[redacted]

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

In Re: OpenAI, Inc. Copyright Infringement Litigation This Document Relates To: All Cases	Case No. 1:25-md-3143 (SHS) (OTV

DECLARATION OF CHRISTOPHER A. BAIL, Ph.D. IN SUPPORT OF MICROSOFT'S MOTIONS FOR SUMMARY JUDGMENT

[redacted]

I, Christopher A. Bail, Ph.D., hereby declare as follows:

1. I am over the age of 18 and am competent to make this declaration.

2. I have been engaged by counsel for Microsoft to analyze and categorize the ways in which U.S. users use consumer Copilot in the English language and assess the prevalence of each use.[1] In support of this assignment, I analyzed a sample of 45,000 conversations that U.S. users had with consumer Copilot between December 2023, when consumer Copilot became generally available in the U.S., and October 2024, when the product was deprecated.[2] I refer to this as the "Copilot Conversations Sample." I designed and supervised a quantitative content analysis conducted by human coders between February and March 2026.

3. I submitted an opening expert report in this case on April 23, 2026, a rebuttal expert report on May 29, 2026, and a reply expert report on June 26, 2026. The opinions below are a subset of the opinions set forth in my Opening and Reply Reports. The statements in my Opening and Reply Reports are true and accurate to the best of my knowledge, information, and belief. Except as otherwise stated herein, I make this declaration based on my personal knowledge and professional expertise, documents produced in this case (as disclosed in my reports), reports by other experts retained by Microsoft in this case (including Professor On Amir and Dr. John Lafferty), and research and analysis conducted by me or under my direction and supervision.

---

[1] My analysis considers only use of the consumer product Bing Chat / Copilot available after its general launch in December 2023 through the product overhaul in October 2024, as I understand from counsel that was the only pool of usage data that could reasonably be made available for study to both sides during the discovery process. Unless I specify otherwise, I use the term "Copilot" to refer solely to the consumer version of the product as defined above. Additionally, I understand from counsel that the only copyrighted works at issue in these cases are U.S. works that were authored in the English language, and that the Copilot product as initially deployed by Microsoft was primarily aimed at English language users. In light of these and other factors, I limited my content analysis to the English language.

[2] *See* Lafferty Declaration, Sections VI (Methodology for Chat Log Data Sample Selection) and VII (Sample Size Robustness Checks) for details on sample selection.

[redacted]

I. Qualifications

4. I am a Professor of Sociology, Computer Science, Political Science, and Public Policy at Duke University, where I also direct the Society-Centered AI Initiative. I lead interdisciplinary teams of researchers who study how technology shapes human behavior using quantitative content analysis techniques, online experiments, and public opinion surveys. Prior to my current role, I held a faculty position at the University of North Carolina, Chapel Hill.

5. I received my B.A. degree in Government and French from Bowdoin College in 2002 and my Ph.D. in Sociology from Harvard University in 2011. I was a visiting scholar at the Fondation Nationale des Sciences Politiques in Paris in 2005 and at the London School of Economics and Political Science in 2008. I completed a post-doctoral fellowship at the University of Michigan from 2011–2013.

6. I have published numerous books and articles within university presses and peer-reviewed journals. My work has appeared in *Science*, *Nature*, and the *Proceedings of the National Academy of Sciences* and has been supported by the National Science Foundation, the National Institutes of Health, the National Security Agency, the Templeton Foundation, and many other funding agencies. Coverage of my work has appeared in *The New York Times*, *The Wall Street Journal*, *The New Yorker*, and *Wired*, on BBC, CBS News, CNN, and NBC *Nightly News,* and many other media outlets.

7. I regularly work with governments, tech companies, and non-profit organizations to evaluate technology's impact on human behavior. I have engaged with many organizations on this topic, including the White House, the Vatican, Google, Meta, OpenAI, the Pew Research Center, and many others.

8. I have received numerous awards for my scholarship, including a Guggenheim Fellowship, a Carnegie Fellowship, and multiple distinctions from the American Sociological Association, the American Political Science Association, and the Association for Research on Non-Profit and Voluntary Organizations. I have served as a keynote speaker at the International Conference on Weblogs and

[redacted]

Social Media, the International Conference on Computational Social Science, and the U.S. National Science Foundation.

9. I am the editor of the Oxford University Press Series in Computational Social Science and served on the Advisory Council to the National Science Foundation's Social, Behavioral, and Economic Sciences Directorate from 2018–2023. I am the Founding Director of the Summer Institutes in Computational Social Science—a program that has trained more than 5,000 students in 52 locations around the world—and designed the curriculum for Duke's Master's Program in Interdisciplinary Data Science. I formerly served as an editor, consulting editor, or editorial board member at the *Proceedings of the National Academy of Sciences*, *Science Advances*, the *American Journal of Sociology*, and other academic journals. I have served in various policy and management committees at Duke University. Finally, I have supervised or served on the dissertation committees of more than thirty doctoral students.

10. My curriculum vitae, which contains further details on my qualifications, publications, and other professional activities, as well as my last four years of expert testimony are attached as **Appendix A**.

II. Summary of Opinions

11. Based on my review of decades of academic literature on how people use web search, the emerging recent literature on how people use LLM technologies, my review and understanding of the results of the survey conducted by On Amir (the "Amir Survey"),[3] my analysis of the Copilot Conversations Sample, my professional expertise, knowledge, and training, and relying on standard methods I employ in my academic work, I have reached the following opinions.

    a. Quantitative content analysis is a well-established, valid, and reliable method for interpreting and classifying textual data and is suitable for

[3] *See* Amir Declaration.

[redacted]

analysis of LLM conversations. To construct my taxonomy, I take into consideration previous research on how people use web search, how they use LLMs, and the findings in the Amir Survey, in addition to my preliminary review of consumer Copilot conversations. I trained and supervised a team of human coders who analyzed the Copilot Conversations Sample and assigned each conversation one of the user intent categories defined in the taxonomy. I discuss this in **Sections III through V**.

b. As reflected in the overall classification of the sample and consistent with my personal review of a sample of conversations from each category, U.S. English-speaking users of consumer Copilot used it for multiple, varied, and distinct purposes. My analysis demonstrates that most users go to Copilot to search for information, to learn new skills or solve problems, to seek opinions or advice, to assist with creating or editing images, to edit text, to revise computer code, or to brainstorm ideas, along with other less common uses. I discuss this in **Section VI.A**.

c. Information Retrieval is the most common use; the subjects about which users inquire are highly varied and indicative of an intent to learn about numerous topics; the degree to which users exhibit the types of uses identified as problematic by the Plaintiffs is infrequent at best. I discuss this in **Section VI.B**.

d. The coding techniques used to categorize usage exhibit high intercoder reliability, and additional analyses corroborate that my findings are reliable. I discuss this in **Section VI.C**.

[redacted]

III. Quantitative Content Analysis Is a Well-Established and Reliable Method for Interpreting and Classifying Textual Data

12. Content analysis is a research method used to make scientifically valid and replicable inferences from unstructured text-based data.[4] It involves the systematic classification of content—by human coders or computational tools—to translate qualitative material into quantitative data that can be analyzed rigorously. Quantitative content analysis emerged in the 1900s, when historians began applying quantitative techniques to study historical documents.[5] The technique became codified in academic literature over the course of the 20th century.[6] Subsequently, content analysis surged in popularity and usage across disciplines, supported by the advent and growth of computing.[7]

13. Quantitative content analysis has been used in a broad range of contexts. For example, content analysis is frequently used to analyze newspaper articles, books, audio transcripts, open-ended survey response data, and many other types of text-based data within psychology, political science, communications, sociology, medicine, and many other fields.[8] Quantitative content analysis is also broadly used within the field of human–computer interaction, and is frequently employed to categorize human behavior in web search data, blog posts, social media

---

4 Neuendorf, Kimberly A., *The Content Analysis Guidebook*, Sage Publications, 2002 ("Neuendorf"), p. 1 ("Content analysis may be briefly defined as the systematic, objective, quantitative analysis of message characteristics.").

5 Neuendorf, pp. 31-32.

6 Neuendorf, p. 30 ("The history of the use of the various techniques called content analysis has been documented only piecemeal, with some histories emphasizing text analysis (e.g., Stone et al., 1966); some, computer text analysis (e.g., Diefenbach, in press); and others, specific applications (e.g., Rosengren, 1981), including communication (Berelson, 1952) and psychological diagnosis (Gottschalk, 1995).").

7 Neuendorf, p. 40 ("With the diffusion of the personal computer, both text analysis and data analysis have been brought to the desktop. PCs have revolutionized the means by which messages are created, stored and archived, disseminated, and organized."). As an example, one study of health messages in U.S. media from 1985 to 2005 identified 441 articles in peer-reviewed journals that involved quantitative content analysis methods, *see* Manganello, Jennifer and Nancy Blake, "A Study of Quantitative Content Analysis of Health Messages in U.S. Media From 1985 to 2005," *Health Communication*, Vol. 25, No. 5, July 30, 2010, pp. 387 - 396.

8 *See, e.g.*, Neuendorf, pp. 191-211.

[redacted]

messages, and, more recently, LLMs.[9] I also use quantitative content analysis in my own research. For example, I performed a content analysis for a study of how cultural tastes and consumer interests spread globally across countries using a large database of internet search terms, as documented in my peer reviewed publication "Prestige, Proximity, and Prejudice."[10]

14. Once a research question has been developed and a sample of relevant content has been identified, quantitative content analysis proceeds through four broad stages: (i) development of the coding taxonomy, a set of categories that will be applied to the content; (ii) development of training materials, followed by recruitment and training of appropriate coders; (iii) application of the final coding taxonomy to the content, enabling rigorous and reliable categorization of a large set of texts by multiple human coders, and (iv) reporting of results and reliability measures.[11]

15. In this matter, my research question is described by my assignment in paragraph 2 above. I describe the content to be studied in my analysis, the Copilot Conversations Sample, in **Section V.A**.

16. After the research question and data source have been identified, a quantitative content analysis proceeds by identifying the categories for which the selected content will be assessed. A common way to identify categories is using previous

---

[9] *See, e.g.*, Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser, "Analyzing Qualitative Data," *Research Methods in Human-Computer Interaction*, 2 Ed., Todd Green, 2017, pp. 300-302; Rose, Daniel E. and Danny Levinson, "Understanding User Goals in Web Search," *Proceedings of the 13th International Conference on World Wide Web 2004*, May 2004, pp. 13 - 19; Bodonhelyi, Anna, et al., "User Intent Recognition and Satisfaction with Large Language Models: A User Study with ChatGPT," *arXiv*, November 19, 2024.

[10] Bail, Christopher A., Taylor W. Brown, and Andreas Wimmer, "Prestige, Proximity, and Prejudice: How Google Search Terms Diffuse Across the World," *American Journal of Sociology*, Vol. 124, No. 5, March 2019, pp. 1496 - 1548. This research has been published in the *American Journal of Sociology*, a top-tier sociology journal.

[11] *See* Neuendorf, pp. 50-51 for a more detailed breakdown of this process.

[redacted]

research and theory.[12] However, when such research is insufficient or does not exist, researchers also frequently create taxonomies in an inductive manner through rigorous analysis of a pilot sample of documents.[13] I describe my approach to category development in my quantitative content analysis in **Section IV**.

17. Once the initial taxonomy has been developed, researchers typically recruit a team of human coders to apply the taxonomy to the set of documents. Coders must be carefully selected and trained to ensure that they have a thorough understanding of the categories and how they should be applied to the documents. Quantitative content analyses typically involve multiple rounds of coder training where the researcher provides exemplars of each category and tests the human coders' capacity to categorize documents in a reliable manner. This phase may also include category refinement, in which coders practice applying the categories to the content, and the researcher can learn about the reliability of the coding taxonomy and make revisions before the final coding process.[14] I describe coder training, category refinement, and pilot testing procedures for my content analysis in **Section V.B**.

18. The final step in operationalizing a quantitative content analysis is for coders to apply the finalized coding taxonomy to the content sample. Once the content analysis is complete, final results can be reported, reliability measures can be assessed, and sensitivity analyses can be performed. I describe the results of my

---

[12] Neuendorf, p. 97 ("The sections that follow describe four recommended techniques for selecting variables for a content analysis: 1. A consideration of universal variables 2. Using theory and past research for variable collection 3. A grounded or "emergent" process of variable identification 4. Attempting to find medium-specific critical variables").

[13] Neuendorf, pp. 102-103 ("When existing theory or research literature cannot give a complete picture of the message pool, the researcher may take a more practical approach. The researcher may need to immerse himself or herself in the world of the message pool and conduct a qualitative scrutiny of a representative subset of the content to be examined. In this way, variables emerge from the message pool, and the investigator is well grounded in the reality of the messages.").

[14] Neuendorf, p. 133 ("Three words describe good coder preparation: Train, train, and train. As part of the training process, the researcher may need to revise the codebook repeatedly until researcher and coders are all comfortable with the coding scheme. Practice coding, called pilot coding, can inform the researchers as to the reliability and overall viability of the coding scheme. Then, revisions may be made before final coding commences.").

[redacted]

content analysis in **Sections VI.A and VI.B**. A quantitative content analysis must include a method to assess the reliability with which different coders apply the same categories to the same content.[15] Reliability of a quantitative content analysis can be measured by assessing the rate of coder agreement and using statistics that adjust for the probability that an agreement occurs by chance, such as Cohen's Kappa.[16] I describe the manner in which multiple coders reviewed content in my analysis, along with intercoder reliability metrics, in **Section VI.C**.

IV. Development of Coding Categories Using Best Practices in Content Analysis

19. In order to develop my taxonomy for analysis of the Copilot Conversations Sample, I started with a review of existing literature on both web search and LLM

[15] Neuendorf, pp. 112, 142-144 ("Reliability is the extent to which a measuring procedure yields the same results on repeated trials.").

[16] Neuendorf, pp. 148-151.

[redacted]

usage.[17] To ensure that the categories identified in the broader literature encompassed the specific uses of Copilot, I also reviewed Microsoft studies that examine Copilot usage and consulted findings from the Amir Survey. Once I developed a preliminary taxonomy, I assessed and iteratively revised definitions and examples in the taxonomy during the category refinement and coder training phase of my content analysis, and confirmed the final taxonomy could be applied reliably through coding of a pilot sample.

A. I Drew Upon Prior Literature on Web Search and LLM Usage for Insights into Consumer Use of Copilot

20. Since LLMs such as OpenAI's ChatGPT, Microsoft's Copilot, Anthropic's Claude, or Google's Gemini started to become widely accessible to consumers in late 2022

[17] I recognize that the literature on how people employ both web search and LLMs cover a wide range of research topics and concepts, in addition to usage or user intent. While for the purposes of my assignment I focused on literature on user intent, my review of the literature surfaced many studies on topics adjacent to – but not directly informative of – user intent. For example, one large body of literature analyzes LLM adoption in the workplace (*see, e.g.,* Bonney, Kathryn, et al., "Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey," *National Bureau of Economic Research*, April 2024; Humlum, Anders and Emilie Vestergaard, "The Unequal Adoption of ChatGPT Exacerbates Existing Inequalities Among Workers," *Proceedings of the National Academy of Sciences*, Vol. 122, No. 1, December 30, 2024; National Academies of Sciences, Engineering, and Medicine, *Artificial Intelligence and the Future of Work*, Washington, DC, National Academies Press, 2025 ("Artificial Intelligence and the Future of Work")). Another prevalent research topic I encountered involves studying the impact of LLM adoption on worker productivity (*see, e.g.*, Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond, "Generative AI at Work," *The Quarterly Journal of Economics*, Vol. 140, No. 2, February 4, 2025, pp. 889 - 942; Cambon, Alexia, et al., "Early LLM-based Tools for Enterprise Information Workers Likely Provide Meaningful Boosts to Productivity," *Microsoft Research*, December 2023; Edelman, Ben, Donald Ngwe, and Sida Peng, "Measuring the Impact of AI on Information Worker Productivity," *SSRN*, November 2023; Noy, Shakked and Whitney Zhang, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," *Science*, Vol. 381, No. 6654, July 14, 2023, pp. 187 - 192; Peng, Sida, et al., "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," *arXiv*, February 13, 2023; "What Can Copilot's Earliest Users Teach Us About Generative AI at Work?," *Microsoft*, November 15, 2023, https://www.microsoft.com/en-us/worklab/work-trend-index/copilots-earliest-users-teach-us-about-generative-ai-at-work). Because these studies do not address user intent, they were not the focus of my targeted literature review.

[redacted]

and throughout 2023,[18] academics and industry professionals have sought to understand how people use these tools in their daily lives.[19] Just over three years later, this body of research is still in the early stages of development, with new research emerging frequently. As a result, I believed it was important to look more broadly at research regarding use of information technologies such as web search. I performed a targeted literature review to collect and review studies that may be informative to classifying usage of LLMs. I start by providing details on my approach to the literature review. I then discuss relevant features of the identified studies, such as data sources, methodologies, and research questions, which informed the design of my analytical framework. I conclude by summarizing existing research that specifically focuses on various aspects of Copilot usage.

1. Targeted Literature Review Methodology

21. Given the relative infancy of research on LLM usage, I began my review by analyzing the well-established literature on how people use web search. See **Appendix C.1** for the core web search literature I reviewed.

22. Research suggests that LLMs' abilities to tailor answers to a particular context, answer more complex questions, synthesize, explain, and analyze information, and provide responses in a more customized and directly usable way allow users to expand usage into new areas that were not available when using traditional

---

[18] *See, e.g.,* OpenAI released ChatGPT to the public in November 2022. History.com Editors, "ChatGPT, the Generative AI Chatbot, is Released," *History*, November 24, 2025, https://www.history.com/this-day-in-history/november-30/chatgpt-released-openai. Claude and Gemini (formerly known as "Bard") were released in March 2023. Wiggers, Kyle, "Anthropic Launches Claude, a Chatbot to Rival OpenAI's ChatGPT," *TechCrunch*, March 14, 2023, https://techcrunch.com/2023/03/14/anthropic-launches-claude-a-chatbot-to-rival-openais-chatgpt/; "An Overview of the Gemini App," *Gemini*, https://gemini.google/overview/, accessed March 11, 2026.

[19] This topic has also been covered by popular press and general-interest media. *See, e.g.,* Zao-Sanders, Marc, "How People Are Really Using GenAI," *Harvard Business Review*, March 19, 2024, https://hbr.org/2024/03/how-people-are-really-using-genai; Merrill, Jeremy B. and Rachel Lerman, "What Do People Really Ask Chatbots? It's a Lot of Sex and Homework," *Washington Post*, August 4, 2024, https://www.washingtonpost.com/technology/2024/08/04/chatgpt-use-real-ai-chatbot-conversations/.

[redacted]

search engines.[20] So while the literature on web search is part of my analytical approach and provides a foundation that the literature on LLM usage builds upon, this literature is insufficient to analyze the prevalence of different uses of Copilot by itself.

23. To gain an understanding of the landscape of how existing literature categorizes LLM uses, I conducted a thorough review of studies of LLM usage that develop taxonomies of user intents based on analyses of chatlog data.[21] My review spanned published, peer-reviewed studies as well as working papers and industry reports in areas of human–computer interaction, information retrieval, and the social impact of artificial intelligence, among others. My review aimed to surface studies that (i) examined real-world chatlogs,[22] (ii) included a taxonomy, and (iii) tried to discern user intent.[23] Accordingly, at my direction, my support team conducted a targeted keyword-based search of academic publications such as the Journal of the ACM (JACM) and ACM Transactions on the Web, along with databases and working paper repositories such as Google Scholar, arXiv, SSRN,

---

[20] *See, e.g.*, Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," March 19, 2024, working paper, p. 8 ("When people use a generative search engine, they are interacting with online information differently than when they use a traditional search engine. They are not simply looking information up but instead doing tasks that involve processing that information, for example by analyzing or evaluating it"); Yu, Yulin, et al., "From Searchable to Non-Searchable: Generative AI and Information Diversity in Online Information Seeking," *CHI EA '26: Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*, April 2026, p. 1 ("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."); Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29 at p. 4 ("We expect chat interactions to exhibit new intents (e.g., creation) compared to traditional search.").

