From Searchable to Non-Searchable: Generative AI and Information Diversity in Online Information Seeking

Yulin Yu
The Northwestern Institute on Complex Systems (NICO)
Northwestern University
Evanston, Illinois, USA
School of Information
University of Michigan
Ann Arbor, Michigan, USA
yulin.yu@kellogg.northwestern.edu

Yizhou Li
School of Information
University of Michigan
Ann Arbor, Michigan, USA
yizhouli@umich.edu

Siddharth Suri
Microsoft Research
Redmond, Washington, USA
suri@microsoft.com

Scott Counts
Microsoft Research
Redmond, Washington, USA
counts@microsoft.com

Abstract

Conversational generative AI systems such as ChatGPT are transforming how people seek and engage with information online. Unlike traditional search engines, these systems support open-ended, conversational inquiry, yet it remains unclear whether they ultimately expand or constrain the diversity of knowledge that users encounter in online search spaces—a primary foundation for knowledge work, learning, and innovation. Using over 200,000 real-world human–ChatGPT interactions, we examine how generative-AI–mediated inquiry reshapes diversity in both user inputs and system outputs through the lens of searchability—whether queries could plausibly be answered by traditional search engines. We find that almost 80% of ChatGPT user queries are non-searchable and span a broader knowledge space and topics than searchable queries, indicating expanded modes of inquiry. However, for comparable searchable queries, AI responses are less diverse than Google search results in the majority of topics. Moreover, the diversity of AI responses predicts subsequent changes in users' inquiry diversity, revealing a feedback loop between AI outputs and human exploration. These findings highlight a tension between expanded inquiry and constrained information exposure, with implications for designing hybrid search and generative-AI systems that better support exploratory knowledge seeking.

CCS Concepts

• **Human-centered computing** → *Empirical studies in HCI*; Natural language interfaces; • **Data mining**;

Keywords

generative AI, ChatGPT, search engines, information diversity, knowledge diversity, exploratory search, conversational search, human–AI interaction, hybrid search, query searchability

ACM Reference Format:
Yulin Yu, Yizhou Li, Siddharth Suri, and Scott Counts. 2026. From Searchable to Non-Searchable: Generative AI and Information Diversity in Online Information Seeking. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3772363.3798802

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
CHI EA '26, Barcelona, Spain
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3798802

1 Introduction and Literature Review

The rapid emergence of conversational generative AI systems such as ChatGPT is fundamentally transforming how people search for and engage with information. Unlike traditional search engines, which require users to translate information needs into keyword-based queries and navigate ranked lists of documents, generative AI supports conversational, iterative, and open-ended inquiry[11]. This shift changes not only the volume and form of information seeking, but also the underlying mechanisms of knowledge access. As a result, these systems may reshape how people formulate questions, interpret answers, and reflect on decisions—core processes of learning and knowledge work that ultimately underpin intelligent behavior and innovation[1, 2, 7, 10, 13–15, 17].

At the core of these processes lies knowledge diversity—the breadth and heterogeneity of information people ask for and receive. Prior research across HCI, information science, and computational social science shows that exposure to diverse knowledge fuels creativity and high-impact innovation, while constrained diversity can reinforce narrow thinking, echo chambers, and polarization [5, 12, 16]. Importantly, both input diversity—the diversity of users' inquiries—and output diversity—the diversity of information returned by systems—play foundational roles in shaping learning and downstream exploration[4, 8, 9]. Many of these studies examine how human–ChatGPT interactions can limit collective opinion formation and creative problem-solving diversity, raising

