GEMINI A.I. TOOL

RECOMMENDATION REPORT

Tamekia Brookins, Julia Lilly

Nancy Nguyen, Kendall Hayes

Table of Contents

Executive Summary

Purpose

To see how a newly branded Generative AI (Gemini) would compute a given task when prompted. Briefly summarize the key points of the report: the problem addressed, methodology, major findings, and your top recommendations.

Methodology

We conducted a “think aloud” with four participants. We then compiled our responses/comments with open coding to see the common problems/comments into three codes, and finally transcribed everything onto excel document.

Findings

Based on the UX research, several errors and areas for improvement were identified in the AI tool Gemini. Key recommendations include broadening the scope of targeted words to better understand context, allowing users to edit bot responses for greater control over the conversation, and defaulting to light mode for improved accessibility. Implementing these changes will enhance Gemini's user experience and conversational intelligence.

Introduction

Our team prepared this UX project and research to complete an in-class college group project. The intention behind the project was to provide a UX study. We used the scenario of using Generative A.I. to develop material for a business, specifically a tri-fold presentation. We chose Gemini by Google Corporation as the A.I. tool to test. This decision was made because it was released close to the start of the project, and we wanted to see how a newly branded A.I. would complete our given task.

Gemini A.I. is a tool created by Alphabet Inc. through their subsidiary Google. Their original A.I. was called Bard but was switched to Gemini in February 2024. It uses LaMDA for the Large Language Model behind it. Gemini has a plethora of applications such as generating and formatting text, preparing code, and looking up information.

This report will cover the procedures and analysis of suggestions our group has found while trying to generate information for our fake business presentation. The things we interact with were only the text generation. Our sample size was only 4, so our data is not the most tested for large feedback amounts. Another limitation we had for data is the time period and content to test it on. All four participants used the same information about a company to gather this information. Different types of information and different goals could yield

unmentioned results. Our data was collected from first impressions, as per a think-aloud protocol, and not focused on product testing the entire artificial intelligence site.

Methodology

Our UX data was collected by using a “think-aloud” protocol. A “think-aloud” protocol is a type of UX testing method where testers record users “out loud thinking” when using a product. Rather than using an alternative method, like say a survey, the “think-aloud” method allows researchers to view and understand users’ honest and open opinions about Gemini. Other testing methods can be limiting the users’ expressions about a product or not properly address a problem a user has faced.

To properly collect data, each participant conducted their own think-aloud protocols in their homes, recording their own session as they proceeded with the protocol. After the think-aloud process, each participant transcribed their own transcript with the use of their recording. As a control for this test, participants used made-up data for a faux pet store company. All participants used the same data and asked Gemini at least one similar request, to create content for a “trifold for new aquarium owners.”

With each finished transcript, the data was divided into two. Participants analyzed their data by using an open coding process. With an open coding process, the data was analyzed line-by-line. Two participants received two transcripts and the other two participants received the other two transcripts. With each line, the participants searched, marked and color coded any common patterns or oddities within each transcript. To ensure that no bias was curated, participants did not receive their own transcript. Each transcript was also viewed twice by two participants to ensure no patterns or oddities were missed.

After the data analysis process, one participant viewed all the transcripts to see what common problems or contentment were seen across the board. The same participant then grouped each commonly seen pattern into three codes, navigation (code 1), prompt production (code 2), and limitations (code 3). Each code also contained whether the data within it was positive or negative. All this data was then transcribed into an excel sheet.

The Code Book

Code 1	Code 2	Code 3
Navigation	**Prompt Production**	**Limitations**
Whether users were able to find and access tools that Gemini provides	User satisfaction with Gemini's results and how much of said product is produced	Any restrictions that limited user from creating desired product

Findings

The findings from participant feedback provide valuable insights into the usability and effectiveness of an AI system. Through a participant evaluation of navigation, prompt production, and AI policy limitations, key strengths and areas for improvement have been identified below.

