Exploring the Influence of User Interface on User Trust in Generative AI

Morteza Ahmadianmanzary and Sofia Ouhbi[ID][a]

Dept. Information Technology, Uppsala University, Uppsala, Sweden

fi

Keywords: Generative AI, Trust, User Interface, Quality in Use, User Experience.

Abstract: As generative AI tools become increasingly integrated into everyday applications, understanding the impact of user interface (UI) design elements on user trust is essential for ensuring effective human-AI interactions. This paper examines how variations in UI design, particularly avatars and text fonts, influence user trust in generative AI tools. We conducted an experiment using the Wizard of Oz method to assess trust levels across three different UI variations of ChatGPT. Nine volunteer university students from diverse disciplines participated in the study. The results indicate that participants' trust levels were influenced by the generative AI tool's avatar design and text font. This paper highlights the significant impact of UI design on trust and emphasizes the need for a more critical approach to evaluating trust in generative AI tools.

1 INTRODUCTION

Generative AI generates new content, such as text, images, video, audio, or other forms of data using generative models, often in response to prompts (Feuerriegel et al., 2024). As these systems become more integrated into daily life, understanding user perceptions and interactions with them becomes critical. Understanding the relationship between AI systems and user trust is important, however, there is currently no standardized approach for measuring trust in AI systems (Ueno et al., 2022). Trust could be defined as "the degree to which a user or other stakeholder has confidence that a product or system will behave as intended" (ISO/IEC 25022, 2016).

Currently, there is a lack of empirical studies examining how the user interface (UI) of AI systems influences user trust (Bach et al., 2024). Several challenges exist in understanding user trust in AI system and its implications for software engineering, particularly in identifying UI elements linked to user trust (Sousa et al., 2023). This study investigates how variations in UI design influence user trust in generative AI tools, with a focus on ChatGPT 3.5. ChatGPT, developed by OpenAI, rapidly surpassed 100 million active users between November 2022 and January 2023 (Baek and Kim, 2023).

Since the inception of ChatGPT, numerous new generative AI tools have been released each month (McKinsey & Company, 2023). Despite the

[a] https://orcid.org/0000-0001-7614-9731

widespread use of ChatGPT (Baek and Kim, 2023) and similar generative AI systems, there remains a research gap regarding the elements influencing user adoption and usage. The motivation for this study arises from reported incidents where interactions with AI chatbots led to harm and distress (Atillah, 2023; Klar, 2023; Fowler, 2023), highlighting the need to understand the factors driving user trust in generative AI technologies.

To address this gap, we conducted an experiment using the *Wizard of Oz* method, involving interaction with three simulated tools controlled by a human operator. These tools share a similar UI structure as ChatGPT but vary in UI elements, particularly avatar design and text font. Nine volunteer participants, all university master's students from various programs, took part in our experiment and interacted with the three tools and ChatGPT in individual sessions.

2 RELATED WORK

A recent literature review by (Bach et al., 2024) underscored the scarcity of research providing an overview of empirical studies focused on the user-AI relationship regarding user trust in AI systems. The review identified three primary factors influencing user trust in such systems: socio-ethical considerations, technical and design features, and user characteristics. However, there is little understanding of how UI design in AI systems affects user trust.

An experiment was conducted by (Seitz et al., 2022) to explore the initial trust-building process with a diagnostic chatbot, identifying software-related, user-related, and environment-related elements that influence trust. The UI plays a pivotal role in shaping the user experience and perceptions of generative AI systems, thereby influencing trust and engagement (Gupta et al., 2022). Usability and aesthetics are highlighted as crucial factors in conveying trustworthiness and enhancing user engagement. (Gupta et al., 2022) found that text-based conversational interfaces are perceived as more trustworthy than web-based graphical UIs. The effect of anthropomorphic features and conversational interfaces on user perceptions has also been investigated (Zierau et al., 2021).

(Yashmi et al., 2020) discovered that a well-designed website garners more attention, trust, and satisfaction. Furthermore, their analysis indicated that visual appeal contributes more to trust than ease of use. A conceptual framework developed by (Yang and Wibowo, 2022) attempts to enhance understanding of user trust in AI, identifying components, influencing factors, and outcomes of user trust. Their proposed conceptual framework contributes significantly to our understanding of user trust in AI.