concerns about the potential narrowing of diversity in generative-AI–mediated interactions. This line of work also motivates further investigation into why such narrowing occurs, particularly during the transition from traditional search-engine–based knowledge access to generative-AI–mediated inquiry [4, 8, 9, 13, 17]. Understanding how the two dimensions—user inputs and ChatGPT outputs—interact under generative-AI mediation, through the lens of searchability (i.e., whether user inputs are traditionally searchable), allows us to assess whether generative-AI systems expand or narrow information diversity in online searching space. This perspective enables a deeper examination of diversity at the level of inputs, outputs, and their interaction. Accordingly, we empirically ask: **What do ChatGPT-like systems contribute to knowledge diversity? Specifically, to what extent—and in what ways—do users engage with more diverse knowledge through interactions with ChatGPT-like systems beyond traditionally searchable queries? How do responses generated by these systems compare with traditional search-engine results, and how do AI responses, in turn, shape and potentially alter users' engagement diversity?** Addressing these questions is critical for evaluating current trends and for informing the design of future knowledge-interaction paradigms that enable users to inquire into, and be exposed to, a broader and more diverse knowledge space.

In this paper, we analyze over 200,000 real-world human–ChatGPT interactions (Chatlog) from 2023 to 2025 to examine how generative-AI–mediated inquiry reshapes knowledge diversity. We distinguish between queries that could plausibly be addressed by traditional search engines and those that could not, and we track how diversity in both user inputs and system outputs evolves over time. By comparing generative-AI responses with ranked results from traditional search engines, we reveal a fundamental tension: while generative AI expands the space of what users can ask, its outputs may inadvertently constrain informational diversity downstream. Our findings illuminate how generative AI reshapes curiosity and learning, and point to design opportunities for hybrid systems that better support diverse, exploratory, and innovation-relevant knowledge seeking.

2 Empirical Setting and Data

We approach this problem through an empirical analysis of observational data capturing how users interact with ChatGPT. We train a model to classify whether a user input resembles a traditional search-engine query or whether it could be effectively answered through Google search. We then define and quantify the knowledge space—i.e., the diversity of content—of both user inputs and the outputs produced by generative AI or Google search, in order to statistically examine our research questions.

2.1 Data Description

2.1.1 Corpus Construction. Our analysis is based on a corpus of 234,839 real-world conversations between humans and ChatGPT collected from 2023 to 2025, drawn from the WildChat-4.8M dataset[18][1] . This dataset is a large-scale collection of real user–ChatGPT interaction logs, where users interacted with ChatGPT via public interfaces operated by the Allen Institute. To mitigate temporal imbalance and manage computational constraints, we construct a temporally balanced analysis subset. Specifically, we group conversations by calendar week (based on UTC timestamps) and uniformly sample up to 2000 conversations per week. In this dataset, a conversation is defined as a sequence of user–assistant message turns occurring within a single chat session. To enable topic-based analysis, we also label user inputs using the prompt-based taxonomy [3].

2.1.2 Fetching Search Engine Results. To compare ChatGPT responses with search-engine results, we retrieve Google search results for a fixed set of searchable user input (defined below) using the Google Custom Search API[2] . For each sampled searchable question, we collect the top 10 results.

2.2 Variable Construction

2.2.1 Input searchability : To operationalize whether a user query is likely to be answerable via a web search engine (i.e., searchable), we adopt a hybrid labeling-and-scaling approach that combines LLM-assisted pre-labeling, human verification, and supervised learning. We randomly sample 3000 non-empty user questions from the study corpus to create an annotation set. Initial labels are generated using a large language model, after which human annotators review and confirm or correct the binary labels. Using this verified set, we train a lightweight text classifier, which is then applied to all user questions in the corpus to infer searchable versus non-searchable queries at scale. A value of 1 indicates a high likelihood of being searchable, while 0 indicates a low likelihood; we use a likelihood threshold of 90% as the model cutoff. Using this model, we additionally create a random validation sample of 100 groups, each with one searchable and one non-searchable instance, which is reviewed by human annotators; the resulting accuracy is 91%.

