Challenges and Opportunities for Prompt Management: Empirical Investigation of Text-based GenAI Users

Nitish Patkar
University of Applied Sciences and
Arts Northwestern Switzerland
Windisch, Switzerland
nitish.patkar@fhnw.ch

Anton Fedosov
University of Applied Sciences and
Arts Northwestern Switzerland
Windisch, Switzerland
anton.fedosov@fhnw.ch

Martin Kropp
University of Applied Sciences and
Arts Northwestern Switzerland
Windisch, Switzerland
martin.kropp@fhnw.ch

Abstract

Generative AI (genAI) tools, like ChatGPT, have become popular not only with everyday users but also with Human-Computer Interaction (HCI) researchers and practitioners. Despite their rapid adoption, there is a lack of studies examining their design, particularly regarding prompt handling, organization, and management. Our empirical survey study, involving 61 genAI tool users, addresses this gap by investigating the usability and user experience of the current features of these tools. We illustrate that advanced search and labeling functionalities and innovative interface designs can significantly enhance user experience as well as aid in reflecting on sustainability when using this technology. As genAI approaches the so-called "Trough of Disillusionment" (in Gartner's Hype Cycle terms),[1] our research aims to guide the design of genAI tools toward a more pragmatic and practical fit to end-user practices, ensuring that technology adoption comes with a deeper understanding of its capabilities and offerings.

CCS Concepts

• **Human-centered computing → Empirical studies in interaction design**.

Keywords

Generative AI, AI chatbot, LLM, Empirical Study, User survey, Prompt management

1 Introduction

Generative AI tools have rapidly gained popularity, with ChatGPT alone attracting 100 million users and 590 million visits just two months after its launch in January 2023 [16]. This surge underscores the transformative potential of the genAI market, which is expected to grow at an annual rate of 24.4% from 2023 to 2030, significantly impacting user interaction paradigms [10, 16]. Against this backdrop, HCI researchers have outlined how designers can effectively employ these tools to support Human-Centered Design

---

[1]https://www.gartner.com/en/research/methodologies/gartner-hype-cycle

processes [7], reflected on how "Large Whatever Models" can transform HCI research cycles [20], and examined their potential and implications for HCI education and pedagogy [11].

Despite widespread adoption and use, there is a notable gap in HCI research focusing on the design and usability of these tools, particularly concerning prompt management, handling, and organization. While recent HCI and computing studies have started to widely explore prompt engineering (*e.g.,* [5, 14, 26]), the practical aspects of how users handle, organize, and manage prompts remain largely under-investigated. In turn, efficient prompt management is crucial for reducing inefficiencies, which can lead to frustrating user experiences, and mitigating environmental costs associated with redundant interactions.

Our empirical survey study addresses this gap by investigating the usability and user experience of current genAI tools. We collected data from 61 text-based genAI tool users (*e.g.,* ChatGPT, Claude, *etc.*) to understand their routines, challenges, and needs regarding prompt management, *i.e.,* concerning organizing past prompts, their overall experience with chatbot interfaces, and the effectiveness of existing features for managing interaction histories. This research aims to guide the design of more intuitive and efficient genAI tools, with the goal of enhancing user experience throughout their use. Additionally, it calls for reflection on supporting sustainable consumption practices due to the growing energy demands of such technology [4]. Consequently, we aim to answer the following research question:

> How do users perceive the current prompt management features in genAI tools, and what specific improvements do they seek to enhance their interaction experience?

The gathered insights help us identify opportunities to make prompt management more intuitive and efficient. Our insights help enhance how genAI tools are designed, operated, and adopted.

