What do Users Really Ask Large Language Models?

An Initial Log Analysis of Google Bard Interactions in the Wild

Johanne R. Trippas[1], Sara Fahad Dawood Al Lawati[1], Joel Mackenzie[2], and Luke Gallagher[3]

1 OVERVIEW

Search engines and LLM systems have similarities; people use these services for information access.
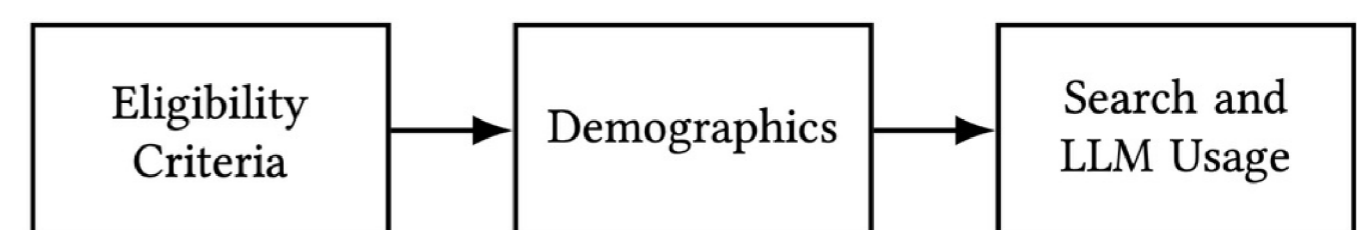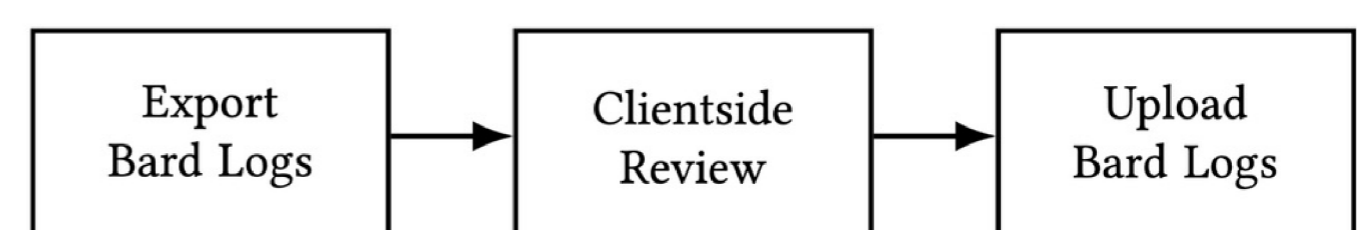
But how do people interact with LLMs?

We crowd-source self-reported "LLM prompt logs" from "real-world" Google Bard (Gemini) users and analyze them.

2 *BIDD* LOG COLLECTION

Survey Stage 1

Eligibility Criteria → Demographics → Search and LLM Usage

Survey Stage 2

Export Bard Logs → Clientside Review → Upload Bard Logs

3 DEMOGRAPHICS (*N*=95)

- 39 years (min. 18, max. 75, std 11)
- 79% male, 18% female, and 4% non-binary/third gender (1% preferred not to say)
- 74% are in the US, and 26% in the UK
- Native English (96%), fluent (2%), and good (2%)
- 66% with a higher degree, 24% had some higher education without a degree, 1% some education, 2% did not disclose

SESSION

1. *How old do you have to be to open a Roth IRA*
2. *I want to start investing on behalf of my nephew in an account for him. How would I do that*
3. *Can I open up a UTMA for my nephew?*
4. *If I wanted to start investing on behalf of my nephew, should I just do so in my own account and transfer to him eventually?*
5. *Tell me everything I need to know about UTMA in a few sentences. Explain as if I'm an average adult. Can I create this account for my nephew?*

4 RESULTS

BIDD LOG DESCRIPTIVES AND USER FEEDBACK

Log statistics	Value
Total logs	15,023
Prompt logs	14,595
Feedback logs	407
Selected-draft logs	21

User and session metrics	Value
Sessions (15 min. pseudo)	4,666
Logs per user	158.1
Avg session per user	49
Avg interactions per session	3
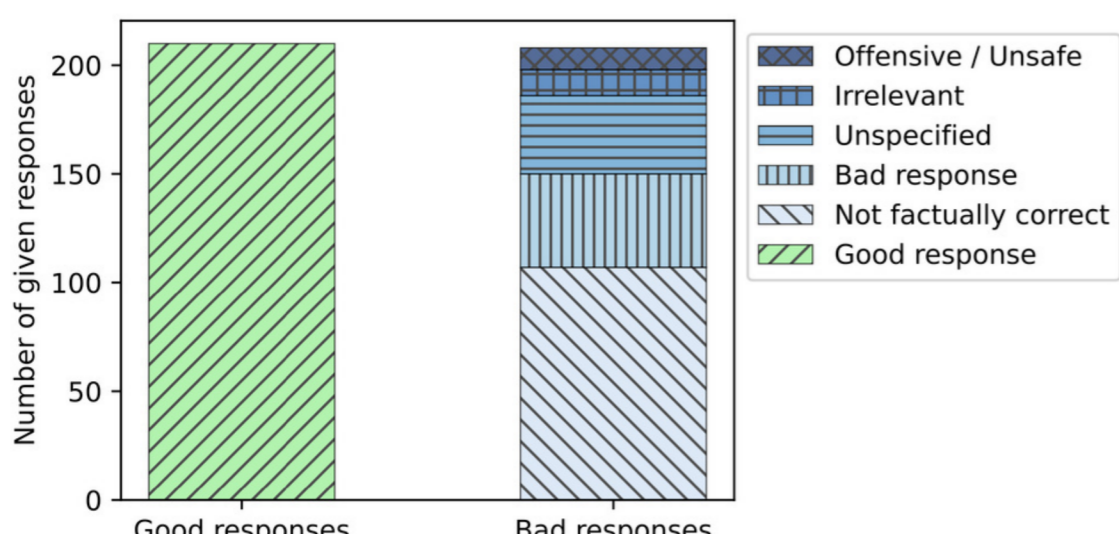
PROMPT DIVERSITY

- 96% of prompts were unique
- Frequently posed prompts
  - More content (0.23%)
  - Reply in English (0.18%)
  - Image generation (0.17%)
  - First-person perspective (0.16%)
- Stop word removal shows *action-oriented* prompts
  - *"make"*, *"write"*, *"answer"*
- Interaction management and navigation
  - *"more"*, *"next page"*
- Personal information management

PROMPT CHARACTERISTICS

BIDD vs MSMARCO
- 43 vs 3 word length
- 1.7% vs 0.004% single-word
- 21% vs 65% start with interrogative words (*what, why, how, when,…*)

User intent strategies
- *Imperatives* (please, write, translate, make, or tell)
- *Auxiliary verbs* (can, is, do, are, does, or will)
- *Pronouns*

Limited prompt engineering
System limitation testing

5 CONCLUSION

- Tasks for *generating* or *extracting* content instead of searching
- Most sessions are short
- 50% of sessions are single-turn
- Prompts are diverse and go beyond keyword search

Limitations
Small sample size from "early adopters" who are on Prolific and willing to share their data