Matteo Bombelli 301550507
Ryan Hargrave 301550214
Miles Grant 301478268
Weiqi Wang 301417203

ChatCompare: Generative-AI Chat Bot Comparison

1. Problem Statement

Our project aims to enhance the efficacy and usability of large language models (LLMs) for professionals working in the workplace and research environments. Despite the abundance of generative AI chatbots available today, there remains a significant gap: users lack an effective method to compare the performance and responses of different LLMs simultaneously.

Our primary target users are workplace professionals and researchers—such as business analysts, data scientists, and academic researchers—who rely on LLMs for informed decision-making, problem-solving, and streamlined workflows.

vs vs

2. Research and Interviews

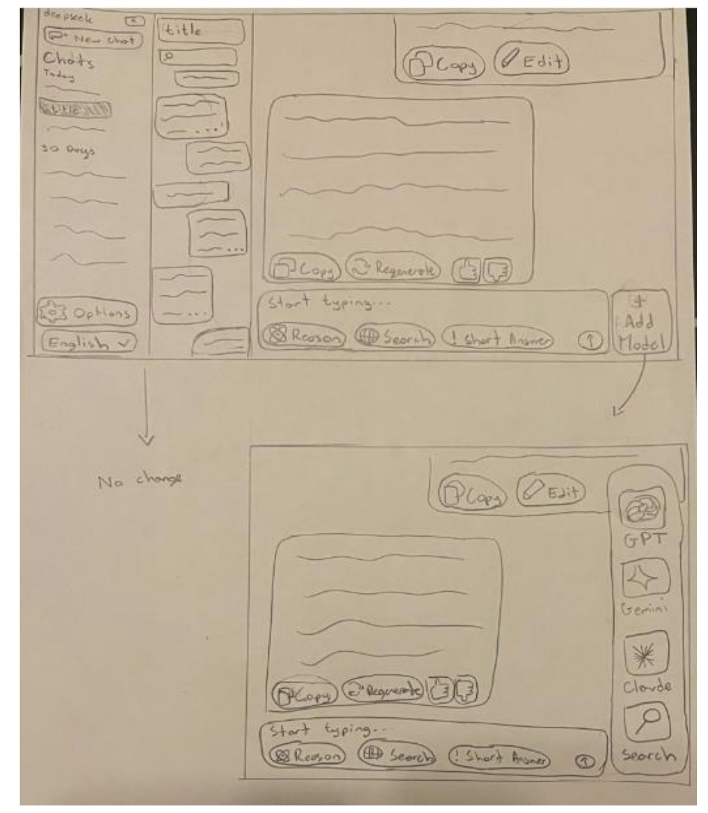
To determine what features we wanted to add to our chat bot, we first performed interviews with users to gain a better understanding of what they like and dislike about the currently available chat bots. Our interviews targeted users who use LLMs on a regular basis to gain better insight as to how we could improve the experience for people who were already familiar with the basic functions of LLMs. These users included software engineers and computer science students.