2 Related work

Our work draws on the related research in chatbot use and design published during the last five years. Maroengsit *et al.* provide a comprehensive review of chatbot evaluation methods, focusing on effectiveness, efficiency, and user satisfaction [15]. Hussain *et al.* discuss the evolution of chatbots from simple scripts to advanced systems using AI and machine learning to enhance NLP capabilities [9]. Singh and Thakur analyzed the development of chatbots, emphasizing the transition to more sophisticated systems [23]. They highlighted the integration of Semantic Nets and Machine Learning to improve chatbots' ability to remember facts from conversations

and accurately identify user intents, which enhances the overall interaction quality. Almansor and Hussain's survey examines the role of AI in chatbots, identifying key challenges and future research areas in HCI: maintaining contextual understanding, generating contextually appropriate responses, and incorporating sentiment analysis [2].

Furthermore, our work is motivated by nascent research in improving the usability of AI chatbots. Akma *et al.* examined design techniques relevant to sectors such as education, healthcare, and customer service [1], while Vishwakarma *et al.* focused on methodologies for constructing chatbots [25]. Borsci *et al.* identified 27 key attributes for user satisfaction in CRM chatbots, aiding in the evaluation and design of chatbot interactions [3]. Among the most relevant to our study are response time, the chatbot's ability to handle multi-thread conversations, and the perceived ease of use. Ren *et al.* conducted a systematic mapping study on the usability of chatbots, particularly personal assistants in healthcare, pinpointing crucial measures such as satisfaction, efficiency, and effectiveness [18]. Despite these recent research efforts, there is a need for more targeted research on the user interface and user experience aspects of genAI tools, which is practically absent at the moment.

3 Study Design

We designed a cross-sectional survey following the guidelines by Kitchenham and Shull [12, 22]. The first author created the survey instrument, while the second author reviewed it. The feedback led to minor corrections in the wording of some questions, as no major problems were identified.

3.1 Survey Instrument

The survey was divided into four sections with a mix of single-select, multiple-choice, and open-text questions:[2]

(1) *User Familiarity and Use Cases*: This section included seven questions to gauge respondents' familiarity with generative AI tools and identify common use cases.
(2) *Prompt Management Challenges*: Four questions aimed to understand the difficulties users face when revisiting past conversations, locating specific information, and managing prompts.
(3) *Satisfaction with Current Features*: Seven questions assessed user satisfaction with existing features of generative AI tools, especially those related to prompt management.
(4) *Desired Features for Improvement*: This section included seven questions soliciting user opinions on potential features that could enhance prompt management.

We collected data for a month, from April 14 until May 15, 2024, and used Google Forms to administer the questionnaire. To ensure completeness, all closed-ended questions were marked as mandatory, minimizing the risk of partial responses.

3.2 Population and sample

The target population comprised individuals with experience using text-based genAI tools (*e.g.,* ChatGPT, Gemini, Claude), without restriction to specific tools or domains of use. The main sampling

---

[2]The questionnaire is available at: https://doi.org/10.6084/m9.figshare.26012959.v1

method was self-recruiting [17]. We published invitations on the social media platform LinkedIn. The secondary sampling method involved sending direct invitations to individuals from our professional networks.

3.3 Data Analysis and Validation

Responses were automatically collected in a Google Sheet, which is available in the online supplementary material. We used frequency analysis to analyze the single- and multiple-choice responses and employed an open coding strategy to analyze responses to open-ended questions, aiming to extract concrete feature requests and identify overarching themes pertinent to the usability of prompt management [22]. Through recurrent meetings, we reviewed the emerging clusters of code groups and conceptualized the themes. We refined them through iterative discussions, resolving ambiguities in code grouping and reaching a consensus on the final naming and composition of the high-level themes.

4 Survey Results

We received a total of 61 responses. In terms of the primary work domain, the majority of respondents (37) were employed in the Technology/IT sector, followed by 13 in Technology/IT with a specific focus on Education. Other domains included Pharma and Healthcare, see Figure 1. A significant number of respondents, especially those in technology and education, interact with genAI tools frequently: 27 individuals use them several times a week and 21 use them daily. Lastly, the most common purposes for using genAI tools were communication and writing (82%) as well as assistance in developing code (80.3%). Other uses of genAI include educational aid, such as support in learning new topics or subjects, including tutoring or explanatory content (63.9%), and data analysis (23%).