2.2.2 Information Diversity-input: We operationalize the *information diversity of inputs* as dispersion in a semantic embedding space, following prior work that uses a similar approach to define knowledge space in science [6]. For a set of user inputs in group g (searchable vs. non-searchable), we encode each input into a unit-normalized embedding vector $\mathbf{x}_i$ (SentenceTransformer; `all-MiniLM-L6-v2`). We compute the group centroid $\bar{\mathbf{x}}_g = \frac{1}{n_g} \sum_{i \in g} \mathbf{x}_i$, and define each input's deviation as the Euclidean distance $d_i^{\text{in}} = \|\mathbf{x}_i - \bar{\mathbf{x}}_g\|_2$. When quantifying the knowledge space of each group (e.g., searchable versus non-searchable inputs), we define the group-level input information diversity as the maximum deviation within the group, $\text{ID}_g^{\text{in}} = \max_{i \in g} d_i^{\text{in}}$ [6].

2.2.3 Information Diversity-output: We operationalize the *information diversity of outputs* following a similar method used to define input information diversity; instead, we use responses from ChatGPT or Google Search. For each group g, we embed every AI response into a unit-normalized vector $\mathbf{y}_i$, compute the centroid $\bar{\mathbf{y}}_g = \frac{1}{n_g} \sum_{i \in g} \mathbf{y}_i$, and define distances $d_i^{\text{out}} = \|\mathbf{y}_i - \bar{\mathbf{y}}_g\|_2$. The size of the knowledge space is defined as [6], $\text{ID}_g^{\text{out}} = \max_{i \in g} d_i^{\text{out}}$.

