ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience

Ruiyun (Rayna) Xu

Assistant Professor in Information Systems

Farmer School of Business

Miami University

Joint work with Yue Katherine Feng (PolyU HK) and Hailiang Chen (HKU)

Research Objective

Investigate how generative AI technology, specifically ChatGPT, impacts user search behaviors and performance compared to traditional search engines like Google Search, across various search tasks.

Our Study

➢ The first comprehensive comparison of performance and user behaviors between ChatGPT and Google Search in the context of information retrieval.

➢ Two randomized online experiments

Experiment 1: Directly compare ChatGPT and the Google search engine in terms of users' search behaviors and task performance.

Experiment 2:

Examine the effectiveness and strategies of using both tools as opposed to using just one.

Explore the mechanisms of using different search tools on task performance.

Summary of Results (1/2)

1. ChatGPT group consistently **spends less time** on all tasks, while there is no significant difference in overall task performance between the groups.

2. ChatGPT offers **better user experiences** in terms of usefulness, enjoyment, and satisfaction, while perceived ease of use is comparable between the two tools.

3. Users spend a similar level of effort in using both tools, but query length is consistently longer in user interaction with ChatGPT.

4. The performance of using ChatGPT vs. Google Search is **task-dependent**.
    - Notably, ChatGPT excels in answering information look-up questions, which have widely available information from its training data, but struggles with **fact-checking** tasks.
    - A higher **belief in the accuracy** of search results negatively impacts the performance when using ChatGPT.
    - Participants' perception of **search fluency, overreliance on search results, and prior experience using ChatGPT** tend to lower their performance on **fact-checking tasks**.

Summary of Results (2/2)

5. Using both tools result in **a more stable performance**, compensating for the limitations of either tool.

6. When both tools are accessible, the outcomes regarding search efficiency and efforts are **intermediate** compared to using each individually.

7. Participants **strategically use both tools**, often choosing the most appropriate one for a specific task.

Implications for search users

◦ Users can leverage generative AI tools for information search to increase their search productivity.

◦ However, users should avoid over-relying on the tools for fact checks. Cross-validation is often needed.

◦ Strategic use of different tools depending on task types:
  ◦ Unfamiliar topics: ChatGPT can be used first as a line of inquiry
  ◦ Personalization and specificity: ChatGPT can provide immediate support
  ◦ Requiring source tracking: Google has advantage in providing references

Implications for search service providers

- Integrating generative AI technologies into traditional search engines strategically:
  - Generally, dual-tool usage generates **more stable results** with intermediate search efficiency.
  - The efficacy of dual-tool usage may **depend on the task's nature**.
    - When a single tool meets a task's demands, such as using Google Search for fact verification or consulting ChatGPT for step-by-step summaries, introducing the secondary tool may not provide added value.
  - Design strategy when combining chat with search functions may involve recommending specific functions to users based on their queries.

- Integrating chat and search functionalities should also consider the behavioral nuances, such as query length and format.

- It is prudent to alert users to possible inaccuracies within responses, especially for fact verification queries.

Thank You

rayna.xu@miamioh.edu