We asked participants about the challenges of reviewing past conversations, locating specific information, finding prompts, and sorting or filtering conversation histories. The responses indicate that most users (59%) rarely or never review their past conversations with genAI tools (see Figure 2). Additionally, 39% of the respondents said they never needed to find a past prompt, while 49% reported that it is difficult to find a past prompt. Lastly, 34% of the respondents never felt a need to sort or filter their past conversations, while 55% do not feel particularly restricted by the lack of sorting/filtering options.

We also asked participants their opinions about a few concrete features related to prompt management. 49% of the respondents said that history sorted by date is not particularly useful, while 42% indicated that a full-text search on past history would be helpful. Lastly, 42% said tagging or labeling of prompts could be very useful, with a total of 64% generally finding it useful.

Finally, we asked participants to list desired features to aid their workflows with the genAI tools. The following features have been specifically requested by our participants:

(1) *Prompt Suggestion System*: Dynamic suggestions for completing prompts based on initial input.
(2) *Conversation Management Options*: Ability to choose whether to save or delete conversations after completion.
(3) *Collaborative Use*: Shared access to AI functionalities among users in the same project.

(a) Primary domain

(b) Primary role

Figure 1: Respondent background

(a) How often respondents revisit history
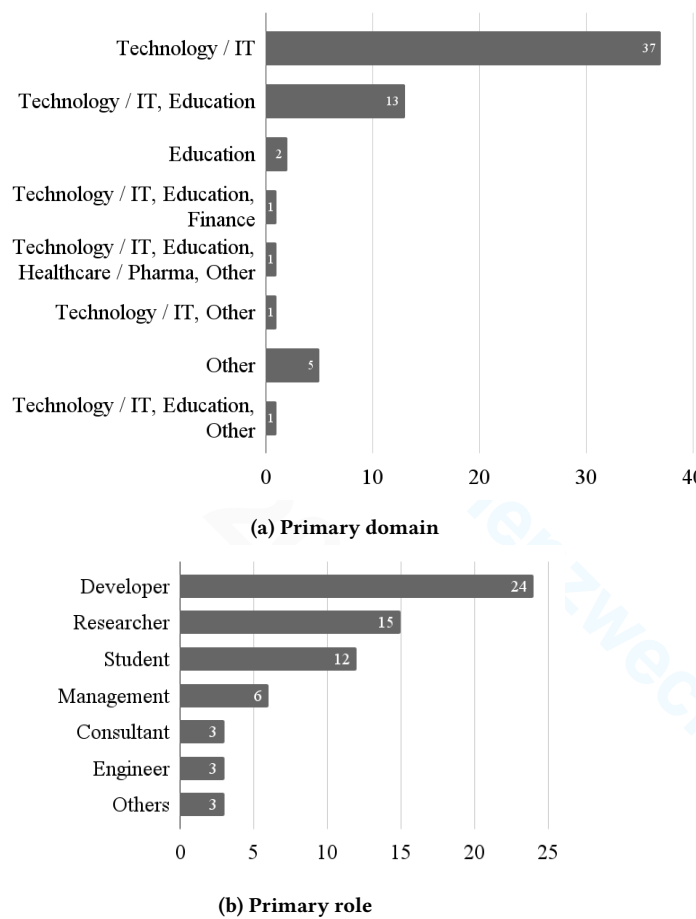

(b) Ease of finding information

Figure 2: Finding information

(4) *AI-assisted Search for Conversations*: Chatbot assistance in finding specific past conversations.
(5) *Chatbot-Specific Prompt Lists*: Lists of effective prompts tailored to each chatbot's capabilities and updates.
(6) *Template System with Placeholders*: Customizable templates for generating content.
(7) *Tagging for Interlinking*: Use of tagging to organize and interlink concepts.

