*The Human Side of Service Engineering,* Vol. 143, 2024, 139–149
https://doi.org/10.54941/ahfe1005092

# Enhancing Trust in LLM Chatbots for Workplace Support Through User Experience Design and Prompt Engineering

**Zhiyun Gong[1], Shivani Birajdar[1], Tam Cao[1], Yi Qing Khoo[1], Chia-Fang Chung[1], Yassi Moghaddam[1], Anbang Xu[2], Hridhay Mehta[2], Aaditya Shukla[2], Zhilin Wang[2], and Rama Akkiraju[2]**

[1]University of California - Santa Cruz, Santa Clara, CA 95054, USA
[2]NVIDIA, Santa Clara, CA 950511, USA

## ABSTRACT

Traditional chatbots have been essential for workplace support. With the rise in Large Language Model (LLM) chatbots and their quick and efficient solution to user queries, this new generation of chatbots will soon enter the field of workplace support to assist with IT, HR, and general workplace queries. However, trust concerns with LLM chatbots, which arise from factual errors, inaccuracies, and suboptimal response formatting, have become prominent and will be particularly critical in professional settings such as employee support within a company. This paper investigates factors influencing user trust in AI chatbots for workplace support, proposing solutions through UX design improvement and prompt engineering experiments. We conducted mixed-method user research to study the impact of response formatting and presentation on user trust and experience. Our qualitative user interviews and contextual inquiries aim to understand users' expectations of these chatbots and their perspective of usage, followed by user surveys that validate users' preferences through quantitative measures. The findings reveal that trust challenges arise from a perceived lack of credibility and transparency as a result of hallucinations, as well as concerns about data privacy. They also show the need for improved chatbot conversational experiences with more human-likeness, better contextual understanding abilities, and higher flexibility in input and output formats. To address these challenges, our research uniquely proposes and implements a solution based on the interception of UX design and prompt engineering. Actionable UX design implications for a trustworthy interface are outlined, along with prompt engineering solutions demonstrated through a prototype. This research contributes to the evolution of AI-driven chatbot technology, aligning with the broader goal of enhancing user satisfaction and trust in automated support systems. This paper provides valuable insights for AI chatbot developers, designers, and researchers to meet the critical need for effective and reliable chatbots tailored to workplace support. This study also points to opportunities for future research topics around trustworthiness in Artificial Intelligence to explore how diversity, technology, research design, and ethical aspects would factor into user trust and experience.

**Keywords:** LLM chatbot, Trustworthy AI, Workplace support, User experience, Prompt engineering

## INTRODUCTION

Chatbots have become an essential part of workplace employee support, and Large Language Model (LLMs) chatbots such as ChatGPT have been demonstrated to provide quick and efficient solutions to users' queries (White

 

*et al.*, 2023). However, these chatbots may make factual mistakes and suffer from inaccuracies, which create trust issues for the users. We uniquely aim to find solutions at the interception of UX design and prompt engineering, followed by a prompt engineering experiment to apply the findings. Through this research, we aim to identify the specific factors that contribute to a user's level of trust (or distrust) in AI chatbots and their expectations for these interactions (Amershi *et al.*, 2019). Ultimately, this will help to inform the development of more effective and trustworthy chatbots for workplace support.

## RELATED WORK

### User Trust in AI Chatbot

Trustworthiness in AI-enabled systems has been discussed in many studies (Bach *et al.*, 2022), and specifically in AI chatbots the most influential trust factors have been found to be the relevancy of responses and whether the problem was resolved (Nicolescu and Tudorache, 2022). Anthropomorphism, or human-like features including verbal communication, expression and gestures (Weitz *et al.*, 2019), has a significant impact on users' trust (Bach *et al.*, 2022; Fan *et al.*, 2021; Weitz *et al.*, 2019), users' responses (Cheng *et al.*, 2022), and thus overall user experience (Nicolescu and Tudorache, 2022) when interacting with AI chatbot. What's more, design principles have been proposed to build and maintain user trust in AI-based chatbots (Zierau *et al.*, 2020).

### Prompt Engineering

Studies have been conducted to improve system prompts for AI chatbots. A prompt pattern catalog was created to enhance output generation and interaction when working with LLMs (Schmidt *et al.*, 2023; White *et al.*, 2023). Chain-of-thought prompting was demonstrated to be a simple and broadly applicable method for enhancing reasoning in language models, through experiments on arithmetic, symbolic, and commonsense reasoning (Wei *et al.*, 2022). The optimization of user prompts for AI chatbots has been another focus of study. Various tools have been developed and tested using visual analytics (Mishra *et al.*, 2023) and automatic prompt editing techniques (Wang, Shen and Lim, 2023).