Navigation

The navigation aspect of an AI system plays a crucial role in facilitating user interaction and ensuring a seamless experience. This section evaluates our participant's feedback regarding the ease of navigation, visual appeal, and accessibility of interface elements.

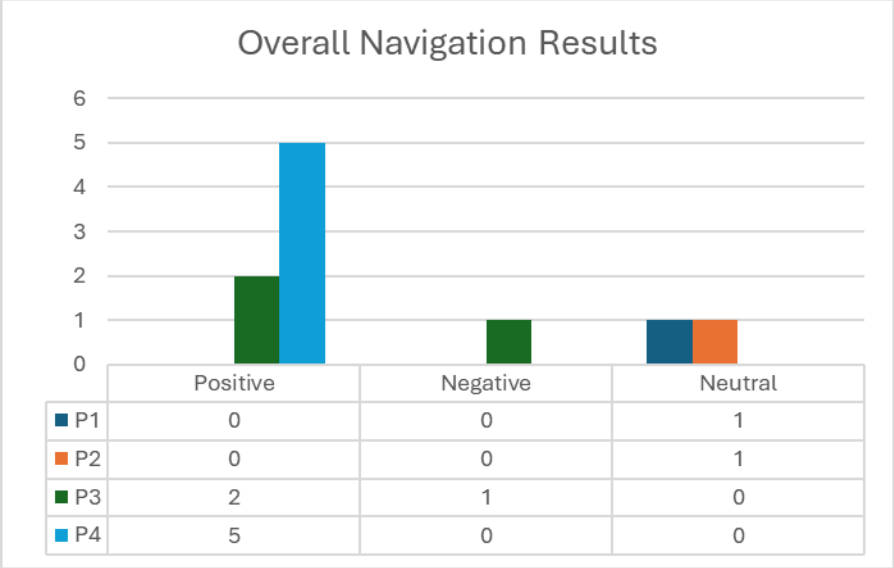

Figure 1. Participant Test Study results on Gemini AI's navigation. Results categorized by positive, negative, and neutral meaning no strong feelings either way.

Significant Positive Findings:

- Participants appreciate the dark/light mode.
- Participants appreciate the presence of helpful tools such as the audio input speaker and Python coding integration.

Significant Negative Findings:

- Participants have concerns regarding the visibility due to a dark background and light text color that potentially hinders quick navigation.

Summary:

While participants had mostly positive interactions with Gemini's navigation and acknowledged the presence of helpful navigation features, such as a speaker for voice prompts and the ability to use specialized coding, concerns regarding visibility issues should be noted. Addressing this single concern should create a more participant-friendly interface and create a good first impression, thus improving participant satisfaction at the onset.

Prompt Production

The prompt production of an AI system is a critical aspect influencing the quality and relevance of generated content. Therefore, it is not surprising that this area of our study received the most participant comments. This section assesses our participant's feedback regarding the clarity, relevance, and inclusiveness of prompts provided by Gemini.

