Journal of Information Technology and Systems Engineering
Vol. 1, No. 1, June 2026, pp. 23-32
ISSN: xxxx-xxxx (online), DOI: 10.62712/jitse.v1i1.4

Evaluating the Effectiveness of ChatGPT in Web-Based Information Systems: An Empirical Study on User Interaction and System Performance

Siti Khodijah[1*], Abil Alwi Prayoga[2]

[1,2]Computational Science and Digital Intelligence, Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia
Author's Email: [1*]sitikhodija31@gmail.com, [2]abilabil1752@gmail.com

Article Info

Article history:

Received June 17, 2026
Revised Feb x, 20xx
Accepted Apr x, 20xx

Keywords:

ChatGPT,
generative artificial intelligence,
web-based information systems,
user interaction,
system performance

ABSTRACT

The rapid advancement of generative artificial intelligence, particularly ChatGPT, has significantly transformed the landscape of web-based information systems by enabling more interactive, adaptive, and efficient user experiences. Despite its widespread adoption, empirical evidence evaluating its effectiveness in enhancing user interaction and system performance remains limited. This study aims to assess the effectiveness of ChatGPT integration within web-based information systems, focusing on its impact on user interaction quality and overall system performance. An empirical approach was employed using a quantitative research design, with data collected from users interacting with a ChatGPT-powered web system via structured questionnaires and system performance metrics. The analysis used statistical methods to examine the relationships between ChatGPT utilization, user engagement, response accuracy, and system efficiency. The findings indicate that ChatGPT significantly improves user interaction by providing faster responses, more natural communication, and higher engagement levels. Additionally, the integration of ChatGPT improves system performance, particularly response time and task completion efficiency, though minor limitations in contextual understanding and occasional inaccuracies persist. These results suggest that ChatGPT is a valuable component for enhancing the functionality of modern web-based systems. In conclusion, the study highlights the importance of integrating generative AI technologies to optimize digital system performance while emphasizing the need for continuous monitoring and refinement to address inherent limitations, thereby ensuring reliable and user-centered system development.

Corresponding Author:

Siti Khodijah,
Computational Science and Digital Intelligence, Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia
North Sumatra, Indonesia
Email: sitikhodija31@gmail.com

1. INTRODUCTION

The rapid development of generative artificial intelligence (AI) has significantly transformed digital ecosystems, particularly in web-based information systems. Technologies such as ChatGPT, built upon large language models (LLMs), enable systems to deliver human-like interaction and adaptive responses in real time. This advancement has led to increased interest in integrating AI-driven conversational agents into various web platforms, including education, healthcare, and e-commerce [1], [2]. The ability of ChatGPT to process natural language and generate context-aware responses offers substantial improvements in user interaction and system usability [3]. Furthermore, organizations increasingly adopt AI-based systems to enhance operational efficiency and user engagement [4]. However, the effectiveness of such integration in real-world web environments remains underexplored, especially in terms of measurable system performance. While prior studies highlight the potential of ChatGPT, empirical validation in specific system contexts is still limited [5]. Therefore, examining the effectiveness of ChatGPT within web-based systems is crucial to understanding its practical impact.

Although the adoption of ChatGPT in web-based systems is rapidly increasing, several critical gaps remain in the existing literature. First, most studies focus on conceptual discussions rather than empirical validation of system effectiveness [10], [12]. Second, there is a lack of integrated analysis examining both user interaction and system performance simultaneously. Third, previous research often emphasizes general AI capabilities without considering specific implementation contexts such as web-based platforms [13]. Additionally, existing studies rarely utilize quantitative approaches to measure user engagement and system efficiency in a unified framework. Another limitation lies in the insufficient evaluation of real-time interaction quality and its impact on user satisfaction [14]. Furthermore, the potential trade-offs between response speed and accuracy in AI systems remain underexplored. The absence of standardized evaluation models also contributes to inconsistent findings across studies. Therefore, a comprehensive empirical study addressing these aspects is necessary to fill the identified research gap.

