

Not All Transparency Is Equal: Source Presentation Effects on Attention, Interaction, and Persuasion in Conversational Search

Jianguen He
School of Information Sciences
University of Tennessee
Knoxville, Tennessee, USA
jianguen@utk.edu

Jiqun Liu
School of Library and Information Studies
University of Oklahoma
Norman, Oklahoma, USA
jiqunliu@ou.edu

Abstract

Conversational search systems increasingly provide source citations, yet how citation or source presentation formats influence user engagement remains unclear. We conducted a crowdsourcing user experiment with 394 participants comparing four source presentation designs that varied citation visibility and accessibility: collapsible lists, hover cards, footer lists, and aligned sidebars. High-visibility interfaces generated more hovering on sources, though clicking remained infrequent across all conditions. While interface design showed limited effects on user experience and perception measures, it significantly influenced knowledge, interest, and agreement changes. High-visibility interfaces initially reduced knowledge gain and interest, but these positive effects emerged with increasing source usage. The sidebar condition uniquely increased agreement change. Our findings demonstrate that source presentation alone may not enhance engagement and can even reduce it when insufficient sources are provided.

CCS Concepts

• Information systems → Search interfaces; • Human-centered computing → User studies.

Keywords

conversational search, large language models, source citation, credibility, transparency, information seeking, attitude change

ACM Reference Format:

Jianguen He and Jiqun Liu. 2026. Not All Transparency Is Equal: Source Presentation Effects on Attention, Interaction, and Persuasion in Conversational Search. In *2026 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '26)*, March 22–26, 2026, Seattle, WA, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3786304.3787874>

1 Introduction

Conversational search and large-language-model (LLM)-enabled intelligent assistants are reshaping how people seek, judge, and use information. Within the field of human-computer interaction and retrieval (HCIR), decades of research show that interaction design, sense-making, and credible source use are central to effective search [13, 17, 20, 27]. In particular, user evaluation studies further

demonstrate that the interface is not a neutral conduit: presentation choices modulate attention, implicit feedback, and verification behaviors [2, 8, 10]. In parallel, credibility research in information science establishes that users apply heuristics and strategies sensitive to cues exposed by the interface [14, 22, 25, 28]. For LLM-based systems, transparency features, such as explanations, provenance, and citations, are repeatedly recommended to support scrutiny and appropriate trust [6, 16, 19]. Yet, real-world studies of generative search caution that users often do not meaningfully inspect sources even when citations are provided, risking shallow verification, mistrust, and homogenized exposure [4, 11, 23]. Together, these strands indicate that *source-presentation design* is a first-order determinant of credible, accountable conversational and generative interactions.

Despite significant progress, two gaps persist. First, most work examines transparency at a conceptual or system-level (e.g., “provide citations”) rather than as a precise interface variable whose visibility and accessibility can be manipulated and measured. We lack controlled evidence isolating how concrete presentation formats (e.g., inline hover cards vs. footers vs. side panels) affect micro-behaviors such as hovering, clicking, and revisiting sources—the very behaviors through which users enact scrutiny and build credibility judgments [2, 9, 10]. Second, prior studies often stop at global perceptions (e.g., trust, satisfaction, engagement) without connecting interface-induced behaviors to downstream knowledge and attitude change [26], which are outcomes long recognized in information-seeking and persuasion research [5, 7, 22, 25]. Closing these gaps is crucial for both research and practice: it sharpens theories linking interface affordances, attention, and credibility assessment, and yields prescriptive guidance for designing systems that promote verification without overloading users—particularly vital in domains like health and public policy where misuse of fluent but weakly grounded responses can have real societal costs and diminish public information integrity [3, 12, 14, 24, 27].

We address these gaps by experimentally varying how citations are presented during generative search sessions. Specifically, we compare four source-presentation designs that systematically vary *source visibility* and *accessibility*—a collapsible list, inline hover cards, a footer list, and an aligned sticky sidebar—and log their effects on inspection behaviors (hovering/clicking), perceived response quality, and changes in knowledge, interest, and attitude on controversial topics. This design operationalizes interface-level transparency recommendations [6] within a conversational and generative IR task framework [1, 20], linking interaction traces to cognitive/attitudinal outcomes grounded in credibility and persuasion theory [18, 22]. To address the gaps above, Our crowdsourcing user experiment answers following research questions:



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHIIR '26, Seattle, WA, USA*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2414-5/26/03
<https://doi.org/10.1145/3786304.3787874>

- **RQ1:** How do different source UI designs influence interaction behaviors with source information?
- **RQ2:** How do source UI designs affect people’s perceptions of AI response quality and their experience?
- **RQ3:** How do source UI designs affect people’s knowledge, attitude, and interest change?

