.43% and 93.75%, respectively, which are competitive with or higher than previously reported results of 75.05% (Amit et al., 2024) and 89.33% (Ali et al., 2024).

| Study                     | JM1   | KC3   | MC2   | MW1   | PC4   | Ant-1.7 | Camel-1.6 | Synapse-1.2 | Prop-6 | Xalan-2.4 |
|---------------------------|-------|-------|-------|-------|-------|---------|-----------|-------------|--------|-----------|
| (Al-Fraihat et al., 2024) | 81.95 | -     | -     | -     | -     | -       | -         | -           | -      | -         |
| (Siva et al., 2023)       | 80.68 | -     | -     | -     | -     | 88.82   | 87.66     | -           | -      | 67.84     |
| (Ali et al., 2024)        | 79.12 | -     | 68.42 | 89.33 | 87.14 | -       | -         | -           | -      | -         |
| (Iqbal et al., 2019)      | 73.96 | -     | 64.86 | 82.66 | 86.08 | -       | -         | -           | -      | -         |
| (Dada et al., 2021)       | 81.11 | 86.84 | -     | -     | -     | -       | -         | -           | -      | -         |
| (Amit et al., 2024)       | -     | -     | 75.05 | -     | -     | -       | -         | -           | -      | -         |
| (Iqbal & Aftab, 2020)     | 62.78 | -     | 62.16 | 77.33 | 74.80 | -       | -         | -           | -      | -         |
| (Iqbal & Aftab, 2020)     | 80.68 | -     | -     | -     | -     | 88.82   | 87.66     | -           | -      | 67.84     |
| (Mcmurray & Sodhro, 2023) | -     | -     | -     | -     | -     | 92.06   | 92.22     | -           | -      | -         |
| (Abubakar et al., 2023)   | -     | -     | -     | -     | -     | 83.20   | 87.80     | -           | -      | -         |

|                           |              |              |              |              |              |              |              |              |              |              |
|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| (Alsghaier & Akour, 2020) | -            | 90.35        | 67.10        | 91.83        | 88.21        | -            | -            | 71.40        | -            | -            |
| <b>Proposed Framework</b> | <b>83.58</b> | <b>92.65</b> | <b>91.43</b> | <b>93.75</b> | <b>96.52</b> | <b>94.58</b> | <b>95.87</b> | <b>81.82</b> | <b>91.18</b> | <b>90.00</b> |

**Table 9.** Comparison of classification accuracy (%) of the proposed stacked ensemble framework with previous studies on NASA and PROMISE datasets.

In the case of PROMISE datasets, the same trend is observed. In Ant-1.7 and Camel-1.6, previous studies reported optimal accuracies of 88.82% (Siva et al., 2023) and 87.80% (Siva et al., 2024), while our model reaches 94.58% and 95.87%, respectively. In particular, for Prop-6, we were unable to find the values of published accuracy metrics in the reviewed studies, but other evaluation metrics were applied. Thus, our result fills this gap and can act as a baseline for future comparative studies. Our accuracy of 91.18% in Prop-6 is a good starting point for future studies. Generally, these results demonstrate that using a stacked ensemble not only improves prediction performance on individual datasets, but also generalizes well to a very diverse set of software projects, and our approach is a proper reference point for future work.

## 5. Conclusion & Future Work

This study presents a stacked ensemble based approach to software bug prediction (SBP), aiming to improve software quality and reduce development costs. The methodology combines the predictive strengths of Random Forest, XGBoost, and CatBoost as base classifiers, with a neural network meta-model to enhance decision-making. Extensive pre-processing steps, including data balancing using ADASYN, feature selection with RFE (Recursive Feature Elimination), were employed to ensure data quality. The model was evaluated on five NASA datasets and five PROMISE datasets, and the results demonstrate that the proposed ensemble approach outperforms individual classifiers in accuracy, precision, recall, and F1 score. The best performance is obtained for PC4 using SMOTE + Boruta at 96.52%, and Camel-1.6 using ADASYN + RFE at 95.87%. The results showed that the performance of the combination using SMOTE was higher than the performance of the combination using ADASYN for the NASA datasets, but the performance was comparable for the selected PROMISE datasets. Most noteworthy is the improvement obtained for the MC2 dataset, where the improvement is as high as 19.64%, thus emphasizing the importance of pre-processing for imbalanced datasets that is one of the major contribution of this work. Although the framework achieves promising results, further research can explore the integration of other advanced machine learning techniques, such as deep learning architectures or transformers, to enhance predictive performance. Additionally, testing the model on larger and more diverse datasets from various domains could generalize its applicability. Finally, incorporating real-time prediction mechanisms and explainable AI (XAI) techniques would increase the usability of the model in practical software development environments.