(Alagarsamy and Mehrolia, 2023) found that incorrect replies from chatbots lead to customer dissatisfaction. (Bae et al., 2023) observed a lack of trust in AI among some users, despite the increasing use of AI services. (Yen and Chiang, 2021) discovered that trust in chatbots is influenced by anthropomorphism, competency, trustworthiness, social presence, and informativeness.

3 METHODOLOGY

This paper attempts to address the following research question: *"Does variation in avatars and text fonts in UI design impact user trust in generative AI tools?"*

To address this question within the available resources and a reasonable timeframe for the first author's master's degree final project, we opted to employ the Wizard of Oz method rather than developing three distinct chatbots using the ChatGPT API.

3.1 The Wizard of Oz Method

The *Wizard of Oz* method is a user-research approach involving interaction with a simulated interface controlled by a human operator (Paul and Rosala, 2024), as shown in Fig. 1. It is particularly useful when technology development costs are limited, offering valuable insights into usability testing at minimal expense.

This study was conducted with two computers, one as the user device and the other as the Wizard system. The Wizard manages the interface and generates responses from ChatGPT to user inputs, simulating a generative AI system's behavior. The Wizard's identity remained undisclosed to participants during interaction to eliminate bias or awareness of human involvement. Ethical considerations dictate revealing the Wizard's presence at the end of the study (Paul and Rosala, 2024), which was the case in this experiment. Participants were also informed that all responses were generated from ChatGPT at the experiment's conclusion. Since no personal information was collected, obtaining ethical approval from the Swedish Ethical Review Authority was not required.
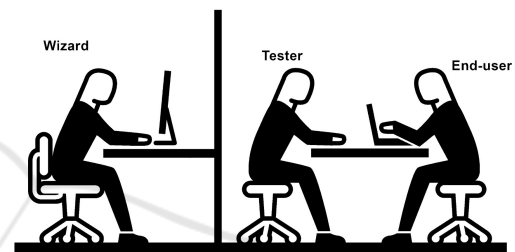

Figure 1: Wizard of Oz Testing (Sara and Maria, 2024).

3.2 UI Design

To address the research question, we designed three UIs similar in construct to ChatGPT 3.5, which UI is shown in Fig. 2, with variations in avatar and font while sharing the same color scheme.
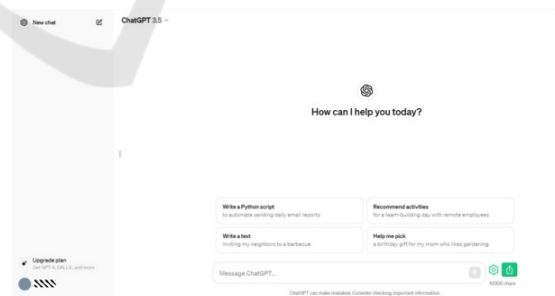

Figure 2: Starting page of ChatGPT 3.5.

Tool 1 features a friendly, anthropomorphic avatar sourced from Freepik and shares the same font, *Inter*, as Tool 3 to isolate the effect of avatar variation on perceived trust, as shown in Fig. 3.

The avatar of Tool 3, sourced from a Google search and modified to be blue, explicitly features "AI" in its design, as shown in Fig. 4, to examine whether this choice influences participants to exercise
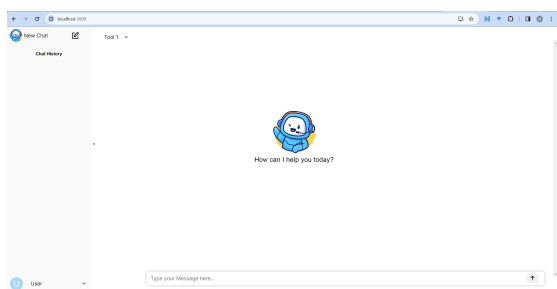
Figure 3: Starting page of Tool 1.
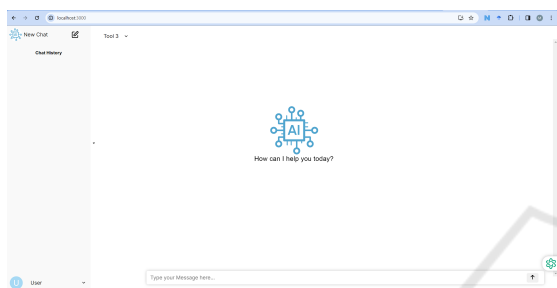
caution when interacting with the tool.