[21] The term "user intent" is commonly employed in this literature to refer to the underlying objective or need that motivates a user's input into a web search engine or an LLM. While neither I nor an LLM can read a user's mind, "user intent" in this context refers to an attempt to use the available input supplied by the user to understand what the user is trying to accomplish.

[22] Real-world chatlogs refers to records of independent interactions between users and LLMs, as opposed to chatlogs generated in a research setting and for the purposes of research.

[23] I focused on studies examining general types of LLM uses, excluding those limited to specific sectors or domains such as healthcare or software development. *See, e.g.*, Paruchuri, Akshay, et al., "'What's Up, Doc?': Analyzing How Users Seek Health Information in Large-Scale Conversational AI Datasets," *Association for Computional Linguistics*, November 2025, pp. 2312 - 2336; Zhang, Binquan, et al., "Decoding Human-LLM Collaboration in Coding: An Empirical Study of Multi-Turn Conversations in the Wild," *arXiv*, December 12, 2025.

[redacted]

NBER, and ScienceDirect.[24] Because this body of literature is still developing and relevant sources in this area of research may come in many forms including conference proceedings or industry studies, my support team also searched proceedings from conferences such as the ICLR (The International Conference on Learning Representations), the ACM CHI conference on Human Factors in Computing Systems, the Conference on Empirical Methods in Natural Language Processing, and the ACM SIGIR Conference on Research and Development in Information Retrieval. I further coupled the targeted keyword search by drawing on academic literature and studies known to me in my capacity as a researcher studying how technology shapes human behavior. Finally, I expanded the set of potentially relevant articles through review of references *cited in* these studies and review of studies that *cited* this literature subsequently.

24. Taken together, the process outlined above yielded hundreds of studies. As an initial step, my support team under my guidance screened abstracts of studies returned in targeted searches or through cross-references in order to identify a smaller subset of candidate articles for in-depth review. The 30 studies on LLM use identified for in-depth review are listed in **Appendix C.2**.

25. Whether studies met only some or all of my criteria, their methodological approaches are nevertheless useful points of reference for developing the research techniques necessary to answer my research question.[25] In the following sections, I discuss the LLM literature listed in **Appendix C.2** with respect to each of the three criteria, describing the range of decisions researchers have made regarding each dimension.

---

[24] My support team under my guidance used a variety of keywords and search terms including but not limited to variations of phrases such as "real world AI use," "how people use LLMs," "user intent taxonomy AI," and Boolean search queries combining terms such as "LLM" and ("user intent" or "usage") and ("taxonomy" or "classification" or "typology" or "prompt analysis").

[25] For example, while Zhao et al. applied a pre-existing taxonomy and classifier to a dataset of real-world chatlogs they did not develop this taxonomy and neither the authors nor the underlying source provide any details on the methodology used to create this taxonomy. I include this study in **Appendix C.2** but because of the lack of methodological detail I do not include it among the nine studies at the center of my user intent taxonomy. *See* Zhao, Wenting, et al., "WILDCHAT: 1M ChatGPT Interaction Logs in the Wild," *ICLR 2024*, May 2, 2024, pp. 3-4, Table 4, note 9.

[redacted]

2. Previous Studies Employ a Broad Range of Data Collection Methods to Study Different Types of LLM Usage

26. Studies considered in my review of literature rely on data that vary widely across many dimensions, including the particular LLM-based product or context in which data were generated or collected, the extent to which behavior was directly measured or self-reported, whether the data capture real-world usage or are synthetic—that is, generated for research purposes—and the population of users represented in the data, among others. Though no single study examines the same data on usage of Copilot that I have been asked to analyze, I rely on these studies in forming my opinions and draw upon methods and conclusions presented therein to the extent applicable to my assignment.

27. Many of the studies I reviewed relied on real-world chatlogs to assess observed usage of LLMs. These observational studies can generally be grouped into two broad categories according to the data being analyzed: (i) ad hoc datasets compiled for research or testing purposes, often collected using incentives to encourage participation, and not necessarily representative of average users' use; and (ii) samples of user chatlogs accessing a particular LLM product, often analyzed by the developer of that product. Both categories can be informative with regards to the methodological approach to analysis of LLM usage. Even if not representative of average U.S. users' use of consumer Copilot, findings about prevalence of different types of LLM uses from these studies can still provide insight as to the types of tasks for which LLMs are used.[26]

[26] There is a tradeoff between "observer bias," present when participants know they are being monitored, and "self-selection bias," present when users can choose ex-post which conversations to share. *See* Yan, Yueru, et al., "SHARECHAT: A Dataset of Chatbot Conversations in the Wild," *arXiv*, January 27, 2026, p. 2 ("[M]ost datasets rely on participants who know they are being watched, introducing the Observer Bias (or Hawthorne Effect) [...]. Users who are aware they are being monitored for research purposes may consciously or unconsciously alter their behavior, which often exhibits higher social desirability compared to authentic, unobserved private usage.") and p. 9 ("[The approach] introduces Self-Selection Bias (Heckman, 1979), because users voluntarily choose which conversations to generate links for, the corpus is likely skewed towards interactions they deem successful, interesting, or novel. Therefore, SHARECHAT may underrepresent mundane utility tasks or failed interactions where the model performed poorly, as users are less motivated to share such content publicly (Hu et al., 2017).").

[redacted]

28. In short, existing studies offer unique insights into how people use LLMs, even if the exact characteristics of underlying data vary. While the findings of these studies may not be generalizable to my study of Copilot usage during the time period of interest and relevant to my assignment, they serve as valuable reference points for development of my own taxonomy.

3. Previous Studies on LLM Usage Rely on Different Methodologies to Create Taxonomies and/or Categorize Use

29. Literature analyzing LLM usage usually proceeds in two steps. In the first step, the researcher develops a set of labels, or a taxonomy, for all potential uses. In the second step, the researcher proceeds to assign each observation, which can be an individual prompt or a full chatlog, one or more labels from the list developed in the first step.

4. Microsoft's Publicly Released Research on LLM Usage

30. Given the focus of my analysis on Copilot and in light of the product-specific differences in usage of LLMs, I reviewed several studies that Microsoft has made public over the past few years relating to usage of its Copilot product. These studies, along with the rest of the literature I reviewed, helped inform development of my taxonomy and informed the types of usage I expected to be present among Copilot users.

31. Suri et al. (2024) compared the ways people use search engines and LLMs.[27] The authors employed an LLM to generate "domain" or subject matter categories among Bing Copilot and Bing Search data collected between May and July

[27] Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," March 19, 2024, working paper.

[redacted]

2023.[28] They found that people used Bing Copilot for more complex tasks and tasks in a domain more likely to be associated with knowledge work.[29] The top three domains they identified were "business and economics," "translation and language learning," and "web development."[30]

32. Shah et al. (2025) proposed a new methodology to categorize Bing Chat usage by English speakers from May to June 2023 that relies on LLMs to generate a user intent taxonomy that is validated and refined by human coders.[31] They found the three most common user intents were "information retrieval," "learning," and "problem solving."[32]

33. Another study that informed my understanding of Copilot usage is Tomlinson et al. (2025), which examined which O*NET work activities users asked Microsoft Bing Copilot to assist with.[33] The authors did not create a taxonomy that was designed to describe the data. Instead, they applied O*NET to compare workforce

---

[28] Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," March 19, 2024, working paper, pp. 2-3 ("We created a dataset of 80000 randomly sampled, de-identified Bing Copilot conversations by selecting 10000 conversations at random for each of eight consecutive weeks (weeks starting May 28, 2023 to ending July 22, 2023)[.] [...] [W]e used GPT-4 to create labels and domain definitions and classify each Bing Copilot conversation into one of 25 topical domains [...]. [W]e note that it is an unsupervised algorithm that generates the taxonomy of these data organically in a bottom-up fashion [...].").

[29] Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," March 19, 2024, working paper, p. 5 ("Thus, not only are people using Bing Copilot for more knowledge work, but they are also using them for more complex tasks.").

[30] Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," March 19, 2024, working paper, p. 4.

[31] Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29 at pp. 6-7.

[32] Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29 at p. 12. The third top user intent category for differs based on the underlying data. In Table 3, among the conversations that were coded by GPT-4, the category with the third largest number of prompts was "Problem Solving." Among conversations that were coded by humans, the category with the third largest number of prompts was tied between "Problem Solving" and "Content Creation."

[33] Tomlinson, Kiran, et al., "Working with AI: Measuring the Applicability of Generative AI to Occupations," *arXiv*, December 22, 2025, p. 8. ("We analyze anonymized and privacy-scrubbed U.S. conversation data from Microsoft Bing Copilot [...] gathered from January 1, 2024 to September 30, 2024. We focus only on conversations in the United States to align with occupation and work activity information from O*NET.").

[redacted]

tasks and LLM tasks.[34] They found users most often use these tools to "gather information," "obtain info about goods/services," and "respond to customer problems / inquiries."[35]

34. Costa-Gomes et al. (2025) examined how people used Copilot between January and September 2025, and how that usage varied by device type and over months and hours within a day.[36] The authors used "machine-based classifiers" to categorize chatlogs by both topic and intent, finding that the top three intents were "searching," "advice," and "creating" and the top three topics were "technology," "work and career," and "health and fitness."[37]

35. While research published by Microsoft is informative about the broad landscape of Copilot uses and offers insights about how to create taxonomies that describe such behavior, none of the existing studies quantifies the frequency of different types of consumer Copilot uses utilizing a dataset that is similar to the Copilot Conversations Sample provided to me for this matter.

B. Development of My User Intent Taxonomy

1. Existing Literature Served as a Starting Point for Development of My User Intent Taxonomy

36. As discussed in **Section IV.A**, existing literature provides a useful point of departure to understand Copilot usage. As a result, I began developing my taxonomy by closely reviewing user intent categories identified by previous studies. I also considered the findings of the Amir Survey, findings from my

---

[34] Tomlinson, Kiran, et al., "Working with AI: Measuring the Applicability of Generative AI to Occupations," *arXiv*, December 22, 2025, p. 8 ("We use O*NET's (16) hierarchical decomposition of occupations into their tasks and work activities.").

[35] Tomlinson, Kiran, et al., "Working with AI: Measuring the Applicability of Generative AI to Occupations," *arXiv*, December 22, 2025, p. 3, Figure 1, Panel A.

[36] Costa-Gomes, Beatriz, et al., "It's About Time: The Copilot Usage Report 2025," *Microsoft*, December 2025, pp. 1-2.

[37] Costa-Gomes, Beatriz, et al., "It's About Time: The Copilot Usage Report 2025," *Microsoft*, December 2025, p. 2 ("Each sampled conversation is labeled by machine-based classifiers that assign the conversation a topic and an intent tag [...].") (emphasis removed), p. 5, Table 3.

[redacted]

preliminary review of Copilot conversations, and my professional expertise to adjust and supplement the list of categories. This approach follows a methodology I commonly apply in my research and peer-reviewed publications.

37. From the literature review described in **Section IV.A**, I identified nine key studies on LLM uses that I found to be the most relevant for my assignment. These studies, listed in **Exhibit 1**, represent a combination of industry and academic research and cover Copilot as well as other LLMs. I selected these studies because they satisfied all of the following criteria: they (i) examined real-world chatlogs, (ii) included a taxonomy, and (iii) tried to discern user intent. These criteria were chosen to identify studies closest to my research question and methodology.

[redacted]

Exhibit 1
Summary of Nine Key Studies Informing Taxonomy

Paper Title	Data Source	Data Sample Selection	Data Sample Time Frame	Taxonomy Development	Coding Process	Manual Validation	Top 3 User Intent Categories
Ouyang, Siru, et al., "The Shifted and The Overlooked: A Task-oriented Investigation of User-GPT Interactions," Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Oct 19, 2023	ChatGPT	User-Provided (ShareGPT)	Unspecified	LLM-Driven	LLM-Driven	☑	• Code generation • Creative writing • Explanation
Wang, Jiayin, et al., "Understanding User Experience in Large Language Model Interactions," arXiv, Jan 16, 2024	ChatGPT	User-Provided (ShareGPT)	Unspecified	Human-Driven	Human-Driven	☑	No quantification of chatlogs
Bodonhelyi, Anna, et al., "User Intent Recognition and Satisfaction with Large Language Models: A User Study with ChatGPT," arXiv, Nov 19, 2024	ChatGPT	User-Provided via HuggingFace	Unspecified	Human-Driven	LLM- and Human-Driven	☑	• Factual queries • Explanatory inquiries • Decision support
Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," ACM Transactions on the Web, Aug 22, 2025	Bing	Proprietary Random Sample	May - June 2023	LLM-Driven	LLM-Driven	☑	• Information retrieval • Learning • Problem solving
Chatterji, Aaron, et al., "How People Use ChatGPT," National Bureau of Economic Research, Sep 2025	ChatGPT	Proprietary Random Sample	May - July 2025	Unclear - Modified from Classifier used by OpenAI research teams	LLM-Driven	☑	• Specific info • Edit or critique provided text • Tutoring or teaching
Shelby, Rene, et al., "Taxonomy of User Needs and Actions," arXiv, Oct 10, 2025	ChatGPT	User-Provided (WildChat and ShareGPT)	Unspecified	Human-Driven	Human-Driven	☑	No quantification
Costa-Gomes, Beatriz, et al., "It's About Time: The Copilot Usage Report 2025," Microsoft, Dec 2025	Copilot	Proprietary Random Sample	January - September 2025	Human-Driven	LLM-Driven	☒	• Searching • Advice • Creating
Ammari, Tawfiq, et al., "How Students (Really) Use ChatGPT: Uncovering Experiences Among Undergraduate Students," arXiv, Jan 19, 2026	ChatGPT	User-Provided	December 2022 - January 2024	LLM- and Human-Driven	LLM- and Human-Driven	☑	• Multiple Choice / Fill-in-the-Blank • Concept Explanation • Rewording
Yan, Yueru, et al., "SHARECHAT: A Dataset of Chatbot Conversations in the Wild," arXiv, Jan 27, 2026	Various	Collected Through Publicly Available Links	April 2023 - October 2025	Human-Driven	LLM-Driven	☑	• Seeking information • Other / unknown • Technical help

[redacted]

38. I began to construct my taxonomy of Copilot usage by identifying overlapping categories across these nine studies. If multiple studies used different labels to describe similar user intents, I grouped those categories and applied a broader label, where applicable. I summarize the resulting taxonomy in **Exhibit 2** below.[38]

[38] Some studies included categories in their taxonomies that could not be cleanly mapped to a user intent taxonomy. These are: "Use through API" in Wang et al. (2024); "Service Utilization," "Societal and Cultural Inquiry," "Existential Questions," and "Moral and Ethical Queries" in Bodonhelyi et al. (2024); "Health, Fitness, Beauty or Self-Care," "Cooking and Recipes," and "Purchasable Products" in Chatterji et al. (2025); "Shared Understanding," "System Management," and "Conversation Management" in Shelby et al. (2025); "Validation or Evaluation of Understanding," "Clarification of Instructions or Tasks," "Uncertainty or Confusion Expression," "Self-Reflection or Learning Awareness," "Follow-Up or Clarification Requests," "Interaction Repair or Prompt Revision," "Appreciation or Gratitude," "Frustration or Dissatisfaction," "Confidence Assessments, "Humor or Anthropomorphism," "Learning Goal Setting," "Apologies," "Stated Limitations," "Unintended or Inconsistent Answers," and "Rhetoric" from Ammari et al. (2026).

[redacted]

Exhibit 2
Summary of User Intent Categories in Key Prior Studies

Ouyang et al., "The Shifted and The Overlooked: A Task-oriented Investigation of User-GPT Interactions," Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Oct 19, 2023	**Wang et al., "Understanding User Experience in Large Language Model Interactions,"** arXiv, Jan 16, 2024	**Bodonhelyi et al., "User Intent Recognition and Satisfaction with Large Language Models: A User Study with ChatGPT,"** arXiv, Nov 19, 2024	**Shah et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies,"** ACM TWEB, Aug 22, 2025	**Chatterji et al., "How People Use ChatGPT,"** NBER, Sep 2025	**Shelby et al., "Taxonomy of User Needs and Actions,"** arXiv, Oct 10, 2025	**Costa-Gomes et al., "It's About Time: The Copilot Usage Report 2025,"** Microsoft, Dec 2025	**Ammari et al., "How Students (Really) Use ChatGPT: Uncovering Experiences Among Undergraduate Students,"** arXiv, Jan 19, 2026	**Yan et al., "SHARECHAT: A Dataset of Chatbot Conversations in the Wild,"** arXiv, Jan 27, 2026	**USER INTENT CATEGORY**
• Advice • Planning • Recommendation • Comparison	• Ask for Advice	• Decision Support • Business and Career Advice • Personal Advice • Planning and Organization				• Getting Feedback or Advice • Planning • Shopping and Product Research		• Practical Guidance	Advice / Recommendation / Seeking Opinions
• Brainstorming	• Seek Creativity	• Idea Generation		• Creative Ideation			• Example Generation • Activity Planning		Brainstorm / Idea Generation
• Character Development	• Leisure	• Conversational Engagement • Reflection and Insight	• Leisure	• Greetings and Chitchat • Games and Role Play • Relationships and Personal Reflection	• Sociability	• Entertainment		• Self-Expression	Companionship from AI Chatbot
• Code Generation • Debugging		• Technical Guidance		• Computer Programming		• Technical Support	• Practical Coding or Technical Help • Clarification of Syntax or Programming Semantics • Code Writing	• Technical Help	Computer Code Generation, Revision, or Debugging
• Design		• Artistic Exploration		• Create an Image • Analyze an Image • Generate or Retrieve Other Media		• Creating Text or Images • Editing Text or Images		• Multimedia	Create, Edit, or Process Images
• Information Retrieval	• Information Retrieval	• Factual Queries • Industry-Specific Inquiries	• Information Retrieval	• Specific Info	• Retrieval • Discovery	• Searching for Information	• Medical Information Seeking • Information Seeking about Cultural or Entertainment Topics • Definitions and Comparisons • Synonyms or Antonyms	• Seeking Information	Information Retrieval – Other; Information Retrieval - Current Events
• Explanation • Clarification • Evaluation		• Tutorial Requests • Learning Support • Explanatory Inquiries • Curricular Planning • Skill Development • Task Automation	• Learning	• Tutoring or Teaching • How-To Advice	• Guidance • Clarification	• Learning	• Concept Explanation • Process Explanation or Workflow Help • Comparison or Contrast • Real-Life Examples or Analogies • Explanation • Step-by-Step Solutions		Practical Guidance / Learning
• Analysis • Table Creation	• Solve Problem in Specialized Areas	• Data Processing • Troubleshooting Assistance	• Problem Solving	• Data Analysis • Mathematical Calculation	• Analysis • Distillation • Execution		• Theory Application • Interpretation of Data or Output • Multiple Choice or Fill-in-the-Blank Questions		Problem Solving / Reasoning
				• Asking About the Model					Questions about AI Capabilities
• Creative Writing • Article Writing		• Content Creation	• Content Creation	• Personal Writing or Communication • Write Fiction	• Generation	• Creating Text or Images	• Outline Generation • Email or Letter Generation	• Writing	Writing - Content Creation
• Summarization • Translation	• Text Assistant			• Edit or Critique Provided Text • Argument or Summary Generation • Translation	• Modification	• Editing Text or Images • Information Summary	• Summarization • Editing or Improving Writing • Rewording • Grammar Check • Translation • Academic Citation • Job Application Content		Writing - Edit User-Provided Text
	• Others		• Other	• Unclear • Other	• Communicative Status	• No Intent		• Other / Unknown	Other

[redacted]

39. As seen in **Exhibit 2** above, information retrieval is one of the most consistently observed categories of user intent across the nine studies, reflecting the close connection between certain types of LLM use and traditional web search.[39] As I described in **Section IV.A**, the literature on web search has an established taxonomy of informational, navigational, and transactional search. The informational portion is covered by "Information Retrieval." However, the relevant LLM literature does not include a category for navigational or transactional search.[40] In order to align with the web search literature, which indicates this is a very common form of behavior, and because I observed some Copilot users attempting to navigate to websites in my preliminary review of Copilot conversations, I included navigational intent as a distinct category within the taxonomy, labeled as "Navigation to Another Webpage."

