How do we make sure LLM prompts are safe

what data is part of the context of the LLM?

How will we update the context?

How to we manage LLM resource cost at scale

How/Will we limit the number of token consumed by each user?

How do we know the AI chat is providing correct information?

How do we empirically measure improvements?

How do we know our users are actually getting value out of this feature and is it worth the time/money?

How do we make sure the model doesn't hallucinate?

How do we ensure we protect against naughty behavior

How can a user tell us that the AI chat gave them incorrect information

protected user data

Copyright laws? (legal issues)

user data privacy should be segregated from other user access

if user demand is variable, how do we scale to accommodate

AI messages to users must be safe guarded from unintended or inappropriate responses

How do we troubleshoot when it breaks?

Externalized costs - power usage, water usage, etc

Can it be hacked for evil purposes?

Tools and patterns change daily

Not much past experience to rely on

How trustworthy is the AI? Can we detect when it hallucinates?

Analytics might be difficult to make useful

The hype/ Expectations of an LLM that can "respond" like a human

Protecting user data if relying on external model

Prompt exploits make it hard to offer too many features safely

LLM providers seem to run into reliability/ uptime issues

LLM services might not follow the same cost reductions as other tech.

Difficulty in recreating experiences with a non deterministic tool

Mismatch between capability and expectations

Can we do everything internally, no outside dependencies?

How to ensure the LLM is 100% honest/correct?

cloud services aren't offered in all regions

we haven't done this before

The model is returning wrong information

client hasn't approved AI usage yet

Are we legally allowed to deploy generated code

we don't know scalability needs yet, so pricing is difficult

limited of access to infrastructure team

At the onset, capture "Known Unknowns." Update this list as "Unknown Unknowns" are discovered

It might get very expensive if the feature gets super successful

Oh shoots I'll need to implement an Oauth Provider to keep things secure