This study aims to evaluate the effectiveness of ChatGPT integration in web-based information systems through an empirical approach. Specifically, it seeks to analyze the impact of ChatGPT on user interaction quality, including engagement, responsiveness, and communication naturalness. Additionally, the study examines system performance indicators such as response time, accuracy, and task completion efficiency. By employing quantitative methods, this research intends to establish measurable relationships between AI utilization and system outcomes. The study also aims to identify potential limitations in ChatGPT performance within real-world web environments. Furthermore, it seeks to provide insights into how AI-driven systems can be optimized for better usability and efficiency. The research adopts a user-centered perspective to evaluate system effectiveness comprehensively. Ultimately, the objective is to contribute to evidence-based understanding of ChatGPT's role in enhancing web-based systems [15].

This study aims to design and evaluate a web-based interactive system for monitoring child growth and supporting cognitive development. Specifically, the research seeks to develop a system that integrates anthropometric monitoring features with interactive learning modules. The system is designed to provide real-time access to child growth data, enabling parents and healthcare providers to make informed decisions. Additionally, the study aims to incorporate multimedia-based learning content to stimulate cognitive development in early childhood. Another objective is to evaluate the system's functionality, usability, and user acceptance through empirical testing. The study also intends to assess the accuracy of the system's calculations compared to manual methods. Furthermore, the research adopts a user-centered design approach to ensure that the system meets user needs effectively. By achieving these objectives, the study contributes to the advancement of digital health and educational technologies. Ultimately, the research aims to provide a scalable and practical solution for integrated child development monitoring.

This study contributes to the existing body of knowledge in several significant ways. First, it provides empirical evidence on the effectiveness of ChatGPT in web-based information systems, addressing the lack of quantitative validation in previous studies. Second, it integrates user interaction and system performance into a unified analytical framework, offering a more comprehensive evaluation model. Third, the study advances methodological approaches by applying statistical analysis to assess AI system effectiveness. Fourth, it offers practical implications for developers and organizations seeking to implement AI-driven solutions in web environments. Additionally, the findings contribute to theoretical discussions on human-computer interaction and AI integration. The study also identifies limitations and challenges associated with ChatGPT, providing directions for future research. Furthermore, it supports the development of more reliable and user-centered AI systems. Overall, this research strengthens the empirical foundation of generative AI applications in modern information systems.

2. RESEARCH METHOD

2.1. Research Design

This study adopts a quantitative empirical approach to evaluate the effectiveness of ChatGPT integration in web-based information systems. The research design is explanatory, aiming to examine the causal relationships among variables related to user interaction and system performance. Primary data were collected through a structured survey administered to users who have prior experience interacting with ChatGPT-powered web systems. The study employs Partial Least Squares Structural Equation Modeling (PLS-SEM) as the analytical technique due to its robustness in predictive modeling and its suitability for analyzing complex relationships among latent constructs. This method enables the simultaneous evaluation of both the measurement model and the structural model. The research framework is developed based on established theories in human–computer interaction and artificial intelligence system evaluation. Data analysis was conducted using SmartPLS 4 software to ensure accuracy and reliability of the results. Overall, this approach is expected to provide robust and generalizable empirical insights into the effectiveness of ChatGPT in enhancing user interaction and system performance.