[1] https://huggingface.co/datasets/allenai/WildChat-4.8M

[2] https://developers.google.com/custom-search/v1/overview

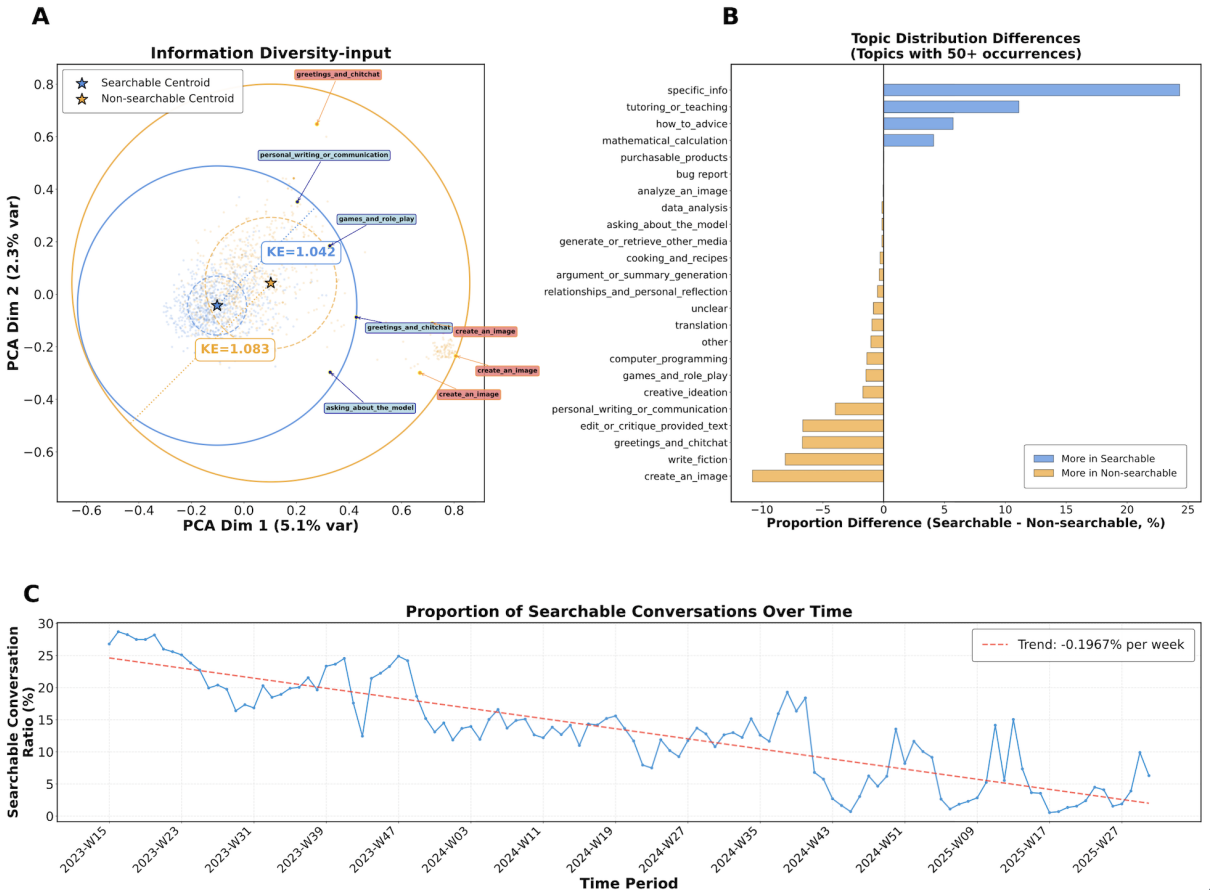

Figure 1: Diversity, topical distribution, and temporal trends in user input searchability on ChatGPT. (A) User inputs are embedded into a 384-dimensional vector space using the SentenceTransformer all-MiniLM-L6-v2, and knowledge extent is measured as the maximum Euclidean distance from the centroid (the embedding-space radius). (B) Topic differences between searchable and non-searchable inputs. Bars show percentage-point differences in topic prevalence (searchable minus non-searchable). Positive values indicate topics more common in searchable inputs; negative values indicate topics more common in non-searchable inputs. (C) Temporal trend in the weekly proportion of searchable conversations. Points show the share of conversations containing searchable inputs in each calendar week; the dashed line is a linear fit summarizing the overall trend (slope annotated in the panel).

3 Findings

3.0.1 Input Diversity. We generate three key findings by analyzing user interactions with ChatGPT. First, by applying a searchability classifier, we find that overall, 21.0% of user inputs are classified as searchable, while 79.0% are non-searchable. As shown in Figure 1C, the proportion of searchable user inputs decreases over time.

Second, Gen-AI–mediated information seeking exhibits much more diverse input diversity in non-searchable chats compared to searchable chats. Figures 1A–B present the diversity metrics. In Figure 1A, we observe that user inputs in non-searchable chats display a broader distribution, with a maxium diversity score of 1.0826, whereas searchable chats show a more concentrated distribution with a maxium of 1.0421 (We conduct a robustness check using the 95th percentile of the values, and the results are qualitatively similar. The 95th percentile diversity score for non-searchable chats is 1.0329, compared to 1.0329 for searchable chats.). Figure 1B shows that only 4 of the 24 topics have more searchable inputs than non-searchable inputs. In Figure 1A, we embed all user inputs into a two-dimensional space, where we visually observe that non-searchable chats expand the knowledge space across a wider range of topics.

Third, we examine the difference in the proportion of searchable versus non-searchable chats across topics to identify which topics exhibit stronger searchability characteristics. We find that the majority of topics skew toward non-searchable interactions, with the exception of specific information retrieval, tourism, advice-seeking, and mathematical calculations. Heavy non-searchable usage is dominated by generative tasks, such as image creation, fiction writing, role-play interactions with chatbots, and editing requests.

3.0.2 Output Diversity. We next examine the diversity of information feedback following user queries, specifically comparing potential ChatGPT outputs with Google Search outputs given the same user input. We observe that, on average, Google Search—whether using the top 1, 3, 5, or 10 results—provides significantly higher information diversity than ChatGPT (1.052 vs. 1.027, $p < 0.001$ for Google Top-5 vs. ChatGPT; results for Top-1/Top-3/Top-10 are qualitatively similar) is qualitatively similar). Figures 2A–B illustrate this difference. We conduct the same analysis across topics. Most topics exhibit the same overall pattern, with one notable exception: in creative topics, ChatGPT outputs display substantially higher diversity than Google Search results (1.190 vs. 1.114, $p < 0.001$). To account for the possibility that extreme maximum values may introduce outliers, we conduct a robustness check using the 95th percentile of the distribution, and the results are qualitatively similar.

