J@TI JURNAL TEKNIK INDUSTRI UNIVERSITAS DIPONEGORO

Vol. 19, No. 3, September 2024

THE INFLUENCE OF DIFFERENT TYPES OF TASKS AND USER EXPERIENCE LEVELS ON HUMAN-COMPUTER TRUST IN THE CHATGPT APPLICATION

Sigit Rahmat Rizalmi*[1], Vridayani Anggi Leksono[1], Abdul Alimul Karim[1], Syarifah Chairunnisaa[1], Putri Gesan Prabawa Anwar[2]

[1]Program Studi Teknik Industri, Jurusan Teknologi Industri dan Proses, Institut Teknologi Kalimantan, Jl. Soekarno-Hatta KM. 15, Balikpapan, Indonesia 76127
[2]Program Studi Teknik Logistik, Jurusan Teknologi Industri dan Proses, Institut Teknologi Kalimantan, Jl. Soekarno-Hatta KM. 15, Balikpapan, Indonesia 76127

Abstract

This study investigates the role of user experience and task type on Human-Computer Trust in the ChatGPT system. The method used in this research is an experiment where respondents are asked to do mathematics, descriptive, translation, and programming tasks on the ChatGPT application. After completing each task given, respondents were asked to fill out the trust questionnaire that had been prepared. Respondents in this study were divided into two categories: novice and expert. The total number of respondents in this study was 32 respondents. The result of this research is that the level of user experience significantly influences human-computer trust in the ChatGPT application in Task 1, Task 2, Task 3, and Task 4. However, there is no significant influence from differences in the types of tasks on human-computer trust in the ChatGPT application.

Keywords: *Human-Computer Trust; ChatGPT; Artificial Intelligence; Type of Task; User Experience*

1. Introduction

Intelligent software and hardware, commonly known as intelligent agents, are increasingly integrated into everyday life due to the increasing use of artificial intelligence (AI). These intelligent agents can perform various tasks, from simple manual work to complex operations. Chatbots are one of the most common examples of AI systems (Ashfaq et al., 2020). They are also the most widespread basic example of Human-Computer Interaction (HCI) using AI systems (Bansal et al., 2018). Although chatbots can simulate human conversation and provide entertainment, the purpose of chatbots is more than that (Iku-Silan et. al., 2023). Chatbots can be used in various fields, such as education, information search, business, and e-commerce (Shawar et al., 2007).

OpenAI is an AI company that launched ChatGPT, a new chatbot version. ChatGPT is a Large Language Model (LLM) that uses machine learning to learn from large text datasets and can produce very sophisticated and intelligent writing. This innovative technology significantly impacts science and society (Van Dis et al., 2023). ChatGPT was launched on November 30, 2022, and has since attracted significant attention; in the first week of launch, more than one million customers used it (Baidoo-Anu et. al., 2023), and to date, there are more than 100 million active ChatGPT users (Baidoo-Anu et al., 2023). With the rapid increase in the use of ChatGPT, especially among pupils and students, ChatGPT is often used to complete academic assignments and help understand the material.

In short, the ChatGPT function shows various ways that can be used to support and complete an activity. General tasks that can be used include (1) providing a summary of texts and literature, (2) providing explanations, clarifications, and additional information regarding the topic you want to know, (3) providing machine translation capabilities in various languages, (4) analyzing large amounts of data. (Frieder et al., 2024; George et al., 2023; Haleem et al., 2022; Kocoń et al., 2023). There are many features or types of tasks that ChatGPT can carry out; there are fundamental things to pay attention to in technological developments, namely, the level of trust.

Trust refers to an individual's belief that a technology or system can be relied upon to work as intended and protect their interests (Falcone et al., 2001; Kesharwani et al., 2012). Trust is an essential factor influencing individual behavior in technology