Based on the literature review, we identified research gaps through our initial understanding of the topic. Despite the number of studies on trust issues in AI chatbots and trust-related UX design features, there have been limited findings in the workplace support context. Prompt engineering appears to be an emerging research topic in literature, and we are interested in learning about its impact on user trust. Therefore, we came up with the following research questions.

- What are users' expectations and current perceptions of chatbots as a workplace support solution?
- What factors contribute to users' trust in workplace support chatbots?

- What type of system prompt patterns are most effective in enhancing trustworthiness in the context of workplace support?
- What UX design suggestions would help build and maintain users' trust over time for workplace support chatbots?

## METHODOLOGY

We applied a mixed methods approach, as shown in Figure 1.

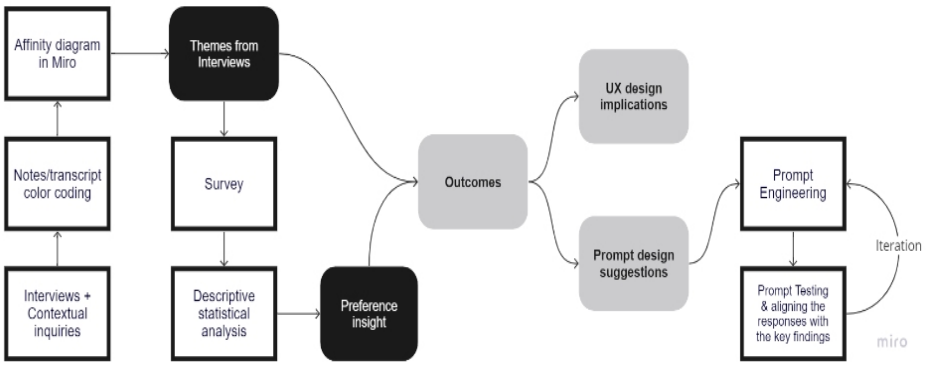


**Figure 1:** Research process.

### Phase 1 – Qualitative Research

Our qualitative research sessions were designed to be around 60 minutes. In the first 30 minutes, we conducted semi-structured interviews to gain insights into users' expectations and perceptions of LLM chatbots, workplace support chatbots, traditional workplace support, and the factors influencing their trust.

The second half of the sessions was used for contextual inquiries to observe user interactions with ChatGPT and an existing non-AI employee support chatbot. Scenarios such as "can't log in to my work email account" and "request for PTO" were provided to simulate a workplace support setting. This approach allowed us to probe perspectives that were not mentioned during the interviews.

We recruited 7 participants through connections within the student team, which was a mix of students and professionals, as well as 5 Nvidia employees. We strongly emphasized recruiting from diverse backgrounds, age groups, and genders (Figure 2), ensuring that different perspectives were represented.

After the sessions, we color-coded all transcripts and used Miro for affinity mapping and thematic analysis.

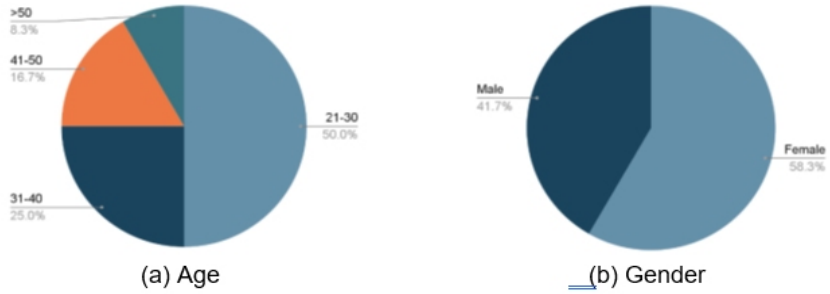


**Figure 2:** Demographics of qualitative research participants.

## Phase 2 - Quantitative Research

In this phase, we launched a user survey to further justify the findings and better inform the design decisions. The design of our survey was guided by the themes from the qualitative interviews (Figure 3(a)). To evaluate user preferences comprehensively, the survey included the following elements (Figure 3(b)):

1. **Likert Scale:** Our survey employed Likert questions inspired by the TrustedUX scale (TrustedUX, no date). We selected questions most pertinent to each topic to avoid a lengthy questionnaire. The Likert scales gauge user preferences based on various aspects, such as whether users wanted to continue the conversation or believed that the chatbot would do its best to help users, aiming to capture nuances beyond mere liking or disliking.
2. **User Preference** (Multiple Choice): In addition to the Likert scale, we included multiple-choice questions to gain a broad overview of the most favored options.
3. **Open-ended comments**: Recognizing the value of qualitative feedback, we included an open-ended comments section as well. This allowed participants to provide further insights and explanations regarding their preferences.