40. In addition to being one of the most consistently represented categories across studies, the literature also indicated that a substantial share of the Copilot Conversations Sample may be classified as information retrieval.[41] To capture additional nuance of usage in this often large category, I refined this category into two separate user intents to distinguish between information retrieval that may include current newsworthy events or involve a temporal element (Information Retrieval – Current Events) and other types of information retrieval without a

---

39 *See, e.g.*, Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29 at p. 17. The authors find that all five of their identified user intents appear in both their web search and LLM chat data. Chatterji, Aaron, et al., "How People Use ChatGPT," *National Bureau of Economic Research*, September 2025, p. 16. The authors recognize that users can seek information and advice (two of the top user intents) from web search engines as well as ChatGPT.

40 Broder, Andrei, "A Taxonomy of Web Search," *ACM SIGIR Forum*, Vol. 36, No. 2, 2022, pp. 3 - 10 at pp. 5-6. Transactional activities can include, for example, online shopping or downloading files such as songs and images.

41 *See, e.g.*, Ammari, Tawfiq, et al., "How Students (Really) Use ChatGPT: Uncovering Experiences Among Undergraduate Students," *arXiv*, January 19, 2026 at pp. 10-12 "Information Seeking" represented 33.4 percent of messages; Costa-Gomes, Beatriz, et al., "It's About Time: The Copilot Usage Report 2025," *Microsoft*, December 2025, p. 4 ("'Searching' is the most common intent across conversations, followed by 'Advice.'").

[redacted]

temporal element (Information Retrieval – Other).[42] I define "Current Events" as prompts that relate to recent or upcoming events or developments that occurred within a week of the date of the prompt.[43,44] For example, a conversation asking about "LA wild fires" while the fires were still raging would be considered to belong to "Information Retrieval – Current Events," but asking the same question two years later would not.[45] In making this distinction, I aim to provide a more transparent and granular analysis of the range of different types of information users aim to retrieve when they use Copilot, recognizing that the claims of the News Plaintiffs are that Copilot will harm them by serving as a substitute for their content. I also created a secondary taxonomy of subject matter, described below in **Section IV.C**, to better understand the topics of information retrieved by Copilot users.

41. Though it is true that there are some "evergreen" stories in the news, the vast majority of news consumption happens within a week. This is borne out in the

---

42 A recent study took a similar approach by including a category, "Knowledge Intent – Contextual," defined as "Information that changes based on time, location, or recency." *See* Lichtenegger, Elsa, Aleksandra Urman, and Aniko Hannak, "A New Taxonomy of Web Search: A User-Centered Framework for Search Intent in the AI Era," *CHI '26: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, April 2026, p. 10.

43 *See* **Exhibit 3** for the full definition of this category and **Appendix E** for training materials that provide hypothetical examples and guidelines around how to identify conversations that should fall under this category.

44 The majority of news stories persist in the public view for approximately one week before or after an event. *See, e.g.,* Mitchell, Amy, Galen Stocking, and Katerina Eva Matsa, "Most Interactions Begin Within One Week of Publication," *Pew Research Center*, May 5, 2016, https://www.pewresearch.org/journalism/2016/05/05/6-most-interactions-begin-within-one-week-of-publication/ ("Indeed, just 6% of short-form interactions and 10% of long-form interactions began more than five days after an article was published."); Horton, Allison, "When Does an Article's Readership Peak?," Memo, https://memo.co/blog/when-does-an-articles-readership-peak/, accessed March 31, 2026 ("An article received 76% of its readership within the first 3 days of publication, on average."); "The Lifespan of News Stories," Schema Design and Google Trends, https://www.newslifespan.com/, accessed March 31, 2026. Using Google Trends, the authors show that searches for a news event peak around the date of the event, with the majority of searches occurring on average within a five-day window on either side.

45 Coders were asked to use the date information in the prompts and responses in conjunction with the timestamps on the conversation to determine recency.

[redacted]

fact that many periodicals such as The New Yorker,[46] The Economist,[47] or The Week[48] come out weekly. Daily newspapers often place more emphasis on their Sunday paper, for example The New York Times offers a Sunday print edition that includes weekly sections like "Book Review," "Sunday Styles," and "Sunday Opinion."[49]

42. The weekly cadence of the news cycle is confirmed by empirical studies that examine public interest in news articles. As shown in **Exhibit 3**, a Pew Research study considering news content accessed via mobile devices found that 66% of visitors discover short-form (and 56% of visitors discover long-form) articles the day the articles are published, with just 6% of interactions with short-form (and 10% with long-form) content being more than five days after an article is published.[50]

---

46 "The Magazine," *The New Yorker*, https://www.newyorker.com/archive, accessed June 19, 2026.

47 "Weekly Edition," *The Economist*, https://www.economist.com/weeklyedition/archive, accessed June 19, 2026.

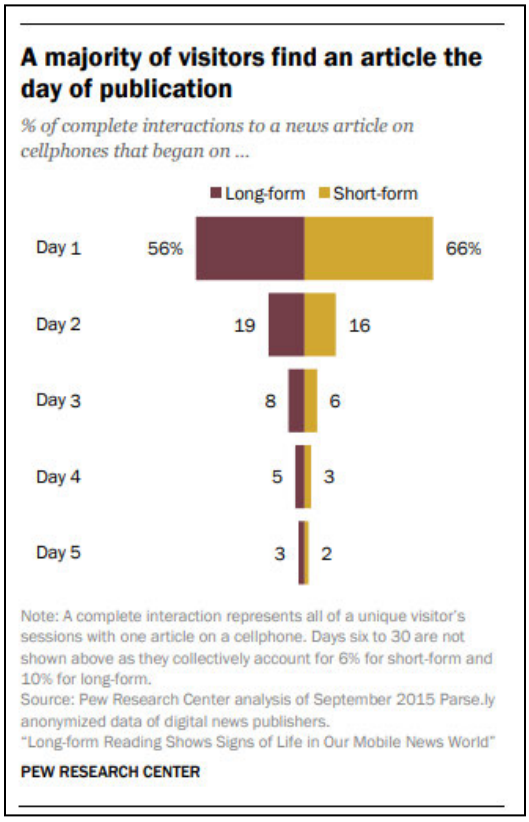
48 "About Us," *The Week*, https://theweek.com/about-us, accessed June 17, 2026.

49 The Sunday edition is the "Most Popular" print delivery option for The New York Times. *See* "The Times in Print. 50% Off Your First Year," *The New York Times*, https://www.nytimes.com/subscription/home-delivery?campaignId=6WFJF&zipCode=02139, accessed June 19, 2026.

50 Amy Mitchell, Galen Stocking, and Katerina Eva Matsa, "Long-Form Reading Shows Signs of Life in Our Mobile News World," *Pew Research Center*, May 5, 2016, https://www.pewresearch.org/wp-content/uploads/sites/8/2016/05/PJ_2016.05.05_LongForm_Final.pdf, p. 25. Chartbeat, a content analytics platform, found that "articles read on Desktop and Mobile both reach 80% of pageviews at 24 hours" and concluded that an article's lifespan was minimally affected by the type of device a reader used. *See* "When Does News Become Old News? How Traffic Source, Topic, and More Impact the Lifespan of an Article," *Chartbeat*, https://web.archive.org/web/20250219173428/https://chartbeat.com/resources/research/whats-the-lifespan-of-an-article/, accessed June 19, 2026.

[redacted]

Exhibit 3
Pew Research Center Analysis[51]



43. Another study performed by the data platform Memo using a sample of 1,110 articles published in March 2023 concluded that "readership typically peaks on the day of publication or the day following before quickly dropping off."[52] On average, the analysis found that articles receive 76 percent of their readership within the first three days of publication (see **Exhibit 4**), with "topical news"[53] readership being even more concentrated with 86 percent of readership in the first three days. Articles included in the analysis were sampled from "national

---

[51] Amy Mitchell, Galen Stocking, and Katerina Eva Matsa, "Long-Form Reading Shows Signs of Life in Our Mobile News World," *Pew Research Center*, May 5, 2016, https://www.pewresearch.org/wp-content/uploads/sites/8/2016/05/PJ_2016.05.05_LongForm_Final.pdf, p. 25.

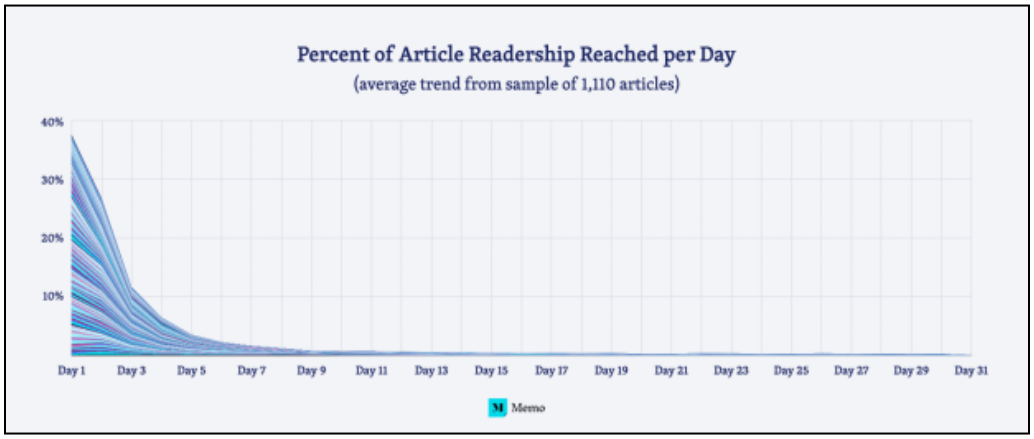
[52] Horton, Allison, "When Does an Article's Readership Peak?," *Memo*, https://memo.co/blog/when-does-an-articles-readership-peak/, accessed March 31, 2026.

[53] Topical news were defined as current events and brand news of immediate interest as opposed to evergreen content. *See* Horton, Allison, "When Does an Article's Readership Peak?," *Memo*, https://memo.co/blog/when-does-an-articles-readership-peak/, accessed March 31, 2026.

[redacted]

and local news sites, lifestyle and entertainment outlets, political news, celebrity news, business and financial press," among others.[54] As the figure shows, the vast majority of readership is captured within the first seven days following an article's publication.

Exhibit 4
Article Readership Over Time Following Publication Day[55]



44. A systematic analysis of Google Trends data also found that public interest in major news events follows a common cycle: attention rises sharply to a peak around the event date, followed by a subsequent decline. This study was the result of collaboration between Google Trends and Schema Design and analyzed a set of 2018 news events curated by the media company Axios. This analysis found that most search interest is concentrated, with the majority of interest in an event occurring within a window of a few days.[56] As **Exhibit 5** shows, this trend was observed across a wide range of topics such as politics & elections, natural catastrophes, environment & science, social issues, violence & war, and obituaries.

---

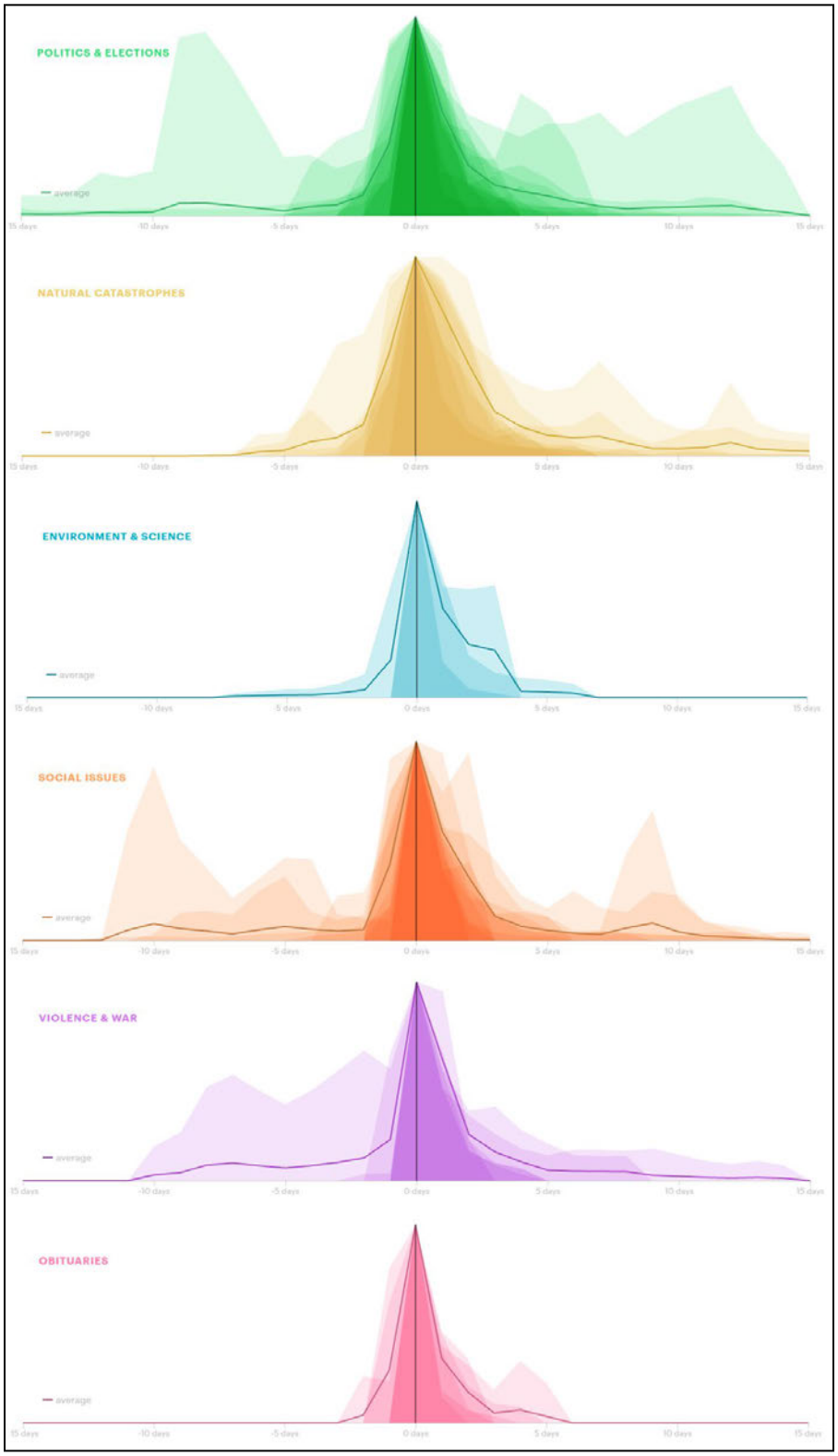
54 Horton, Allison, "When Does an Article's Readership Peak?," *Memo*, https://memo.co/blog/when-does-an-articles-readership-peak/, accessed March 31, 2026.

55 Horton, Allison, "When Does an Article's Readership Peak?," *Memo*, https://memo.co/blog/when-does-an-articles-readership-peak/, accessed March 31, 2026.

56 "The Lifespan of News Stories," *Schema Design and Google Trends*, https://www.newslifespan.com/, accessed March 31, 2026.

[redacted]

Exhibit 5
Lifespan of News Stories by Topic[57]



[redacted]

45. Academic research further confirms these findings. A 2017 study of media attention in news articles across 196 countries presented at the Conference on Web and Social Media organized by the Association for the Advancement of Artificial Intelligence found substantial turnover in the trending topics section at a news site. Notably, the authors found that if a topic appeared among a news site's "top 10 trending topics" on a given day, the probability that it would remain in the top ten the following day was only 30.7%."[58] This suggests that public attention to news topics is often short lived and can be characterized by rapid turnover.

46. Research on news cycles suggests that the pace at which attention shifts from one topic to another may have increased over time as news information is distributed more rapidly and audiences are exposed to a greater volume of content through social media and other sources of digital news content. As noted by a faculty member at the Poynter Institute, a global nonprofit journalism school and media institute, the traditional "24-hour news cycle has turned into a 24-second news cycle" where "every refresh of a browser, app or notification screen brings another wave of headlines."[59] In the words of a former TIME magazine Editor in Chief and professor at the Harvard Kennedy School, "[i]nformation and attention are the battlefields of this century" and "[e]verything is 'breaking news' designed

---

57 This graphic shows data for individual events layered over each other, with darker areas indicating more density, and with a solid line showing the average for the category. *See* "The Lifespan of News Stories," *Schema Design and Google Trends*, https://www.newslifespan.com/, accessed March 31, 2026 ("This visualization depicts United States search metrics on a set of 2018 news events, curated by Axios. Each event was represented by a unique search term (or terms). The Google Trends API was used to obtain the date at which search interest peaked for each event throughout the year. A separate Google Trends API request, using the same search term(s) was used to quantify the evolution of search interest across a 60-day window surrounding each event, centered at the event's peak date.")

58 Jisun An, Haewoon Kwak, "What Gets Media Attention and How Media Attention Evolves Over Time – Large-Scale Empirical Evidence from 196 Countries," *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, 2017, pp. 464-467 at p. 466 ("Of all the attention spans, the proportion of the attention span of 1 day is 69.3%. This indicates that if we see a topic in 'the top 10 trending topics' section at a news site today, the probability of seeing the same topic tomorrow is only 1−0.693=0.307.").

59 Elkins, Tony, "The 24-second News Cycle is Exhausting Journalists. Here's How They Cope," *Poynter*, March 16, 2026, https://www.poynter.org/business-work/2026/how-to-handle-news-fatigue/.

[redacted]

to grab shrinking attention spans."[60] Real-time news delivery and social media have accelerated attention shifts from one topic to another, with shorter attention cycles driven primarily by increasing information flows and the rapid production and consumption of content.[61]

47. Taken together, this evidence consistently demonstrates that public interest in news stories often only persists for a few days and diminishes relatively quickly, supporting the reasonableness of applying a one-week time frame to the "Information Retrieval – Current Events" category. I have not seen any evidence that would suggest that any conversation in the Copilot Conversations Sample relied upon this particular article. Plaintiffs' experts do not provide any definition of news and none of them offers a systematic analysis of news-related conversations in the Copilot Conversations Sample that involves events that occurred more than a week ago.

48. Even if a conversation involves Copilot searching articles published more than seven days earlier on one of News Plaintiffs' websites, such a conversation may also involve current events and be classified as such. For example, **Exhibit 6** shows a conversation where a user asks Copilot in March 2024 for news about the war in Ukraine.[62] While generating a response, Copilot searches—among several other sources—an article published in December 2022 by The New York Times.[63] Even though the article was published more than a year earlier, the

---

60 Hughes, Susan A., "In Today's Media Landscape, Attention Is a Battlefield," *Harvard Kennedy School*, January 6, 2026, https://www.hks.harvard.edu/faculty-research/policy-topics/media/todays-media-landscape-attention-battlefield.

61 Philipp Lorenz-Spreen, Bjarke Mørch Mønsted, Philipp Hövel, Sune Lehmann, "Accelerating Dynamics of Collective Attention," *Nature Communications*, 2019 ("With news pushed to smart phones in real time and social media reactions spreading across the globe in seconds, the public discussion can appear accelerated and temporally fragmented. […] Our modeling suggests that the accelerating ups and downs of popular content are driven by increasing production and consumption of content, resulting in a more rapid exhaustion of limited attention resources. In the interplay with competition for novelty, this causes growing turnover rates and individual topics receiving shorter intervals of collective attention.").