5 Discussion

Our findings highlight improvements in the design of generative AI tools. Additionally, they open discussions around the impact these aforementioned features have on user behavior and sustainability.

5.1 Redesigning for Efficient Prompt Management

Our survey revealed that users rarely review past conversations due to ineffective search functionalities in genAI tools, which prioritize new prompts over searching history. This encourages users to re-prompt rather than search past interactions.
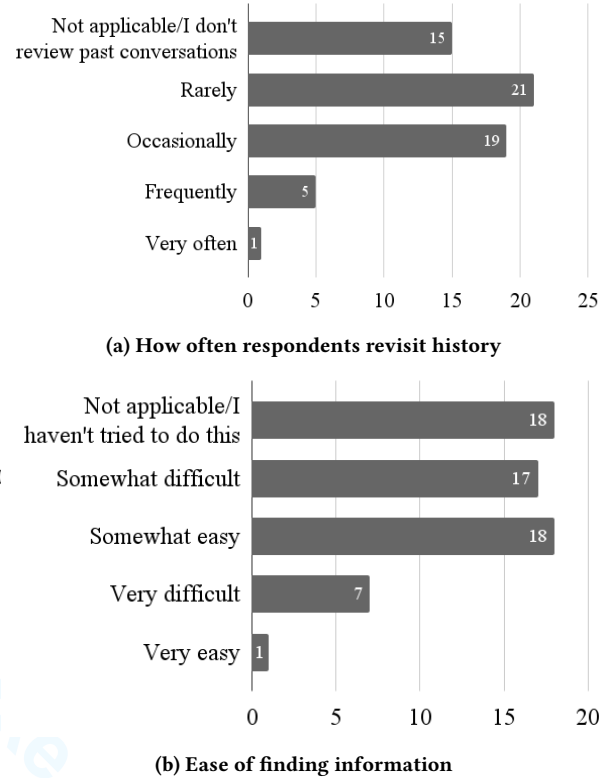
Respondents did not use typical tasks like searching, sorting, and filtering because these features were unavailable. This highlights the need for genAI tool designs that integrate these tasks to enhance user workflows. GenAI tools and instant messaging (IM) apps like WhatsApp rely on conversation as the main interaction type, though they are organized differently. For instance, unlike IM tools, genAI tools involve users "searching" for an answer and following up with prompts until they are satisfied with the response. Additionally, IM apps support multimodal interactions (voice, images, video), while genAI tools, like ChatGPT, primarily focus on textual questions and answers. Despite these differences, features like searching and tagging are still missing in genAI tools.

We speculate that our respondents rarely search their history, potentially due to the suboptimal and immature design of contemporary genAI tools, even though this conceptual model is common in text-centric interfaces. Improving search functionalities and organizing conversation history can enhance the effectiveness and efficiency of user interactions with genAI tools. This involves technical and UX challenges, such as deciding if searches should cover entire conversations or specific prompts and how tagging can improve efficiency. Finding specific past prompts would enable new interaction possibilities, such as initiating a new conversation thread from an existing one, while also presenting challenges, such as preserving and maintaining context. We envision these improvements addressing current limitations and creating more user-friendly and sustainable genAI tools.

5.2 The Environmental Toll of Prompting

Our participants expressed a clear preference for re-prompting rather than searching through history: "If I need a response again, I just ask the same question. It's faster for me than searching for an answer. Also, I'm basically never looking for an answer to a question I asked in the past," and "I'm not sure why to store or search history when you can always ask again and get the answer. You can also get a summary at the end by asking for it." This routine of re-prompting highlights a major design issue and leads to significant environmental costs.

Recent research highlights the significant energy demands of large language model (LLM) chatbots and AI in general [4], which require GPUs or TPUs for each new prompt. While the energy use and emissions from a single inference are minor, the cumulative environmental impact of global chatbot services is substantial. For instance, the monthly energy and carbon footprints of these services are comparable to those from their final training sessions [10].

