"I Didn't Share Anything Personal": Privacy Perceptions and Behaviors of Low-Income, Low-Digital-Literacy ChatGPT Users in India

Jasmeet Kaur
CISPA Helmholtz Center for Information Security

Jane Im
CISPA Helmholtz Center for Information Security

1 Introduction

With the increasing pervasiveness of Large Language Models (LLMs), marginalized populations, particularly those having limited digital literacy and living in low-resource settings, are also becoming exposed to these technologies [5, 7]. While LLMs offer significant benefits, they also introduce privacy issues, including memorization risks and the potential use of user inputs for model training [4, 6, 9, 12]. Prior research has shown how different populations perceive data storage, personally identifiable information (PII) and inferences in LLM interactions [2, 3, 11, 13, 14]. However, work at the intersection of LLMs and user privacy in the context of marginalized populations is underexplored. This gap is especially pronounced for women with limited digital literacy belonging to low-income communities and living in patriarchal contexts, which shape their use of technologies [1, 10]. Therefore, we used a human-centered approach to understand these women's privacy behavior in their interactions with LLMs and identify insights for designing systems that better support their privacy. Specifically, our research questions are:

1. How do users with limited digital literacy perceive sensitive information and privacy risks in LLM interactions?

2. How do users manage privacy risks in conversational AI systems, and what support do they want for safer use?

In our study, we selected ChatGPT due to its widespread adoption in India, having over 100 million weekly active users.[1] This poster presents preliminary findings showing how participants' privacy literacy and usage preferences shape their interactions with LLMs, and how their privacy behavior evolves with awareness of privacy risks.

2 Methodology

Our target population comprised low-income Indian women with limited digital literacy. We recruited 15 participants, who self-expressed having low digital literacy, through social networks and snowball sampling, leveraging the first author's five years of trusted community engagement with marginalized populations [8]. The study received institutional IRB approval.

We first conducted an introductory Zoom session on using ChatGPT, including basic querying practices and suggested participants to check out the settings without explicitly mentioning privacy settings. Participants then used ChatGPT for five days and completed daily diary entries via Google Forms to capture their information-sharing practices and emerging privacy concerns. The diary study included prompts such as "Did you have any questions regarding sharing information with ChatGPT?"

We then conducted semi-structured Zoom interviews, where we examined participants' understanding of how ChatGPT memorizes and uses data for training, personalization, and potential privacy risks. Next, we explored participants' existing privacy settings to assess how they balanced personalization with privacy. To probe understanding of PII and inferences, we developed and used a prototype mimicking ChatGPT that highlighted PII in a sample query and demonstrated possible inferences, along with data minimization options such as deletion and generalization [11]; participants were asked to identify PIIs, inferences, risks, and evaluate the options. Interview questions also included: "How would you feel if ChatGPT remembers information about you and creates

USENIX Symposium on Usable Privacy and Security (SOUPS) 2024.
August 11–13, 2024, Philadelphia, PA, United States.

[1] https://www.socialsamosa.com/news-2/india-100-million-weekly-active-chatgpt-users-11121146

a profile of you over time?", "Why have you set these particular privacy settings?" Finally, participants completed a second five-day diary study to capture changes in their interactions after the interview.

We analyzed Hindi interview recordings and Hinglish [2] diary entries using inductive thematic analysis. We coded interviews directly from recordings and translated only the excerpts selected as quotations into English. Codes from both datasets were iteratively refined into themes.

3 Preliminary Findings

We present insights from interviews and diary studies showing participants' personalization preferences, privacy understanding, and emerging data-sharing practices with ChatGPT.

3.1 Strong Preference for Personalization and Utility Over Privacy Boundaries

Participants primarily approached ChatGPT as a search-oriented tool, similar to Google Search, expecting direct and task-focused responses and valued its free access, and on-demand availability in their low-income contexts. Participants strongly preferred personalized outputs, driven by the need to avoid repeatedly entering the same information, reduce incomplete queries, and obtain more accurate and contextually relevant answers. Furthermore, they rarely articulated cases in which personalization might become intrusive, and prioritized functional benefits over privacy risks. As a result, participants tended to rely on default privacy settings to support personalization and improve response quality.

Participants' risk perceptions varied by data type. They seemed to largely perceive text inputs as low-risk, leading to frequent sharing of sensitive health information, including medical test reports. However, textual queries containing financial information were consistently perceived as highly sensitive due to concerns about fraud. Participants viewed face-visible images as risky, particularly due to their prior exposure to deepfake news on social media, yet such images were still commonly shared in practice. Similarly, intimate or personal conversations (e.g., about relationship with husband) were acknowledged as potentially risky (e.g., due to risk of blackmail) by a few participants, while some participants normalized such conversations.

3.2 Limited Understanding of PIIs, Inferences, and Privacy Controls

Participants expressed confidence that they do not share personal information in their queries. For instance, P02 said, "*I do not give any personal information. I just ask general things.*" However, this confidence reflects a limited and fragmented understanding of privacy. Most participants were unaware of how their data may be used for training, storage, or memorization, and showed little familiarity with privacy settings or consent mechanisms.