62 Conversation ID: 3EA377E7946DC33A5DF5D106E56F11937EF6692092FAA67A4B124F0B31EB8B10.

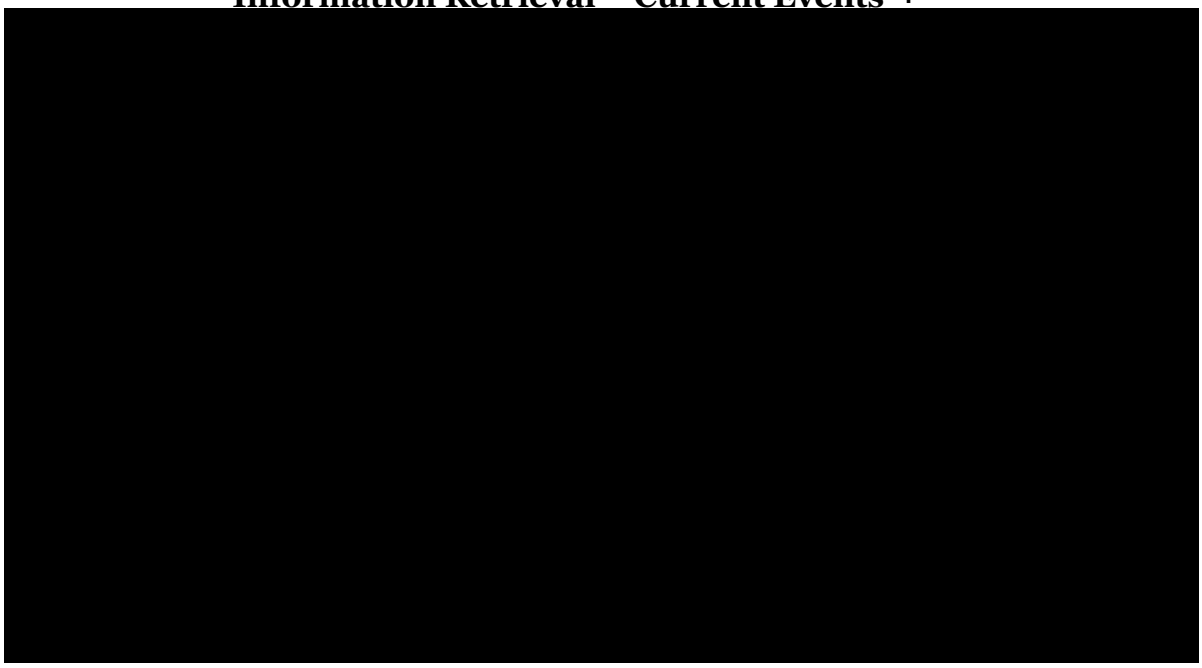
63 The article searched by Copilot is: https://www.nytimes.com/live/2022/12/05/world/russia-ukraine-war-news. *See* **Appendix H.**

[redacted]

user's request for "news" indicates intention to seek information about current events; this conversation was classified by coders as "Information Retrieval – Current Events." This directly refutes the argument that news-related conversations relying on sources published more than a week ago would be miscategorized because of the "artificially limiting" choice of one week to define current events.

Exhibit 6

Information Retrieval – Current Events[64]



49. Nonetheless, even under these conservative assumptions regarding the boundaries between the Information Retrieval categories, conversations seeking current events information remain a small share of overall Copilot usage.

50. The nine studies that met my criteria do not include categories for illicit uses of LLMs, but several studies in my literature review report instances of illicit

[64] Conversation ID: 3EA377E7946DC33A5DF5D106E56F11937EF6692092FAA67A4B124F0B31EB8B10.

[redacted]

behavior.[65] Because these use cases have been identified by other studies, I include a separate category to capture user attempts to bypass Copilot's safeguards, or "jailbreak behaviors."[66]

51. I also added three categories for conversations for which coders were not able to classify user intent. "Unable to Classify – Requires Translation" includes conversations where the English-speaking coders determined that they could not reliably classify user intent because of foreign-language content in the user prompts.[67] "Unable to Classify – Redacted" includes conversations where privacy redactions (described below) prevented the coder from determining what the user was attempting to search or ask.[68] Finally, "Unable to Classify – Insufficient

---

[65] *See, e.g.,* Zhao, Wenting, et al., "WILDCHAT: 1M ChatGPT Interaction Logs in the Wild," *ICLR 2024*, May 2, 2024, pp. 5-7; Zheng, Lianmin, et al., "LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset," *ICLR 2024*, March 10, 2024, p. 4; Trippas, Johanne R., et al., "What do Users Really Ask Large Language Models? An Initial Log Analysis of Google Bard Interactions in the Wild," *SIGIR '24: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 2024, pp. 2703 - 2707 at p. 2706.

[66] I define "Jailbreak behaviors" as prompts that are designed to bypass the safeguards and moderation measures of an AI chatbot, regardless of whether they succeed or not.

[67] I focused on classifying conversations in English since the allegations in this matter relate to news articles and other copyright-protected content which I understand from counsel were authored in English. This approach is supported by the literature, with several studies only classifying English conversations. *See, e.g.,* Ouyang, Siru, et al., "The Shifted and the Overlooked: A Task-Oriented Investigation of User-GPT Interactions," *Association for Computional Linguistics*, December 2023, pp. 2375 - 2393 at p. 2378 ("Since our investigation focuses on the English language and ShareGPT contains only English samples, datasets collected from Huggingface also follow this paradigm."); Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29 at p. 7 ("We ensured that these conversations were in English, however, some of the user requests were interspersed with non-English words. We do not believe this impacted any text processing by GPT-4 for our purposes."); Zheng, Lianmin, et al., "LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset," *ICLR 2024*, March 10, 2024, p. 3 ("We conduct a topic distribution analysis on user prompts by applying a clustering algorithm. From 100K randomly sampled English conversations, we extract user prompts, which include both the initial and follow-up turns."). Furthermore, I performed additional analyses to confirm that the presence of non-English conversations in the Copilot Conversations Sample does not alter my opinions, as discussed below in **Section VI.C.3**.

[68] I discuss these privacy redactions in more detail below in **Section VI.A**.

[redacted]

Content" includes conversations where the user's prompts were either too short to classify or otherwise incomprehensible.[69]

2. Results from the Amir Survey Informed Development of My User Intent Taxonomy

52. I also reviewed results from the Amir Survey to ensure that categories with a high share of usage were represented in my taxonomy.[70] Given that the Amir Survey relied on self-reported survey data and asked respondents about their broader LLM usage, the categories include a level of subjective and overarching detail on usage that an external third-party coder might not be able to determine from review of an individual LLM conversation. For example, the Amir Survey included both "Research or search about other personal topics of interest" and "Research or search about other professional topics of interest."[71] While the respondents in the Amir Survey can self-report whether they use Copilot for professional or personal uses, the human coders reviewing the Copilot Conversations Sample in context of my report do not know the user's profession and therefore cannot accurately determine if someone searching for, say, bank processing times is conducting that search for professional or personal use. However, I found that all prominent usage categories found in the Amir Survey were represented in my taxonomy.

3. Findings from Category Refinement and Feedback from Coder Training Further Informed My User Intent Taxonomy

53. As I explain further in **Section V.B**, in tandem with my team of coders, I reviewed a sample of the data to ensure that the category labels were clear and

---

69 Other studies exclude messages that are too short or have incoherent text. *See, e.g.*, Zheng, Lianmin, et al., "LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset," *ICLR 2024*, March 10, 2024, p. 3 ("We remove prompts that are [...] too short (fewer than 32 characters)[.]"); Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," March 19, 2024, working paper, p. 2 ("Of the 80000 conversations we sampled, 112 conversations [were] unusable, mostly due to incoherent text, leaving 79888 conversations.").

70 Amir Declaration, Figure 1, Exhibit 3.

71 Amir Declaration, Figure 1, Exhibit 3.

[redacted]

comprehensible and could be applied reliably to the data.[72] Based on feedback from coders and my own review of the data, I made small adjustments to the descriptions of certain categories and combined "Quantitative Data Analysis" with "Problem Solving / Reasoning," as reliably distinguishing between these two uses proved challenging. These changes are detailed in **Appendix D**.

54. The final taxonomy of 18 user intent categories, including the three additional user intent categories ("Navigation to Another Webpage," "Jailbreak Behaviors," and "Information Retrieval – Current Events") and the three "Unable to Classify" categories along with category definitions is shown in **Exhibit 7** below.

[72] This process involved 2,000 conversations displayed to coders during the Training and pilot coding stages I describe in more detail below.

[redacted]

Exhibit 7
User Intent Categories

User Intent	Definition
Advice / Recommendation / Seeking Opinions	Capture prompts where user is asking for help in making a choice or seeking to form an opinion. Relies on the AI Chatbot's ability to provide personalized recommendations / output. Includes forward-looking predictions of future outcomes and answers to questions with subjective, opinionated answers. Does not include seeking help on homework multiple-choice questions.
Brainstorm / Idea Generation	Capture prompts where user is asking AI Chatbot to help generate creative ideas or a hypothetical list of options. Includes requests for quiz questions. Unlike "Advice / Recommendation / Seeking Opinions," user should not be seeking any ranked or sorted "best" results.
Companionship from AI Chatbot	Capture cases where user is using the AI Chatbot as a confidant, friend, therapist, or role-play partner.
Computer Code Generation, Revision, or Debugging	Capture prompts where user is asking AI Chatbot to help write, debug, or edit computer code. If the AI Chatbot responds with functions or computer code, it should fall under this category.
Create, Edit, or Process Images	Capture prompts where user is asking AI Chatbot to generate or edit an image, or prompts that ask AI Chatbot to read, process, or explain an image file. As a reminder, images will not be displayed.
Information Retrieval - Current Events	Captures prompts that request factual information about events or developments that occurred within a week of the date of the prompt (before or after). This would include information that one may expect to find in local, regional, or national news sources. You can use AI Chatbot's responses and the date of the conversation to help you determine whether the user is seeking recent or more historical information.
Information Retrieval - Other	All other information retrieval queries that do not fall within "Information Retrieval - Current Events."
Jailbreak Behaviors	Capture prompts that are designed to bypass the safeguards and moderation measures in AI Chatbot, regardless of whether they succeed or not.
Navigation to Another Webpage	Capture prompts where the user is trying to navigate to another specific, often common, website or attempting to log in to a website or portal.
Practical Guidance / Learning	Capture prompts focused on learning/understanding a concept or acquiring a new skill. Relies on AI Chatbot's ability to tutor or deliver tailored explanations or synthesize information. Does not include quiz or multiple choice questions (see "Problem Solving / Reasoning") unless user engages to actively learn the material, rather than to simply obtain the correct answer from a multiple-choice question.
Problem Solving / Reasoning	Capture prompts where user wants AI Chatbot to not only find information but also apply it to a given situation or engage in logical reasoning. Relies on the ability of the AI Chatbot to process information and construct a response based on available information and what the user has provided. Include quiz or multiple choice questions, potentially from a homework assignment. Includes prompts that ask AI Chatbot to process and make inferences from a set of data or to perform mathematical computations.
Questions about AI Capabilities	Capture prompts where the user is asking AI Chatbot about the AI Chatbot itself, what it was trained on, what it can help with, or what it is.
Writing - Content Creation	Capture prompts where user is asking the AI Chatbot to write new text without any provided text to modify. Includes requests for both fictional and non-fictional written work product.
Writing - Edit User-Provided Text	Capture prompts where user is asking the AI Chatbot to transform user-provided text. This can include changing the form, editing for grammar or spelling, translating, summarizing text into bullets, or expanding on a certain point.
Other	This category is for the unlikely scenario that a conversations involves a user intent that does not fall into any of the categories above. This category should only be selected when you understand what the user is trying to accomplish but the user intent is outside of the pre-defined categories above. This category should not be selected for conversations that are too short to understand intent; unintelligible; in languages another than English that you cannot understand without help of a translator; or have been redacted in a way that obscures user intent.
Unable to Classify - Insufficient Content	Capture prompts that are unclear, because the user has submitted text that is too short, did not provide sufficient context, or the prompt is riddled with errors such that it is not clear what the user was hoping to accomplish. Any prompt where the AI Chatbot does not understand what the user is asking for should fall under this category.
Unable to Classify - Redacted	Capture instances where redaction results in insufficient context in the user prompt(s) to reliably classify user intent.
Unable to Classify - Requires Translation	Capture prompts that require translation of the user prompt(s) to determine user intent.

[redacted]

C. Development of My Subject Matter Taxonomy

55. My secondary taxonomy of subject matter, which I apply to conversations classified in the "Information Retrieval – Current Events" or "Information Retrieval – Other" categories, was informed by three studies on LLM usage among those in my **Appendix C.2** that examined real-world chatlogs using a taxonomy of subject matters, as well as my preliminary analysis of Copilot conversations and feedback from coders during the training and category refinement phase of my content analysis. I began to construct my taxonomy of subject matters of Copilot conversations by identifying overlapping categories across the three studies. If multiple studies used different labels to describe similar subject matters, I grouped those categories and applied a broader label, where applicable. I summarize the resulting taxonomy in **Exhibit 8** below.[73]

[73] I note that the three studies include a subject matter that involves "Education." I decided not to include this as a subject matter category because in my review of conversations, education appeared in context of one or more of my other subject matter categories like "Science / Math" or "History / Social Sciences / Politics." Other subject matters in the three studies are "Autos and Vehicles," "Pets and Animals," and "Home and Garden" in Costa-Gomes et al. (2025) and "Small Talk and chatbot" and "Data Analysis and Visualization" in Suri et al. (2024).

[redacted]

Exhibit 8
Summary of Subject Matter Categories in Key Prior Studies

Costa-Gomes, Beatriz, et al., "It's About Time: The Copilot Usage Report 2025," Microsoft, Dec 2025	Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," Microsoft Research, March 19, 2024	Ouyang, Siru, et al., "The Shifted and The Overlooked: A Task-oriented Investigation of User-GPT Interactions," Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Oct 19, 2023	SUBJECT MATTER CATEGORY
• Art and design • Books and literature • Language learning and translation	• Translation and language learning • Linguistics and language analysis • Creative writing and editing • Academic writing and editing	• Literature • Language	Arts / Literature / Language
• Business and industry • Money • Work and career	• Business and economics • Marketing and sales • Accounting and finance	• Business • Finance	Business / Finance
• Internet and social media • Programming • Technology	• Programming and scripting • Web development • Machine learning and AI	• Social media • Technology	Computing Devices / Software
• Entertainment • Hobbies and leisure • Games	• Gaming and entertainment		Entertainment (Celebrities, Movies, Concerts, Games)
• Beauty and fashion	• Fashion and beauty		Fashion / Beauty
• Food and drink	• Food and drink		Food / Cooking / Recipes
• Health and fitness		• Health	Health / Fitness
• Society, culture, and history	• History and culture	• History	History / Social Sciences / Politics
	• Law and politics	• Legal	Law / Tax
• Local • News and current events			Other Local, Regional, National or International Events
• Personal growth and wellness • Relationships		• Relationship	Personal Growth / Family
• Religion and philosophy	• Religion and philosophy		Religion
• Math and logic • Science	• Mathematics and logic • Physics and chemistry • Engineering and design • Biology and medicine	• Mathematics • Science • Biology	Science / Math
• Sports	• Sports and fitness	• Sports	Sports
• Travel	• Travel and tourism		Travel / Geography
• No topic	• Other		Other

[redacted]

56. In addition to the categories found in the literature, I included a subject matter category for Consumer Products and Services. Both Costa-Gomes et al. (2025) and Chatterji et al. (2024) include categories in their user intent taxonomies that refer to consumer products and services. Costa-Gomes et al. (2025) include the category "Shopping and Product Research" and Chatterji et al. (2024) include the category "Seeking Information – Purchasable Products."[74] The final taxonomy of 17 subject matter categories along with their definitions is shown in **Exhibit 9** below.

[74] Costa-Gomes, Beatriz, et al., "It's About Time: The Copilot Usage Report 2025," *Microsoft*, December 2025, p. 4; Chatterji, Aaron, et al., "How People Use ChatGPT," *National Bureau of Economic Research*, September 2025, p. 14.

[redacted]

Exhibit 9
Subject Matter Categories

Subject Matter	Definition
Arts / Literature / Language	Information on artistic works, literary texts, authors, artists, or any analysis related to visual, written, or performing arts. Includes prompts that seek to identify synonyms, ask for meaning of word, or understand the meaning of a text in other languages.
Business / Finance	Information about financial markets, investing, personal finance, or business operations.
Computing Devices / Software	Information related to computing devices, mobile devices, software, digital systems, or other technical concepts.
Consumer Products and Services	Information on buying, comparing, reviewing, or using consumer goods and services, or information about consumer brands, except prompts that fall within other named subject matter categories.
Entertainment (Celebrities, Movies, Concerts, Games)	Information about entertainment topics such as celebrities, movies, TV, and recreational activities such as attending concerts, live events, hobbies, table-top games, and other media and recreational experiences.
Fashion / Beauty	Information on clothing, style trends, personal grooming, skin and hair care including general guidance on styling or beauty routines.
Food / Cooking / Recipes	Information on food, restaurants, cooking methods, or recipes, excluding nutrition or healthfulness of foods (which belong to Health / Fitness).
Health / Fitness	Information on physical and mental health, medical conditions, symptoms, treatments, nutrition, exercise, lifestyle practices or fitness routines.
History / Social Sciences / Politics	Information related to historical events or figures, civilizations and cultural developments, as well as topics in social science including sociology, political science, anthropology, or economics. Includes both historical and contemporary questions about society, cultures, institutions, political processes, public policy, or political figures.
Law / Tax	Information related to tax information, past or current legal systems, laws, institutional regulations, and citizen rights.
Other Local, Regional, National, or International Events	Information related to specific, identifiable events or incidents that occur at the local, regional, national, or global level, other than events that fall within other named subject matter categories.
Personal Growth / Family	Information related to personal development, self-improvement, wellbeing, relationships, and family dynamics
Religion	Information about the history and evolution of religions, belief systems, spiritual practices, religious texts, rituals, festival and their dates, places of worship as well as notable religious figures. Include prompts that ask the AI Chatbot to analyze, interpret or compare religious texts.
Science / Math	Information about scientific concepts, natural phenomena, mathematical principles, formulas, and other academic literature and topics across fields other than events that fall within other named subject matter categories.
Sports	Information about trending or past sport events, notable personalities in sports, sports teams, game rules and as well analysis of games, performance or sports-related events.
Travel / Geography	Information on countries, regions, landmarks or geographic features as well as logistical information on available routes, transportation and destination guidance.
Other	Information on any other subject matter that does not fall into any category listed above.

[redacted]

V. Quantitative Content Analysis Framework for Interpreting Copilot Conversation Intent and Subject Matter

57. In this section, I outline the quantitative content analysis framework I developed to understand and classify user intent in the sample produced by Dr. Lafferty.[75] First, I provide an overview of the Copilot Conversations Sample that is the subject of my analysis. Second, I describe the recruitment and training of the human coders, instructions provided to them, and the category refinement process undertaken to ensure reliability of my methodology.

A. Copilot Conversations Sample

58. My analysis focuses on the Copilot Conversations Sample, a sample of 45,000 conversations that users had with consumer Copilot that occurred between December 10, 2023 and September 30, 2024 in the United States, as described in the Lafferty Report.[76] A "conversation" in the current context is defined as a set of user prompts and Copilot responses associated with the same *Conversation ID*.[77] I, along with members of my support team working at my direction, analyzed the different fields available in the output logs provided by Dr. Lafferty and his team and determined that the following six fields were relevant to assessing user intent: *ConversationId*, *Author*, *EventInfo_Time*, *Timestamp*, *MessageType*, and *Text*.

59. The screenshots of conversations I include in this report are from the user interface of a proprietary coding platform developed under my direction by my support team. This platform was hosted in a secure Azure cloud environment to

---

75 *See* Lafferty Declaration.

76 I understand that the Lafferty Declaration explains the creation of this sample and concludes that an appropriate minimum sample size for evaluating the general, real-world usage patterns of consumer Copilot is 30,000 conversations. *See* Lafferty Declaration, Sections VI (Methodology for Chat Log Data Sample Selection) and VII (Sample Size Robustness Checks) for discussion of data filtrations, sampling methodology, and determination of appropriate sample size.

77 The *Conversation ID*s referenced in my report are obtained from the variable *ConversationId* in the raw chatlog data provided to me. *See* Lafferty Declaration, Section V.C (Data Fields) for definitions of data fields. *See* **Appendix F** for details regarding additional processing steps I undertake to facilitate my content analysis.

[redacted]

ensure privacy and security of the data in the sample and to facilitate a consistent and reliable classification process.[78] The term "User" describes the author of the prompt (displayed in blue); Copilot responses are listed with "Bot" as the author and with content of the responses left-aligned (and displayed in green). The prompts and responses are sorted in chronological order, with the first message sent displayed at the top and the most recent prompt or response at the bottom. **Exhibit 10** below depicts an example of a conversation where a user asks Copilot to suggest itinerary for a vacation.

