Citation Design in AI Overviews: Effects on User Reliance and Source Engagement

Kshitijaa Jaglan[a], Elsa Lichtenegger[a], Aleksandra Urman[a] and Aniko Hannak[a]

[a] *University of Zurich, Zurich, Switzerland*

1. Introduction

Web search engines play a central role in shaping how users access and evaluate information. They actively curate information through ranking, summarization, and interface design, influencing what users see and perceive as relevant. A large body of research shows that users disproportionately trust and select top-ranked results, often relying on rank as a proxy for credibility (e.g., position bias; Joachims et al., 2007), with evidence that exposure to highly ranked results can even influence users' attitudes toward the viewpoints they present (Epstein & Robertson, 2015). Beyond ranking, modern search engine results pages (SERPs) include features that further shape user attention and interpretation. Prior work shows that direct answers in featured snippets attract disproportionate attention, reduce search effort, and increase perceived usefulness (Wu et al., 2020). At the same time, users tend to overestimate the credibility of such summaries, and exposure to them can influence both opinions and reasoning processes, particularly for uncertain or debated topics (Bink et al., 2022, 2023). Despite search engines' power to steer user perceptions, attitudes, and behaviors, they are not neutral intermediaries. Prior work shows that algorithmic, data-driven, and user-related biases can systematically shape which information users are exposed to, ultimately influencing what users perceive as relevant and credible (Makhortykh et al., 2020; Paramita et al., 2021; Urman et al., 2022).

With the introduction of AI-generated summaries, such as Google's AI Overviews, these dynamics take a new form. AI-generated summaries synthesize information from multiple sources into a single, coherent-sounding answer box, prominently positioned on the SERP. Emerging evidence suggests that such summaries can significantly influence users' attitudes and beliefs (Xu et al., 2025), raising concerns given hallucinated or misleading content and the limited transparency of AI-generated outputs (Bender et al., 2021). Further, studies show that LLM-based conversational search systems can reinforce existing biases more strongly than traditional search engines (Sharma et al., 2024), and users tend to exhibit greater overreliance on such systems (Spatharioti et al., 2025). To address these risks, search systems increasingly incorporate **transparency mechanisms**, such as citations linking summaries to underlying sources. However, these mechanisms vary widely in design: citations may be presented as icons, numeric markers, or explicit domain names. Such differences may appear superficial, but prior research suggests that interface cues strongly influence user trust, attention, and decision-making, particularly under conditions of limited cognitive effort (Sundar, 2008). As a result, certain citation designs may not promote meaningful verification, but instead contribute to an illusion of transparency, where information is perceived as well-supported without users engaging with the underlying sources. Against this background, it remains unclear whether and how the design of citation cues and the reliability of the sources they reference influence user behavior in AI-augmented search.

To investigate how citation design shapes user behavior in AI-augmented search, we conduct a controlled experiment using a mock SERP. Specifically, we ask: How do citation attribution design, source reliability, and their interaction in AI-generated search summaries influence (a) users' engagement with underlying sources, operationalized through clicks on cited sources, and (b) users' reliance on AI-generated summaries, operationalized through interactions such as hovering, scrolling behavior, and engagement with the summary itself? We hypothesize that making source information more salient (e.g., through inline domain names) would encourage greater attention to source reliability, particularly when lower-credibility sources are presented. In contrast, less informative citation formats (e.g., icon-based citations), would lead users to rely more heavily on the AI-generated summary.

SEASON 2026, September 15–17, 2026, Hamburg, Germany
DOI: https://doi.org/10.48441/4427.3735

2. Experiment Design

We conducted a controlled user experiment using a custom web application that replicated the visual structure and interaction logic of Google's SERP. Participants answered six short factual questions (e.g., "Which food produces the highest methane emissions per kg?") using the mock search interface (median task duration = 44.7 s). The experiment followed a 3 × 3 between-subjects design manipulating (1) citation attribution design (icon, numeric, domain) and (2) source reliability (low, baseline, high). Citation designs differed in the source information they provided: the icon condition showed no visible source information unless participants opened a sidebar; the numeric condition displayed indexed superscripts with hover previews; and the domain condition displayed inline root domains with hover previews. Source reliability was manipulated by changing citation targets while keeping the summary text constant. High-reliability citations are linked to authoritative institutional domains, whereas low-reliability citations are linked to low-credibility domains. We recruited 496 UK-based participants via Prolific (52% men, 46% women; 93% regular Google Search users). After quality filtering, the final dataset comprised 2,917 task observations. Interaction logs recorded cursor position, hover events by SERP region, clicks, scrolling, and timestamps at 100 ms resolution.