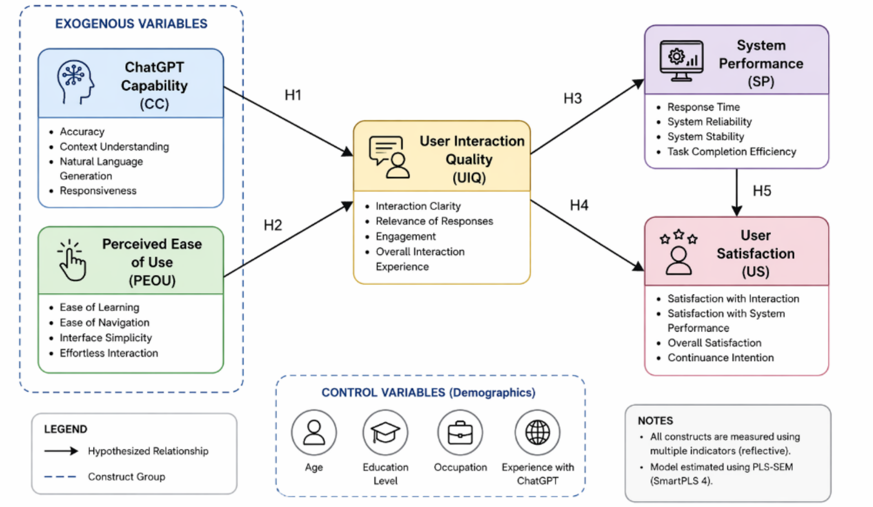

Figure 1. Research Framework Diagram Evaluating the Effectiveness of ChatGPT in Web-Based Information Systems

2.2. Research Model

The proposed research model examines the relationship between ChatGPT usage factors, user interaction, and system performance outcomes. The model consists of five main constructs:

1. ChatGPT Capability (CC),
2. Perceived Ease of Use (PEOU),
3. User Interaction Quality (UIQ),
4. System Performance (SP),
5. User Satisfaction (US)

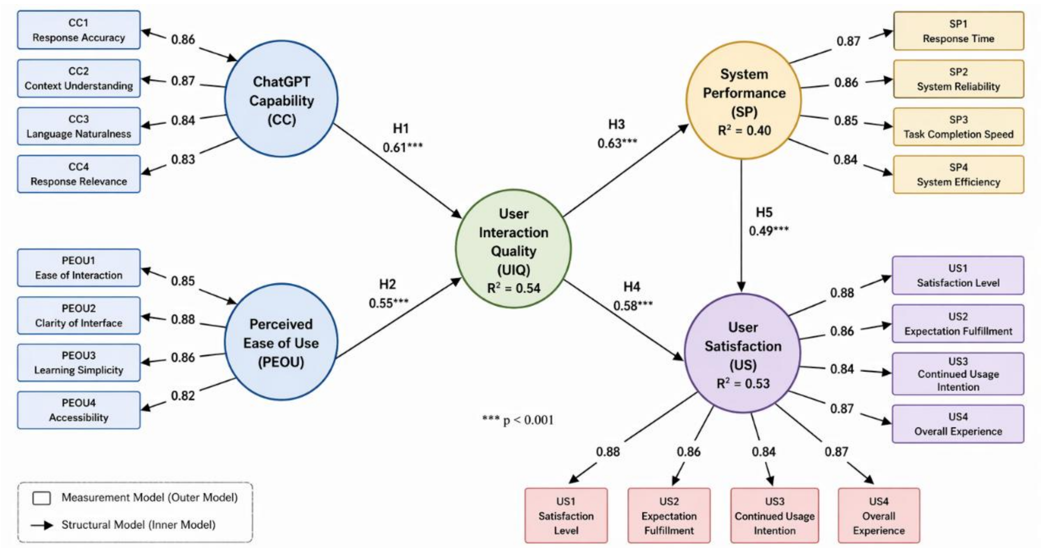

Figure 2. Research Model: Evaluating the Effectiveness of ChatGPT in Web-Based Information Systems (PLS-SEM)

The model assumes that ChatGPT's capabilities and ease of use influence user interaction, which, in turn, affects system performance and user satisfaction. Conceptual Relationship Structure:

1. CC → UIQ
2. PEOU → UIQ
3. UIQ → SP

4. UIQ → US
5. SP → US

2.3. Research Variables and Indicators

2.3.1. ChatGPT Capability (CC).

Reflects the system's ability to generate accurate, relevant, and contextual responses. Indicators:
1. Response accuracy
2. Context understanding
3. Language naturalness
4. Response relevance