Exhibit 10
Screenshot of a Conversation[79]



60. The conversations included in the Copilot Conversations Sample were redacted to prevent disclosure of potentially identifiable or sensitive information,[80] as further

---

[78] This may differ from the way the information was originally displayed in the Copilot product. Note that this excludes internal processing steps. I describe this platform further in **Section V.B.1**.

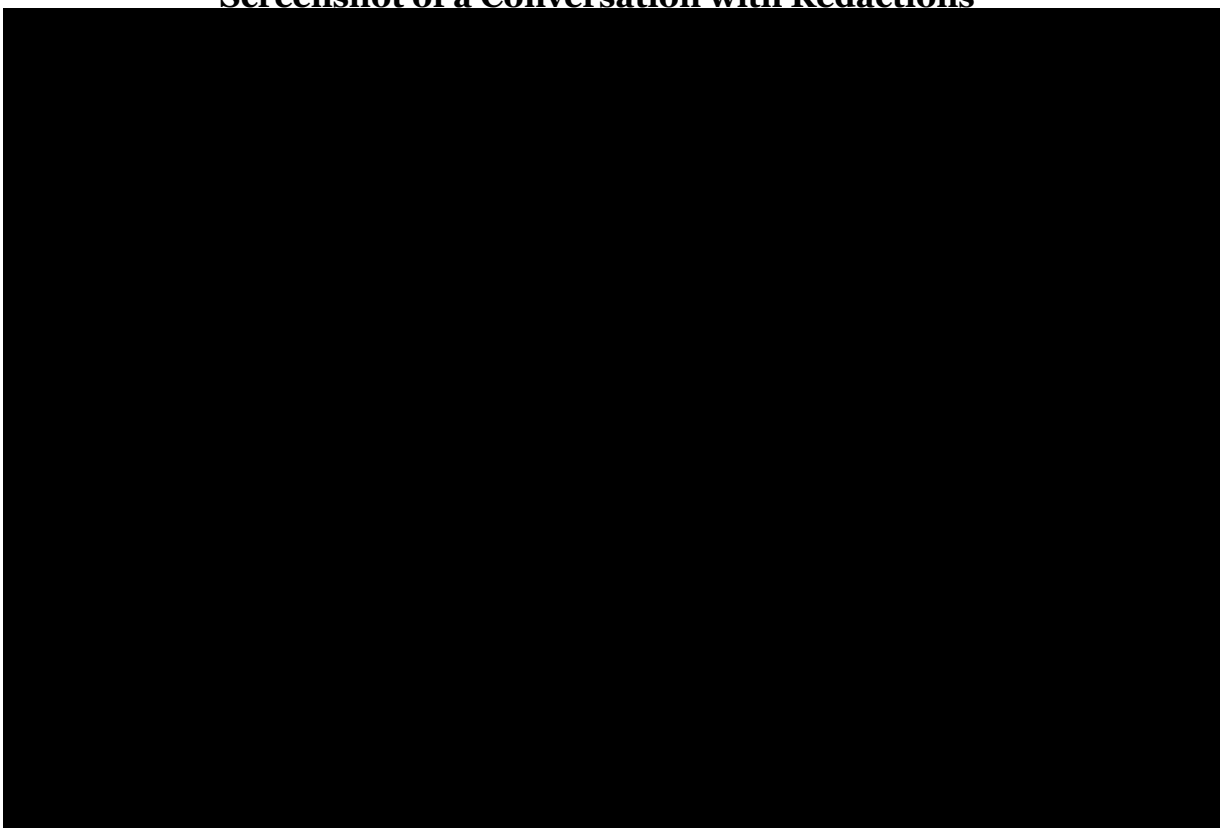
[79] *See* Conversation ID: DA009A1E74F5E6FA06B3130C51966BECEE80ADD32D6892BD77D91C408C480E4B.

[80] This is consistent with approach in the literature. *See, e.g.*, Ammari, Tawfiq, et al., "How Students (Really) Use ChatGPT: Uncovering Experiences Among Undergraduate Students," *arXiv*, January 19, 2026, p. 7 ("[T]he content of the conversations could potentially reveal aspects of a participant's identity. To mitigate this risk, we removed all identifiable user information during the data cleaning process.").

[redacted]

described in the Lafferty Declaration.[81] **Exhibit 11** below depicts an example of a conversation with such information redacted.

Exhibit 11
Screenshot of a Conversation with Redactions[82]



61. Every conversation in the Copilot Conversations Sample is made up of one or more "turns," where a "turn" is one prompt from a user and the associated response—if any—produced by Copilot. A "user turn" is the user prompt part of the turn, while a "system turn" is the AI system's response part of the turn.[83] For instance, the example conversation in **Exhibit 11** above depicts one turn. The average conversation in the Copilot Conversations Sample contains 2.2 user turns

[81] *See* Lafferty Declaration, Section VIII.C (Redaction for Personally Identifiable Information (PII)) for details on redaction methodology.

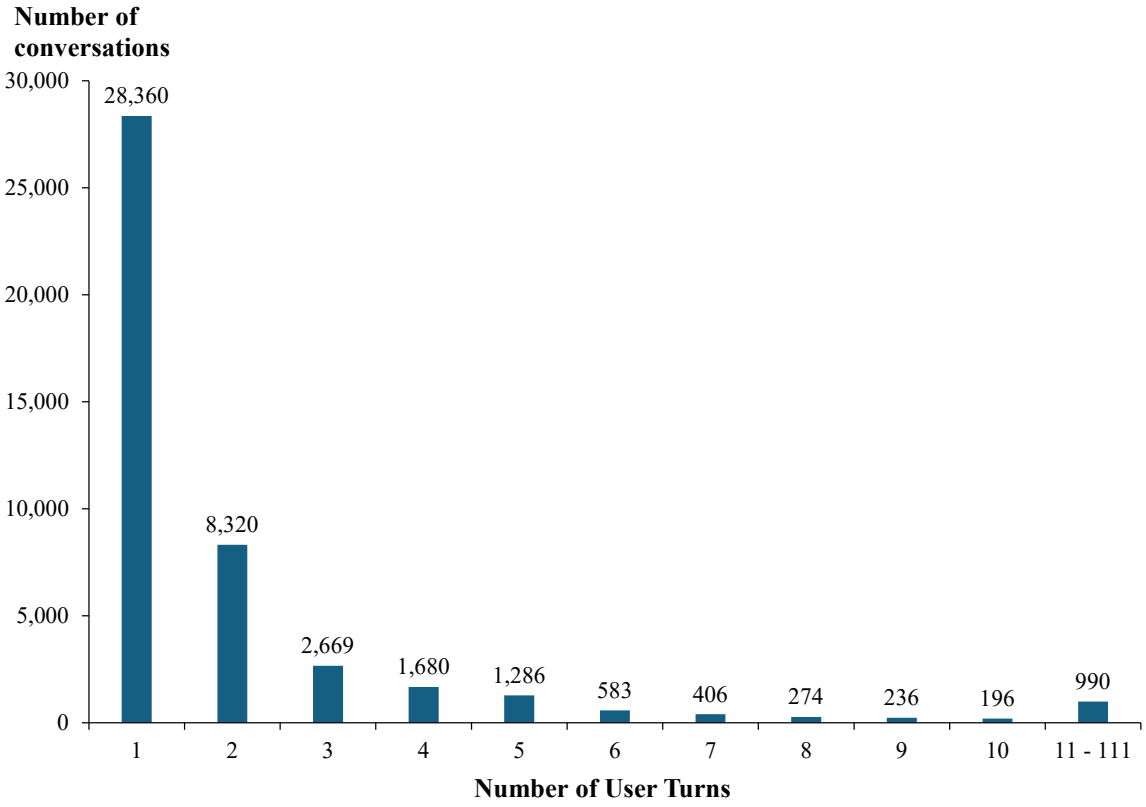
[82] Conversation ID: 022735674BD0F1E10C73DB811306D3569EB373B5E93D317C5950B4EB3B63D199.

[83] In some instances, I do not see a system turn associated with every user turn. Additionally, some conversations start with a system turn rather than a user turn.

[redacted]

and the median conversation consists of one user turn.[84] **Exhibit 12** below shows a histogram of the number of user turns per conversation in the Copilot Conversations Sample.

Exhibit 12
Distribution of User Turns Per Conversation



B. A Human-Coded Quantitative Content Analysis Pipeline Was Used to Assign Categories to Copilot Conversations

62. In order to categorize each Copilot conversation using my taxonomy, I first developed a multi-stage training and coding pipeline, illustrated below in **Exhibit 13**, in which a set of trained coders, blind to the purpose of the analysis, independently reviewed the conversations. I broke down the 45,000 conversations in the Copilot Conversations Sample into three buckets: (i) 1,000

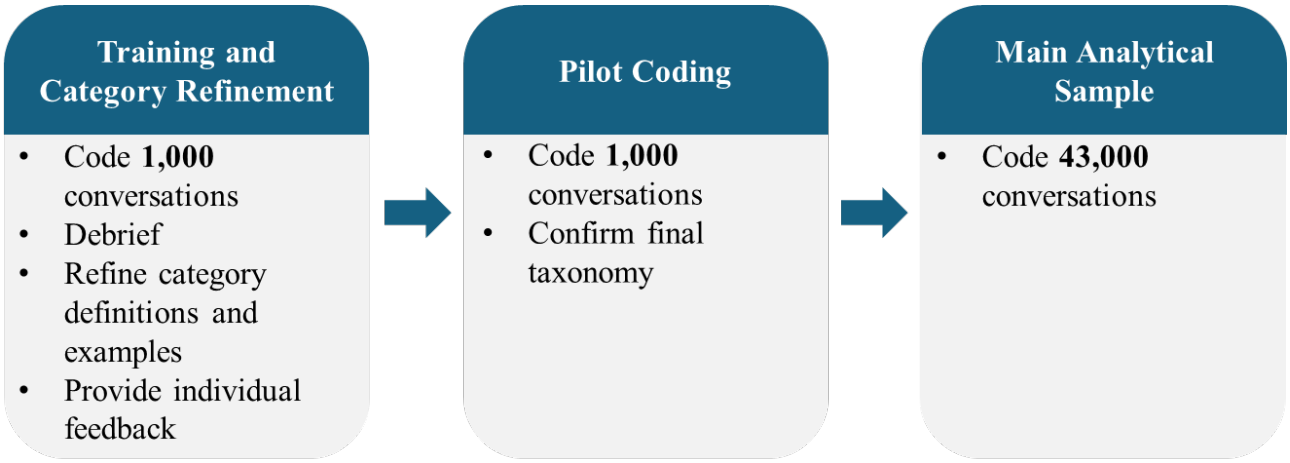
[84] This is consistent with the literature. For example, Zhao et al. found that the median conversation consisted of one turn and the average conversation consisted of 2.5 turns. *See* Zhao, Wenting, et al., "WILDCHAT: 1M ChatGPT Interaction Logs in the Wild," *ICLR 2024*, May 2, 2024, pp. 3-4.

[redacted]

randomly selected conversations reserved for coder training and category refinement, (ii) 1,000 randomly selected conversations reserved for pilot coding, and (iii) the 43,000 remaining conversations reserved for my final analysis. I refer to this set of 43,000 conversations as my Main Analytical Sample. In **Section VI** below, I report results only for my Main Analytical Sample. The 1,000 conversations reserved for coder training and category refinement and the 1,000 conversations reserved for pilot coding are not included in the reporting of my results.

63. To categorize conversations in the Main Analytical Sample, each conversation was initially reviewed by two different coders, unaware of each other's category selections. If the two coders disagreed, the conversation was then displayed to a third coder, who did not know the category selections of the initial two coders, and was not even aware that he or she was serving the role of the third reviewer. If all three coders chose different user intent categories, the conversation was categorized as "Unclear Principal Use."

Exhibit 13
Outline of the Training, Pilot Coding, and Classification Process



1. Coding Platform Interface

64. The proprietary coding platform allowed the coders to review user prompts and bot responses in an easy-to-digest format in the left pane and provided a convenient set of "radio buttons" that describe each category in my taxonomy. As described above, some of the conversations had information redacted for data

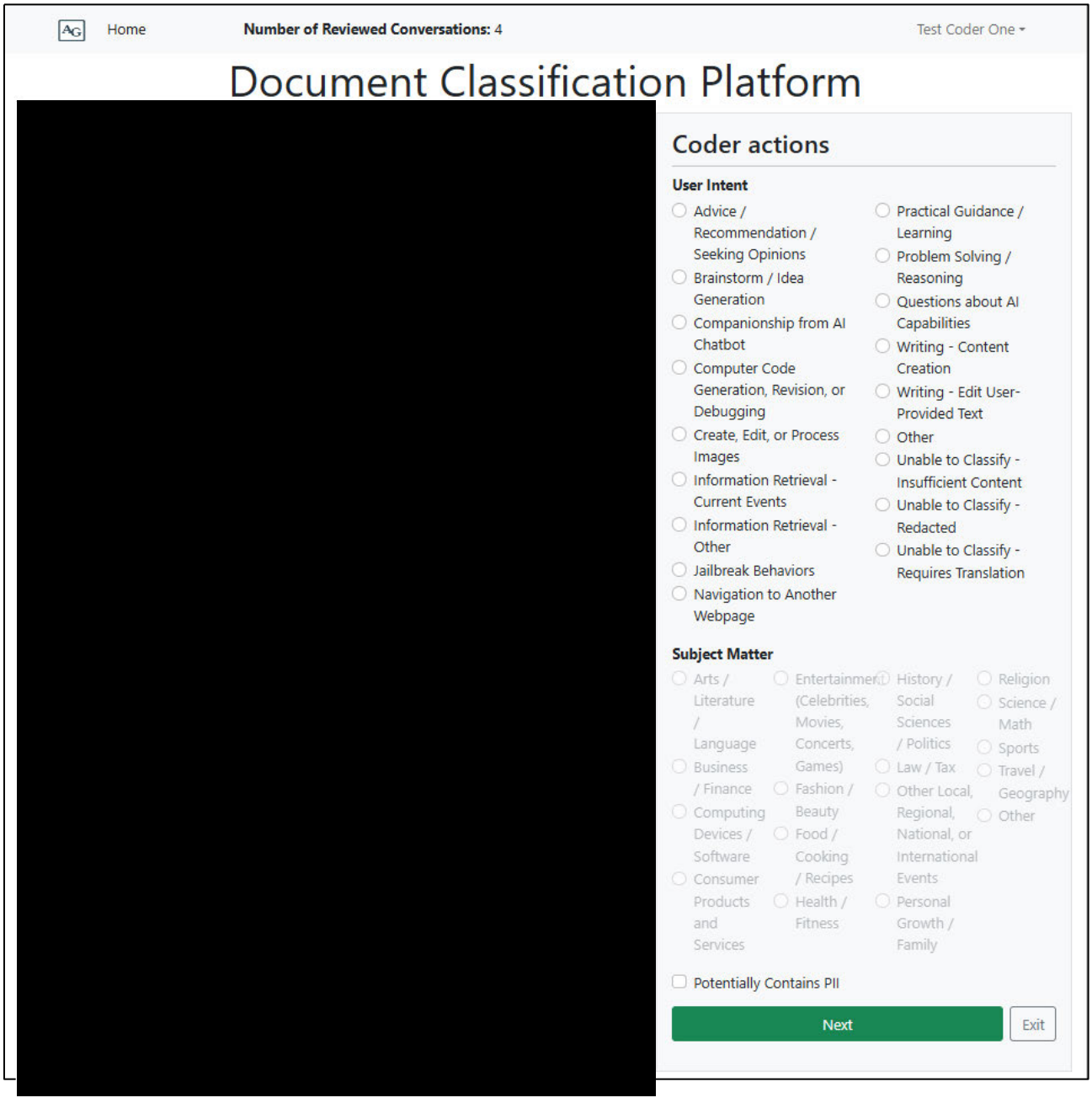
[redacted]

privacy reasons. Coders were only able to see a scrambled *Internal ID*, which is an arbitrary number assigned to uniquely identify each conversation in the sample. In an effort to recreate the Copilot user experience, coders were not shown any internal processing or reasoning steps used by the model. The content of any files or images uploaded by the user or created by Copilot, to the extent such files may have been part of a conversation, were not included in the data and therefore not shown to the coders. See **Exhibit 14.A** below for a view of the platform interface.[85]

[85] While coders were reviewing redacted versions of chatlogs, out of an abundance of caution they were asked to select the "Potentially Contains PII" checkbox to flag conversations that, despite redactions, may have included unique details that could create potential for user reidentification. *See* **Appendix E**, p. E-6. The list of flagged conversations was provided to counsel for consideration for further redactions.

[redacted]

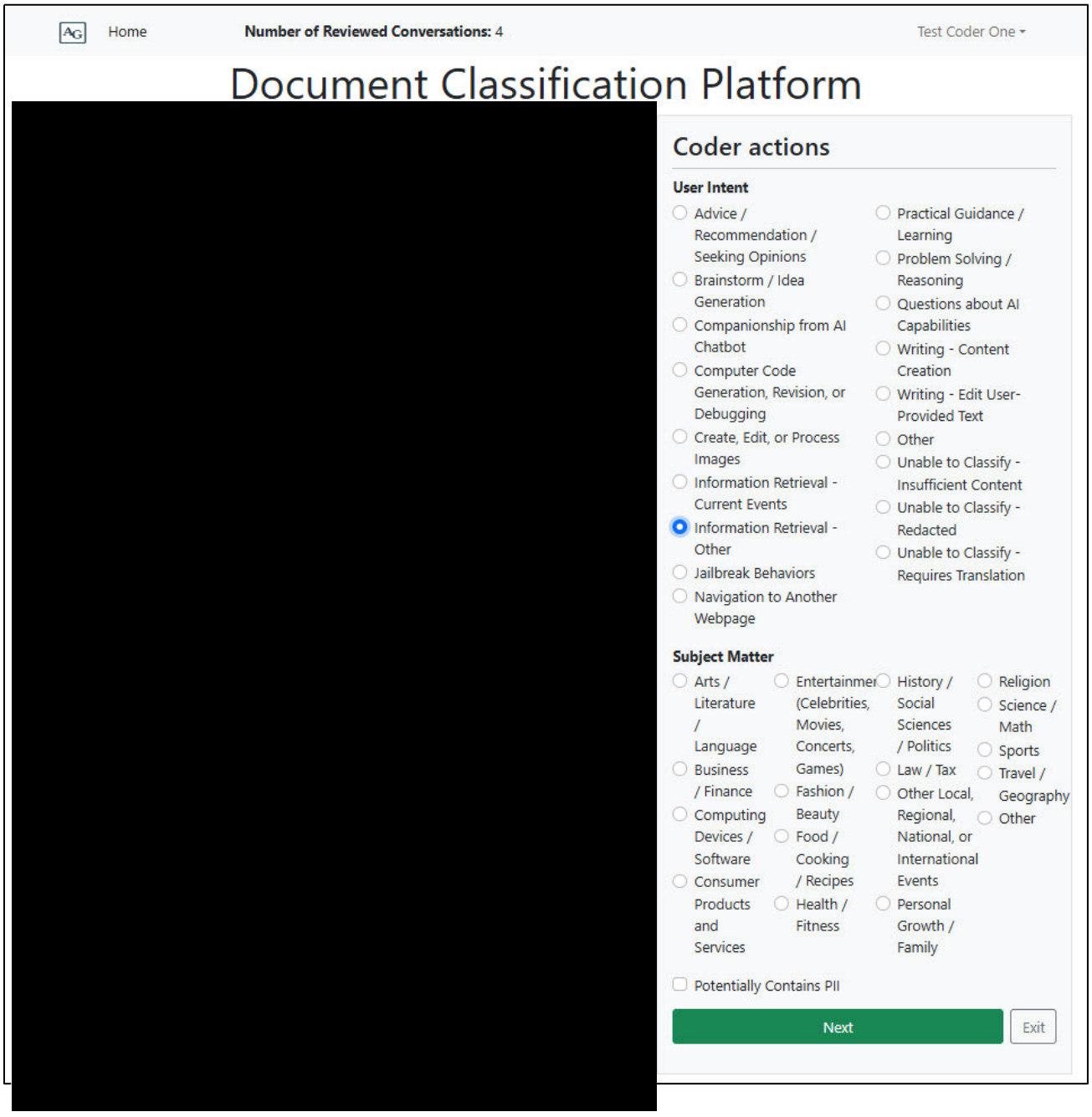
Exhibit 14.A
Coding Platform Interface[86]



[86] Conversation ID: F1DE58D38B05AB311CE83D3E7C9D151B341CD0572729E22A7FAE1F4163CCC634.