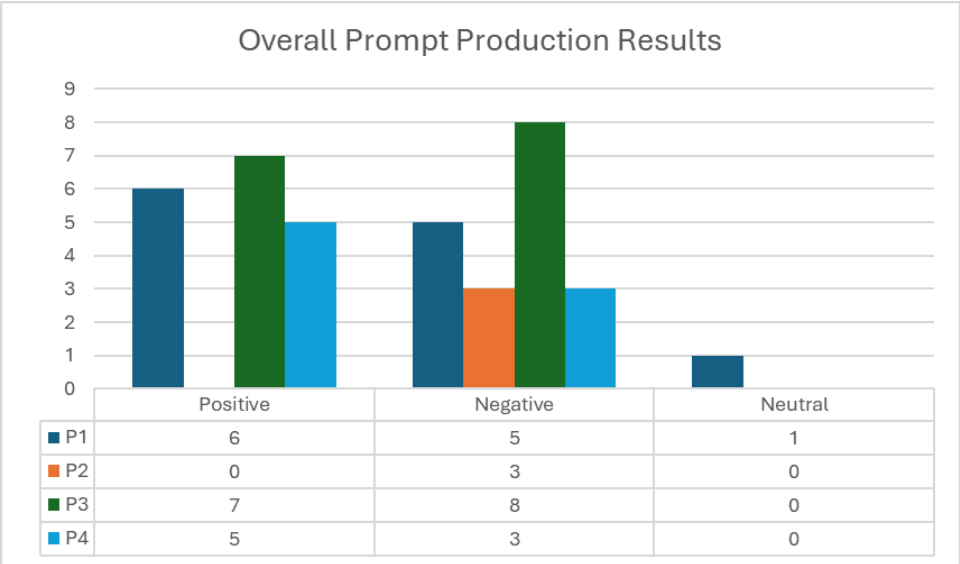

Figure 2. Participant Test Study results on Gemini AI's prompt production. Results categorized by positive, negative, and neutral meaning no strong feelings either way.

Significant Positive Findings:

- Participants appreciate the AI's ability to understand prompts and return relevant information.
- Participants appreciate the AI's intuitiveness when returning results for a specific layout.
- Participants appreciate the AI's inclusion of additional elements such as recommended sections and content based on traditional layouts. For example, one participant received a section to enter her company name and address.

Significant Negative Findings:

- Participants have concerns about autonomy due to the AI's guidance function, feeling it was overly instructive in some cases and lacked guidance or explanation in other cases.
- Participants have concerns about the AI's bulkiness in the amount of information returned.
- Participants have concerns about the AI's ability to remain consistent with the participants preferences when returning results on follow up / editing prompts.

Summary:

While participants appreciate the AI's ability to understand prompts and provide relevant information, as well as its intuitiveness in returning results for specific layouts, significant concerns arise regarding autonomy due to the AI's overly instructive guidance function. Participants also express reservations about the AI's bulkiness, as it returns a large amount of information, and its inconsistency in adhering to participant's preferences during follow-up or editing prompts. Additionally, while participants value the AI's inclusion of additional elements such as recommended sections and content based on traditional layouts, these positive aspects are tempered by concerns about usability and user control. These findings emphasize the need to refine the AI's guidance function, improve consistency, and optimize the amount of information returned to enhance overall user satisfaction and usability.

AI Policy Limitations

The policy limitations built into an AI system play an important role in the user's trust of the generated results. While it is important to build reasonable restraints to foster healthy learning and output by the AI, a policy that is too strict will create challenges that the user may not be able to overcome. This section examines our participant's

feedback regarding warnings, limitations, and technical issues experienced throughout the process.

Figure 3. Participant Test Study results on Gemini AI's Policy limitations. Results categorized by positive, negative, and neutral meaning no strong feelings either way.

Significant Positive Findings:

- Participants appreciate the AI system putting in safeguards when encountering prompts with consistent negative word input in one prompt but find it necessary to consider a work-around such as defining key terms that may trigger the AI policy.

Significant Negative Findings:

- Participants have concerns with the system's inability to respond to standard word count requests in prompts.
- Participants have concerns with the system generating prompts back to the user rather than writing content as requested.
- Participants have concerns with the system not being able to contextualize standard corporate jargon thus triggering the AI policy to shut down the prompt.

Summary:

The analysis of Limitations reveals a range of technical challenges and constraints experienced by participants during interaction with the AI system. Issues such as prompt loss, formatting difficulties, and constraints in participant control highlight areas for improvement in system reliability and participant experience. Addressing

these limitations will contribute to a smoother and more seamless interaction, enhancing overall participant satisfaction and usability.

Finding Analysis

The analysis of participant feedback on navigation, prompt production, and AI policy limitations provides valuable insights into the strengths and weaknesses of the Gemini AI system.