For RQ1, we examine hovering and clicking patterns across interface conditions to understand how visibility and accessibility affect participants’ engagement with sources during information-seeking tasks. To answer RQ2, this study investigates whether different source presentation formats influence perceived credibility, relevance, confidence, satisfaction, and engagement with the AI agent. As a response to RQ3, we explore whether the effectiveness of sources in influencing users’ knowledge, interest, and agreement depends on how those sources are presented in the interface.

2 Methods

2.1 Interface Design

To investigate how different source presentation formats influence user experience and chat experience, we implemented five distinct interface conditions (see Figure 1). Except the baseline interface, the four interfaces with source presentation presents the same source details (title, domain name with favicon, and text snippet), but each interface varies in how sources are displayed and accessed.

2.1.1 No Sources Interface. The no sources interface condition serves as a control condition in which no source badges or source cards are shown. Sources are used in generating the responses like the other four conditions, but participants have no access to the underlying sources used to generate the information.

2.1.2 Collapsible Interface. In the collapsible interface condition, source badges appear inline within the text as clickable links. A collapsible “Used Sources” button displays at the beginning of each AI response. Users can click this button to expand and reveal the complete list of source cards containing source details

2.1.3 Hover Card Interface. In the hover card interface condition, when users hover over a citation badge, a card appears displaying the source details. This design provides on-demand access to citation details, allowing users to quickly preview sources while maintaining their position in the text.

2.1.4 Footer Interface. In the footer interface condition, all source cards are grouped at the end of each response message. All source cards are displayed by default without the need to expand, compared to the collapsible interface. This approach mimics bibliography formatting, consolidating all reference information in a location.

2.1.5 Aligned Sidebar Interface. The aligned sidebar interface condition presents source cards in a persistent sidebar adjacent to the main content. Source cards are dynamically positioned to align vertically with their corresponding in-text citations. The interface employs the bounded isotonic regression to compute optimal non-overlapping positions for the cards. The display of source card updates dynamically in response to scrolling and window resizing. When citations appear multiple times within the text, positioning

is weighted toward the first occurrence using exponential decay to prioritize initial encounters.

We use the four interface conditions to manipulate the source visibility and accessibility. Visibility refers to the explicit display of source details, while accessibility refers to the proximity of source details to the in-text citations. These dimensions form four interface conditions: a collapsible list (low visibility, low accessibility), a footer (high visibility, low accessibility), a hover preview (low visibility, high accessibility), and an aligned sidebar (high visibility, high accessibility). The AI agent was powered by Perplexity AI’s sonar-pro model, configured with a temperature of 0.7. The model utilized web search capabilities with medium search context size and user location set to ‘US’ (United States) to retrieve real-time information from the internet and generate responses with inline citations.

2.2 Experiment Procedure

The experimental procedure consisted of three stages: a pre-study survey, the main information-seeking and essay-writing task, and a post-study survey. The pre-study survey asked participants to rate their prior experience with and attitudes toward conversational AIs, their confidence in evaluating information, and their intellectual humility [cf. 21]. Participants also rated their familiarity with, interest in, and attitudes toward the controversial topic assigned to them (described below). For the main task, participants were randomly assigned to one of five interface conditions. They received instructions for interacting with the chatbot interface and information about sources, and they were instructed to search for and use credible, accurate, and useful information to write an essay on the assigned topic with at least 50 words. Copying and pasting were allowed during the main task. To encourage richer chat interactions [15], we asked participants to engage in at least four turns of conversation. The post-task survey asked participants to rate their perceptions of the AI agent’s credibility, relevance, confidence, satisfaction, response quality, and engagement, as well as their knowledge of, interest in, and attitude toward the topic.