3.0.3 Impact of input on users. We then examine conversational dynamics, focusing on longer conversations (i.e., multi-turn user–AI interactions), and analyze how the diversity of user inputs changes over time. Figure 3A shows that input diversity steadily decreases as the conversation progresses. On average, the diversity score decreases by 4.6% from turn 0 to the 70th turn. Based on qualitative inspection of randomly sampled conversations, we observe that users tend to ask follow-up questions that are strongly conditioned on the AI's previous responses.

We next ask whether the diversity of AI outputs is associated with the diversity of user inputs in the subsequent turn. To test this, we estimate an OLS regression in which the dependent variable is the user's next-turn input diversity score and the independent variable is the diversity of the AI's focal output (controlling for the user's input diversity at the focal turn). We additionally control for the user's input diversity in the focal turn, since user input diversity may confound AI output diversity. We find that a one SD increase in AI output diversity is associated with an 0.50 percent increase in user input diversity in the following turn.

4 Discussion and Future Work

By conducting a quantitative empirical study of over 200,000 human–ChatGPT interaction logs, we find that this new mode of knowledge-seeking engagement enabled by generative AI expands the range of inquiries beyond what traditional search engines can readily support. However, the resulting information returned by generative AI can be narrower in scope. We further show that the diversity of AI responses is associated with subsequent changes in users' inquiry diversity. Together, our results highlight ChatGPT's potential to broaden curiosity fulfillment, while also revealing limitations and opportunities for improvement in response diversity.

Our finding that majority ChatGPT queries are non-search-engine–style conveys two key points. First, the 79% of non-searchable queries indicate that users are primarily using ChatGPT in new modes of information inquiry that go beyond traditional search, and that the knowledge space covered by these non-searchable queries is substantially broader than that of searchable queries. At the same time, the 21% of searchable queries shows that some conventional search habits persist, even in a Gen-AI–dominant environment. Users continue to rely on familiar information-seeking patterns when interacting with ChatGPT, similar to how they use Google for search. However, the declining proportion of searchable queries over time suggests that users may gradually adapt to new ways of inquiry—shifting from keyword-based search habits to more conversational, exploratory interactions with ChatGPT. Moreover, the topic-level variation in searchability indicates that search habits are closely associated with subject matter. Topics that are more "searchable" tend to be those that users have historically searched for frequently on traditional search engines, and these entrenched habits may therefore take longer to change.

Comparing ChatGPT's outputs for searchable user queries with Google search results shows that, on average, Google provides more diverse information. This adds an important layer to our findings: while Gen-AI appears to expand users' curiosity and encourage new ways of formulating questions, the answers it provides may be more narrowing in scope. Our third analysis further shows that the diversity of Gen-AI responses is associated with the diversity of users' subsequent queries. This suggests that users' information-seeking behavior is shaped by AI outputs, and that increasing the diversity of AI responses may inspire users to explore knowledge across a broader conceptual space. Our results resonate with prior literature examining diversity in settings such as opinion generation and creative problem solving[4, 8, 9, 13], which generally finds that AI systems provide narrower perspectives across most topics. However, our study further shows that for certain tasks—such as brainstorming—AI can generate more diverse outputs than Google Search. More importantly, we find that the diversity of AI responses can meaningfully influence the diversity of users' subsequent information seeking.

The study is not without limitations and also motivates several directions for future work. Unfortunately, there is currently no publicly available dataset that allows us to directly compare search-engine user interactions and ChatGPT user interactions within the same time period, which constitutes a limitation of this study. Future work could replicate our analysis using real-world search-engine interaction data and directly compare it with ChatGPT interaction data. More importantly, many search engines are now integrating generative AI and traditional search modes. To better understand human curiosity and the structure of knowledge exploration, it is therefore critical to study how these hybrid systems reshape search behavior, what new possibilities they unlock, and what unintended consequences may emerge. Another promising direction for future research is longitudinal user studies that examine how sustained interaction with new information-inquiry systems influences users' engagement with online knowledge over time, and how these changes may ultimately affect decision-making and innovation. We hope our study opens up a new approach for quantitatively understanding how information diversity in generative-AI–mediated conversational search compares with that of traditional search-engine–based search.