Figure 4: Starting page of Tool 3.

Tool 2 has an avatar, designed by the first author, that resembles two bubble messages with a mechanical appearance and a robotic-style font, *Space Grotesk*, as shown in Fig. 5. The selection of the avatar and font underwent three iterations, incorporating feedback from two other master's students who were not involved in the experiment.
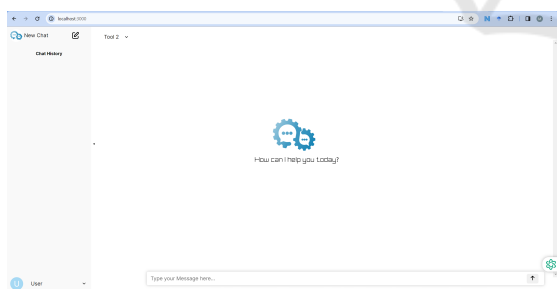

Figure 5: Starting page of Tool 2.

Each UI features a conversational chat format with unique avatar and font, potentially leading users to think they are interacting with three generative AI tools.

3.3 Experiment

Nine volunteers were recruited to engage with the designed UIs during a user testing phase. All participants had prior experience using ChatGPT before tak-

ing part in the experiment. To safeguard participant privacy, the experiment guaranteed anonymity, and all collected data were anonymized to ensure confidentiality. Participants were assured the right to withdraw from the study at any point without needing to provide a reason. Participants provided verbal consent after reading a consent letter before the experiment. Upon completion of the experiment, participants were informed that the responses were generated by Chat-GPT with the assistance of a Wizard.

Participants were given the option to select one of two scenarios:

1. Preparing for a trip in an unfamiliar country.

2. Preparing for a cultural event about a foreign culture.

Focusing on unfamiliar topics in the scenarios aimed to ensure a more objective evaluation of the responses generated by the tools, which originate from ChatGPT. If participants were familiar with a topic, their judgments might be influenced by their own knowledge, potentially biasing their perception of the accuracy and quality of the provided information.

3.3.1 Questionnaire 1

Following their scenario choice, each participant completed a questionnaire using a 5-point Likert-type scale (ranging from 1 for "Strongly Disagree" to 5 for "Strongly Agree") after interacting with each tool.

- I find this this tool trustworthy.

- The tool user interface inspires confidence in its responses.

- I am comfortable conversing with this tool.

3.3.2 Questionnaire 2

Participant filled out also another questionnaire at the end of their interaction with all tools. These questionnaires were presented in a paper format.

1. Which of the following tools is the most trustworthy or inspires the highest level of trust? Why?

○ Tool 1
○ Tool 2
○ Tool 3
○ ChatGPT (GPT 3.5)

2. Which of the following tools is the least trustworthy or inspires the lowest level of trust? Why?

○ Tool 1
○ Tool 2
○ Tool 3
○ ChatGPT (GPT 3.5)

3. I believe a conversational interface builds user trust in generative AI responses to a high extent.

○ Strongly disagree   ○ Disagree   ○ Neither agree nor disagree   ○ Agree   ○ Strongly agree

4. Which user interface elements make you trust a Generative AI tool more while you are working with it?
   o Avatar or Icon
   o Text font
   o Message display area
   o Text input field
   o Typing indicators (e.g. three dots or a pulsating icon)
   o Others:..............

4 RESULTS

The results of the first questionnaire are shown in Table 1, while the results of the second questionnaire are shown in Table 2. Two responses were excluded from the analysis due to issues encountered during the testing phase. Responses of participant 3 were discarded because of a mistake by the Wizard, resulting in duplicate messages being sent. Responses of participant 8 were also discarded due to issue with the chat history of a tool that failed to display the entire conversation.

Participants provided valuable insights into the factors shaping their perceptions of trustworthiness. Tool 1 received positive feedback for its user-friendly logo and font style, which contributed to higher trust scores. Tool 2 consistently received the lowest trust ratings among all tools, with its mechanical logo and robotic-like font style detracting from the perceived trustworthiness.

Results show also that ChatGPT emerged as the most trusted tool among participants, with six out of seven expressing high levels of trust in its responses. Some participants found that others tools trustworthy as well. Two of participants alongside ChatGPT found Tool 1 trustworthy, and only one mentioned Tool 3. Tool 2 was also reported as trustworthy by one participant due to its responses. However, five out of seven identified Tool 2 as the least trustworthy among the others, while this number reduces to two participants for Tool 1 and only one participant for Tool 3. These findings highlight the impact of UI design on trust levels, suggesting varying degrees of trust across different UI designs.