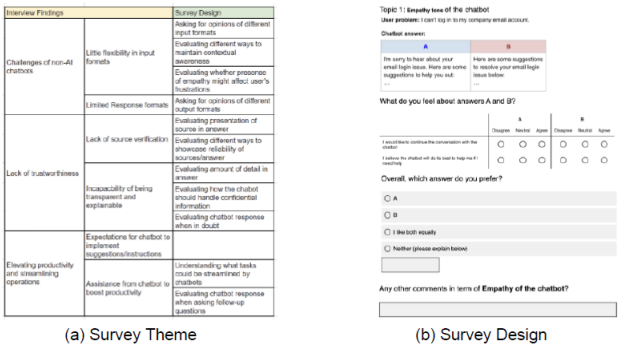


**Figure 3:** Design of the survey based on interview findings.

We received responses from 42 participants, 14 of whom did not fully complete the survey. This group also included both external users and Nvidia employees.

To analyze the survey data, we first translated all data into numeric form. We assigned the value of 1, 2, 3 to "Disagree", "Neutral", and "Agree" in the Likert responses, and calculated the percentages in the preference questions. Then we calculated the mean values and the standard deviation of all these data, which we used to triangulate and validate our qualitative findings.

## RESULTS

Combining the findings from our interviews and survey data, we have identified 3 key themes in terms of users' unmet needs and preferences.

### Theme 1: Need for Better Conversational Experiences

Chatbots have transformed the landscape of human-computer conversational interactions. However, they still face challenges that need to be addressed to ensure better conversational experiences.

- **Lack of Human-likeness:** Chatbots are often ill-equipped to identify human emotions and understand complexity and nuances in human language, further diminishing user trust in their capabilities and increasing the distance between users and chatbots. Participant 6 liked the empathy they felt when interacting with ChatGPT, *"it's good that when I said it, it's able to empathize with me and say, OK, I'm really sorry, even though it's a very small issue, saying I'm sorry, they're having trouble.*" In the survey, 67% of the respondents voted for a response that was more empathetic and tended to agree that "I would like to continue the conversation with the chatbot" (Mean = 2.36/3.00, SD = 0.75). Users also expressed frustration when the chatbot was unable to solve the problems and it would not allow them to speak to human agents. Participant 6 shared a personal experience during the interview, *"I had this experience with one of the food delivery apps...I had one specific question, but the chatbot couldn't respond to me because, first of all, it didn't connect me to a human…and that was very annoying because I had a very difficult time to go through whatever answer I needed. Because I'd already placed the order, I couldn't cancel the order.*"
- **Lack of Contextual Understanding:** Users may need to repeat information or clarify their intent, leading to a fragmented and disjointed experience. This lack of contextual understanding can decrease trust as users don't feel they are engaged in a meaningful conversation. Participant 6 mentioned how past topics could be easily picked up in ChatGPT, *"chats get saved automatically, which is great"*. Furthermore, 80% of the survey participants prefer to have a feature that saves chat history for them to go back to the topics.
- **Lack of Flexibility:** Users may feel a sense of limited adaptability when interacting with the chatbot, which can negatively impact their experience. The lack of variety in input and output formats was brought up in the interviews. Participant 3 shared *"I'm used to messaging friends using speech-to-text. It helps me chat faster without having to type with my fingers. I hope I can talk to a chatbot like that.*" In the survey, we also found out that, apart from text, the users would also like to be able to speak to the chatbot (71%) or upload an image (42%) to the conversation as

alternative ways to input. In return, they would like to receive audio messages (38%) and images (38%) from the chatbot as well. In addition, users prefer to get links (42%) in the responses.

- **Uncertainty in Data Privacy:** Users may be concerned about the privacy of their conversations with AI chatbots. Participant 7 believes that *"you need a high level of trust because people are going to tell you like sensitive information."* This was further validated in the survey where all the respondents preferred the chatbot to decline to answer any questions about confidential/private information, e.g. login credentials.