[redacted]

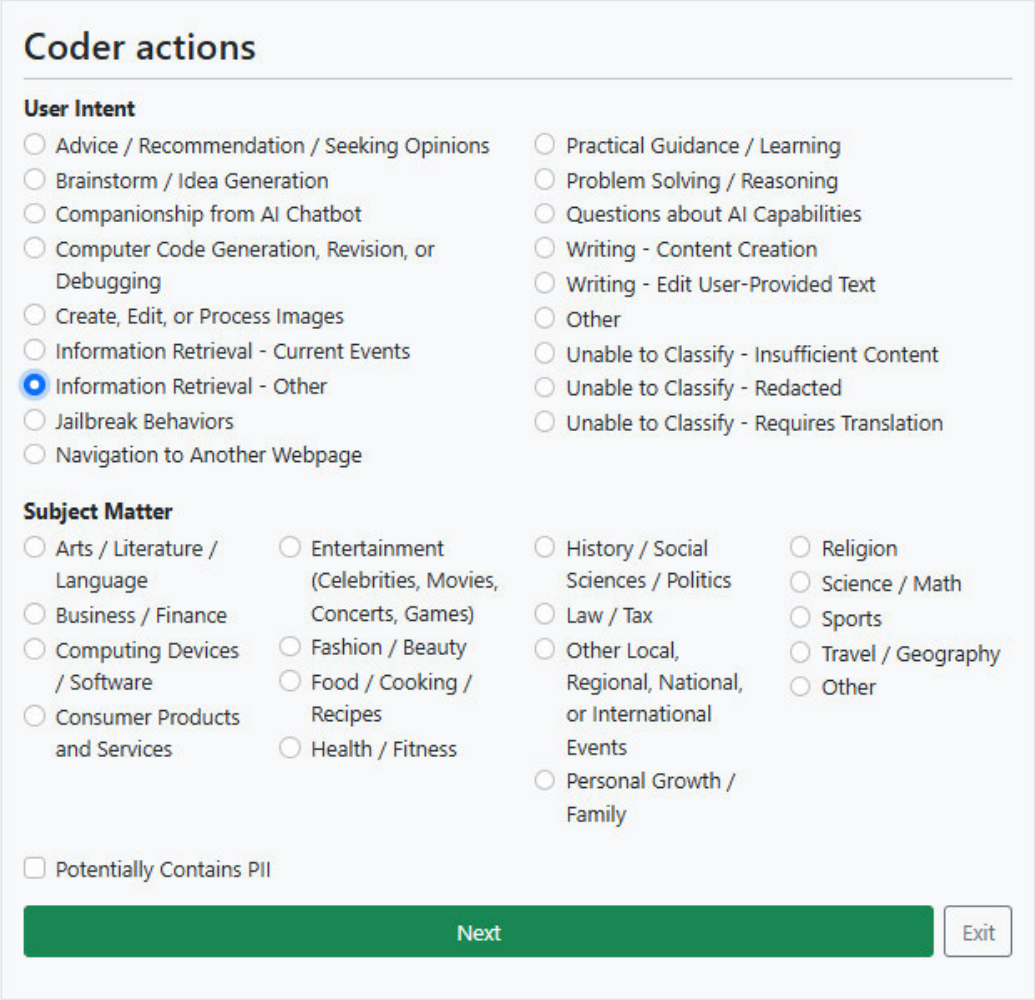
Exhibit 14.B
Coding Platform Interface
Example of "Information Retrieval – Other" Selected[87]



[87] Conversation ID: F1DE58D38B05AB311CE83D3E7C9D151B341CD0572729E22A7FAE1F4163CCC634.

[redacted]

Exhibit 14.C
Coding Platform Interface – Detail View of Coder Actions
Example for User Intent "Information Retrieval – Other" Selected



65. Using this platform, coders were asked to select the user intent that best fit the first substantive request in the conversation by selecting exactly one of the radio buttons pictured in **Exhibit 14.A**. The first substantive request was described to the coders during training as possibly consisting of one or more user messages. Coders were provided hypothetical examples that explained that any introductory greetings should not be considered a substantive request. The end of the first substantive request was defined as either the end of the conversation or the point

[redacted]

in the conversation at which user made an abrupt switch to an unrelated topic.[88] See **Appendix E** for instructions on this topic.

66. If the coder selected "Information Retrieval – Other" or "Information Retrieval – Current Events," the coder was also prompted to select a subject matter by choosing one of the radio buttons depicted in the bottom right of **Exhibit 14.B** above, with detail shown in **Exhibit 14.C**. Coders were not able to move on to the next conversation until they had selected a user intent (and subject matter if the user intent selected was "Information Retrieval"). If coders selected one of the other categories of user intent, they were not allowed to select a subject matter category.

67. The same coding platform interface was used throughout the coder training and pilot coding phase as well as for coding of the Main Analytical Sample.

2. Coder Selection, Training, and Category Refinement Phase

68. At the start of the coding process, coders were selected according to best practices in quantitative content analysis. Coders were chosen based on qualifications that reflect high reading comprehension and analytical rigor by reviewing candidate resumes and recommendations based on prior candidate performance. Coders were never informed: (i) that this work was in the context of a litigation, which litigation was supported by this work, or on behalf of which client they worked, (ii) the research question addressed by the coding exercise, (iii) that the LLM

[88] As I described above, the average number of user turns per conversation is 2.2 and the median number of user turns is one. Over 60 percent of conversations included only one user turn, and therefore could not have had more than one substantive request. There is no consensus in the literature on using the first or the last substantive request, or allowing for multiple uses. I chose to focus on the first request because coders were more consistently able to identify the first request given that humans read from top to bottom. *See, e.g.,* Liu, Jiqun, et al., "Task, Information Seeking Intentions, and User Behavior: Toward A Multi-Level Understanding of Web Search," *CHIIR '19: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, March 2019, pp. 123 - 132 at p. 130 ("The early stages of a search session can be a good time for task prediction [… F]irst query measures can be at least as good as, and sometimes better than, whole session measures for task type predictions.").

[redacted]

data they were seeing was generated by Copilot,[89] or (iv) any of the results of their coding or those of the other coders.

69. Coders were trained according to content analysis best practices so that they could categorize conversations accurately and reliably by following standardized definitions and decision rules (see **Appendix E** for details on training materials). Each coder received the same set of standardized instructions that detailed the definition and hypothetical examples for each user intent and subject matter category and attended common training sessions to ensure coders approached categorization in a consistent manner. Training consisted of an initial set of written instructions and a multi-day coder training curriculum:

    a. Step 1: Introduction, Technical Orientation and IT set-up

    b. Step 2: User Intent Categories Overview and Subject Matter Categories Overview

    c. Step 3: Refresher on Guidelines and Hypothetical Examples

70. Once the coders' technical setup was complete, they were introduced to the task at hand – to categorize user intent captured in interactions with an LLM. Coders were then presented with each user intent category and each subject matter category, its definition, and a few hypothetical examples.[90] This presentation was delivered over a Microsoft Teams call, and coders were encouraged to ask questions and discuss how to further clarify the definition of categories as a group. At regular intervals, coders were given a break and asked to complete a worksheet or quiz to allow for processing of the instructions, to test

[89] While coders were never informed that the data they were reviewing was generated solely by Copilot, I note that some conversations include boilerplate language indicating that the data originated from Microsoft Copilot. *See, e.g.*, Conversation ID AD6DA1F827AEA3A034653F4A54EF03A12E0BE170D9F68D972EB17C3C7FE43E21 ("I'm the new AI-powered chat mode of Microsoft Bing…"). However, coders were not informed whether the data originated from a single or multiple LLM products.

[90] *See* **Appendix E**, pp. E-13 – E-84.

[redacted]

understanding, and to allow coders to internalize the information they learned by immediately applying it to a task.[91]

71. After coders became familiar with the user intent and subject matter topics, they were trained on general guidelines including how to handle foreign language content, redactions, conversations with no bot response, and conversations with multiple substantive requests.[92] Finally, coders were given a demonstration of how to use the coding platform interface.

72. Following the multi-day instructional training, coders moved on to a practical training phase designed to further familiarize them with the interface of the classification tool and allow them to apply the instructional training in practice. As an initial step, all coders were provided with the same set of 250 conversations randomly selected from the 1,000 conversations reserved for coder training and category refinement and asked to independently categorize them. Afterwards, I, along with members of my support team working at my direction, conducted a focus group with the coders to understand any confusion or ambiguities in category definitions or other challenges coders faced using the coding platform interface.[93]

73. In the next step in the training process, I provided coders with the 750 remaining conversations from the 1,000 conversations reserved for coder training and category refinement to analyze. The purpose of this phase was to obtain their input on category definitions to allow for any revisions necessary to allow coders to clearly distinguish categories of user intent and subject matter. I also used this phase to confirm that all coders were able to follow instructions and execute the task reliably and consistently. Again, I, along with members of my support team working at my direction, conducted focus groups with the coders to understand

---

91 *See* **Appendix E**, pp. E-104 – E-116.

92 As discussed above, coders were asked to select the user intent that best fit the *first* substantive request.

93 I recruited eleven coders, but one was removed during the training stage due to failure to meet deadlines and follow general instructions. This left a total of ten coders who completed training and proceeded to analyze conversations in the Main Analytical Sample.

[redacted]

any challenges related to the coding process or the category definitions. Additionally, individualized feedback was provided to coders who were diverging from the group consensus and categorizing conversations in a manner inconsistent with the instructions and provided definitions.[94]

3. Pilot Coding

74. After coder training and category refinement concluded, I performed a final step of testing the refined category definitions to confirm no further adjustments were necessary before proceeding to the analysis of the full sample. At this point, the document classification platform changed from showing the same conversation to all coders to instead showing one conversation to two coders, and then, if they selected different options, showing the conversation to a third coder, simulating the process as later deployed for the Main Analytical Sample. The coders used this environment to test the categories revised through the iterative category refinement process in order to ensure no further changes were needed. They performed this test on the 1,000 conversations reserved for pilot coding. Following this pilot coding process, I did not observe that additional category changes were required. At this point, I considered my taxonomy finalized and ready for application to the Main Analytical Sample.

4. Main Analytical Sample Coding

75. Once the pilot coding was complete, the coders began analyzing the 43,000 conversations in my Main Analytical Sample.[95] As I alluded to above, each conversation was independently reviewed by two coders chosen at random and who were not able to view the decisions of the other coder. If the two coders

---

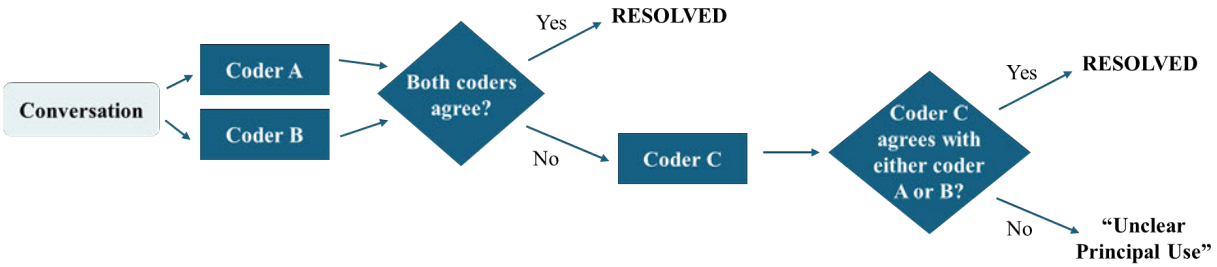
94 This process of individualized feedback is standard practice. *See, e.g.*, Neuendorf, pp. 133-134. The feedback given was regarding the differences between "Information Retrieval – Other" and "Practical Guidance / Learning." Individualized feedback was only provided if coders were diverging from group consensus, and these were the only categories that we observed this behavior.

95 To summarize, out of the 45,000 conversations provided to me as the Copilot Conversations Sample, 1,000 were used for training, and category refinement, and 1,000 were used for pilot coding, which left 43,000 for my Main Analytical Sample.

[redacted]

applied the same category to the document, their coding decisions were logged into the system. If the two original coders did not choose the same category, a third coder was assigned to review the conversation.[96] This third coder was not able to view the coding decisions of the two coders who originally classified the conversation. Conversations where all three coders selected different categories on user intent were marked "Unclear Principal Use."

Exhibit 15
Main Analytical Sample User Intent Analysis



76. As described above, coders who chose "Information Retrieval – Other" or "Information Retrieval – Current Events" for a conversation also had to select a subject matter category. To facilitate my analysis of subject matters, conversations where two coders agreed on either "Information Retrieval – Other" or "Information Retrieval – Current Events" category but selected two different subject matter classifications were also assigned to a final coder for resolution.[97] Conversations where three coders selected different subject matters were marked "Unclear Principal Subject Matter."

77. Once coders began analyzing the Main Analytical Sample, they were not able to ask me or members of my support team questions about how to classify user intents or subject matters of specific conversations, but were allowed to

---

96 In instances where the first two coders both selected Information Retrieval as the user intent, but different subject matters, the conversation was shown to a third coder. However, these conversations are considered resolved at the user intent level.

97 In order to ensure that "Information Retrieval" conversations could reliably be classified by majority rule into subject matter categories, conversations where two of three coders agreed on an "Information Retrieval" user intent category but disagreed on subject matter (with third coder selecting a non-"Information Retrieval" user intent category) were shown to an additional coder who only coded subject matter.

[redacted]

communicate any logistical challenges they experienced while using the proprietary coding platform interface.

78. I monitored coder performance over the course of coding the Main Analytical Sample to ensure that coders were able to maintain consistent performance. I reviewed various metrics, including average number of seconds per conversation reviewed, total number of conversations reviewed, agreement rates, and Cohen's Kappa. Over the course of coding the Main Analytical Sample, two coders left of their own accord for reasons unrelated to the work they were asked to perform.[98] I present summary data about coder performance in **Appendix G**.

VI. Copilot Is Used for a Wide Range of Purposes and to Retrieve Information About a Breadth of Subject Matters in My Main Analytical Sample

79. Following coders' completion of their review of the Main Analytical Sample, I analyzed the coders' classifications and measured intercoder reliability. In the sections that follow, I detail the results and reliability of these analyses. Overall, I find that Copilot is used for a wide range of purposes and users retrieve information about a breadth of subject matters. In support of this conclusion, I personally reviewed at least 100 random conversations classified in each user intent and subject matter category (or all conversations in categories with fewer than 100 total conversations).

A. My Analysis of User Intent Categories Observed in My Main Analytical Sample Reveals That Copilot Is Used for a Wide Range of Purposes

80. To understand usage of Copilot, I developed user intent categories that coders assigned to each conversation in the Main Analytical Sample, as described in **Section IV.B**. These categories include 15 user intent categories that are

[98] The first coder left after coding 1,570 conversations and the second left after coding 3,619 conversations. **Appendix G** includes as a sensitivity analysis the distributions of user intent and subject matter categories excluding any conversations reviewed by the two coders who left before the coding process concluded. These results do not alter my opinions.

[redacted]

informative about Copilot usage,[99] and three categories specifying reasons why the conversation could not be classified.[100] **Exhibit 16** below summarizes the distribution of user intent categories across my Main Analytical Sample, as well as the average number of user turns present in the conversations in each category.

Exhibit 16
Distribution of User Intent Categories[101]

User Intent Category	N	Percent	Average User Turns
Information Retrieval - Other	15,578	36.2%	1.6
Practical Guidance / Learning	3,792	8.8%	2.7
Problem Solving / Reasoning	2,746	6.4%	4.5
Writing - Content Creation	1,799	4.2%	2.1
Advice / Recommendation / Seeking Opinions	1,702	4.0%	1.8
Create, Edit, or Process Images	1,648	3.8%	3.2
Writing - Edit User-Provided Text	1,372	3.2%	2.4
Computer Code Generation, Revision, or Debugging	906	2.1%	3.7
Navigation to Another Webpage	871	2.0%	1.2
Information Retrieval - Current Events	547	1.3%	1.5
Brainstorm / Idea Generation	450	1.0%	2.3
Companionship from AI Chatbot	203	0.5%	3.0
Questions about AI Capabilities	140	0.3%	3.0
Other	29	0.1%	1.7
Jailbreak Behaviors	4	0.0%	3.0
Unable to Classify - Requires Translation	4,660	10.8%	2.4
Unable to Classify - Redacted	2,726	6.3%	1.5
Unable to Classify - Insufficient Content	1,400	3.3%	1.4
Unclear Principal Use	2,427	5.6%	2.6
Total	**43,000**	**100%**	**2.2**

[99] There are: "Advice / Recommendation / Seeking Opinions," "Brainstorm / Idea Generation," "Companionship from AI Chatbot," "Computer Code Generation, Revision, or Debugging," "Create, Edit, or Process Images," "Information Retrieval – Current Events," "Information Retrieval – Other," "Jailbreak Behaviors," "Navigation to Another Webpage," "Practical Guidance / Learning," "Problem Solving / Reasoning," "Questions about AI Capabilities," "Writing – Content Creation," "Writing – Edit User-Provided Text," and "Other."

[100] These are: "Unable to Classify – Requires Translation," "Unable to Classify – Redacted," and "Unable to Classify – Insufficient Content."

[101] In instances where the first two coders disagreed on user intent, but the third coder agreed with one of the first two, that conversation is classified per majority 2/3 rule. I define user turns as the number of distinct user messages within a conversation.

[redacted]

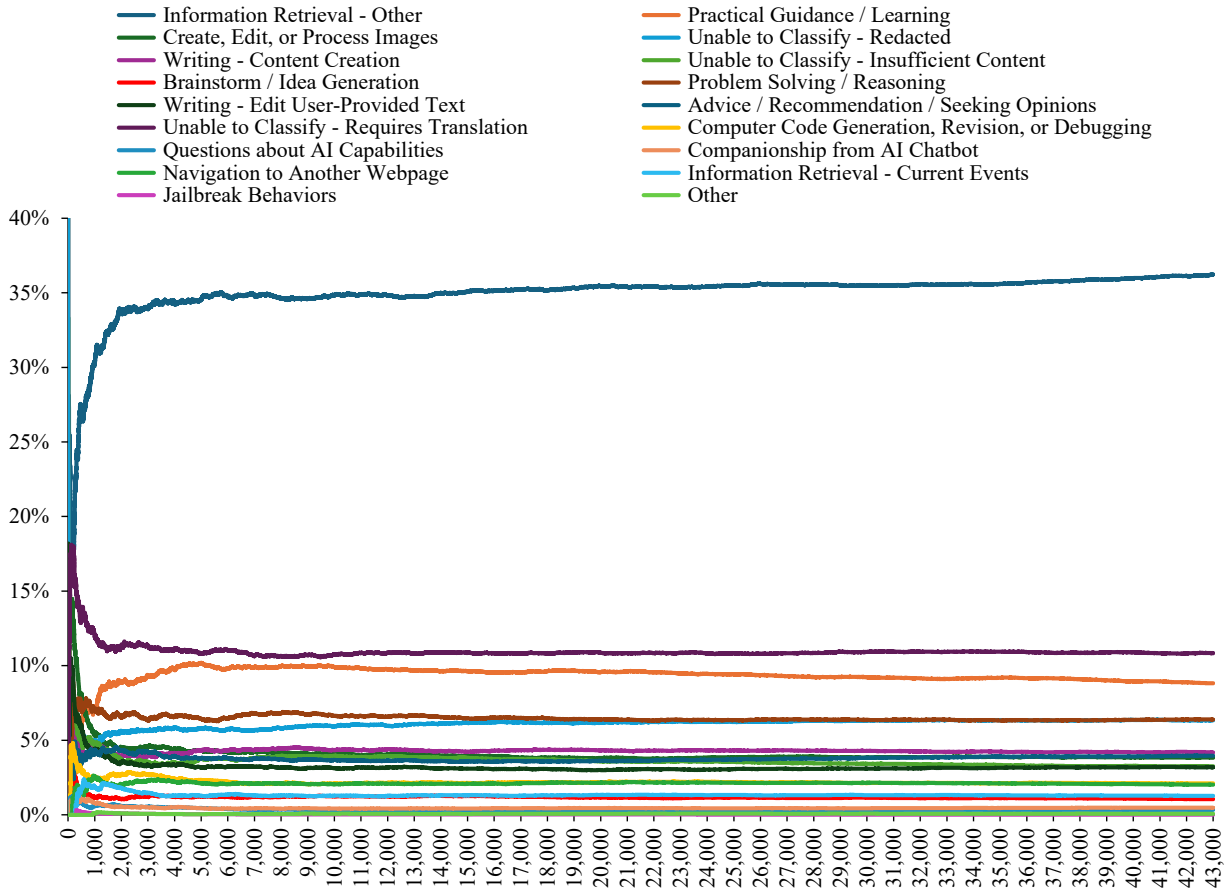
81. Conversations span all 15 of my user intent categories and all three of my "Unable to Classify" categories. There is substantial use of Copilot for a wide range of intents, with at least one percent of conversations in my Main Analytical Sample categorized into eleven user intent categories.