Moreover, participants demonstrated a relatively low understanding of PII. Six out of 15 participants identified one or no PII items, while only five participants were able to identify all seven. PII perception was narrowly defined, largely restricted to financial details, with minimal awareness of risks arising from data aggregation.

Participants' understanding of inferences was also limited. While some participants (8/15) could identify an income-related inference, about half of the participants (7/15) could not identify even one. Importantly, none recognized the possibility of a demographic inference (e.g., gender). We also note that once becoming aware of inferences, many participants predominantly viewed them as beneficial, rather than as a potential privacy risk, because it enabled improved responses.

3.3 Emergence of Informed and Controlled Data-Sharing Practices

Based on the diary entries, we observed that before the interviews, participants assumed sharing information with ChatGPT was safe, leading to unfiltered and often risky disclosures. Afterward, they became more selective and adopted basic anonymization strategies (e.g., removing identifiable details or sharing partial information), though most continued using default privacy settings. Some also began to question whether ChatGPT employees might access user conversations for system improvement and potentially misuse personal data, indicating a growing but still limited awareness of privacy risks.

Further, participants predominantly viewed data protection as an individual responsibility, emphasizing the need to limit what they share rather than relying on systemic safeguards. At the same time, when viewing our potential designs, they expressed clear appreciation for features that could actively support safer interactions, such as alerts for potential oversharing, clearer indications of risk, and tools that help balance personalization with privacy. Participants also strongly desired greater transparency in data practices, alongside preferences for enhanced control mechanisms, including auto-deletion of sensitive chats, time-bound memory retention, and clearer visibility and control over stored data.

4 Future Work

Our analysis of this work is ongoing and upon completion, we aim to provide an empirical understanding of LLM-related privacy literacy among low-digitally literate populations. We also aim to offer specific design implications for developing privacy mechanisms that better support such populations.

[2] A hybrid language that blends Hindi and English

References

[1] Mahnaz Akhter. Digital divide and its impact on rural and tribal women in india: A gendered analysis of access, education, and empowerment.

[2] Sumit Asthana, Jane Im, Zhe Chen, and Nikola Banovic. " i know even if you don't tell me": Understanding users' privacy preferences regarding ai-based inferences of sensitive information for personalization. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2024.

[3] Rahime Belen Saglam, Jason RC Nurse, and Duncan Hodges. Privacy concerns in chatbot interactions: When to trust and when to worry. In *International Conference on Human-Computer Interaction*, pages 391–399. Springer, 2021.

[4] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. Extracting training data from large language models. In *30th USENIX security symposium (USENIX Security 21)*, pages 2633–2650, 2021.

[5] Roshini Deva, Dhruv Ramani, Tanvi Divate, Suhani Jalota, and Azra Ismail. "kya family planning after marriage hoti hai?": Integrating cultural sensitivity in an llm chatbot for reproductive health. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.

[6] Huseyin A Inan, Osman Ramadan, Lukas Wutschitz, Daniel Jones, Victor Rühle, James Withers, and Robert Sim. Training data leakage analysis in language models. *arXiv preprint arXiv:2101.05405*, 2021.

[7] Jasmeet Kaur and Pushpendra Singh. Exploring the potential of chatgpt for supporting family planning needs of postpartum women in resource-constrained areas of india. *Proceedings of the ACM on Human-Computer Interaction*, 9(7):1–26, 2025.

[8] Jasmeet Kaur and Pushpendra Singh. Exploring the potential of peer support group for family planning needs of women in resource-constrained settings in india. *Proceedings of the ACM on Human-Computer Interaction*, 9(2):1–25, 2025.

[9] Milad Nasr, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A Feder Cooper, Daphne Ippolito, Christopher A Choquette-Choo, Eric Wallace, Florian Tramèr, and Katherine Lee. Scalable extraction of training data from (production) language models. *arXiv preprint arXiv:2311.17035*, 2023.

[10] Parnika Saha, Ajay Kumar Prusty, and Chinmaya Nanda. Extension strategies for bridging gender digital divide. *Journal of Applied Biology & Biotechnology*, 12(4):76–80, 2024.

[11] Synthia Wang, Sai Teja Peddinti, Nina Taft, and Nick Feamster. Beyond pii: How users attempt to estimate and mitigate implicit llm inference. *arXiv preprint arXiv:2509.12152*, 2025.

[12] Chiyuan Zhang, Daphne Ippolito, Katherine Lee, Matthew Jagielski, Florian Tramèr, and Nicholas Carlini. Counterfactual memorization in neural language models. *Advances in Neural Information Processing Systems*, 36:39321–39362, 2023.

[13] Zhiping Zhang, Michelle Jia, Hao-Ping Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. "it's a fair game", or is it? examining how users navigate disclosure risks and benefits when using llm-based conversational agents. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–26, 2024.

[14] Jijie Zhou, Eryue Xu, Yaoyao Wu, and Tianshi Li. Rescriber: Smaller-llm-powered user-led data minimization for llm-based chatbots. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–28, 2025.