## References

- Jardas Anonic, J., Srok, A. & Vretenar, N. (2024). Blended Learning in the Spotlight of Educational Management – A Scientometric Analysis, 48(2), 279-294.
- Abubakar, H., Umar, K., Auwal, R., Muhammad, K., & Yusuf, L. (2023). Developing a machine learning-based software fault prediction model using the improved whale optimization algorithm. *Engineering Proceedings*, 56(1), 334. <https://doi.org/10.3390/ASEC2023-16307>.
- Albattah, W., & Alzahrani, M. (2024). Software defect prediction based on machine learning and deep learning techniques: An empirical approach. *AI*, 5(4), 1743–1758. <https://doi.org/10.3390/ai5040086>.
- Al-Fraihat, D., Sharrab, Y., Al-Ghuwairi, A. R., Alshishani, H., & Algarni, A. (2024). Hyper-parameter optimization for software bug prediction using ensemble learning. *IEEE Access*, 12, 51869–51878. <https://doi.org/10.1109/ACCESS.2024.3380024>.
- Ali, M., Mazhar, T., Arif, Y., Al-Otaibi, S., Ghadi, Y. Y., Shahzad, T., Khan, M. A., & Hamam, H. (2024). Software defect prediction using an intelligent ensemble-based model. *IEEE Access*, 12, 20376–20395.
- Alsghaier, H., & Akour, M. (2020). Software fault prediction using particle swarm algorithm with genetic algorithm and support vector machine classifier. *Software: Practice and Experience*, 50(4), 407–427. <https://doi.org/10.1002/spe.2784>.