Similar to trust-supporting design elements in online platforms (*e.g.,* [8]), developing interface features that encourage the reuse of past interactions can reduce redundant prompts and lower the carbon footprint associated with genAI use [10, 13, 24]. We hypothesize that emphasizing the effects of each prompt/transaction using simple graphical representations in the UI, such as potential CO2 emissions generated or signifiers alike, would increase end-user understanding of the effects of genAI usage on the environment. We draw inspiration from a study by Abbing [19], where the author discusses a solar-powered website built as a low-tech solution, demonstrating how design can play a pivotal role in reducing the environmental footprint of digital technologies. The author employed principles from degrowth [21] and sustainable HCI [6] practices to build a static site structure with minimal client-side computation, significantly minimizing energy use. We envision that such an innovative approach could be uniquely tailored to guide the design of sustainable genAI tools.

6 Conclusion

Our study explores the challenges and opportunities in prompt management for genAI tools, focusing on improving user experiences with them. The findings indicate that the lack of effective searching and intuitive organization of prompts and conversations leads users to repeatedly prompt for similar queries, which is environmentally inefficient and contributes to a suboptimal user experience.

To aid usability and sustainability, future designs should include novel interaction capabilities, such as adequate search functionalities to facilitate better organization and retrieval of past interactions and 'environment-aware' signifiers to reduce re-prompting. Implementing features like dynamic prompt suggestions, prompt templates, conversation management options, and AI-assisted search can significantly enhance end-user workflows and aid in the sustainability of such systems.

In our further work, we aim to explore the ecological impact of generative AI tools and propose strategies of reducing redundant prompts through efficient prompt management features. We look forward to feedback from HCI researchers during the workshop on their prompt management practices with genAI, beyond text-based tools, which was the primary subject of our inquiry.

References

[1] Nahdatul Akma Ahmad, Mohamad Hafiz Che, Azaliza Zainal, Muhammad Fairuz Abd Rauf, and Zuraidy Adnan. 2018. Review of chatbots design techniques. *International Journal of Computer Applications* 181, 8 (2018), 7–10.

[2] Ebtesam H Almansor and Farookh Khadeer Hussain. 2020. Survey on intelligent chatbots: State-of-the-art and future research directions. In *Complex, Intelligent, and Software Intensive Systems: Proceedings of the 13th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2019)*. Springer, 534–543.

[3] Simone Borsci, Alessio Malizia, Martin Schmettow, Frank Van Der Velde, Gunay Tariverdiyeva, Divyaa Balaji, and Alan Chamberlain. 2022. The chatbot usability scale: the design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and ubiquitous computing* 26 (2022), 95–119.

[4] Alex de Vries. 2023. The growing energy footprint of artificial intelligence. *Joule* 7, 10 (2023), 2191–2194.

[5] Paul Denny, Viraj Kumar, and Nasser Giacaman. 2023. Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 1136–1142.

[6] Carl DiSalvo, Phoebe Sengers, and Hrönn Brynjarsdóttir. 2010. Mapping the landscape of sustainable HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1975–1984. https://doi.org/10.1145/1753326.1753625

[7] Passant Elagroudy, Jie Li, Kaisa Väänänen, Paul Lukowicz, Hiroshi Ishii, Wendy E Mackay, Elizabeth F Churchill, Anicia Peters, Antti Oulasvirta, Rui Prada, et al. 2024. Transforming HCI Research Cycles using Generative AI and "Large Whatever Models"(LWMs). In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–5.

[8] Anton Fedosov, Liudmila Zavolokina, Sina Krumhard, and Elaine M Huang. 2023. "This Could Be The Day I Die": Unpacking Interpersonal and Systems Trust in a Local Sharing Economy Community. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[9] Shafquat Hussain, Omid Ameri Sianaki, and Nedal Ababneh. 2019. A survey on conversational agents/chatbots classification and design techniques. In *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019) 33*. Springer, 946–956.