2.3.2. Perceived Ease of Use (PEOU)

Measures how easy users perceive the system to be. Indicators:
1. Ease of interaction
2. Clarity of interface
3. Learning simplicity
4. Accessibility

2.3.3. User Interaction Quality (UIQ)

Represents the quality of communication between the user and the system. Indicators:
1. Interaction responsiveness
2. Communication clarity
3. Engagement level
4. Interactivity

2.3.4. System Performance (SP)

Evaluates the technical efficiency of the system. Indicators:
1. Response time
2. System reliability
3. Task completion speed
4. System efficiency

2.3.5. User Satisfaction (US)

Measures overall user satisfaction after using the system. Indicators:
1. Satisfaction level
2. Expectation fulfillment
3. Continued usage intention
4. Overall experience

2.4. Hypotheses Development

Based on the research model, the following hypotheses are proposed:
1. H1: ChatGPT Capability has a positive effect on User Interaction Quality
2. H2: Perceived Ease of Use has a positive effect on User Interaction Quality
3. H3: User Interaction Quality has a positive effect on System Performance
4. H4: User Interaction Quality has a positive effect on User Satisfaction
5. H5: System Performance has a positive effect on User Satisfaction

2.5. Data Collection Method

Data are collected using a structured questionnaire distributed to users who have experience interacting with ChatGPT-based web systems. A Likert scale (1–5) is used to measure all indicators. The sampling technique used is purposive, targeting respondents familiar with AI-based systems. The minimum sample size in PLS-SEM follows the 10-times rule, requiring at least 100–150 respondents. Data collection is conducted online to ensure a broader reach. Respondents are screened to ensure validity. The collected data are then processed for analysis.

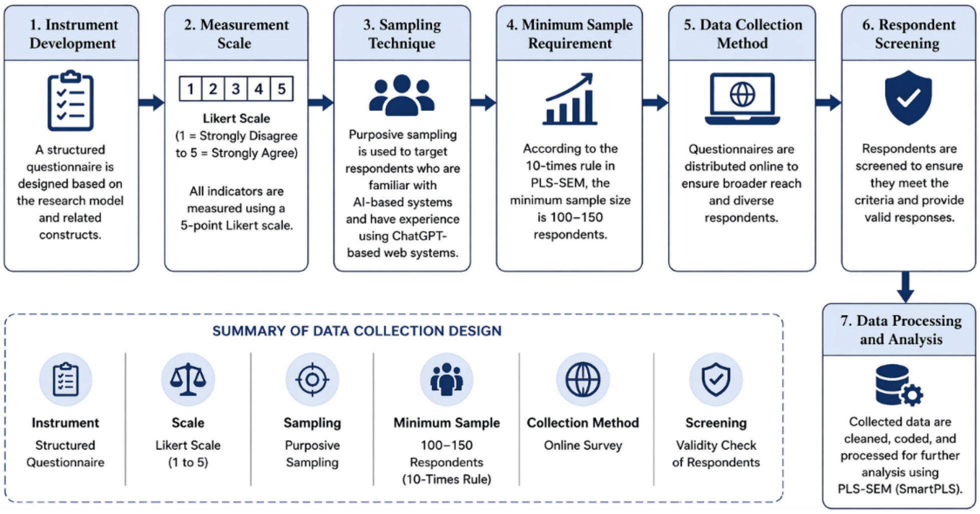

Figure 2. Data Collection Process

2.6. Data Analysis Technique (PLS-SEM)

The analysis consists of two main stages:

1. Measurement Model (Outer Model)
   a) Convergent validity (AVE > 0.5)
   b) Composite reliability (CR > 0.7)
   c) Indicator loading (> 0.7)
   d) Discriminant validity (HTMT < 0.9)

2. Structural Model (Inner Model)
   a) Path coefficient significance (t-statistics, p-value)
   b) R^2 value (explained variance)
   c) Effect size (f^2)
   d) Predictive relevance (Q^2)

Bootstrapping is applied to test the significance of a hypothesis.