- Amit, K. S., Aditya, K. N., Ankita, N., & Debasish, P. (2024). A NASA dataset based automated software bug prediction model using feature reduction techniques. *International Journal of Engineering Research & Technology (IJERT)*, 13(2), 1–6. <https://doi.org/10.17577/IJERTV13IS020071>.
- Balogun, A. O., Basri, S., Mahamad, S., Abdulkadir, S. J., Capretz, L. F., Imam, A. A., Almomani, M. A., Adeyemo, V. E., & Kumar, G. (2021). Empirical analysis of rank aggregation-based multi-filter feature selection methods in software defect prediction. *Electronics*, 10(2), 179. <https://doi.org/10.3390/electronics10020179>.
- Borandag, E. (2023). Software fault prediction using an RNN-based deep learning approach and ensemble machine learning techniques. *Applied Sciences*, 13(3), 1639. <https://doi.org/10.3390/app13031639>.
- Dada, E. G., Oyewola, D. O., Joseph, S. B., & Duada, A. B. (2021). Ensemble machine learning model for software defect prediction. *Advances in Machine Learning and Artificial Intelligence*, 2, 11–21.
- Das, H., Das, S., Gourisaria, M. K., Khan, S. B., Almusharraf, A., Alharbi, A. I., & Mahesh, T. R. (2024). Enhancing software fault prediction through feature selection with spider wasp optimization algorithm. *IEEE Access*, 12, 20376–20395. <https://doi.org/10.1109/ACCESS.2024.3435333>.
- Daza, A. (2025). Software defect prediction based on a multiclassifier with hyperparameters: Future work. *Results in Engineering*, 25, 104123. <https://doi.org/10.1016/j.rineng.2025.104123>.
- Ferenc, R., Gyimesi, P., Gyimesi, G., Tóth, Z., & Gyimóthy, T. (2020). An automatically created novel bug dataset and its validation in bug prediction. *Journal of Systems and Software*, 169, 110691.
- Gray, D., Bowes, D., Davey, N., Sun, Y., & Christianson, B. (2013). The misuse of the NASA metrics data program data sets for automated software defect prediction. In *Proceedings of the 15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2013)* (pp. 96–103). <https://doi.org/10.1049/ic.2011.0012>.
- Gupta, A., Sharma, S., Goyal, S., & Rashid, M. (2020). Novel XGBoost tuned machine learning model for software bug prediction. In *Proceedings of the 2020 International Conference on Intelligent Engineering and Management (ICIEM)* (pp. 376–380). <https://doi.org/10.1109/ICIEM48762.2020.9160152>.
- Hammouri, A., Hammad, M., Alnabhan, M., & Alsarayrah, F. (2018). Software bug prediction using machine learning approach. *International Journal of Advanced Computer Science and Applications*, 9(2), 78–83.
- Bala, Y. Z., Samat, P. A., Sharif, K. Y., & Manshor, N. (2022). Current software defect prediction: A systematic review. *Proceedings of the 2022 Applied Informatics International Conference (AiIC)*, 117–121. <https://doi.org/10.1109/AiIC54368.2022.9914586>.
- Iqbal, A., & Aftab, S. (2020). A classification framework for software defect prediction using multi-filter feature selection technique and MLP. *International Journal of Modern Education and Computer Science*, 12(1), 18–25. <https://doi.org/10.5815/ijmecs.2020.01.03>.
- Iqbal, A., Aftab, S., Ali, U., Nawaz, Z., Sana, L., Ahmad, M., & Husen, A. (2019). Performance analysis of machine learning techniques on software defect prediction using NASA datasets. *International Journal of Advanced Computer Science and Applications*, 10(5), 1–6. <https://doi.org/10.14569/IJACSA.2019.0100538>. 12(1), 18–25. <https://doi.org/10.5815/ijmecs.2020.01.03>.
- Khleel, N. A. A., & Nehéz, K. (2023). Overview of modern software bug prediction approaches. *Doktoranduszok Fóruma*, 55, 55–61.
- Mahfoodh, H., & Obediat, Q. (2020). Software risk estimation through bug reports analysis and bug-fix time predictions. In *Proceedings of the 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)* (pp. 1–6). <https://doi.org/10.1109/3ICT51146.2020.9312003>.
- Mcmurray, S., & Sodhro, A. H. (2023). A study on ML-based software defect detection for security traceability in smart healthcare applications. *Sensors*, 23(7), 3470. <https://doi.org/10.3390/s23073470>.
- Meenakshi, D., & Singh, S. (2019). Software bug prediction using machine learning approach. *International Journal of Advanced Computer Science and Applications*, 10(5), 1–6.
- Naidu, G. G., Balamurugan, M. S., & Savitha, M. S. (2022). A novel XGBoost tuned machine learning model for software bug prediction. *International Journal for Research Trends and Innovation*, 7(6), 764–768.
- Puranik, S., Deshpande, P., & Chandrasekaran, K. (2016). A novel machine learning approach for bug prediction. *Procedia Computer Science*, 93, 924–930. <https://doi.org/10.1016/j.procs.2016.07.271>.

Rizwan, S., Tiantian, W., Xiaohong, S., & Salahuddin. (2017). Empirical study on software bug prediction. In *Proceedings of the 2017 International Conference on Software and e-Business* (pp. 55–59). <https://doi.org/10.1145/3178212.3178221>.

Sharma, T., Jatain, A., Bhaskar, S., & Pabreja, K. (2023). Ensemble machine learning paradigms in software defect prediction. *Procedia Computer Science*, 218, 199–209. <https://doi.org/10.1016/j.procs.2023.01.002>.

Shin, J., Aleithan, R., Nam, J., Wang, J., & Wang, S. (2021). Explainable software defect prediction: Are we there yet? *arXiv*. <https://doi.org/10.48550/arXiv.2111.10901>.