82. The conversations vary in length, with "Problem Solving / Reasoning" conversations averaging 4.5 user turns, while "Navigation to Another Webpage" conversations average only 1.2 user turns.

83. **Exhibit 17** below plots the distribution of user intent category assignments over the course of the coding process. This graph shows that the user intent category percentages stabilize early in the coding process after a short period when the sample size was very small, and remain consistent thereafter, indicating that the user intent category distribution converges well before the full dataset is reviewed. This suggests that the sample size is more than sufficient to capture the underlying patterns in the data.

[redacted]

Exhibit 17
User Intent Category Assignments over the Course of the Coding Process



84. To illustrate how these user intent categories differ in practice, I provide excerpts of conversations classified into selected categories below. These examples highlight the defining characteristics that distinguish one user intent category from another, according to the user intent category definitions I provided to the coders.

85. The most common use of Copilot, "Information Retrieval – Other," accounted for 36.2 percent of conversations. This category represents Copilot queries that are similar to a Google or Bing search, such as those requesting simple facts or information about broad topics. **Exhibit 18** below shows an example of a conversation coded into "Information Retrieval – Other." As described in more detail in **Section VI.B**, users retrieved information about a variety of subject matters, such as health and fitness, consumer products, and entertainment.

[redacted]

Exhibit 18
Information Retrieval – Other Example[102]

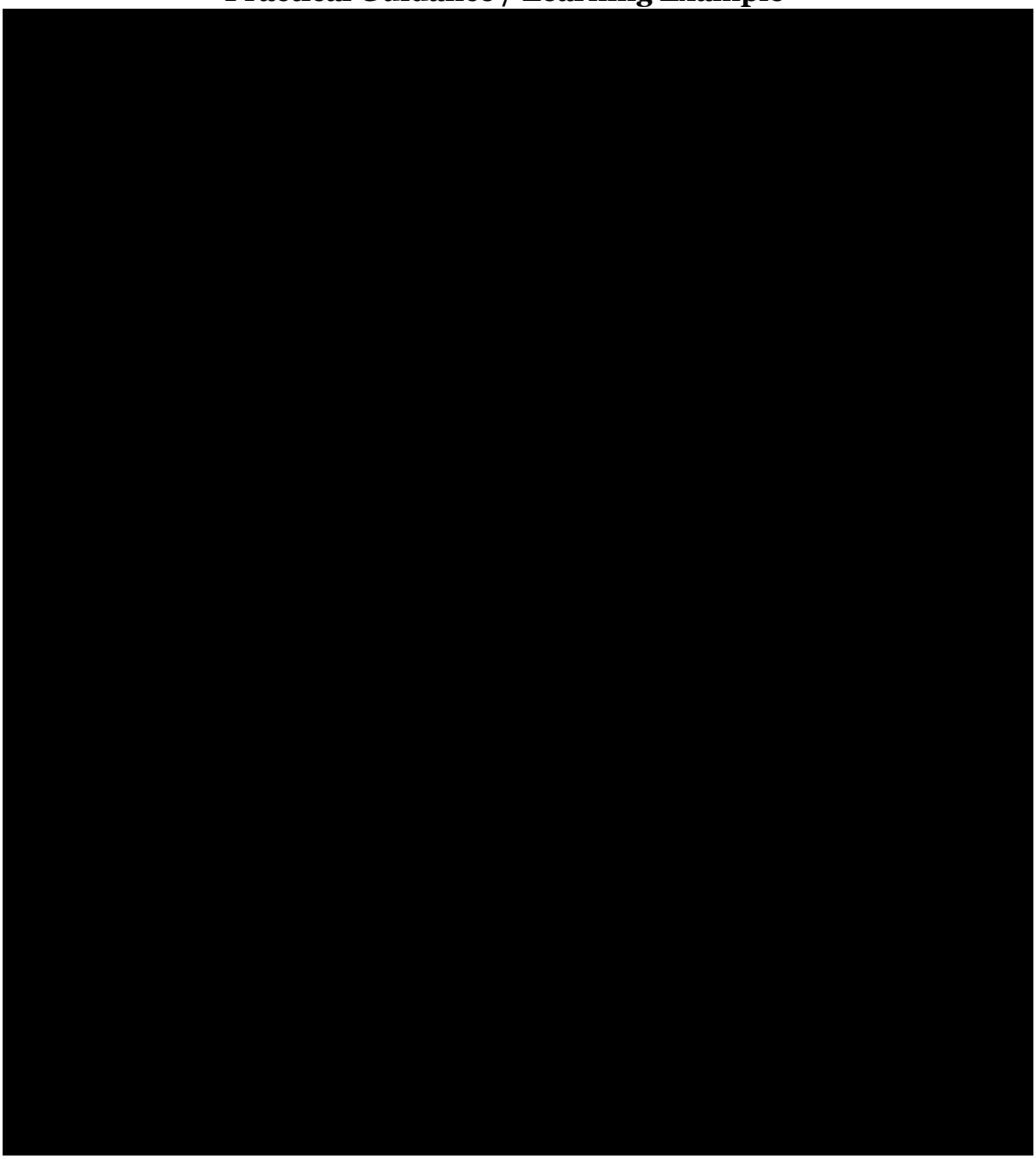


86. The next most common use of Copilot was "Practical Guidance / Learning," at 8.8 percent of conversations. These are conversations where the user appears to focus on learning and understanding a concept or seeking to acquire a new skill. I observed that these conversations included users asking Copilot how to perform a task or seeking to understand a topic of interest. **Exhibit 19** below shows example of a conversation coded as "Practical Guidance / Learning."

[102] Conversation ID: B614F91C6B4F7978AB1577EFFBBF470B90BC493715B9D30C1CFAA6A6AC461F34.

[redacted]

Exhibit 19
Practical Guidance / Learning Example[103]



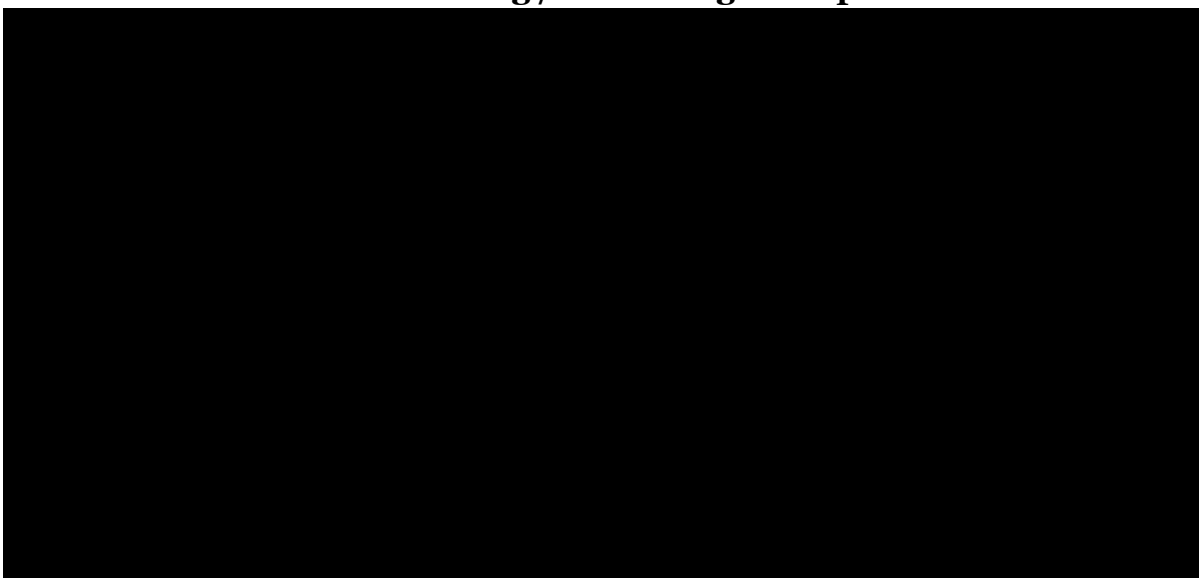
87. The third most common use of Copilot was "Problem Solving / Reasoning," representing 6.4 percent of the conversations. This category captures prompts

[103] Conversation ID: F284FA364F5772CEBA805436BB42E94338347B5B637A1C6B0D47334119E680DA.

[redacted]

where Copilot processes information and applies it to a given situation. My review of conversations in this category showed that these conversations often involved users asking Copilot to solve multiple choice or other homework-style questions. **Exhibit 20** below shows example of a conversation coded into "Problem Solving / Reasoning."

Exhibit 20
Problem Solving / Reasoning Example[104]



88. Additional examples of the broad use cases for which Copilot is used are present across other categories as well. For example, some users asked Copilot to draft written material or create images. 4.2 percent of conversations were classified as "Writing – Content Creation," where users asked Copilot to assist in creating written materials (see **Exhibit 21** for an example) and 3.2 percent of conversations were classified as "Writing – Edit User-Provided Text," in which case users provided draft language and asked Copilot to help revise the text (see **Exhibit 22** for an example). My review of conversations coded into each of these categories revealed writing tasks including drafting emails, summarizing or rephrasing user-provided text, correcting grammar, and assisting in writing user reviews, among others.

[104] Conversation ID: C4DC95156CB50F63A281458364CD9C4C4D1E8DC31C090A4D1F5518C278BF46D9.

[redacted]

Exhibit 21
Writing – Content Creation Example[105]



Exhibit 22
Writing – Edit User Provided Text Example[106]



[105] Conversation ID: BED776E09A47F7A35227D79943560F0FDA52560E4BF76DD81F90A9ED0392237A.

[106] Conversation ID: B15894DBE2B86B41AF150A0B1A0FF7F59AA61EDBF63F7123E51F4B1FF43E4ED4.

[redacted]

89. Additionally, 3.8 percent of conversations were classified as "Create, Edit, or Process Images" (see **Exhibit 23** for an example). The conversations in this category included requests for image generation based on a user-provided description and follow-up adjustments to the images, as well as help reviewing or understanding images uploaded by the user.

Exhibit 23
Create, Edit, or Process Images Example[107]

[redacted]

90. Other users prompted Copilot to help with technical tasks, with 2.1 percent of conversations being classified as "Computer Code Generation, Revision, or Debugging" (see **Exhibit 24**). I observed that these conversations included, for example, creating computer code to perform a particular task or requests for help editing provided code, across a wide variety of programming languages and technical tools such as HTML, Java, JavaScript, C++, Python, SQL, and others.

[107] Conversation ID: 5D665A4B904D556AF31C0FBFDB1B6441AF3A13ED4793597DAC8A886BE0AF2037.

[redacted]

Exhibit 24
Computer Code Generation, Revision, or Debugging Example[108]

[redacted]

[108] Conversation ID: F02515B21DEF4EDFC3F81143A3DE36C827B9AD835A07C3461704A9088E8BE35B.

[redacted]

91. Yet other users requested an opinion or sought advice from Copilot. 4.0 percent of conversations were coded as "Advice / Recommendation / Seeking Opinions," which captured prompts where the user was asking for help in making a choice or seeking to form an opinion (see **Exhibit 25**). I observed that these conversations included, for example, users asking for suggestions for trip itineraries, recommendations regarding interior design, or advice on obtaining discounts. Another 0.5 percent of conversations were coded as "Companionship from AI Chatbot" (see **Exhibit 26**), which captured conversations where the user was using Copilot as a confidant, friend, therapist, or role-play partner. I observed that these conversations included examples where users would greet and casually converse with Copilot in the style of an everyday conversation or ask for jokes.

[redacted]

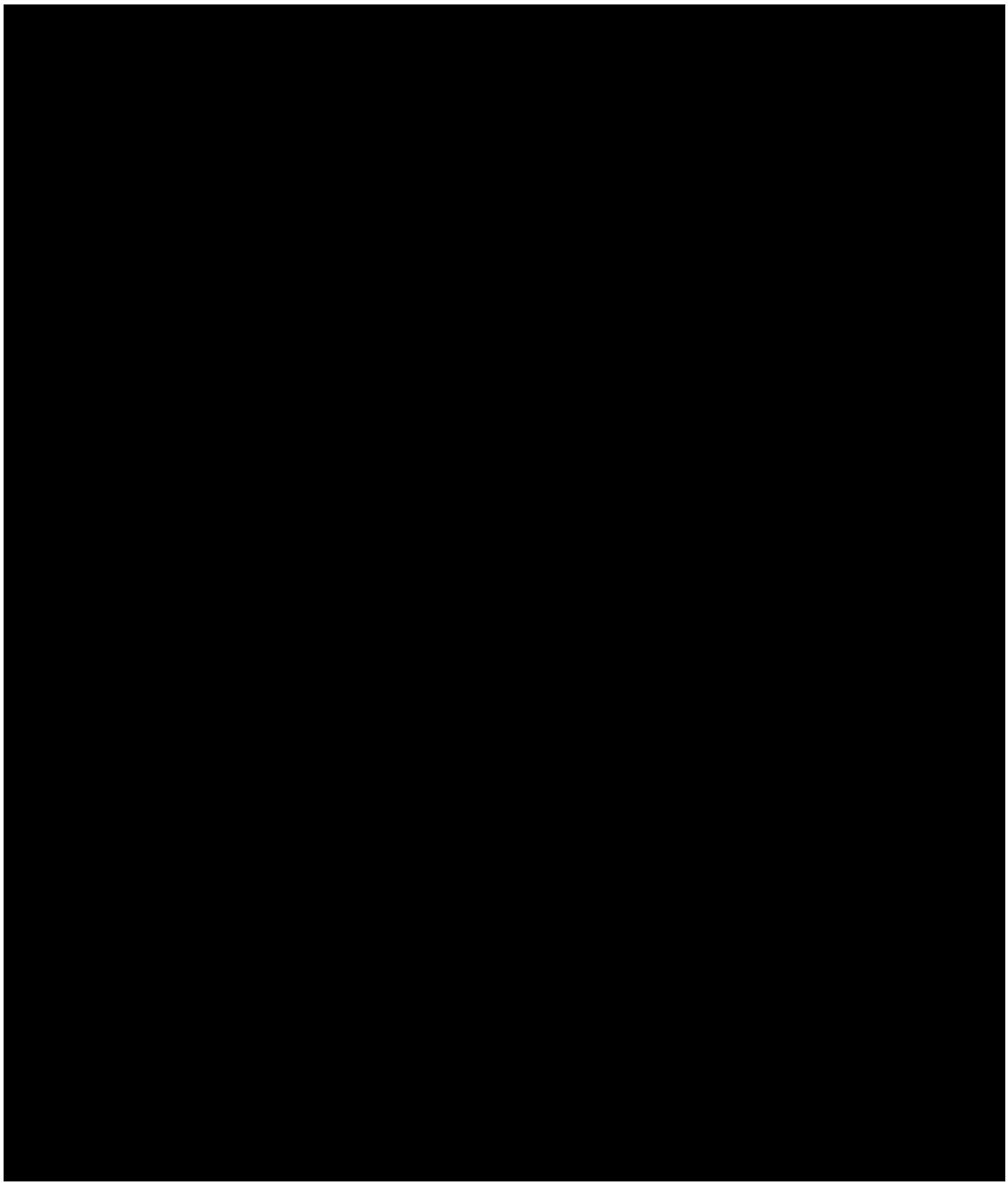
Exhibit 25
Advice / Recommendation / Seeking Opinions Example[109]

[redacted]

[109] Conversation ID: DD6894D4AEBB2EEF65C209DB577F988A680161ECF683E2985B96782B27719253.

[redacted]

Exhibit 26
Companionship from AI Chatbot Example[110]



92. In addition to my 15 user intent categories, conversations could be coded into three categories representing conversations that could not be classified into a

[110] Conversation ID: FE54463B52E061F309576A8748F01EABF28A8A7C1EECBBF2E5B1806836422DE9.

[redacted]

specific user intent: "Unable to Classify – Requires Translation," "Unable to Classify – Redacted," and "Unable to Classify – Insufficient Content." These categories represent conversations for which coders reliably identified that insufficient information existed in the conversation for them to be able to make a determination as to what the intent of the conversation was.

93. "Unable to Classify – Requires Translation" (10.8 percent of conversations) captures conversations where the English-speaking coders could not make a determination of the user intent behind the conversation because of foreign language content in the user prompts (see **Exhibit 27** for an example). I discuss this category of conversations further in **Section VI.C.3**.

Exhibit 27
"Unable to Classify – Requires Translation" Example[111]

[redacted]

94. "Unable to Classify – Redacted" (6.3 percent of conversations) captures conversations where the coders could not make a determination regarding the user's intent because the user's prompts were wholly or substantially redacted (see **Exhibit 28** for an example).[112] In instances where users' prompts were substantially redacted but Copilot's responses offered non-redacted content, coders were asked to refrain from inferring user intent solely from the responses.

---

[111] Conversation ID: 45E13CB52D558475EC2068F3424720BDB1F661EEF296D73F4D07AEEA1FF803F7.

[112] I understand that the Lafferty Declaration, Section VIII.C (Redaction for Personally Identifiable Information (PII)) explains that the Copilot Conversations Sample was redacted using Azure AI Language (PII Named Entity Recognition) to remove users' personally identifiable information. *See* "What is PII Detection in Azure Language?," *Microsoft*, March 30, 2026, https://learn.microsoft.com/en-us/azure/ai-services/language-service/personally-identifiable-information/overview.

[redacted]

While existing literature on LLM usage that analyzes proprietary data sources has included data passed through privacy filters, these analyses used LLM-driven methodologies that classified the unfiltered data without human review.[113] As my analysis relied on human coders who could use their judgement as to whether the conversation was classifiable, "Unable to Classify – Redacted" represents unclassifiable content beyond what is typically reported in the literature. Notably, on average, 46.0 percent of the words in the user turns of conversations coded as "Unable to Classify – Redacted" consisted of the text string "[Redacted]" whereas, on average, only 8.5 percent of the words in the user turns of all conversations in the Main Analytical Sample were "[Redacted]."

Exhibit 28
"Unable to Classify – Redacted" Example[114]

[redacted]

95. "Unable to Classify – Insufficient Content" (3.3 percent of conversations) captures conversations where the coders could not make a determination as to the intent behind the conversation because the user prompt did not contain enough identifiable content (see **Exhibit 29** for an example). Existing literature on LLM usage makes use of similar categories when a conversation cannot be classified due to, for example, length. Costa-Gomes et al. (2025) includes a

[113] *See, e.g.*, Chatterji, Aaron, et al., "How People Use ChatGPT," *National Bureau of Economic Research*, September 2025, p. 6; "Clio: A System for Privacy-Preserving Insights into Real-World AI Use," *Anthropic*, December 12, 2024, https://www.anthropic.com/research/clio.

[114] Conversation ID: DF2C74427EB1242BE0A834AF00F8B4870C0A87ACF199268D2EF333394BACE547.

[redacted]

category called "No Topic / No Intent," and described these conversations as "vague" and "short."[115] Zheng et. al. (2024) instead removes prompts that are fewer than 32 characters altogether from their analysis.[116] Notably, conversations coded as "Unable to Classify – Insufficient Content" on average contained 8.0 words in user turns, whereas the average number of words in the user turns of all conversations in the Main Analytical Sample was 65.2 words.