*Corresponding Author

E-mail: sigit.rahmat@lecturer.itk.ac.id

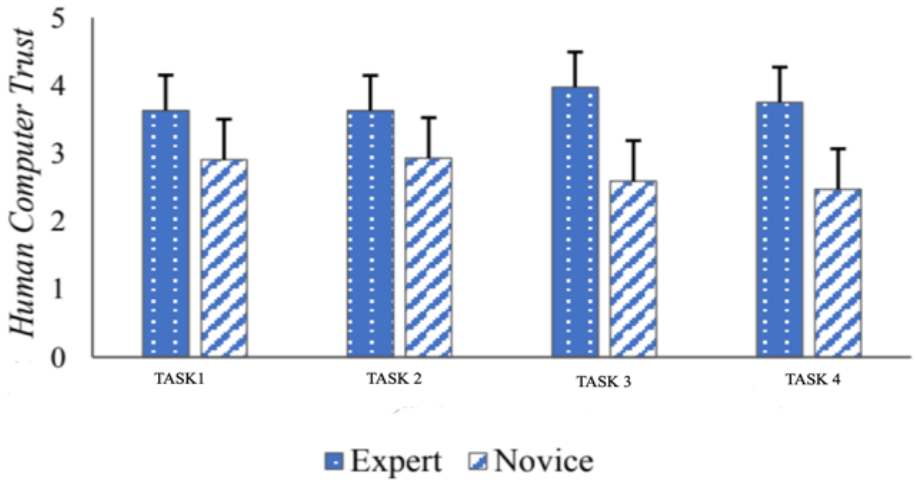

Figure 1. Human-Computer Trust (HCT) Score

use (Kim et al., 2016). If users trust a system to provide accurate and valuable information, users will develop a positive attitude towards the system. So, in the end, it positively impacts purchasing and revisiting the system (Limbu et al., 2012). According to Falcone et al. (2001), trust is foundational for individuals to delegate tasks to AI systems. Trust is necessary for users to engage with or rely on these systems, even if they are technically capable. In the HCI context, trust bridges the gap between the complexity of AI systems and the user's willingness to adopt them. Studies like Kim and Gambino (2016) highlight that trust influences attitudes toward technology, shaping user behavior such as continued use and task reliance.

Several previous studies have examined human-computer interaction related to trust in AI systems. However, several previous studies focused more on the characteristics of conversations in AI systems (for example, tone of voice, voice accent, and gender) (Chang et al., 2022; Edward et al., 2019; Niculescu et al., 2013; Tamagawa et al., 2011; Tore et al., 2020). Several studies also discuss listening to conversations in AI systems when testing user perception (Lortie et al., 2011). No research examines the differences in the types of tasks that can be carried out on AI systems and the level of user experience (novice and expert), which is a gap or update in this research. This research aims to determine the influence of different types of tasks on Human-Computer Trust and the influence of user experience level on Human-Computer Trust. Cognitive ergonomics and statistical approaches will be used to answer the research objectives.

2. Methods

2.1 Respondents

The study had 32 respondents. Respondents were divided into novice and expert categories based on their experience level using ChatGPT. The expert category is determined based on the amount of time spent using ChatGPT, with a minimum usage of 4 hours in one week. Meanwhile, the novice category includes respondents who have never used ChatGPT. The number of respondents in the novice category was 16, and the other 16 were in the expert category. All respondents in this study were active students at the Kalimantan Institute of Technology, with an average age of 21.28 years (± 0.29).

2.2 Research Procedures

This research uses a computer device to operate the ChatGPT system. Respondents will be given instructions regarding the scenario that will be carried out. Respondents were asked to do four tasks: mathematics, programming, descriptive, and translation. After completing the task and finding answers via ChatGPT, respondents were asked to fill out the Human-Computer Trust questionnaire. This questionnaire was developed by Madsen and Gregor (2000) to measure respondents' level of trust. This questionnaire consists of perceived understandability, perceived technical competence, perceived understandability of faith, personal attachment, and risk perception. This questionnaire will contain 30 questions about the value of trust in carrying out types of tasks. Each task will include 1 question and given a scale statement (1) strongly disagree (2) disagree (3), neutral (4), agree (5), strongly agree. Reliability testing yielded a Cronbach's alpha score of 0.91, confirming high consistency. The validation process included pilot testing with 10 participants.

2.3 Statistical Data Analysis

The Independent Sample T-Test was used to determine the effect of differences in user experience levels, namely expert and novice respondents because the data was normally distributed. Meanwhile, the Repeated Measure ANOVA test was used to determine the differences in the four types of tasks. The significant level is determined if $p < 0.05$. Testing uses JASP 0.17.3.