### Theme 2: Challenges in AI Hallucination

AI chatbot hallucination refers to a situation where the chatbot generates responses that are incorrect, nonsensical, or unrelated to the input provided by the user. Hallucinations in AI chatbots can occur due to various reasons, such as users asking ambiguous or unclear questions, users' questions being outside of the chatbot's dataset, or a lack of logical reasoning/chain of thoughts. As a result, users usually feel that there are:

- **Lack of Credibility:** Consistent errors or hallucinations can undermine the chatbot's credibility and discourage users from relying on its information. Participant 5 said during the interview, *"if it gives a disclaimer saying that the answers you need to still verify on your own, then how do I trust the system where I see that disclaimer on the first screen itself?"* According to Participant 1, *"Ultimately, you can't really trust the answers further than you could like a cursory Google search without actually doing a lot of verification and validation. You should always be taking it kind of with a grain of salt."* In our surveys, we explored users' preference for the display of sources, where 74% prefer links at the bottom of chatbot responses and 38% also like the idea of having citations among the text. Another important aspect of hallucinations was how the chatbot reacts when in doubt, i.e. the question lacks clarity or is outside of its dataset. Participant 9 felt that the bot *"should not jump to the answers right away…I would say ask (for) more context from the user in order to give them the right answer"* after being frustrated by the employee support chatbot during the contextual inquiry. To validate, we explored users' preferences in the survey. 44% preferred the chatbot to apologize and decline to answer, 38% wanted the chatbot to ask for clarification and provide a best guess, and users tend to disagree (Mean = 1.76/3.00, SD = 0.73) that "the chatbot is interested in understanding my needs and preferences" if the chatbot directly answers with a best guess.
- **Lack of Transparency:** Users often find it difficult to determine how an AI chatbot arrives at its responses. Participant 7 expressed in our conversation that *"this is right now a gray area wherein the user doesn't exactly know that which part is accurate and which part is not accurate."* Participant 2 referred to the concept of "Chain of Thoughts" and said, *"I think I would trust an answer more if the bot actually lays out all of the reasoning and details so that I can see how it arrives at the solution."* Building off of these insights, we found out in the survey that 2 completely different styles, "Concise" (43%) and "Detailed with keywords highlighted"

(46%), received similar votes, but users perceive that the detailed chatbot (Mean = 2.57/3.00, SD = 0.69) is more likely to "do its best to help me if I need help" than the concise chatbot (Mean = 2.20/3.00, SD = 0.71). Another aspect of transparency is the reliability indicators. A few users expressed confusion about the "percentage match" data in the responses during the contextual inquiries with the employee support chatbot. Therefore, we set out to look for a better representation, and we found in the survey that users are more likely to trust a response if they see the number of users that liked it (57%) or if it has been approved by experts (43%).

### Theme 3: Elevating Productivity and Streamlining Operations

Users typically aim to maximize efficiency and reduce manual tasks, and they anticipate that AI chatbots will play a key role in enhancing productivity and streamlining various aspects of their workflows and operations.

- **Task execution:** Users not only look to AI chatbots for guidance but also expect them to actively implement the suggestions and instructions they provide. As Participant 5 mentioned in the interview, *"...so in short, you would expect the chatbot to process all those requests for you, for example, reconnecting or change password...the chatbot (to) be able to run a script at the back end once I report it and then it should ask me to try and reconnect."* In the survey, 90% of the respondents showed a preference for the chatbot to be able to generate hyperlinks in the responses to accelerate user actions or to provide a set of related questions to stimulate further discussion.
- **Boosting work productivity:** Users seek more than just advice related to IT/HR issues, which is what most workplace support chatbots currently offer; they desire to rely on them as valuable allies to enhance productivity and efficiency. Participant 12 pointed out *"I guess if it could...look at the transcript of meeting notes and then update ticket items automatically."* Participant 10 felt that *"I think documenting your code, that would be a helpful feature because I know for a fact what my function is doing, but I don't want to spend time writing that down."* We then collected more users' needs in terms of productivity boosting features in the survey, which included email summary and drafting, daily work scheduling, coding documentation/debugging, etc.

## UX DESIGN SUGGESTIONS

Based on our results, we propose UX design suggestions to make the chatbot more trustworthy in a workplace support context (middle column in Table 1).

1. Chat history: Include a feature that allows reviewing past work support conversations, retrieving key information or references, and continuing the chats
2. Multimedia input formats: Allow users to input with spoken language or attach files, e.g. images, to expedite workplace support conversations.