References

[1] Michelle Brachman, Amina El-Ashry, Casey Dugan, and Werner Geyer. 2025. Current and future use of large language models for knowledge work. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–24.

[2] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction* 5,

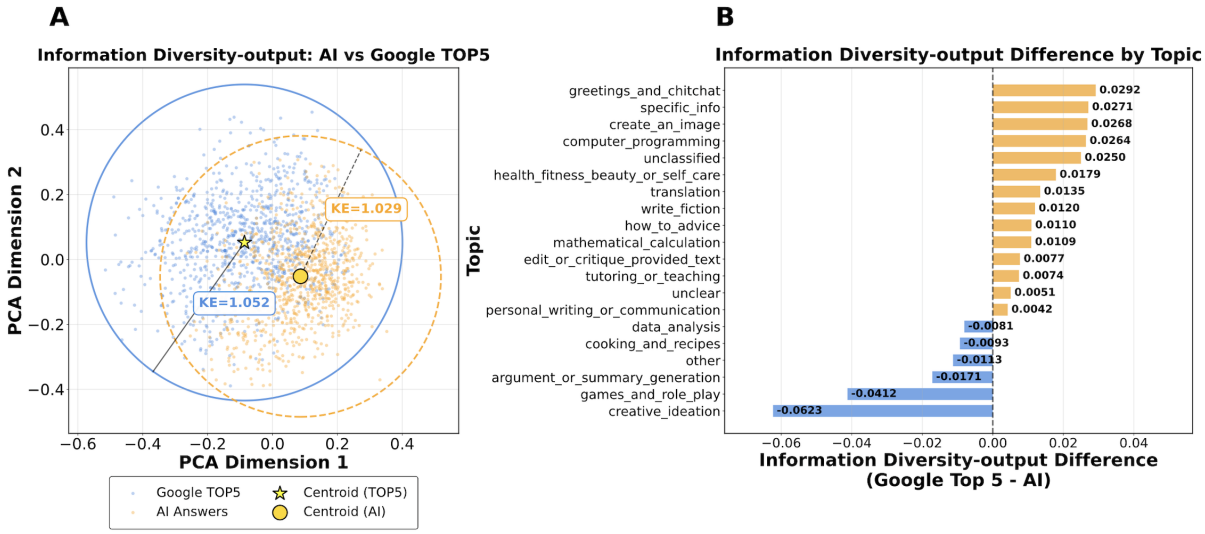

Figure 2: Diversity of ChatGPT answers versus Google Search results for the same searchable inputs. (A) Responses from ChatGPT (one per query) and Google Search (top five results) are embedded into a 384-dimensional vector space using the all-MiniLM-L6-v2 model, and knowledge extent is defined as the maximum Euclidean distance from the centroid. (B) Topic-level differences in knowledge extent between Google Top-5 results and Chatgpt answers (Google minus Chatgpt), sorted by effect size. Positive values indicate broader semantic coverage for Google; negative values indicate broader coverage for Chatgpt.