3. Results and Discussion

3.1 Human Computer Trust score

The human-computer trust (HCT) score is based on measurements of perceived reliability, technical competence, perceived understandability, personal attachment, faith, and risk perception. **Figure 1** shows

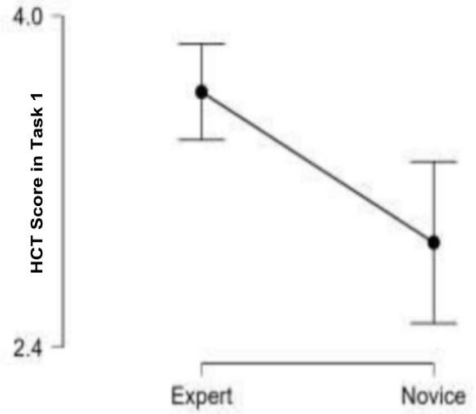

Figure 2. Human-Computer Trust Score in Task 1

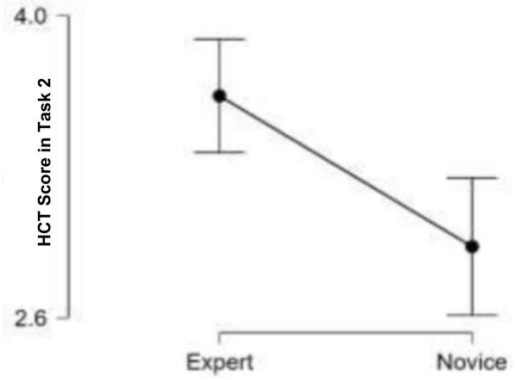

Figure 3. Human-Computer Trust Score in Task 2

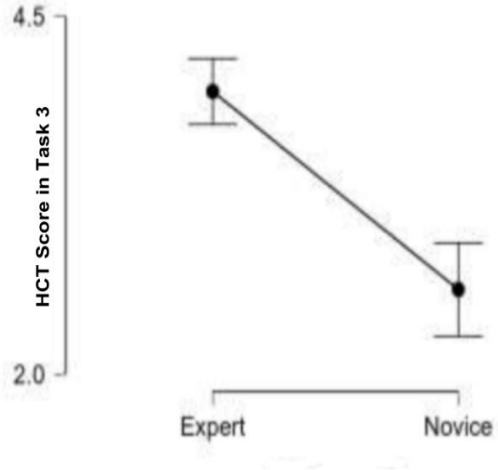

Figure 4. Human-Computer Trust Score in Task 3

the average human-computer trust scores for the types of tasks for expert and novice users.

Figure 1. shows the comparison of the average Human Computer Trust scores on Task 1 (Mathematics), Task 2 (Descriptive), Task 3 (English Translation), and Task 4 (Programming) for expert users and novice users. The HCT value in task 1 for expert users has an average score of 3.632, which is higher than that of novice users, who have an average score of 2.906. The HCT value in task 2 for expert users has an average score of 3.627, higher than novice users with an average score of 2.929. The HCT value in task 3 for expert users has an average score of 3.973, higher than novice users with an average score of 2.591. The HCT value in task 4 for expert users has an average score of more than 3,750, higher than novice users with an average score of 2.471.

3.2 Task 1

Task 1 is to do a composition function mathematics assignment. The human-computer trust score in task 1 is presented in **Figure 2.** Human-computer trust scores on the level of user experience in task 1 differ between expert and novice users. Expert users give higher trust scores than novice users. The independent sample t-test shows that in task 1, There is a significant difference in expert and novice use of the Chat GPT application with a P-value of 0.002 and Cohen's d 1.208, which explains that there is a relatively significant influence on the level of user experience on mathematics tasks.

3.3 Task 2

Task 2 is a descriptive task looking for information explaining technological developments. Human-computer trust score in task 2 is presented in **Figure 3.** Human-computer trust scores on the level of user experience in task 2 differ between expert and novice users. Expert users give higher trust scores than novice users. Independent samples t-test shows a significant difference between expert and novice use of the Chat GPT application, with a P-value of 0.001 and Cohen's d of 1.276, which explains that there is a relatively significant influence on the level of user experience on descriptive tasks.