Inspired by the research of Sharma et al. [23], we searched for topics in healthcare, science, and news from ProCon.org¹ to engage participants’ interest and cognitive load. We selected three topics: (1) *News*: “Should the U.S. government impose stricter regulations on social media platforms to curb misinformation?” (2) *Science*: “Should the U.S. permit clinical use of germline gene editing in humans?” (3) *Healthcare*: “Should the U.S. Government Provide Universal Health Care?”

2.3 Participants

We recruited 394 participants (195 male, 199 female) with a mean age of 43.63 years (SD = 13.45) on Prolific². The inclusion criteria were fluent English-speaking participants from the United States. We also ran an evenly split quota sampling for gender and political spectrum (conservative, moderate, liberal). The median completion time was 17.77 minutes. All the participants were recruited through Prolific and were paid \$3.00 for each session. The study was approved by the Institutional Review Board of the first author’s institution.

¹<https://www.procon.org/>

²<https://www.prolific.co/>

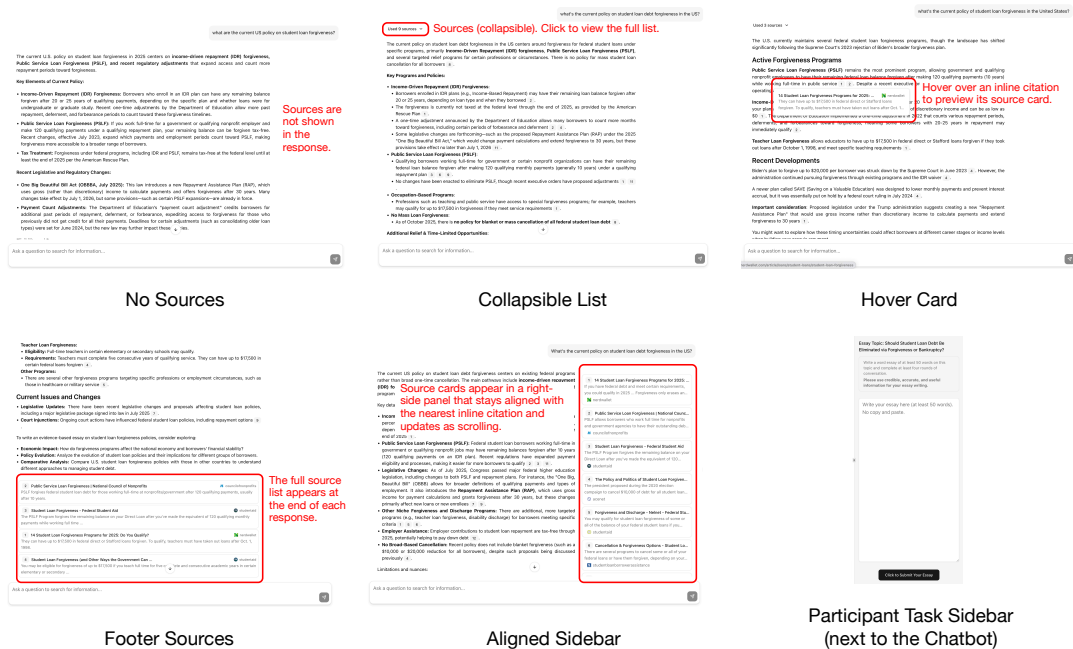


Figure 1: Five user interfaces/conditions.

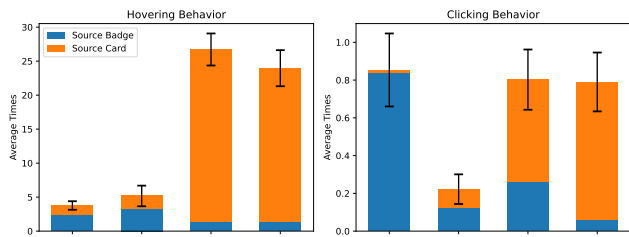


Figure 2: Hovering and clicking by interface condition.

3 Results

3.1 Behavior Analysis (RQ1)

3.1.1 *Hovering and Clicking.* We tracked user interactions with source information by logging hover and click events on source cards and citation badges. A hover event was counted once per continuous hover interaction. Figure 2 presents the mean hovering and clicking times per participant across the four interfaces.

