

User Issues and Concerns in Generative AI: A Mixed-Methods Analysis of App Reviews

Vanessa Bracamonte¹, Sascha Loebner², Frederic Tronnier², Ann-Kristin Lieberknecht², and Sebastian Pape²
¹KDDI Research, Inc., ²Goethe University Frankfurt
 va-bracamonte@kddi-research.jp

Introduction

Generative AI models such as ChatGPT and Stable Diffusion have become easily available to end users through various apps. Research has identified several safety risks and limitations of generative AI, but the experiences and issues faced by real users of this technology in the wild have not been systematically investigated. In this study, we investigate the opinions, complaints and concerns found in user reviews of use generative AI-based apps. Our objective to understand the types of issues that real users encounter when interacting with generative AI through these apps.

Method

Data Collection / Processing

- We targeted US Google Play Store apps with a description related to AI text (chatbot) and image generation. We identified 62 Google Play apps (35 image, 27 text including the ChatGPT app) and 249,482 user reviews were collected by November 9, 2023. The data was cleaned and processed for analysis.

Topic Modeling

- We first conducted topic modeling analyses, separately for the image, text and ChatGPT app user reviews.
- The appropriate number of topics was identified by evaluating topic coherence values on models with 2 to 20 topics. The selected models consisted of 9 topics for the image generation apps, 8 topics for the text generation apps and 7 topics for the official ChatGPT app user reviews.

Example of topics and top keywords

In Image generation apps	
Output errors	put word type thing prompt face draw stuff show turn simple character people search picture write description
Output quality	prompt result style option lot add feature generation user perfect choose bit detail imagine render realistic
In Text generation apps	
AI issues and limitations	chat gpt people update talk conversation feel friend fun real bot language human world fact put program datum
Answer performance	question answer give write information provide essay story accurate point simple wrong topic interesting
In the ChatGPT app	
AI issues	great thing time problem nice write review word day change solve bad improve save math month reply stuff
Data-specific limitations	information update knowledge free time datum access world 2021 people real human september info pay ad

Qualitative Analysis

- We then conducted a qualitative analysis, using an inductive approach to identify themes in the user reviews.
- We randomly sampled 20 user reviews from each topic, for each type of app. The total sample was 480 user reviews for the qualitative analysis.
- Each review was independently coded.

Results

Each identified theme was framed according to the AI trustworthiness dimensions of validity, reliability, safety, security, privacy, transparency and bias. We grouped the themes in the user reviews into Performance, Threat concern, Personalization and Censorship and bias.

Theme	Framing	Img	Txt	Chat GPT
Performance				
• Incorrect output	Validity / Transparency	✓	✓	✓
• Unasked for output	Reliability	✓	✓	
• AI limitations	Validity / Transparency	✓	✓	✓
• Issues with AI model used	Validity / Transparency	✓	✓	✓
Threat concern				
• Threat from others using the AI	Security	✓	✓	✓
• Threat from the AI itself	Security	✓	✓	✓
Personalization				
• Would give user information for personalization	Privacy		✓	✓
• Benefited by sharing user information	Privacy	✓	✓	✓
Censorship and bias				
• General prompt considered inappropriate	Safety	✓	✓	✓
• Inappropriate content from a general prompt	Safety	✓	✓	
• Opinion on censorship	Safety	✓	✓	✓
• Censorship options	Safety	✓	✓	✓
• Censorship rules	Safety / Transparency	✓	✓	✓
• Bias	Bias	✓	✓	✓

“... Overall, very limited information to be given by the AI, but still extremely useful in its own ways. It definitely has its drawbacks with its capabilities, like the usage of websites and info from 2022-2023...”

“...just finishing reading how this program can literally be used by people to write malware even with no programming experience...”

“...but I do feel like it is dumb that ChatGPT forgets previous chats because of privacy. The user should be able to change it in settings, it is their choice to have more privacy or a better more personal experience...”

“...It's bizarre to me that certain words are prohibited in making the pictures but when benign words and phrases are used, pornographic pictures are being created.”

Conclusions

Overall, our findings indicate that users have expectations of generative AI that do not align with the current actual capabilities of these models. In addition, mentions of issues that can be framed as related to AI safety and validity were frequently found. On the other hand, issues and concerns which can be framed as related to security and privacy were not as prevalent.

References

- Liesenfeld, A., Lopez, A., Dingemans, M.: Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators. In: Proceedings of the 5th International Conference on Conversational User Interfaces (2023)
- Kim, Y., Lee, J., Kim, S., Park, J., Kim, J.: Understanding Users' Dissatisfaction with ChatGPT Responses: Types, Resolving Tactics, and the Effect of Knowledge Level. In: Proceedings of the 29th International Conference on Intelligent User Interfaces. pp. 385–404 (2024)
- Zhang, Z., Jia, M., Lee, H.P., Yao, B., Das, S., Lerner, A., Wang, D., Li, T.: "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 CHI Conference on Human Factors in Computing Systems. pp. 1–26 (2024)