Siva, R., Hariharan, B., & Premkumar, N. (2023). Automatic software bug prediction using adaptive artificial jelly optimization with long short-term memory. *Wireless Personal Communications*, 132(3), 1975–1998. <https://doi.org/10.1007/s11277-023-10694-9>.

Siva, R., Kaliraj, S., Hariharan, B., & Premkumar, N. (2024). Automatic software bug prediction using adaptive golden eagle optimizer with deep learning. *Multimedia Tools and Applications*, 83(1), 1261–1281. <https://doi.org/10.1007/s11042-023-16666-2>.

Tong, H., Lu, W., Xing, W., Liu, B., & Wang, S. (2022). SHSE: A subspace hybrid sampling ensemble method for software defect number prediction. *Information and Software Technology*, 142, 106747.

Wang, S., Wang, J., Nam, J., & Nagappan, N. (2021). Continuous software bug prediction. In *Proceedings of the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (pp. 1–12). <https://doi.org/10.1145/3475716.3475790>.

## Contact

Igor Balaban  
University of Zagreb  
Faculty of Organization and Informatics  
Pavlinka 2, HR-42000 Varaždin, Croatia  
Tel.: + 385 42 390 832  
Fax: + 385 42 213 413  
E-mail: igor.balaban@foi.hr  
E-mail: jios@foi.hr (Correspondence Mail)  
jios-pub@foi.hr (Publishing Board)  
URL: www.jios.foi.hr

JIOS is published with financial support of the Ministry of Science, Education and Youth of the Republic of Croatia.

## JIOS is indexed in the following repositories and portals

Academic Journals Database  
Academic Resource Index  
Advanced Science Index  
BASE  
Central and Eastern European Online Library (CEEOL)  
CORE  
Crossref  
CWTS Journal Indicators  
Dimensions  
Directory of Open Access Scholarly Resources (ROAD)  
Directory of Open Access Journals (DOAJ)  
Directory of Research Journals Indexing (DRJI)  
Elektronische Zeitschriftenbibliothek  
ERIH PLUS  
EuroPub – Directory of Academic and Scientific Journals  
Genamics JournalSeek  
Google Scholar  
Hrčak - Central portal of Croatian scientific journals  
Index Copernicus  
Internet Archive  
Information Matrix for the Analysis of Journals (MIAR)  
International Journal Impact Factor (IJIFACTOR)  
J-Gate  
JournalTOCs  
LENS.ORG  
Microsoft Academic  
OALster  
Open AIRE  
Open Academic Journals Index (OAJI.net)  
PRADEC  
Scientific Indexing Services (SIS)  
Scilit  
Sherpa Romeo  
Ulrich's Periodicals Directory  
VINITI RAN  
WorldCat  
World Journal Clout Index Report (WJCI)

[www.jios.foi.hr](http://www.jios.foi.hr)

## Proofreading

Authors

## Copyeditor and Computer Type Set

Goran Hajdin  
Mladen Konecki

## Cover and Layout Design

ri mo dizajn

## JIOS is indexed in the following databases

SCOPUS  
Emerging Sources Citation Index (ESCI) - WoS  
Academic Search Complete (EBSCO)  
Academic Search Elite (EBSCO)  
Academic Search Premier (EBSCO)  
Academic Search Ultimate (EBSCO)  
Aerospace Database - ProQuest  
Civil Engineering Abstracts - ProQuest  
Central & Eastern European Academic Source (CEEAS) - EBSCO  
Communication Abstracts - ProQuest  
EBSCO Essentials  
INSPEC  
Library and Information Science Abstracts (LISA) - ProQuest  
Library & Information Science Source - EBSCO  
Library, Information Science & Technology Abstracts (LISTA) - EBSCO  
Library Literature & Information Science - EBSCO  
Metadex – ProQuest

## Frequency

JIOS is published twice per year.