The footer and sidebar interfaces generated substantially more hovering activity on source cards (26.7 and 24.0 times respectively) compared to the collapsible (3.8 times) and hover card (5.2 times) conditions. It reflects the visibility of source details in source cards in these designs. In contrast, clicking behavior is less frequent. Most of participants did not click the sources to read the original sources. Surprisingly, the collapsible interface showed highest click rates (0.85 times) when participants needed to actively expand the source list to access citation details. Conversely, the hover card interface showed the lowest click rates (0.32 times) when participants could preview the source details on hover. The high availability of source

details in the hover card interface might lead to participants not clicking the sources to read the sources.

3.1.2 *Participant Prompts and Essay writing.* Table 1 presents descriptive statistics for participant prompts and essay characteristics across the five interface conditions. Kruskal-Wallis H tests revealed no significant differences across interface conditions for any of the metrics. However, pairwise comparisons using Mann-Whitney U tests revealed significant differences in total message composition time between the no sources condition and several interface conditions with citations. Specifically, the no sources condition differed significantly from the Hover card ($U = 3339.0, p = 0.040$), Footer ($U = 3455.0, p = 0.013$), and Sidebar ($U = 3747.0, p < 0.001$) conditions, with participants in the no sources condition spending more time composing prompts.

3.2 Perception (RQ2)

We used ordinal logistic regression models to analyze perception and user-experience outcomes because they were measured on 5-point Likert scales. Each model included the UI conditions (using the no-sources interface as the reference category), an interaction term between UI condition and citation count, demographic variables (age, gender, education), pre-study measures (LLM usage frequency, information-evaluation confidence, intellectual humility, prior knowledge, familiarity, and interest), and behavioral metrics (source hover times, source click times, number of prompts, prompt word count, length of LLM responses, prompt typing time, and total task time). This comprehensive modeling approach allows us to isolate the effects of UI design while accounting for individual differences and engagement patterns.

Table 1: Participant Prompts and Essay Writing Statistics (mean and standard deviation) by Interface Condition

Metric	No sources	Collapsible	Hover Card	Footer	Sidebar	Kruskal-Wallis H	
						H	p
Prompt Count	4.4 (0.9)	4.7 (1.4)	4.8 (1.4)	4.7 (1.5)	4.6 (1.1)	2.269	0.519
Prompt Composition Time (s)	203.7 (199.6)	171.7 (179.8)	165.9 (180.9)	153.4 (162.6)	137.8 (153.2)	4.393	0.222
Prompt Word Count	72.1 (58.7)	65.9 (48.3)	58.7 (36.2)	61.3 (48.6)	73.8 (96.8)	0.170	0.982
Essay Word Count	97.7 (46.1)	97.5 (40.8)	101.1 (47.2)	89.4 (33.9)	103.3 (50.7)	4.705	0.195
Essay Writing Time (s)	325.6 (222.4)	311.0 (226.1)	336.5 (243.1)	310.9 (188.5)	340.0 (249.0)	1.256	0.740

Table 2: Regression Analysis Results for Perception Measures

Predictor	Cred	Rel	Conf	Sat	Use	Qual
UI Conditions (Reference: No sources)						
Collapsible	0.28	1.31	-1.09	-0.56	0.70	-1.46
Hover	0.49	1.01	-0.86	0.51	-0.47	-0.71
Footer	-0.70	-1.17	-1.92*	-0.45	-1.02	-1.21
Sidebar	-1.44	-0.22	-0.98	-0.13	0.42	-1.17
UI × Citations Interactions						
Colla. × Citations	-0.01	-0.05	0.02	0.02	-0.03	0.04
Hover × Citations	-0.01	-0.03	0.04	-0.01	0.01	0.02
Footer × Citations	0.03	0.04	0.07*	0.02	0.04	0.05*
Sidebar × Citations	0.07*	0.02	0.03	0.02	-0.01	0.07

Cred = Credibility, Rel = Relevance, Conf = Confidence, Sat = Satisfaction, Use = Usefulness, Qual = Quality. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2 presents the key regression coefficients for UI conditions and their interactions with citation count across six outcome measures. Coefficients are reported with standard errors and significance levels. Control variables are not displayed in the table but were included in all models. The regression analysis revealed several effects related to UI design and citation presentation. For credibility, a significant Sidebar × Citations interaction ($\beta = 0.07$, $p < 0.05$) indicated that sidebar presentation becomes more effective with increasing citation counts. The Footer condition showed a negative main effect on confidence ($\beta = -1.92$, $p < 0.05$) but positive interactions with citation count for both confidence and quality ($\beta = 0.07$ and 0.05 , respectively, $p < 0.05$). No significant effects emerged for relevance, satisfaction, or usefulness.