Participants also identified several UI elements influencing their trust in generative AI tools, including avatars or icons, text fonts, message display areas, and typing indicators. Since all tools displayed text generated from ChatGPT, the results indicate that UI design significantly influences user trust in generative AI systems, with specific design elements exerting more pronounced effects on trust levels.

Upon the experiment's conclusion, when participants were informed about the wizard's role in generating responses from ChatGPT for the simulated tools, the majority were surprised. Interestingly, some participants were unaware of the option to prompt ChatGPT to provide more concise responses. This lack of awareness led them to believe that the tools were not directly linked to ChatGPT. Upon learning the source of the text, all participants acknowledged to the tester the influence of the UI on shaping their perception of trust regarding the simulated tools.

5 DISCUSSION

5.1 Main Findings

Understanding trust dynamics in the context of generative AI is essential for promoting user acceptance and adoption of these technologies (Yen and Chiang, 2021). The findings of this study underscore the significant role of UI design in shaping user trust in generative AI systems. The identified UI elements in this study align with prior research emphasizing the role of UI design in shaping user confidence and trust in AI systems (Zierau et al., 2021). The influence of conversational interfaces on trust is further supported by research highlighting the role of relational conversational style and avatars in building user trust (Zierau et al., 2021; Gupta et al., 2022). For instance,

Table 1: Questionnaire 1 Results.

	Questions	Participant 1 - man	Participant 2 - man	Participant 3* - woman	Participant 4 - woman	Participant 5 - woman	Participant 6 - woman	Participant 7 - man	Participant 8* - woman	Participant 9 - man
Tool 1	I find this tool trustworthy	3	4	4	4	4	3	4	4	4
	The tool user interface inspires confidence in its responses	2	4	3	3	3	4	4	4	4
	I am comfortable conversing with this tool	4	4	4	4	4	4	4	4	5
	Average score	3	4	3.6	3.6	3.6	3.6	4	4.3	4.3
Tool 2	I find this tool trustworthy	4	4	4	4	3	4	4	5	4
	The tool user interface inspires confidence in its responses	3	4	2	2	4	4	2	3	2
	I am comfortable conversing with this tool	4	3	4	3	2	4	4	2	4
	Average score	3.6	3.6	3.3	3	3	4	3.3	3.3	3.3
Tool 3	I find this tool trustworthy	4	4	4	4	4	4	4	3	3
	The tool user interface inspires confidence in its responses	3	4	5	4	4	4	4	4	2
	I am comfortable conversing with this tool	4	4	5	4	4	4	4	4	4
	Average score	3.6	4	4.6	4	4	4	4	3.6	3

Excluded responses, not considered in the analysis of the results.

Table 2: Questionnaire 2 Results. *Acronym: Participant (P), Conversational Interfaces (CI), and Generative AI (GAI).*

P.	Most trustworthy tools	Why?	Least trustworthy tools	Why?	CI influence on GAI trust	UI elements influencing GAI trust
P1	ChatGPT 3.5	Its answers are more detail oriented	Tool 1	[Its] answer doesn't have enough explanation or reasons to help me decide confidently	4	Avatar or Icon
P2	ChatGPT 3.5	I didn't get any difference between those three tools, and the only reason I chose ChatGPT is that it is more famous	Tool 2	I cannot mention any specific reason, but its appearance was less attractive [to] me	3	Text font
P3*	Tool 3 and ChatGPT 3.5	To develop ChatGPT a big language database has been used then the given information are based on more data	Tool 2	Because it repeats the one answer twice, and its interface was not as attractive as the other ones	3	Text font, Message display area, and Typing indicators
P4	Tool 1	I liked the logo. It was user friendly, also the font style was good	Tool 2	I didn't like the font style. Also at first I thought the logo was two gears and after a while I saw the message bubbles	5	Avatar or Icon, Text font, and Message display area
P5	ChatGPT 3.5	Its used by million of people, so it is the most trustworthy among these four	Tool 2	The logo and font [are] not good and it make you feel uncomfortable with reading responses	4	Avatar or Icon, Text font, and Message display area
P6	Tool 2 and ChatGPT 3.5	ChatGPT is already approved and Tool 2 gave me more trustable answers (more specific)	Tool 1	I could not trust the responses	4	Message display area and Typing indicators
P7	Tool 3 and ChatGPT 3.5	Because I had some similar experience with ChatGPT and I think the response is more near to realistic	Tool 2	Because I expected that it recommends Iran as a good country for tourists. But it does not recommend Iran as 7 top countries in this regard. Also, the font was not favorite for me	4	Avatar or Icon, Text font, Message display area, and Typing indicators
P8*	Tool 1 and ChatGPT 3.5	ChatGPT: it is the famous AI. Tool 1: the appearance and responses is the same as ChatGPT	Tool 2 and Tool 3	I see something weird in the user interface such as first sentence, that I typed first in the other tool	3	Avatar or Icon and Typing indicators
P9	Tool 1 and ChatGPT 3.5	Because ChatGPT is very common and famous in addition Tool 1 has good responses and I have better emotion with it	Tool 2 and Tool 3	I believe these logos [look] like artificial and I do not have good feeling [about] them. But totally the responses looks similar	5	Avatar or Icon, Text font, and Typing indicators