## Access to JIOS is free of charge through following sources:

Journal of Information and Organizational Sciences  
Hrčak - central portal of Croatian scientific journals  
Directory of Open Access Journals (DOAJ)  
OALster



## Creative Commons License

Journal of Information and Organizational Sciences by Faculty of Organization and Informatics is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.  
<http://creativecommons.org/licenses/by-nc-nd/4.0/>



## CONTENTS

|                                                                                                                                                                                                                                                                            |         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| <b>Mateja Gorenc and Janez Likozar:</b><br><b>Optimization of Batch Processing Through Advanced Management in Public Institution X</b><br>Original Scientific Paper                                                                                                        | 1-15    |
| <b>Ivana Petkovski and Petar Vranić:</b><br><b>Evaluation of Digital Development Indexes Using MEREC and Hybrid MCDM Methods</b><br>Original Scientific Paper                                                                                                              | 17-43   |
| <b>Raghad Baker Sadiq, Nur Fazidah Elias and Hazura Mohamed:</b><br><b>The Meta-analysis based Modified Unified Theory of Acceptance and Use of Technology Model: A Systematic Review</b><br>Original Scientific Paper                                                     | 45-64   |
| <b>Shofia Amin, Akhmad Habibi, Siti Hajar Halili, Hamdy Abdullah, Muthia Shahnaz, Rafiqi Rafiqi and Nela Safelia:</b><br><b>Factors Affecting Behavior, Perceived Impact of AI on Work Engagement, and AI Application: Indonesian Context</b><br>Original Scientific Paper | 65-82   |
| <b>Matija Novak and Marija Maček:</b><br><b>AI Tools in Programming Education: Student Perspectives and Usage Trends</b><br>Original Scientific Paper                                                                                                                      | 83-104  |
| <b>Yudhy Setyo Purwanto, Rahmat Gernowo and Dinar Mutiara Kusumo Nugraheni:</b><br><b>User Experience in Mobile Language Learning Applications (MLLA): Insight from Mixed-Methods Triangulation Study</b><br>Original Scientific Paper                                     | 105-127 |
| <b>Raghad Baker Sadiq, Nur Fazidah Elias and Hazura Mohamed:</b><br><b>Factors Affecting Lecturers' Intention to Adopt Artificial Intelligence Tools for Assessing Students' Performance in Malaysian Universities</b><br>Original Scientific Paper                        | 129-155 |
| <b>Matea Cvjetković:</b><br><b>Why Communication Dimensions Matter: Change Readiness Under Workplace Stressors In SMEs</b><br>Original Scientific Paper                                                                                                                    | 157-175 |
| <b>Vu Nguyen Huynh Anh, Nguyen Ton Nu Khoa and Vy Nguyen Dinh Nhat:</b><br><b>Institutional and Resource-Based Drivers of E-Commerce Adoption in Relationship-Embedded Wholesale Markets: A Business Model Component Analysis</b><br>Original Scientific Paper             | 177-193 |
| <b>Cemil Ceylan, Caner Tacoglu and Gozde Koc:</b><br><b>Consultant Competencies and ERP Project Success: The Mediating Role of ERP Implementation Critical Success Factors</b><br>Original Scientific Paper                                                                | 195-209 |
| <b>Candraningrat, Nimas A. Pratiwi and Azmil Chusnaini:</b><br><b>Smart Agriculture in Southeast Asia: Digitization of Smallholders through Organizational Commitment and Collective Ambition</b><br>Original Scientific Paper                                             | 211-232 |
| <b>Rosalina, Noor Lees Ismail, Genta Sahuri and Joseph Tedja Nugraha Wibawa:</b><br><b>Embedding Retrieval-Augmented Generation into SME Customer Communication Work Practices</b><br>Original Scientific Paper                                                            | 233-246 |
| <b>Sabrina Nasrin Saima, Abdus Salam, Mir Md. Kawsur and K. M. Imtiaz-Ud-Din:</b><br><b>An Advanced Stacked Ensemble Framework for Enhanced Software Bug Predictions</b><br>Original Scientific Paper                                                                      | 247-262 |