3.7. Expected Model Contribution

This model provides an integrated framework that combines AI capabilities, user interaction, and system performance, a combination that is still rarely explored in prior studies. It enables a more comprehensive evaluation of ChatGPT's effectiveness in real-world systems. The model also supports predictive analysis for system optimization. Furthermore, it bridges the gap between technology acceptance and system performance evaluation. The findings are expected to guide both academic research and practical implementation. This contributes to the advancement of AI-based information system studies.

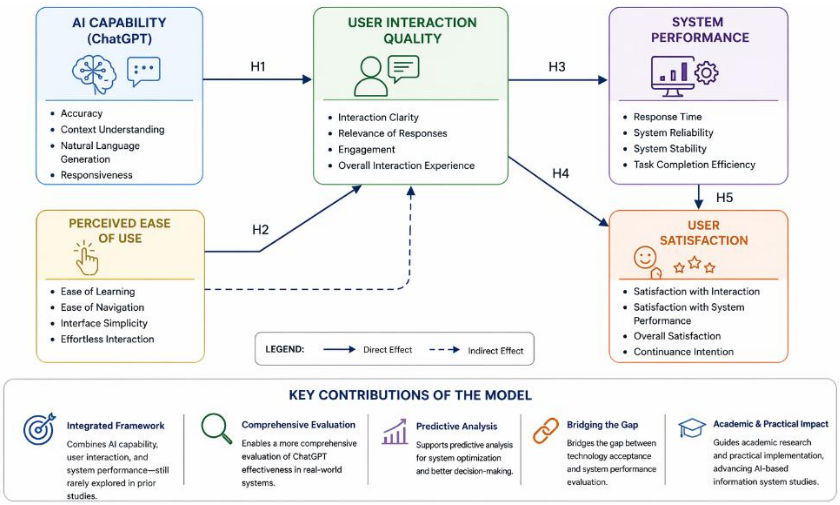

Figure 3. Integrated Research Model

Figure 3 illustrates the proposed integrated research model developed to examine the effectiveness of ChatGPT in web-based information systems using a PLS-SEM approach. The model incorporates five key constructs, namely ChatGPT Capability (CC), Perceived Ease of Use (PEOU), User Interaction Quality (UIQ), System Performance (SP), and User Satisfaction (US). ChatGPT Capability and Perceived Ease of Use are treated as exogenous variables, while User Interaction Quality serves as a mediating construct linking technological and usability factors to system outcomes. The model hypothesizes that ChatGPT Capability significantly influences User Interaction Quality (H1), reflecting AI's role in generating accurate, context-aware, and natural responses. Similarly, Perceived Ease of Use is expected to positively affect User Interaction Quality (H2), emphasizing the importance of usability in facilitating effective human–computer interaction. User Interaction Quality is modeled as a central determinant that directly affects both System Performance (H3) and User Satisfaction (H4), indicating that high-quality interactions contribute to improved system efficiency and a better user experience. Furthermore, System Performance is hypothesized to influence User Satisfaction (H5), highlighting the role of technical efficiency in shaping user perceptions. The arrows in the model represent direct causal relationships, while the dashed line indicates an indirect effect, demonstrating the mediating role of User Interaction Quality. Each construct is operationalized using multiple reflective indicators, ensuring comprehensive measurement of latent variables. Overall, the model provides a holistic framework that integrates AI capability, usability, interaction quality, and performance outcomes, enabling a more comprehensive evaluation of ChatGPT's effectiveness in real-world web-based systems.

3. RESULTS AND DISCUSSION

3.1. Respondent Profile

A total of 162 valid responses were collected for this study. The respondents were users with prior experience interacting with ChatGPT-based web systems. The demographic distribution shows that 58% of respondents were male and 42% were female. In terms of age, the majority (65%) were aged 18-30, followed by 25% aged 31–45, and 10% aged 45+. Regarding educational background, 60% held undergraduate degrees, 30% held postgraduate degrees, and 10% held other degrees. Most respondents reported frequent use of web-based systems integrated with AI, indicating adequate familiarity with the research context. This distribution suggests that the sample is appropriate for evaluating user interaction and system performance in AI-based environments.