3.4 Task 3

Task 3 is to do an English translation assignment. Human-computer trust scores in task 3 are presented in **Figure 4.** Human-computer trust scores on

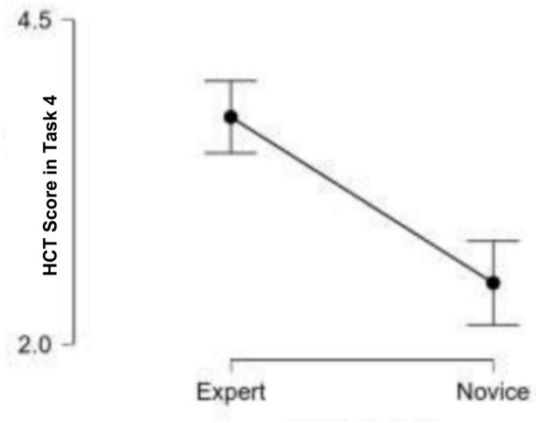

Figure 5. Human-Computer Trust Score in Task 4

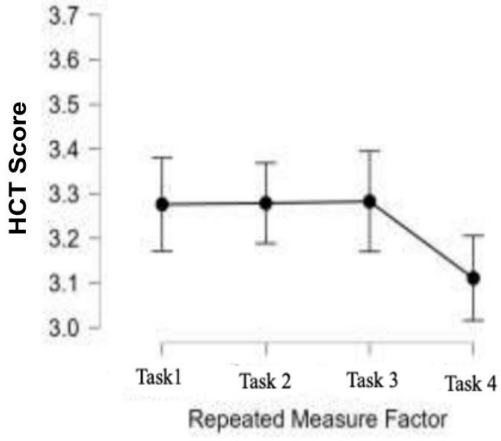

Figure 6. Human-Computer Trust Scores for Each Type of Task

the level of user experience in task 3 differ between expert and novice users. Expert users give higher trust scores than novice users. Independent samples t-test shows that there is a significant difference in expert and novice use of the Chat GPT application with a P-value <0.001 and Cohen's d of 2.614, which explains that there is a relatively significant influence on the level of user experience on English translation assignments

3.5 Task 4

Task 4 is to do programming assignments in Python. Human-computer trust scores in task 4 are presented in **Figure 5.** Human-computer trust scores on the level of user experience in task 4 differ between expert and novice users. Expert users give higher trust scores than novice users. Independent samples t-test shows significant differences in expert and novice use of the GPT Chat application with a P-value <0.001 and Cohen's d of 2.269, which explains that there is a significant influence on the level of user experience on programming tasks in Python.

The findings indicate that user experience significantly shapes trust towards ChatGPT, with expert users consistently showing higher trust scores across all task types. This may be attributed to their familiarity with the system's capabilities and limitations, likely reducing uncertainty and enhancing confidence in its outputs. Conversely, novice users may exhibit lower trust due to unfamiliarity or a lack of understanding of ChatGPT's potential and scope.

This aligns with previous research by Kim and Gambino (2016), highlighting that user experience in interacting with technology enhances their perception of system reliability and competence, thereby strengthening trust.

Experienced users possess a deeper understanding of how to utilize system features optimally. For instance, expert users may feel more confident in evaluating the accuracy and relevance of code generated by ChatGPT in programming tasks. This is consistent with the study by Baidoo-Anu and Ansah (2023), which emphasizes the role of user experience in maximizing the potential of AI for complex tasks.

Furthermore, the study by Lortie and Guitton (2011) indicates that trust in technology is also influenced by users' perceptions of system transparency and technical competence. Due to their lack of experience, Novice users often exhibit more significant uncertainty regarding the system's capabilities, negatively impacting their trust levels. In contrast, experienced users, through repeated interactions, develop a better understanding of the system's limitations and strengths, enabling them to manage their expectations more realistically.

This study also supports the notion that user experience directly enhances trust and mitigates perceived technological risks. According to Falcone and Castelfranchi (2001), trust in human-computer interaction is often shaped by direct experiences that reinforce users' confidence in the system's reliability and integrity.

3.6 Type of Task

Human-computer trust scores for each type of task are presented in **Figure 6.** The average score on Task 1 was 3.276; Task 2 amounted to 3,279; Task 3 amounted to 3,283; and Task 4 was 3,111. A Repeated Measurement ANOVA test was conducted to compare the types of human-computer trust tasks. The test results show p-value = 0.146; $X2 = 5.386$ at $\alpha = 0.05$, indicating no difference in HCT for each type of task.

Although this study demonstrates that task types such as mathematics, description, translation, and programming do not significantly impact human-computer trust, it contributes to discussing how ChatGPT is utilized across various contexts. Previous literature, such as the study by Frieder et al. (2024), highlights ChatGPT's capability to analyze large datasets and provide topic clarification, frequently leveraged for academic and professional tasks and may enhance user trust in specific scenarios.