In the navigation section, participants appreciate the system's features like dark/light mode and helpful tools such as the audio input speaker and Python coding integration. However, concerns about visibility issues due to a dark background and light text color were noted, suggesting a need for improvement in this aspect to enhance user satisfaction.

Moving on to prompt production, participants commend the AI's ability to understand prompts and return relevant information, as well as its intuitiveness in returning results for specific layouts. However, significant concerns arise regarding the AI's overly instructive guidance function, bulkiness in returning information, and inconsistency in adhering to participant preferences during follow-up or editing prompts. While participants value the AI's inclusion of additional elements such as recommended sections, concerns about usability and user could negate these positive aspects, emphasizing the need for refinement.

In terms of AI policy limitations, participants appreciate the system's safeguards against negative word input but express concerns about its inability to respond to standard word count requests and its tendency to generate prompts back to the user rather than writing content as requested. Additionally, challenges in contextualizing standard corporate jargon raise concerns about the AI policy's strictness and its impact on user interaction.

Overall, addressing these technical challenges and constraints, such as prompt loss and formatting difficulties, is crucial to enhancing system reliability and participant experience and improving overall user satisfaction and usability.

Recommendations

Improve Accessibility

- Defaulting on light mode for the background would likely improve accessibility and readability for most users. Dark mode is popular but can cause issues with contrast and readability. Making light mode the default ensures the broadest accessibility while allowing a dark mode option.

Enable Conversational Threading

- Enabling Gemini to respond to follow-up questions in a conversational thread would make interactions feel more natural and user-friendly. Currently, each question is treated independently, but having the context of the conversation flow would improve coherence. This could be implemented by having a session ID to link sequential questions.

Allow Editing/Removing Previous Response

- Adding the ability to go back and edit or remove previous responses would give users more control over the conversation. Sometimes an AI response may need to be clarified or retracted, and allowing editing lets users guide the conversation better. This could work similarly to message editing in chat apps.

Expand Contextual Language Understanding

- Broadening the scope of targeted words/phrases and learning context would help Gemini understand nuance and avoid overly literal interpretations. Language is very contextual, so understanding slang, cultural references, sarcasm, etc. based on the broader conversational context would make Gemini more adaptable. This requires large diverse training datasets.

Conclusion

Through the comments from the participants while attempting to navigate the generative A.I. recently released as Gemini, our group was able to category the comments into navigating the site, the limitations of the tool, and how it produced the prompt. We further split the categories into tone of the comments with Positive, Neutral, and Negative codes. Our analysis showed that the most positive topic on the website was the ability to navigate easily. The most negatively viewed topic was the limitations that impede the user from completing their desired tasks. Prompt Production had roughly equal amounts of negative and positive comments.

We made a few suggestions that were of utmost importance to consider. The first suggestion is to talk with the Diversity and Inclusion department and programming department to broaden the scope of what terms are marked for harassment or derogatory searches. It will allow Gemini to become more adaptable and allow the user to have a more friendly experience with the tool. The next suggestion we would like to make is to switch the default viewing mode to light mode to allow easier reading of the site.

Google should take these recommendations seriously to avoid the potential risks of excluding certain groups by being too proactive with their stance on diversity and inclusion. Although our group was able to find a few problems and recommendations, our group's scope was small. We only had four participants and the direction of our research was the same for each person. Future research on these cases should expand in areas we could not. The number of participants should be increased alongside different demographic makeup. These changes to future research will allow the comments to have more variability to spot different suggestions others could miss.

Appendices

Appendix A: Think Aloud Response by Participant

- Participant 1
- Participant 2
- Participant 3
- Participant 4

Appendix B: Think Aloud Summary

- Participant Think Aloud Summary

Appendix C: Participant Code Book

- Finding Nemo Code Book

Appendix D: Additional Resources

- Pichai, Sundar, and Demis Hassabis. Introducing Gemini: Our Largest and Most Capable AI Model. Google, 2023. (Full Article)