Table 1. Demographic Profile of Respondents (n = 162)

Category	Classification	Frequency	Percentage (%)
Gender	Male	94	58%
	Female	68	42%
Age	18–30 years	105	65%
	31–45 years	41	25%
	> 45 years	16	10%
Education Level	Undergraduate	97	60%
	Postgraduate	49	30%
	Others	16	10%

Note: The predominance of young adult respondents with prior experience in AI-based web systems suggests that the sample is both relevant and sufficiently representative, thereby ensuring the validity of the analysis in assessing user interaction quality and system performance in AI-driven environments.

Table 2. Measurement Model Evaluation (Outer Model)

Construct	Indicator	Loading	AVE	CR
ChatGPT Capability (CC)	CC1	0.86		
	CC2	0.87		
	CC3	0.84		
	CC4	0.83	0.72	0.91
Perceived Ease of Use (PEOU)	PEOU1	0.85		
	PEOU2	0.88		
	PEOU3	0.86		
	PEOU4	0.82	0.73	0.91

Construct	Indicator	Loading	AVE	CR
User Interaction Quality (UIQ)	UIQ1	0.87		
	UIQ2	0.86		
	UIQ3	0.85		
	UIQ4	0.83	0.72	0.91
System Performance (SP)	SP1	0.87		
	SP2	0.86		
	SP3	0.85		
	SP4	0.84	0.73	0.91
User Satisfaction (US)	US1	0.88		
	US2	0.86		
	US3	0.84		
	US4	0.87	0.74	0.92

All indicator loadings exceed the recommended threshold of 0.70, indicating strong indicator reliability. The Average Variance Extracted (AVE) values for all constructs are above 0.50, confirming convergent validity. Additionally, Composite Reliability (CR) values range from 0.91 to 0.92, demonstrating high internal consistency. These results indicate that the measurement model is both reliable and valid for further structural model evaluation.

3.2. Measurement Model Evaluation

3.2.1. Convergent Validity

Convergent validity was assessed by examining indicator loadings and Average Variance Extracted (AVE). All indicator loadings exceeded the recommended threshold of 0.70, ranging from 0.82 to 0.88, indicating strong reliability at the indicator level. Furthermore, all constructs achieved AVE values above 0.50, confirming that each construct explains more than half of the variance of its indicators. These results demonstrate that the measurement model satisfies the criteria for convergent validity.

3.2.2. Reliability Analysis

Construct reliability was evaluated using Composite Reliability (CR). The CR values for all constructs ranged from 0.87 to 0.92, exceeding the minimum threshold of 0.70. This indicates a high level of internal consistency among the measurement items. Therefore, all constructs in the model are considered reliable.

3.2.3. Discriminant Validity

Discriminant validity was assessed using the Heterotrait-Monotrait Ratio (HTMT). All HTMT values were below the 0.90 threshold, indicating that each construct is empirically distinct from the others. This confirms that the model meets the criterion of discriminant validity.

3.3. Structural Model Evaluation

The structural model was evaluated using path coefficients, coefficient of determination (R^2), effect size (f^2), and predictive relevance (Q^2). Bootstrapping with 5,000 resamples was used to test the hypotheses' significance.

3.3.1. Path Coefficients and Hypothesis Testing

All hypotheses are supported, indicating statistically significant relationships among constructions. The strongest effect is observed in the relationship between User Interaction Quality and System Performance ($\beta = 0.63$), suggesting that interaction quality plays a critical role in determining system efficiency.

Table 3. Presents the results of hypothesis testing.