3. Analysis Approach

User interaction behavior is analyzed at the task level using mixed-effects regression models, accounting for the repeated-measures structure of the data. Models include fixed effects for citation design, source reliability, task stakes, and relevant demographic covariates. Interaction effects between the three experimental factors are examined across behavioral outcomes. The analyses reported here are preliminary, a full reanalysis is planned prior to the conference.

4. Preliminary Results and Future Analysis

Our preliminary analysis hints that citation design emerges as the dominant predictor of AI overview reliance. Icon attribution is associated with the highest levels of overview engagement (through show more), leading users to extract more from the summary itself. Number attribution, by contrast, appears to impose effort costs without corresponding transparency benefits. When sources are visible (domain condition), users engage more efficiently but do not necessarily evaluate sources more critically; preliminary patterns hint that reputable sources increase reliance on the AI summary. Source reliability effects appear contingent on source visibility, and task stakes show no discernible influence on behavior. Across all conditions, there is no indication that any citation format meaningfully increases outbound traffic to cited sources. These patterns will be subject to full reanalysis before the conference.

Prior to the conference, we plan to: (1) fully rerun and verify all statistical analyses; (2) examine task-level heterogeneity more systematically, including whether topic complexity or query ambiguity moderates citation design effects; (3) explore whether demographic factors (age, education, search frequency) moderate AI overview reliance. These more sophisticated analyses will contribute to empirical accounts of how citation attribution design shapes user reliance in AI-augmented search.

5. Acknowledgements

This work was supported by funding from the Swiss National Science Foundation (Grant number 215354).

6. References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Bink, M., Schwarz, S., Draws, T., & Elsweiler, D. (2023). Investigating the Influence of Featured Snippets on User Attitudes. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, CHIIR '23*, 211–220. https://doi.org/10.1145/3576840.3578323

Bink, M., Zimmerman, S., & Elsweiler, D. (2022). Featured snippets and their influence on users' credibility judgements. *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, 113–122.

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, *112*(33), E4512–E4521.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., & Gay, G. (2007). Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search. *ACM Trans. Inf. Syst.*, *25*(2), 7-es. https://doi.org/10.1145/1229179.1229181

Makhortykh, M., Urman, A., & Ulloa, R. (2020). How search engines disseminate information about COVID-19 and why they should do better. *Harvard Kennedy School Misinformation Review*, *1*. https://doi.org/10.37016/mr-2020-017

Paramita, M. L., Orphanou, K., Christoforou, E., Otterbacher, J., & Hopfgartner, F. (2021). Do you see what I see? Images of the COVID-19 pandemic through the lens of Google. *Information Processing & Management*, *58*(5), 102654.

Sharma, N., Liao, Q. V., & Xiao, Z. (2024). Generative echo chamber? Effect of llm-powered search systems on diverse information seeking. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–17.

Spatharioti, S. E., Rothschild, D., Goldstein, D. G., & Hofman, J. M. (2025). Effects of LLM-based Search on Decision Making: Speed, Accuracy, and Overreliance. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–15.

Sundar, S. S. (2008). *The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility*.

Urman, A., Makhortykh, M., & Ulloa, R. (2022). The matter of chance: Auditing web search results related to the 2020 US presidential primary elections across six search engines. *Social Science Computer Review*, *40*(5), 1323–1339.

Wu, Z., Sanderson, M., Cambazoglu, B. B., Croft, W. B., & Scholer, F. (2020). Providing Direct Answers in Search Results: A Study of User Behavior. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20*, 1635–1644. https://doi.org/10.1145/3340531.3412017

Xu, Y., Dash, S., Kang, S., Liao, W., & Spiro, E. (2025). *AI summaries in online search influence users' attitudes*. https://doi.org/10.48550/arXiv.2511.22809