Excluded responses, not considered in the analysis of the results.

ChatGPT does not have an anthropomorphic avatar to maintain a neutral and universal appearance, and its text-only interface allows users to imagine it persona based on its responses (Liu and Siau, 2023; Nowak and Rauh, 2005).

The results emphasize the critical necessity of enhancing user awareness during interactions with generative AI, owing to several significant limitations. Among these, *hallucination* stands out as a prominent issue, wherein erroneous information is presented as factual, which could have a serious impact on the well-being of individuals, especially vulnerable ones. Additionally, there are risks associated with the *halo effect*, wherein individuals are inclined to trust conversational AI's polished and authoritative language. The *bias* problem within generative AI exacerbates these concerns, as these systems tend to replicate and potentially amplify biases present in the training data, resulting in unfair or discriminatory outputs (Milmo, 2024).

Ethical dilemmas arise from the potential misuse of generated content for malicious purposes, including the creation of deepfakes, dissemination of fake news, or spread of misinformation (Logan, 2024). Moreover, the poor quality of data used in training can lead to misleading answers. It is noteworthy that ChatGPT, for instance, has been trained on data from web crawling, Reddit posts with three or more up-votes, Wikipedia, and internet book collections (Walsh, 2024). Therefore, it is imperative to regulate these tools appropriately to mitigate the adverse consequences of misplaced trust.

5.2 Limitations

This study might have several limitations such as:

- The use of the Wizard of Oz technique, while effective for simulating generative AI functionality, may not fully replicate real-world interactions with generative AI tools.

- The participants consisted solely of Master's students, potentially limiting the generalizability of the findings to other groups with varying educational backgrounds. Future research could address these limitations by including more diverse participant groups.

- The variation in UI design was limited to the avatar and text font. Exploring additional design variables could further enhance understanding of the relationship between UI design and user trust in AI systems.

6 CONCLUSION AND FUTURE PLAN

This study's findings highlighted that avatars are the most influential UI element affecting user trust in generative AI. Results showed also that participants demonstrated sensitivity to text font variations. Interestingly, the results showed that despite participants interacting with the same source of outputs, variations in UI led to differing perceptions of trust, emphasizing the role of UI design in shaping trust in generative AI responses. These results emphasize the importance for designers and developers to exercise caution when designing UI, guiding users to avoid placing excessive trust in unregulated generative AI systems. Users should not be misled by UI design choices into increasing their trust in such systems.

For future research, we plan to conduct a larger experiment with participants from diverse user groups, varying in educational background and familiarity with generative AI. We also aim to explore the nuanced interactions between UI design and user trust in generative AI tools by considering additional UI elements, such as color, which were not a factor studied in this experiment as the three UI shared the color blue.

REFERENCES

Alagarsamy, S. and Mehrolia, S. (2023). Exploring chatbot trust: Antecedents and behavioural outcomes. *Heliyon*, 9(5).

Atillah, I. E. (31 March 2023). Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change. https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-.

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., and Sousa, S. (2024). A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human–Computer Interaction*, 40(5):1251–1266.

Bae, S., Lee, Y. K., and Hahn, S. (2023). Friendly-bot: The impact of chatbot appearance and relationship style on user trust. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 45.