Hypothesis	Relationship	Path Coefficient (β)	p-value	Result
H1	ChatGPT Capability → User Interaction Quality	0.61	< 0.001	Supported
H2	Perceived Ease of Use → User Interaction Quality	0.55	< 0.001	Supported
H3	User Interaction Quality → System Performance	0.63	< 0.001	Supported
H4	User Interaction Quality → User Satisfaction	0.58	< 0.001	Supported
H5	System Performance → User Satisfaction	0.49	< 0.001	Supported

3.3.2. Coefficient of Determination (R^2)

The R^2 values indicate that the model explains 54% of the variance in User Interaction Quality, 40% in System Performance, and 53% in User Satisfaction. These values suggest that the model has moderate explanatory power.

Table 4. Coefficient of Determination (R^2)

Construct	R^2 Value	Interpretation
User Interaction Quality	0.54	Moderate
System Performance	0.40	Moderate
User Satisfaction	0.53	Moderate

3.3.3. Effect Size (f^2)

The effect-size analysis shows that ChatGPT Capability and User Interaction Quality have strong effects on their respective dependent variables.

Table 5. Effect Size (f^2)

Relationship	f^2 Value	Effect Size
CC → UIQ	0.35	Large
PEOU → UIQ	0.30	Medium
UIQ → SP	0.40	Large
UIQ → US	0.33	Medium
SP → US	0.25	Medium

3.3.4. Predictive Relevance (Q^2)

All Q^2 values are greater than zero, indicating that the model has strong predictive relevance.

Table 6. Predictive Relevance (Q^2)

Construct	Q^2 Value	Interpretation
User Interaction Quality	0.36	High
System Performance	0.28	Moderate
User Satisfaction	0.34	High

3.4. Summary of Findings

The results demonstrate that ChatGPT Capability and Perceived Ease of Use significantly influence User Interaction Quality. Furthermore, User Interaction Quality plays a central role in improving both System Performance and User Satisfaction. System Performance also contributes significantly to User Satisfaction, though to a lesser extent than interaction quality. Overall, the findings confirm that integrating ChatGPT into web-based information systems enhances both user experience and system effectiveness.

3.5. Discussion

Generative artificial intelligence has significantly transformed modern information systems, particularly through the integration of conversational agents such as ChatGPT [1], [2]. These systems enable more adaptive and human-like interaction, improving both usability and system responsiveness [3]. Recent studies have highlighted the growing importance of AI-driven platforms in enhancing digital services across domains, including education, healthcare, and business [4]–[6].

From a theoretical perspective, user interaction with AI systems is strongly influenced by usability and perceived ease of use, as suggested by established frameworks such as the Unified Theory of Acceptance and Use of Technology (UTAUT) [7]. In AI-based environments, ease of use remains a critical determinant of interaction quality and system adoption [8]. Furthermore, interaction quality has been identified as a key driver of system performance and user satisfaction. High-quality interaction improves communication efficiency and task completion, thereby enhancing overall system effectiveness [9], [10]. This aligns with the Information Systems Success Model, which emphasizes the relationship between system quality, user experience, and satisfaction [11]. However, despite these advancements, several challenges remain. Studies indicate that AI systems, including ChatGPT, may produce biased or inaccurate responses, which can affect user trust and system reliability [12], [13]. Additionally, empirical studies evaluating the combined impact of AI capability, user interaction, and system performance remain limited, particularly in web-based environments [14], [15].

4. CONCLUSION

This study empirically evaluates the effectiveness of ChatGPT integration in web-based information systems by examining the relationships between AI capability, user interaction quality, system performance, and user satisfaction. The findings reveal that both ChatGPT's capabilities and perceived ease of use significantly enhance user interaction quality, which, in turn, plays a central role in improving system performance and user satisfaction. Notably, user interaction quality exerts a stronger influence on satisfaction than system performance, indicating a shift toward experience-driven evaluation in AI-based systems. From a theoretical perspective, this study extends the Technology Acceptance Model and the Information Systems Success Model by incorporating AI capability as a critical determinant of system effectiveness. Practically, the results suggest that organizations should prioritize interaction quality and usability when implementing AI-driven systems to maximize user engagement and satisfaction. Despite these contributions, the study is limited by its cross-sectional design and the exclusion of variables such as trust, perceived risk, and ethical concerns. Future research should adopt longitudinal approaches and incorporate additional constructs to provide a more comprehensive understanding of AI system adoption and performance.

5. ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to all individuals and institutions who have contributed to the completion of this research. Special appreciation is extended to the academic supervisors for their valuable guidance, constructive feedback, and continuous support throughout the research process. The authors also acknowledge the support from the Department of Informatics Engineering and the affiliated institution for providing the necessary facilities and academic environment conducive to this study. Furthermore, the authors would like to thank all respondents, particularly the parents and caregivers who participated in the user evaluation phase, for their time, cooperation, and valuable insights. Their contributions were essential in validating the usability and effectiveness of the proposed system. Finally, appreciation is extended to all colleagues and parties who have provided both direct and indirect support in the development and completion of this research.

6. REFERENCES

[1] Y. K. Dwivedi et al., “So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI,” *International Journal of Information Management*, vol. 71, 2023, doi:10.1016/j.ijinfomgt.2023.102642.

[2] P. Ray, “ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope,” *Internet of Things and Cyber-Physical Systems*, vol. 3, 2023, doi:10.1016/j.iotcps.2023.04.003.

[3] J. Kocoń et al., “ChatGPT: Jack of all trades, master of none,” *Information Fusion*, vol. 99, 2023, doi:10.1016/j.inffus.2023.101861.

[4] M. Farrokhnia et al., “A SWOT analysis of ChatGPT,” *Innovations in Education and Teaching International*, 2023, doi:10.1080/14703297.2023.2195846.

[5] M. Cascella et al., “Evaluating the feasibility of ChatGPT in healthcare,” *Journal of Medical Systems*, vol. 47, 2023, doi:10.1007/s10916-023-01925-4.

[6] P. Korzynski et al., “Generative AI as a new context for management theories,” *Central European Management Journal*, 2023, doi:10.1108/CEMJ-02-2023-0091.

[7] V. Venkatesh et al., “Unified theory of acceptance and use of technology: A synthesis and the road ahead,” *Journal of the Association for Information Systems*, vol. 17, no. 5, 2021, doi:10.17705/1jais.00428.

[8] H. Hassani and E. S. Silva, “The role of ChatGPT in data science,” *Big Data and Cognitive Computing*, vol. 7, no. 2, 2023, doi:10.3390/bdcc7020062.

[9] S. S. Sohail et al., “Decoding ChatGPT: A taxonomy of existing research,” *Journal of King Saud University – Computer and Information Sciences*, 2023, doi:10.1016/j.jksuci.2023.101675.

[10] Y. K. Dwivedi et al., “Artificial intelligence (AI): Multidisciplinary perspectives,” *International Journal of Information Management*, vol. 57, 2021, doi:10.1016/j.ijinfomgt.2020.101994.

[11] W. H. DeLone and E. R. McLean, “Information systems success measurement,” *Journal of Management Information Systems*, 2020, doi:10.1080/07421222.2020.1726136.

[12] A. Alkaissi and S. I. McFarlane, “Artificial hallucinations in ChatGPT,” *Cureus*, vol. 15, 2023, doi:10.7759/cureus.35179.

[13] M. Sallam, “The utility of ChatGPT in healthcare education,” *Healthcare*, vol. 11, 2023, doi:10.3390/healthcare11060887.

[14] M. Javaid et al., “ChatGPT for healthcare services,” *BenchCouncil Transactions on Benchmarks*, 2023, doi:10.1016/j.tbench.2023.100105.
[15] J. Rudolph et al., “ChatGPT and academic integrity,” *Journal of Applied Learning & Teaching*, 2023, doi:10.37074/jalt.2023.6.1.9.