[10] Peng Jiang, Christian Sonne, Wangliang Li, Fengqi You, and Siming You. 2024. Preventing the Immense Increase in the Life-Cycle Energy and Carbon Footprints of LLM-Powered Intelligent Chatbots. *Engineering* (2024).

[11] Ahmed Kharrufa and Ian G Johnson. 2024. The Potential and Implications of Generative AI on HCI Education. *arXiv preprint arXiv:2405.05154* (2024).

[12] Barbara A Kitchenham, Shari Lawrence Pfleeger, Lesley M Pickard, Peter W Jones, David C. Hoaglin, Khaled El Emam, and Jarrett Rosenberg. 2002. Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on software engineering* 28, 8 (2002), 721–734.

[13] Baolin Li, Yankai Jiang, Vijay Gadepally, and Devesh Tiwari. 2024. Toward Sustainable GenAI using Generation Directives for Carbon-Friendly Large Language Model Inference. *arXiv preprint arXiv:2403.12900* (2024).

[14] Vivian Liu and Lydia B Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–23.

[15] Wari Maroengsit, Thanarath Piyakulpinyo, Korawat Phonyiam, Suporn Pongnumkul, Pimwadee Chaovalit, and Thanaruk Theeramunkong. 2019. A survey on evaluation methods for chatbots. In *Proceedings of the 2019 7th International conference on information and education technology*. 111–119.

[16] Dan Milmo. 2023. ChatGPT reaches 100 million users two months after launch. *The Guardian* 2 (2023).

[17] Teade Punter, Marcus Ciolkowski, Bernd Freimut, and Isabel John. 2003. Conducting on-line surveys in software engineering. In *2003 International Symposium on Empirical Software Engineering, 2003. ISESE 2003. Proceedings*. IEEE, 80–88.

[18] Ranci Ren, Mireya Zapata, John W Castro, Oscar Dieste, and Silvia T Acuña. 2022. Experimentation for chatbot usability evaluation: A secondary study. *IEEE Access* 10 (2022), 12430–12464.

[19] Roel Roscam Abbing. 2021. 'This is a solar-powered website, which means it sometimes goes offline': a design inquiry into degrowth and ICT. In *LIMITS'21, June 14-15 2021, Virtual workshop*. PubPub.

[20] Albrecht Schmidt, Passant Elagroudy, Fiona Draxler, Frauke Kreuter, and Robin Welsch. 2024. Simulating the Human in HCD with ChatGPT: Redesigning Interaction Design with AI. *Interactions* 31, 1 (jan 2024), 24–31. https://doi.org/10.

1145/3637436

[21] Vishal Sharma, Neha Kumar, and Bonnie Nardi. 2023. Post-growth Human–Computer Interaction. *ACM Trans. Comput.-Hum. Interact.* 31, 1, Article 9 (nov 2023), 37 pages. https://doi.org/10.1145/3624981

[22] Forrest Shull, Janice Singer, and Dag IK Sjøberg. 2007. *Guide to advanced empirical software engineering.* Springer.

[23] Siddhant Singh and Hardeo K Thakur. 2020. Survey of various AI chatbots based on technology used. In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO).* IEEE, 1074–1079.

[24] Jovan Stojkovic, Esha Choukse, Chaojie Zhang, Inigo Goiri, and Josep Torrellas. 2024. Towards Greener LLMs: Bringing Energy-Efficiency to the Forefront of LLM Inference. *arXiv preprint arXiv:2403.20306* (2024).

[25] Ashutosh Vishwakarma and Ankur Pandey. 2021. A review & comparative analysis on various chatbots design. *International Journal of Computer Science and Mobile Computing* 10, 2 (2021), 72–78.

[26] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–21.