3. Diverse output formats: Provide the flexibility to receive workplace chatbot responses in various formats, including images, audio, and links to accommodate different user needs and enhance information accessibility.
4. Verbosity option: Enable employees to choose between a concise response and a detailed one with key info highlighted given their changing needs.
5. Reliability indicators: Provide information such as "X% of users liked" and "expert approved" to showcase to employees the reliability of the responses.
6. Task execution: Conclude interactions by directly executing workplace support tasks or providing hyperlinks to facilitate seamless transitions to the next steps.
7. Transactional queries: Streamline work through transactional queries. For example, generate meeting notes automatically; offer comprehensive coding support, including auto-generation of documentation, comments, and assistance with debugging tasks; enable users to initiate software installations through chatbot interactions; facilitate the automation of tasks such as sending automatic emails, requesting Paid Time Off (PTO), thus streamlining administrative processes; allow the chatbot to create and update meeting invitations, ensuring all stakeholders are informed and involved.

**Table 1.** Design recommendations by theme.

| Theme | UX Design Suggestions | Prompt Design Suggestions |
|---|---|---|
| **Need for Better Conversational Experiences** | 1. Chat history<br>2. Speech-to-text input<br>3. Multimedia output formats | 1. Human-like features (e.g. empathy, emotional sentiment, connecting to human agents)<br>2. Handling data privacy |
| **Challenges in AI Hallucination** | 4. Option for verbosity<br>5. Reliability indicators | 3. Source links for the responses<br>4. Handling lack of information<br>5. Chain of thoughts |
| **Elevating Productivity and Streamlining Operations** | 6. Task execution<br>7. Transactional queries | 6. Follow-up questions and relevant links |

## PROMPT ENGINEERING

We also tried to address the issue of trust through a prototype utilizing Prompt Engineering, which is the practice of designing inputs for generative AI tools to produce optimal outputs (McKinsey & Company, 2023). Below are the key relevant discoveries in a workplace support context (right column in Table 1).

1. Human-likeness: the chatbot should speak in an empathetic tone and be able to recognize and respond to employees' emotions and sentiments. They should also make sure to connect with a human agent when necessary.
2. Data privacy: the chatbot should decline to provide information or carry out any queries involving confidential or private data in a work setting.

3. Source display: present the information sources at the end, e.g. a link to the company benefit webpage, with the option of having citations among the text.
4. Handling lack of information: never give wild guesses without asking for clarification. Acknowledge any lack of data or understanding to maintain transparency in employee-AI interaction.
5. Chain of Thought Display: consider displaying the workplace assistant's chain of thought, illustrating how it arrived at a particular response. Using formats such as bullet points to help with readability as well.
6. Follow-up questions and relevant links: enhance employee engagement with the chatbot by proposing next steps through follow-up questions, related questions, or links, to facilitate problem-solving in workplaces.

We designed the prompt to address the above findings for employee support. One such example is stated below in Figure 6, where we included prompts such as:

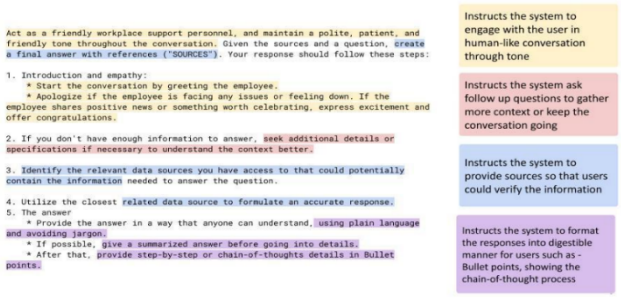


**Figure 4:** Prompt design example.

- "Act as a friendly workplace support personnel, and maintain a polite, patient, and friendly tone throughout the conversation."
- "Apologize if the employee is facing issues or feeling down. If the employee shares positive news or something worth celebrating, express excitement and offer congratulations."
- "Start the conversation by greeting the employee"

Similarly, other prompts were also added to cater to other key findings, and we were able to train the chatbot for all the themes after 8 rounds of iterations. Figure 7 showcases the result for one such scenario tested on the chatbot prototype. In this example, the chatbot is expressing excitement, empathy, and emotions as the user has shared news on the upcoming baby. The chatbot has provided the solution succinctly in bullet point paragraphs. The chatbot asks follow-up questions and references the source of its information.