3.3 Attitude Changes (RQ3)

Table 3 presents regression results for three attitude change measures using the same model specification as Table 2. The findings reveal more pronounced effects of citation UI design on attitudinal outcomes compared to perception measures. For knowledge gain, both the Footer and Sidebar conditions showed significant negative main effects ($\beta = -1.28$ and -1.93 , respectively, $p < 0.05$), suggesting reduced self-reported learning when sources were explicitly displayed (high visibility). However, these conditions also exhibited significant positive interactions with citation count ($\beta = 0.05$ and 0.08 , $p < 0.01$ and $p < 0.01$), indicating that their negative impact reversed as citation density increases. Interest in the topic was significantly lower across multiple UI conditions: Collapsible ($\beta = -2.45$, $p < 0.01$), Hover ($\beta = -2.84$, $p < 0.01$), and Sidebar ($\beta = -2.69$, $p < 0.01$). Nevertheless, positive interactions with citation count emerged for Hover, Footer, and Sidebar conditions

Table 3: Regression Analysis Results for Attitude Change

Predictor	Knowledge	Interest	Agreement
UI Conditions (Reference: No sources)			
Collapsible	-0.80	-2.45**	-1.08
Hover	0.09	-2.84**	-1.83
Footer	-1.28*	-0.98	0.59
Sidebar	-1.93*	-2.69**	1.89*
UI × Citations Interactions			
Collapsible × Citations	0.04	0.06	0.03
Hover × Citations	0.00	0.07*	0.05
Footer × Citations	0.05**	0.04*	-0.01
Sidebar × Citations	0.08**	0.07*	-0.06

($\beta = 0.07$, 0.04 , and 0.07 , respectively, all $p < 0.05$), suggesting that positive effects of source presentation requires sufficient number of citations. Agreement with the topic claims followed a different pattern: the Sidebar condition increased agreement change ($\beta = 1.89$, $p < 0.05$). This indicates that the positive effect of the Sidebar on agreement change does not require many citations to be presented.

4 Conclusion and Future Work

This study examined how different source citation presentation designs influence user behavior, perception, and attitude change in conversational AI-assisted information seeking. Through a controlled experiment with five interface conditions (four citation designs that systematically varied visibility and accessibility, plus a no-sources control), we found that interface design significantly affects user engagement with sources and downstream outcomes. While we observed limited behavioral differences and effects on perception measures such as credibility and satisfaction, the citation presentation format showed more pronounced effects on attitude changes. Specifically, interfaces with higher citation visibility showed negative main effects on knowledge gain and interest, but these effects were reversed with increasing citation density, suggesting that the effectiveness of transparent source presentation depends on sufficient evidence support. The Sidebar condition uniquely increased agreement with topic claims, indicating that persistent, spatially aligned citation display may enhance persuasive impact. In future work, we will analyze the content and construction patterns of participant prompts. A systematic analysis of the essays produced, particularly examining critical analysis and evidence usage, would clarify whether and how citation interfaces influence the quality and rigor of knowledge integration.