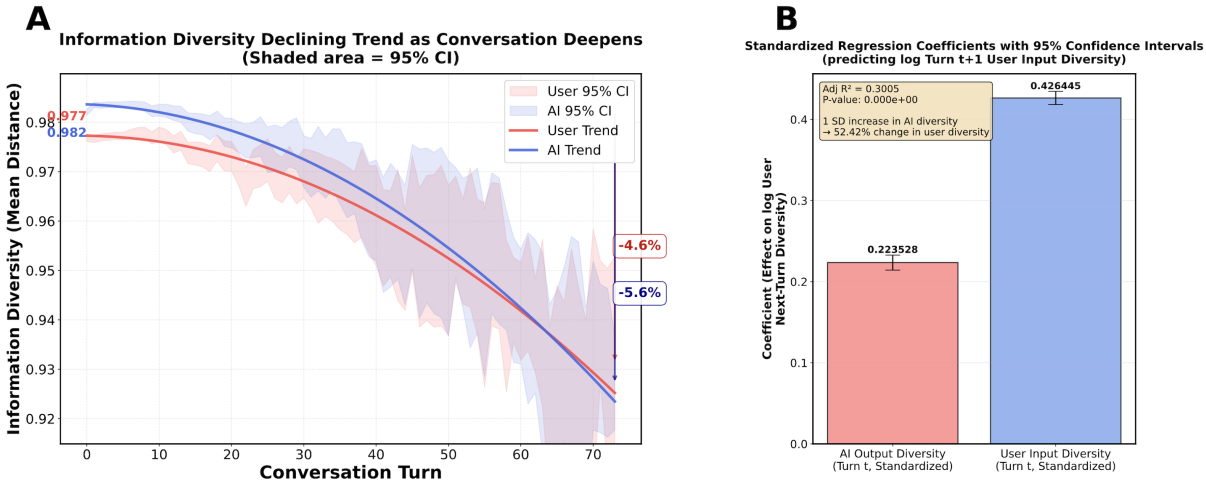

Figure 3: User input diversity over the course of the conversation. (A) Mean input diversity across conversation turns in user–ChatGPT interactions. Shaded regions indicate 95% confidence intervals. (B) Regression coefficients from a model predicting the diversity of the next user input (turn t+1) using (i) Chatgpt-output diversity at turn t and (ii) user-input diversity at turn t. Bars show point estimates with 95% confidence intervals.

CSCW1 (2021), 1–21.

[3] Aaron Chatterji, Thomas Cunningham, David J Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. *How people use chatgpt.* Technical Report. National Bureau of Economic Research.

[4] Anil R Doshi and Oliver P Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science advances* 10, 28 (2024), eadn5290.

[5] Rui Fang et al. 2014. Diversity in collaborative innovation. *Information Systems Research* 25, 4 (2014), 725–744.

[6] Qianyue Hao, Fengli Xu, Yong Li, and James Evans. 2026. Artificial intelligence tools expand scientists' impact but contract science's focus. *Nature* (2026), 1–7.

[7] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference.*

[8] Harsh Kumar, Jonathan Vincentius, Ewan Jordan, and Ashton Anderson. 2025. Human creativity in the age of llms: Randomized experiments on divergent and convergent thinking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems.* 1–18.

[9] Byung Cheol Lee and Jaeyeon Chung. 2024. An empirical investigation of the impact of ChatGPT on creativity. *Nature Human Behaviour* 8, 10 (2024), 1906–1914.

[10] Jialu Liu, Yongfeng Zhang, Min Chen, et al. 2023. Conversational search and recommendation: State of the art and future directions. *Comput. Surveys* (2023).

[11] Rock Yuren Pang, Hope Schroeder, Kynnedy Simone Smith, Solon Barocas, Ziang Xiao, Emily Tseng, and Danielle Bragg. 2025. Understanding the LLM-ification of CHI: Unpacking the Impact of LLMs at CHI through a Systematic Literature Review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[12] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you.* Penguin.

[13] 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.

[14] Ben Shneiderman. 2022. Human-centered AI: A framework for responsible systems. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[15] Siddharth Suri, Scott Counts, Leijie Wang, Chacha Chen, Mengting Wan, Tara Safavi, Jennifer Neville, Chirag Shah, Ryen W White, Reid Andersen, et al. 2024. The use of generative search engines for knowledge work and complex tasks. *arXiv preprint arXiv:2404.04268* (2024).

[16] Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. Atypical combinations and scientific impact. *Science* 342, 6157 (2013), 468–472.

[17] Ryan Yen, Nicole Sultanum, and Jian Zhao. 2024. To search or to gen? Exploring the synergy between generative AI and web search in programming. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[18] Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. 2024. Wildchat: 1m chatgpt interaction logs in the wild. *arXiv preprint arXiv:2405.01470* (2024).