Regarding task types, the study revealed no significant differences in trust levels across mathematics, descriptive, translation, and programming tasks. This uniformity suggests that ChatGPT's performance is perceived as consistent across these domains. However, user expectations, task

complexity, or the specific nature of the task interactions might influence trust in ways not captured in this study. Future research could delve deeper into these nuances, examining how different task characteristics interact with user experience to shape trust.

The lack of significant influence of task types on trust could be attributed to the consistent performance of ChatGPT across tasks or the respondents' familiarity with the functions. Future studies should explore task-specific trust dynamics and consider task complexity and user expectations.

4. Conclusion

The human-computer trust score for expert users in the GPT Chat application has an average Task 1 score of 3.63; Task 2 was 3.62; Task 3 was 3.97; and Task 4 was 3.75. For the human-computer trust value of novice users of the ChatGPT application, the average score for task 1 was 2.90; task 2 was 2.92; task 3 was 2.59; and task 4 was 2.47. The trust score of expert users is higher than that of novice users.

The level of user experience significantly influences human-computer trust in the Chat GPT application in Task 1, Task 2, Task 3, and Task 4. However, there is no significant influence from differences in the types of tasks on human-computer trust in the Chat GPT application.

These findings suggest the need for tailored training programs to familiarize novice users with ChatGPT, thereby enhancing trust. Developers should also consider incorporating user feedback to improve system features and align them with diverse user needs.

Acknowledgment

Thanks are expressed to Direktorat Riset, Teknologi dan Pengabdian Masyarakat, Direktorat Jenderal pendidikan Tinggi, Riset dan Teknologi, Kementerian Pendidikan kebudayaan, Riset dan Teknologi, as providers of research funding in 2024.

5. References

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics*, *54*, 101473.

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52-62.

Bansal, H., & Khan, R. (2018). A review paper on human computer interaction. *International Journal of Advanced Research in Computer Science and Software Engineering*, *8*(4), 53.

Chang, Y., Lee, S., Wong, S. F., & Jeong, S. P. (2022). AI-powered learning application use and gratification: an integrative model. *Information Technology & People*, *35*(7), 2115-2139.

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior*, *90*, 357-362.

Falcone, R., & Castelfranchi, C. (2001). Social trust: A cognitive approach. *Trust and deception in virtual societies*, 55-90.

Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P., & Berner, J. (2024). Mathematical capabilities of chatgpt. *Advances in neural information processing systems*, *36*.

George, A. S., & George, A. H. (2023). A review of ChatGPT AI's impact on several business sectors. *Partners universal international innovation journal*, *1*(1), 9-23.

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil transactions on benchmarks, standards and evaluations*, *2*(4), 100089.

Kesharwani, A., & Singh Bisht, S. (2012). The impact of trust and perceived risk on internet banking adoption in India: An extension of technology acceptance model. *International journal of bank marketing*, *30*(4), 303-322.

Kim, J., & Gambino, A. (2016). Do we trust the crowd or information system? Effects of personalization and bandwagon cues on users' attitudes and behavioral intentions toward a restaurant recommendation website. *Computers in Human Behavior*, *65*, 369-379.

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., ... & Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*, *99*, 101861.

Iku-Silan, A., Hwang, G. J., & Chen, C. H. (2023). Decision-guided chatbots and cognitive styles in interdisciplinary learning. *Computers & Education*, *201*, 104812.

Limbu, Y. B., Wolf, M., & Lunsford, D. (2012). Perceived ethics of online retailers and consumer behavioral intentions: The mediating roles of trust and attitude. *Journal of Research in Interactive Marketing*, *6*(2), 133-154.

Lortie, C. L., & Guitton, M. J. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PLoS One*, *6*(9), e25085.

Madsen, M., & Gregor, S. (2000, December). Measuring human-computer trust. In *11th australasian conference on information systems* (Vol. 53, pp. 6-8).

Niculescu, A., Van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making social robots more attractive: the effects of voice pitch, humor and empathy. *International journal of social robotics*, *5*, 171-191.

Shawar, B. A., & Atwell, E. (2007). Chatbots: are they really useful?. *Journal for Language Technology and Computational Linguistics*, *22*(1), 29-49.

Tamagawa, R., Watson, C. I., Kuo, I. H., MacDonald, B. A., & Broadbent, E. (2011). The effects of synthesized voice accents on user perceptions of

robots. *International Journal of Social Robotics*, *3*, 253-262

Torre, I., Goslin, J., & White, L. (2020). If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior*, *105*, 106215.

Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, *614*(7947), 224-226.