References

- [1] Marwah Alaofi, Negar Arabzadeh, Charles LA Clarke, and Mark Sanderson. 2024. Generative information retrieval evaluation. In *Information access in the era of generative ai*. Springer, 135–159.
- [2] Georg Buscher, Susan T Dumais, and Edward Cutrell. 2010. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. 42–49.
- [3] Jeffrey Dalton, Chenyan Xiong, and Jamie Callan. 2020. TREC CAsT 2019: The conversational assistance track overview. *arXiv preprint arXiv:2003.13624* (2020).
- [4] Fernando Diaz, Bhaskar Mitra, Michael D Ekstrand, Asia J Biega, and Ben Carterette. 2020. Evaluating stochastic rankings with expected exposure. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 275–284.
- [5] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. 2021. This is not what we ordered: Exploring why biased search result rankings affect user attitudes on debated topics. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 295–305.
- [6] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing transparency design into practice. In *Proceedings of the 23rd international conference on intelligent user interfaces*. 211–223.
- [7] Brian Hillgoss and Soo Young Rieh. 2008. Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information processing & management* 44, 4 (2008), 1467–1484.
- [8] Chenhao Hu, Shuhua Huang, Yansen Zhang, and Yubao Liu. 2022. Learning to infer user implicit preference in conversational recommendation. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*. 256–266.
- [9] Jeff Huang, Ryen W White, and Susan Dumais. 2011. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1225–1234.
- [10] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2017. Accurately interpreting clickthrough data as implicit feedback. In *Acm Sigir Forum*, Vol. 51. Acm New York, NY, USA, 4–11.
- [11] Jeena Joseph, Binny Jose, and Jobin Jose. 2025. The generative illusion: how ChatGPT-like AI tools could reinforce misinformation and mistrust in public health communication. *Frontiers in Public Health* 13 (2025), 1683498.
- [12] Yubin Kim, Hyewon Jeong, Shan Chen, Shuyue Stella Li, Mingyu Lu, Kumail Alhamoud, Jimin Mun, Cristina Grau, Minseok Jung, Rodrigo Gameiro, et al. 2025. Medical hallucinations in foundation models and their impact on healthcare. *arXiv preprint arXiv:2503.05777* (2025).
- [13] Jiqun Liu. 2021. Deconstructing search tasks in interactive information retrieval: A systematic review of task dimensions and predictors. *Information Processing & Management* 58, 3 (2021), 102522.
- [14] Jiqun Liu and Jianguan He. 2025. Boundedly Rational Searchers Interacting with Medical Misinformation: Characterizing Context-Dependent Decoy Effects on Credibility and Usefulness Evaluation in Sessions. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval*. 154–167.
- [15] Jiqun Liu, Jamshed Karimnazarov, and Ryen W White. 2025. Trapped by Expectations: Functional Fixedness in LLM-Enabled Chat Search. *arXiv preprint arXiv:2504.02074* (2025).
- [16] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.
- [17] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.
- [18] Richard E Petty and John T Cacioppo. 2012. *Communication and persuasion: Central and peripheral routes to attitude change*. Springer Science & Business Media.
- [19] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–52.
- [20] Filip Radlinski and Nick Craswell. 2017. A theoretical framework for conversational search. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*. 117–126.
- [21] Alisa Rieger, Frank Bredius, Mariët Theune, and Maria Soledad Pera. 2024. From potential to practice: Intellectual humility during search on debated topics. In *Proceedings of the 2024 conference on human information interaction and retrieval*. 130–141.
- [22] Soo Young Rieh and David R Danielson. 2007. *Credibility: A multidisciplinary framework*. (2007).
- [23] Nikhil Sharma, Q Vera Liao, and Ziang Xiao. 2024. Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 1033, 17 pages. doi:10.1145/3613904.3642459
- [24] Nikhil Sharma, Q Vera Liao, and Ziang Xiao. 2024. Generative echo chamber? effect of llm-powered search systems on diverse information seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [25] Ben Wang and Jiqun Liu. 2024. Cognitively Biased Users Interacting with Algorithmically Biased Results in Whole-Session Search on Debated Topics. In *Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval*. 227–237.
- [26] Ran Yu and Jiqun Liu. 2025. Chat as Learning in Interactive Information Retrieval and Generation. In *ACM SIGIR Forum*, Vol. 59. ACM New York, NY, USA, 1–19.
- [27] Hamed Zamani, Johanne R Trippas, Jeff Dalton, Filip Radlinski, et al. 2023. Conversational information seeking. *Foundations and Trends® in Information Retrieval* 17, 3-4 (2023), 244–456.
- [28] Yan Zhang, Jiaying Liu, and Shijie Song. 2023. The design and evaluation of a nudge-based interface to facilitate consumers' evaluation of online health information credibility. *Journal of the Association for Information Science and Technology* 74, 7 (2023), 828–845.