Baek, T. H. and Kim, M. (2023). Is ChatGPT scary good? how user motivations affect creepiness and trust in generative artificial intelligence. *Telematics and Informatics*, 83:102030.

Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, 66(1):111–126.

Fowler, G. A. (10 Aug 2023). AI is acting 'pro-anorexia' and tech companies aren't

stopping it. [Accessed on April 2024]. www.washingtonpost.com/technology/2023/08/07/ai-eating-disorders-thinspo-anorexia-bulimia/.

Gupta, A., Basu, D., Ghantasala, R., Qiu, S., and Gadiraju, U. (2022). To trust or not to trust: How a conversational interface affects trust in a decision support system. In *Proceedings of the ACM Web Conference 2022*, pages 3531–3540.

ISO/IEC 25022 (2016). ISO/IEC 25022:2016. Systems and software engineering – Systems and software quality requirements and evaluation (SQuaRE) – Measurement of quality in use.

Klar, R. (08 July 2023). AI chatbots provided harmful eating disorder content: report. [Accessed on April 2024]. https://thehill.com/policy/technology/4141648-ai-chatbots-provided-harmful-eating-disorder-content-report/.

Liu, Y. and Siau, K. L. (2023). Human-ai interaction and AI avatars. In *International Conference on Human-Computer Interaction*, pages 120–130. Springer.

Logan, N. (4 Feb 2024). This article is real — but AI-generated deepfakes look damn close and are scamming people. [Accessed on April 2024]. https://www.cbc.ca/news/canada/deepfake-ai-scam-ads-1.7104225.

McKinsey & Company (25 Aug 2023). What's the future of generative AI? an early view in 15 charts. [Accessed on December 2024]. https://www.mckinsey.com/featured-insights/mckinsey-explainers/whats-the-future-of-generative-ai-an-early-view-in-15-charts#/.

Milmo, D. (22 Feb 2024). Google pauses AI-generated images of people after ethnicity criticism. [Accessed on April 2024]. www.theguardian.com/technology/2024/feb/22/google-pauses-ai-generated-images-of-people-after-ethnicity-criticism.

Nowak, K. L. and Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11(1):153–178.

Paul, S. and Rosala, M. (2024). The Wizard of OZ method in UX. [Accessed on December 2024]. https://www.nngroup.com/articles/wizard-of-oz/. Nielsen Norman Group.

Sara, R. and Maria, R. (19 April 2024). The Wizard of Oz Method in UX. [Accessed on April 2024]. https://www.nngroup.com/articles/wizard-of-oz/.

Seitz, L., Bekmeier-Feuerhahn, S., and Gohil, K. (2022). Can we trust a chatbot like a physician? a qualitative study on understanding the emergence of trust toward diagnostic chatbots. *International Journal of Human-Computer Studies*, 165:102848.

Sousa, S., Cravino, J., and Martins, P. (2023). Challenges and trends in user trust discourse in AI popularity. *Multimodal Technologies and Interaction*, 7(2):13.

Ueno, T., Sawa, Y., Kim, Y., Urakami, J., Oura, H., and Seaborn, K. (2022). Trust in human-AI interaction:

Scoping out models, measures, and methods. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7.

Walsh, M. (22 April 2024). ChatGPT Statistics — The Key Facts and Figures. [Accessed on April 2024]. https://www.stylefactoryproductions.com/blog/chatgpt-statistics.

Yang, R. and Wibowo, S. (2022). User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets*, 32(4):2053–2077.

Yashmi, N., Momenzadeh, E., Anvari, S. T., Adibzade, P., Moosavipoor, M., Sarikhani, M., and Feridouni, K. (2020). The effect of interface on user trust; user behavior in e-commerce products. In *Proceedings of the Design Society: DESIGN Conference*, volume 1, pages 1589–1596. Cambridge University Press.

Yen, C. and Chiang, M.-C. (2021). Trust me, if you can: a study on the factors that influence consumers' purchase intention triggered by chatbots based on brain image evidence and self-reported assessments. *Behaviour & Information Technology*, 40(11):1177–1194.

Zierau, N., Flock, K., Janson, A., Söllner, M., and Leimeister, J. M. (2021). The influence of AI-based chatbots and their design on users' trust and information sharing in online loan applications. In *Hawaii International Conference on System Sciences (HICSS).-Koloa (Hawaii), USA*.