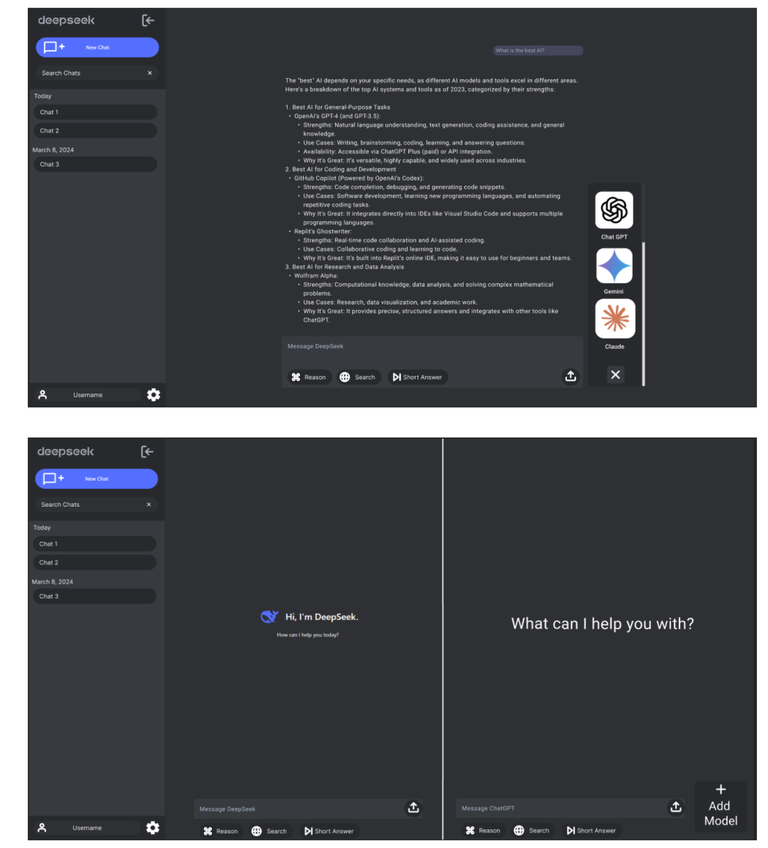
3. User Requirements

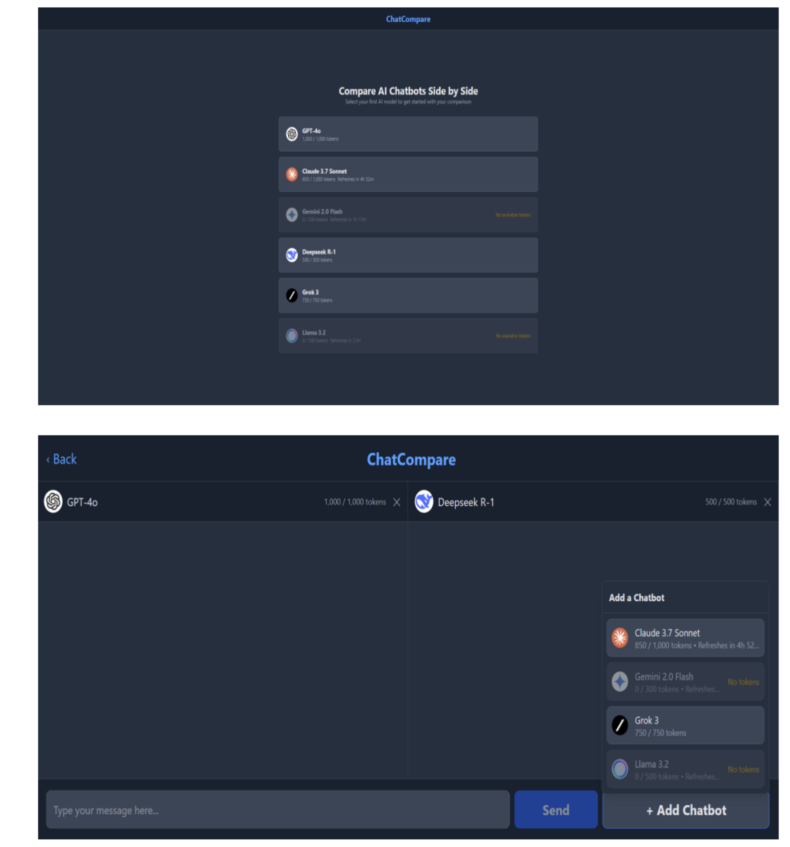
From our interviews, we formed a list of user requirements and determined that the chatbot comparison feature was at the top of our user's list for "must include" features. As such, the chat comparison feature became our priority for the project. We found users also had other features we "could include", however these were not viable to include with respect to the scope of this project. Some of those features were:

- A "quick response" toggle for the chatbots
- A more in-depth and responsive message history or "table of contents"
- Better language settings – multiple languages

4. Prototyping







LFP → **MFP** → **Final**

LFP: We used the Low Fidelity Prototypes (LFPs) to iterate on different ideas, and eventually landed on a design consistent with modern generative AI chatbot systems that included our add model feature as seamlessly as possible in the starting screen, using a popup menu to display our active features when selected

MFP: In our Medium Fidelity Prototype (MFP) we refined and improved our LFP design, getting a sense of how it would look in an implemented application. At this stage, we decided how different models would look side-by-side, choosing to display each model in its own vertical panel and have messages concurrently send to each panel.

Final: Our final design included some simplifications and changes, including a homepage where users select their first model and more detailed information about token limits on the homepage and in the menus. We also opted to include more information about each model, some chat suggestions to get users started testing the models, and a simple rating system (not shown here).

5. Design Challenges

1. **Token Limit Error Prevention**: Since different APIs for different chatbots have different token limits, we needed to maintain visibility of system status and prevent errors through visible word counts and strict disabling rules for temporarily unavailable features.
2. **Visual Space Management:** The need to keep the design consistent and approachable for new users meant our model selection menu was cramped. We had to keep our design clean without oversimplifying.
3. **Contrast and Visual Distinction:** Having side-by-side chats from several different models on one screen appear instantly visually distinct and recognizable was difficult, requiring the use of recognizable icons and distinct background colours without compromising contrast, colour theming, and text visibility.

6. Future Improvements

In the future we would like to add sign in options for users with premium accounts and more of the features that our users requested in the interview. Due to the scope of this project, we were unable to introduce some of the less requested features such as the "quick response" or "improved chat history" feature.