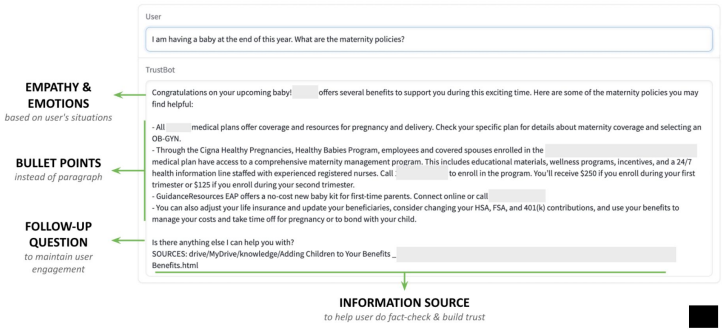


**Figure 5:** Prompt design result.

## LIMITATIONS AND FUTURE RESEARCH

The research acknowledges several limitations that may affect the generalizability of findings and points to avenues for future research, including incorporating diverse and inclusive user samples, exploring improved prompt engineering techniques for more natural conversations, conducting longitudinal studies to track changes in user perceptions over time, and investigating ethical considerations such as biases in chatbot responses.

## CONCLUSION

In conclusion, this paper highlights the importance of understanding and enhancing user trust in AI chatbots for workplace support, sheds light on the factors that influence users' trust or distrust in these chatbots, and proposes how UX design and prompt engineering can help address the pain points. The ultimate goal is to create more effective and reliable chatbots that align with user expectations, contribute to the advancement of AI-driven technology, and foster greater user satisfaction and trust in automated support systems. As businesses increasingly rely on AI chatbots, this research becomes increasingly relevant, offering valuable insights for both developers and users of these systems.

## REFERENCES

Amershi, S., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S. and Bennett, P. N. (2019). Guidelines for Human-AI Interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19. [online] doi: https://doi.org/10.1145/3290605.3300233.

Bach, T. A., Khan, A., Hallock, H., Beltrão, G. and Sousa, S. (2022). A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. International Journal of Human–Computer Interaction, pp. 1–16. doi: https://doi.org/10.1080/10447318.2022.2138826.

Cheng, X., Zhang, X., Cohen, J. and Mou, J. (2022). Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. Information Processing & Management, 59(3), p. 102940. doi: https://doi.org/10.1016/j.ipm.2022.102940.

Fan, M., Zou, F., He, Y. and Jiang, X. (2021). Research on Users' Trust of Chatbots Driven by AI: An Empirical Analysis Based on System Factors and User Characteristics. In: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE). doi: https://doi.org/10.1109/iccece51280.2021.9342098.

McKinsey & Company (2023). What Is Prompt Engineering? [online] www.mckinsey.com. Available at: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-prompt-engineering.

Mishra, A., Soni, U., Arunkumar, A., Huang, J., Kwon, B. C. and Bryan, C. (2023). PromptAid: Prompt Exploration, Perturbation, Testing and Iteration using Visual Analytics for Large Language Models. doi: https://doi.org/10.48550/arxiv.2304.01964.

Nicolescu, L. and Tudorache, M. T. (2022). Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review. Electronics, [online] 11(10), p. 1579. doi: https://doi.org/10.3390/electronics11101579.

Schmidt, D., Spencer-Smith, J., Fu, Q. and White, J. (2023). Cataloging Prompt Patterns to Enhance the Discipline of Prompt Engineering. [online] Available at: https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf.

TrustedUX. (n.d.). [online] Available at: https://www.trustux.org/ [Accessed 21 Jan. 2024].

Wang, Y., Shen, S. and Lim, B. Y. (2023). RePrompt: Automatic Prompt Editing to Refine AI-Generative Art Towards Precise Expressions. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. doi: https://doi.org/10.1145/3544548.3581402.

Wei, J. Z., Wang, X., Schuurmans, D., Bosma, M., Chi, E. H., Le, Q. V. and Zhou, D. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In: Advances in Neural Information Processing Systems 35 (NeurIPS 2022). [online] doi: https://doi.org/10.48550/arxiv.2201.11903.

Weitz, K., Schiller, D., Schlagowski, R., Huber, T. and André, E. (2019). 'Do you trust me?' In: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. doi: https://doi.org/10.1145/3308532.3329441.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. and Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. [online] Available at: https://arxiv.org/abs/2302.11382.

Zierau, N., Hausch, M., Bruhin, O. and Söllner, M. (2020). Towards Developing Trust-Supporting Design Features for AI-Based Chatbots in Customer Service. 41th International Conference on Information Systems.