Reflection

[+1] Look back at your assumptions (in the "Before you get started with users" section). Has your user research contradicted or challenged any of these assumptions? If so, which ones? How?

- Our primary assumption of users zoning out after some intermediate duration was not a universal or close-to-universal experience. This is potentially because of the nature of 'think-aloud' studies or being forced into an environment where you cannot interact colloquially with your workflow. Furthermore, the novelty of interacting with a system such as DeepThink (which most users have had no previous exposure to) incentivizes engagement with higher latency systems. Beyond confounders such as novelty or study anxiety, user engagement generally remained due to a desire to get a workflow done efficiently and because the rate of response from agents such as ChatGPT (and even the streaming of thoughts from DeepThink) never allowed an engaged user an awkward pause to become disengaged. If these problems presented by each user are worth solving, then their solutions are inherently interesting enough to maintain engagement.

[+1] What does your user research tell you about what kinds of interactive systems/solutions are unlikely to be useful/successful? Why not? Any hypotheses about what features of an interactive system will be useful? Why?

- Fundamentally, we cannot guarantee a correct solution within the bounds of a high latency without fundamentally creating a new LLM/changing its design; this is implausible within the timeline of the project. From our research, we've seen that it's unlikely that we revolutionize the way users interact with LLMs overnight; there are certain expectations that exist due to their preexposure to these systems, and endeavoring to 'flip the table' seems like the designer asserting their desires onto the user.

  However, what our research reveals is that there are many points of friction between users and their interaction with LLMs. The duration spent prompt engineering, the duration spent iterating on queries, the lack of trust, and high probability of hallucination/falsehood/misunderstanding all promote user behaviors that may not be the most efficient. Any interactive solution targeting one of these points of friction is likely to be successful, as a common thread we extrapolated between users is that they value any increase in efficiency. In terms of specific features we believe that users being able to choose the level of depth for their responses/having agency in the amount of time taken to generate a response may alleviate some frustration when using these systems.

[+1] What questions do you have about your users? What hunches or hypotheses do you have about promising solutions? What is one thing you can do to begin answering these questions?

- One of our lingering questions was about latency perception: had we removed/minimized the "thinking" text box of deepseek, would those who found the longer wait tolerable still hold that opinion? By removing this stimulus, they'd likely feel that the response took longer or become more disengaged with the LLM. Perhaps by doing another round of interviews with this feature disabled, we could glean some insight on this.

  Furthermore, we were unsure about how much of the engagement with DeepSeek's response was due to the novelty of the DeepThink mechanism. We believe that by encouraging users to

continue to use the model for their tasks (rather than ChatGPT), we can obtain a more nuanced/mature understanding of the role that novelty played in user interactions from our past interviews.

Did you use a generative AI tool for any part of this assignment? If so, which one and how?

- Yes, since our entire methodology involved evaluating user's reactions to different latencies with LLMs. As listed above, we used ChatGPT and Deepseek. However, we wrote our responses to this assignment by ourselves.

How much time did you spend on this assignment as a group? Individually?

- As a group we spent 18 hours cumulatively. Individually we spent about 6 hours each.