Exhibit 29
"Unable to Classify – Insufficient Content" Example[117]

[redacted]

96. The results of my quantitative content analysis show that Copilot is used for a wide breadth of purposes, as demonstrated in the variety of user intent categories. Across the coded categories, usage spans both functional and creative tasks, with meaningful representation in each, suggesting that engagement is not concentrated in a single primary use case. The distribution of themes reinforces that users turn to Copilot for everything from productivity and information-seeking to ideation and personal assistance. As reflected in the overall classification of the sample and consistent with my personal review of a random

---

[115] Costa-Gomes, Beatriz, et al., "It's About Time: The Copilot Usage Report 2025," *Microsoft*, December 2025, p. 2.

[116] Zheng, Lianmin, et al., "LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset," *ICLR 2024*, March 10, 2024, p. 3.

[117] Conversation ID: 88F881EC0BAB7BC2AF03C5ACB8C381FFE85A6B31BF21D51259077754F085DCF8.

[redacted]

sample of at least 100 conversations from each category (or all conversations in categories with fewer than 100 total conversations),[118] people infrequently used Copilot in relation to current events. Further, among the set of conversations I personally reviewed, I did not observe any instances of users requesting verbatim text of news articles, instances of users purposely trying to access content behind a paywall, nor did I observe users requesting verbatim text of existing literary works, writing long form literary works or sequels derived from existing long-form literary works.

B. My Analysis of Subject Matters Present in My Main Analytical Sample Reveals that Users Use Copilot to Retrieve Information on a Breadth of Subject Matters

97. As described in **Section VI.A** above, the largest use of Copilot was for information retrieval, with 36.2 percent of conversations coded as "Information Retrieval – Other," and 1.3 percent of conversations coded as "Information Retrieval – Current Events," which represents a similar use of Copilot for events or developments that appeared to occur within a week of the date of the prompt, based on the wording of the prompt and bot responses in combination with timestamps of when conversations occurred.

98. As shown in **Exhibit 30** below, of the 16,125 conversations coded into one of the Information Retrieval categories, 547 (3.4 percent) were categorized as "Information Retrieval – Current Events" and 15,578 (96.6 percent) were categorized as "Information Retrieval – Other."

Exhibit 30
Information Retrieval

Information Retrieval Category	N	Percent
Information Retrieval - Other	15,578	96.6%
Information Retrieval - Current Events	547	3.4%
Total	**16,125**	**100%**

[118] This is in addition to my review of a sample of 10 percent of conversations in each of the two groups "Unclear Principal Use" and "Unclear Principal Subject Matter."

[redacted]

99. As I described above, coders were asked to select the subject matter of conversations categorized as "Information Retrieval – Current Events" or "Information Retrieval – Other." **Exhibit 31** below summarizes the subject matters selected by coders for the information retrieval conversations.[119] I include the conversations for which there is no majority consensus among the three coders in the category "Unclear Principal Subject Matter."

Exhibit 31
Distribution of Subject Matter Classifications

	Information Retrieval - Other		Information Retrieval - Current Events	
Subject Matter Category	**N**	**Percent**	**N**	**Percent**
Consumer Products and Services	1,755	11.3%	5	0.9%
Health / Fitness	1,669	10.7%	2	0.4%
Entertainment (Celebrities, Movies, Concerts, Games)	1,644	10.6%	42	7.7%
Computing Devices / Software	1,545	9.9%	6	1.1%
Travel / Geography	1,107	7.1%	38	6.9%
Science / Math	1,025	6.6%	20	3.7%
Arts / Literature / Language	1,010	6.5%	1	0.2%
Business / Finance	917	5.9%	88	16.1%
Law / Tax	781	5.0%	4	0.7%
Other	680	4.4%	51	9.3%
History / Social Science / Politics	632	4.1%	24	4.4%
Food / Cooking / Recipes	560	3.6%	2	0.4%
Sports	347	2.2%	134	24.5%
Religion	335	2.2%	1	0.2%
Fashion / Beauty	155	1.0%	2	0.4%
Other Local, Regional, National, or International Events	108	0.7%	98	17.9%
Personal Growth / Family	79	0.5%	0	0.0%
Unclear Principal Subject Matter	1,229	7.9%	29	5.3%
Total Information Retrieval Conversations	**15,578**	**100%**	**547**	**100%**

100. As shown in **Exhibit 31**, there is substantial variation in the subject matters that Copilot users requested information about. For "Information Retrieval – Current Events," the most common subject matters were "Sports" (see **Exhibit 32** for an example), "Other Local, Regional, National or International Events" (see **Exhibits 33a and 33b** for examples), and "Business / Finance" (see **Exhibit**

[119] In instances where the first two coders disagreed on subject matter, but the third coder agreed with one of the first two, that conversation is classified per majority 2/3 rule.

[redacted]

34 for an example). The subject matter of "Other Local, Regional, National, or International Events" included, for example, users asking about severe weather events, political developments, accidents, fires, and other emergencies.

Exhibit 32
Information Retrieval – Current Events
Sports[120]



[120] Conversation ID: A73AB4DC8E8005564866EA4479009970ED5E79A104AFC7D3395CDEC3182C0689.

[redacted]

Exhibit 33a
Information Retrieval – Current Events
Other Local, Regional, National or International Events[121]

[redacted]

Exhibit 33b
Information Retrieval – Current Events
Other Local, Regional, National or International Events[122]

[redacted]

[121] Conversation ID: 64AA2FD0D59965D50862C9F8B0BD61C0EF5B747FE3B20A47B55852AF8C0F3707.

[122] Conversation ID: 070E342C3F1D8C81BB1697604CD23E7CECC04B32B1D8D66D176A144DC2035DCD.

[redacted]

Exhibit 34
Information Retrieval – Current Events
Business / Finance[123]

[redacted]

101. For "Information Retrieval – Other," there was a more uniform distribution of subject matters, with 15 of the 17 categories accounting for at least one percent of the total "Information Retrieval – Other" user intent. The most common subject matters in "Information Retrieval – Other" were "Consumer Products and Services" (see **Exhibit 35** for an example), "Health / Fitness" (see **Exhibit 36** for an example), and "Entertainment (Celebrities, Music, Concerts, Games)" (see **Exhibit 37** for an example). The subject matter of "Consumer Products and Services" included a breadth of inquires ranging from very specific questions about product attributes, pricing, or features (*e.g.*, options for a bike tube of a particular size with a removable valve or features of a car model) to broad questions about product categories (*e.g.*, cost of replacing a garage door).

[123] Conversation ID: 036705E550323B09EAB98B48EBBCA0E9AE689D1AF9934159856C90F75D0A06ED.

[redacted]

Exhibit 35
Information Retrieval – Other
Consumer Products and Services[124]



Exhibit 36
Information Retrieval – Other
Health / Fitness[125]



[124] Conversation ID: C8748AB7FD5FDCF1CA3C315C3B33F80225F959A2323E7E7AFE12A672CC107743.

[125] Conversation ID: F1DE58D38B05AB311CE83D3E7C9D151B341CD0572729E22A7FAE1F4163CCC634.

[redacted]

Exhibit 37
Information Retrieval – Other
Entertainment (Celebrities, Music, Concerts, Games)[126]

[redacted]

102. In addition, coders identified conversations that belong in an "Other" subject matter for both types of Information Retrieval. For example, conversations coded into the "Other" subject matter for "Information Retrieval – Current Events" included conversations about time and weather (see **Exhibits 38 and 39**).

Exhibit 38
Information Retrieval – Current Events
"Other" Subject Matter
Time Example[127]

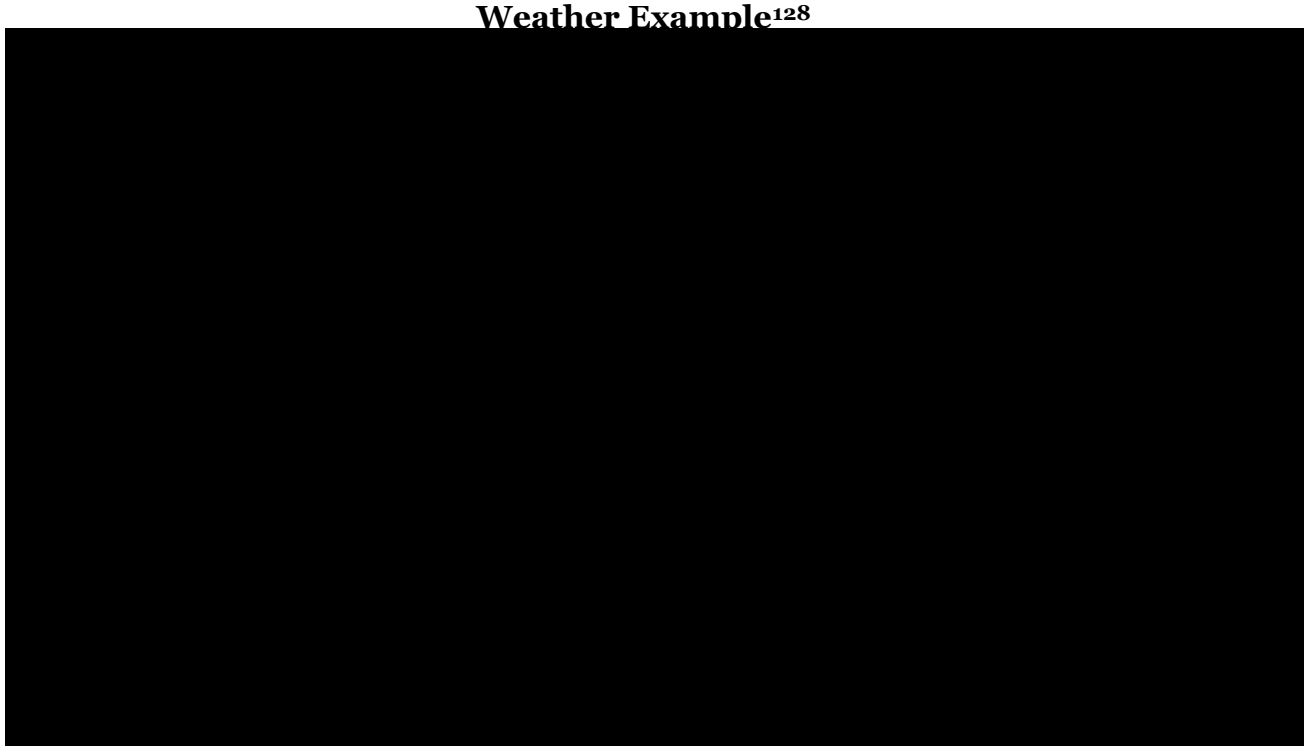
[redacted]

[126] Conversation ID: 3FF9536725B55C8291BBEE0B871DF4768C045349FE076B19CFF80EAE07FF7F12.

[127] Conversation ID: 12136C9015E62AA4C7C077DAF8BED359156A8E522FDD0C0DE06E2C6E886202AD.

[redacted]

Exhibit 39
Information Retrieval – Current Events
"Other" Subject Matter
Weather Example[128]



103. Similarly, conversations coded into the "Other" subject matter for "Information Retrieval – Other" included conversations about real estate and animals (see **Exhibits 40 and 41**).

[128] Conversation ID: 175BF3DDAA495E6258751D943441E0560C5EADA2C76C045F891D9C22B87F9859.

[redacted]

Exhibit 40
Information Retrieval – Other
"Other" Subject Matter
Real Estate Example[129]

[redacted]

[129] Conversation ID: BF5E4F6D9654198B46CCD3DF2565A852543AF1DE2D57F218FCAC79E40118D428.

[redacted]

Exhibit 41
Information Retrieval – Other
"Other" Subject Matter
Animals Example[130]

[redacted]

104. The results of my quantitative content analysis show that Copilot is used to retrieve information about a wide breadth of subject matters. The distribution of themes across subject matters reinforces that users turn to Copilot for requests related to a plethora of topics they encounter in their daily lives.

C. Web Domains Visited by Copilot While Generating Responses to User Requests Do Not Necessarily Capture User Intent

105. The taxonomy I developed categorizes user intent based on what the user sought to accomplish, not based on whether the information a user requests could, in theory, appear in a newspaper. Any attempt to map newspaper sections to my

[130] Conversation ID: C7C9229CF0E933A8E310129E6AE50FEB6AC763C9E4272B47F94936F88E250426.

[redacted]

user intent categories would be methodologically unsound. The mere fact that information could appear in a particular newspaper section does not mean that Copilot relies on that information when responding to user queries. My taxonomy, by contrast, measures what it was designed to measure: user intent as expressed in the user's own words.

106. While reviewing the websites Copilot visited in formulating its response may be helpful in understanding whether information related to the user's query is available on those websites, Copilot's internal processing and searches would not be more informative than the user intent expressed in the user's prompt to understand what the user might have done in the absence of Copilot.

107. Take for example a Copilot conversation from the Bail Conversations Sample in which a user asked Copilot to create an image of someone stealing a pair of boxer briefs.[131] Copilot, in creating that image, visited one of the News Plaintiffs' webpages: https://www.nytimes.com/wirecutter/reviews/best-mens-underwear/. The user was not seeking any information from The New York Times, let alone information on the best men's underwear. Copilot's processing of the user request in this instance happened to involve visiting a News Plaintiff's webpage (which reviews different brands of men's underwear) but this was quite distinct from what the user tried to accomplish by visiting Copilot (i.e., create an image of someone stealing a pair of underwear). I have not seen any evidence that would show that, in absence of Copilot, this user would have visited the Wirecutter page or any other News Plaintiffs' website. The user intent expressed in the prompt is informative to the question of whether the Copilot response could have substituted for a visit to the original source.

108. **Exhibit 42** below shows an example of a Copilot user looking for information about "golf." As part of the RAG process, Copilot visited ten websites. See **Appendix H**. Again, one of these ten was a News Plaintiff's website.[132] Even

[131] Conversation ID: 9A9B6C791B237213E447244FDA590A1BD4EB3F22310411EE31D70319812778BA.

[132] The News Plaintiff's website was: https://www.nytimes.com/athletic/golf/. *See* **Appendix H** for all websites Copilot visited.

[redacted]

though Copilot responded by providing information about recent golf events, nothing in the user prompt indicated that the user was actually looking for news or current events involving golf rather than information about golf in general. Plaintiffs' experts fail to show how the fact that one of the websites searched by Copilot was a News Plaintiff's website changes the user's intent, or that the user would have otherwise searched a News Plaintiff's website if not for the answer provided by Copilot.

Exhibit 42
Information Retrieval – Other[133]

[redacted]

109. **Exhibit 43** shows the breakdown of the conversations generated by Plaintiffs that involved Copilot searching at least one News Plaintiff's website into six user intents that Plaintiffs identified as potentially having a connection to News Plaintiffs' content. This universe of conversations includes instances where the user did not see the News Plaintiff URL or where multiple websites, including those unrelated to News Plaintiffs, were searched. As this exhibit shows, only 0.4 percent of all conversations included at least one News Plaintiff website RAG

[133] Conversation ID: 39383780CFE7D80BE06961A137D3AF625430CE7E7FA324B1A02B9E8354C92DCB.

[redacted]

request (potentially as part of many other non-news related RAG requests). Even without any further analysis, this metric by itself illustrates how infrequently Copilot searches News Plaintiffs' websites when generating responses to user queries. In 99.6 percent of conversations,[134] Copilot generated responses without searching any News Plaintiffs' website.

Exhibit 43
Summary of Conversations with News Plaintiff RAG Request

Bail Report User Intent	Total Number of Conversations (Bail Copilot Content Analysis)	News Plaintiffs' Domains Used for RAG		
		Number of Conversations	Conversations with RAG (%)	Conversations without RAG (%)
Advice / Recommendation / Seeking Opinions	1,702	24	1.4%	98.6%
Brainstorm / Idea Generation	450	1	0.2%	99.8%
Information Retrieval - Current Events	547	15	2.7%	97.3%
Information Retrieval - Other	15,578	56	0.4%	99.6%
Practical Guidance / Learning	3,792	8	0.2%	99.8%
Writing - Content Creation	1,799	4	0.2%	99.8%
All Other User Intents	21,132	77	0.4%	99.6%
Total	**45,000**	**185**	**0.4%**	**99.6%**

D. The Two Categories Most Relevant to Plaintiffs' Content Include Many Conversations Unrelated to Plaintiffs' Content

1. Information Retrieval – Current Events

110. As made clear in the definition and through examples provided to the coders, this category was not restricted to news, but included any "factual information about events or developments that occurred within a week of the date of the prompt."[135]

[134] Calculated as 100% minus 0.4%.

[135] Appendix D at D-2.

[redacted]

Such information could include factual information like the current price of a stock or the weather.[136]

111. 97.3 percent of these conversations did not search News Plaintiffs' websites. So, by Plaintiffs' experts' own measures, many of these conversations did not rely on News Plaintiffs' published content. And nonetheless, the existence of an article on a News Plaintiff's website on a similar topic to one of my user intent categories does not imply that a user would navigate to that article in the absence of Copilot.

112. The following two exhibits show examples of conversations that involve users requesting current factual information, but information one would not necessarily expect to find in a newspaper. Consistent with this expectation, Copilot did not search any News Plaintiffs' websites while generating its response. See **Appendix H**. In **Exhibit 44a**, the user is inquiring about the price of gold in Egypt. In **Exhibit 44b**, the user is asking for the current time.

Exhibit 44a

Example of Conversation Unrelated to News Classified as "Information Retrieval – Current Events"[137]

[redacted]

[136] Appendix D at D-2. Hypothetical examples provided to coders included the following: "What is the price of Apple's stock?"; "weather in NYC"; "sunrise today."

[137] Conversation ID: 91F131EDCDD2DA84191675A1AA0C0BA5255AA4822E82A19E47BA629EE24D49EF.

[redacted]

Exhibit 44b
Example of Conversation Unrelated to News
Classified as "Information Retrieval – Current Events"[138]

[redacted]

2. Writing – Content Creation

113. The category "Writing – Content Creation" was not defined to capture only generation of novels and other books and there is no evidence that conversations in this category solely capture Copilot users engaging in such activities. There are many "Writing – Content Creation" prompts that do not involve a user asking Copilot to generate a novel or a book, making 4.2 percent a theoretical upper bound for this activity. See, for example, **Exhibits 45a** and **45b**. **Exhibit 45a** shows a user asking Copilot to create a post about a brand, while **Exhibit 45b** shows a user asking Copilot to help draft a response to praise. This suggests that all or most uses in this user intent category do not involve writing a novel or competing with works published by the fiction and nonfiction authors which constitute the Class Plaintiffs. Additionally, in this category, conversations have on average 2.1 turns, with 68.9 percent of conversations consisting of a single turn, and 87.4 percent consisting of three or fewer turns.[139] A user could not accomplish writing a novel that could compete with Plaintiffs' works using only a handful of turns.

[138] Conversation ID: 12136C9015E62AA4C7C077DAF8BED359156A8E522FDD0C0DE06E2C6E886202AD.

[139] *Supra* Exhibit 16.

[redacted]

Exhibit 45a
Example of Conversation Unrelated to a Novel or Book Classified as "Writing – Content Creation"[140]

[redacted]

[redacted]

Exhibit 45b
Example of Conversation Unrelated to a Novel or Book Classified as "Writing – Content Creation"[141]

[redacted]

E. My Content Analysis Results Are Reliable and Additional Analyses on Conversation Subsets Do Not Alter My Opinions

1. My Content Analysis Results Exhibit Substantial Intercoder Reliability That Is Consistent with the Literature

114. As described in **Section III**, a quantitative content analysis can be evaluated as to the reliability of its process. The overall pairwise agreement rate of coders on our research team was 72.5 percent. This represents the average rate at which the two coders who viewed the same conversation selected the same user intent category. By layering on a third coder to review conversations where the first two coders selected different user intent categories, my content analysis achieved a majority agreement rate—reflecting the rate at which at least two coders selected the same user intent category—of 94.4 percent.

115. Another alternative measure of intercoder reliability is Cohen's Kappa, which corrects the agreement rate by accounting for agreement that would be expected to occur purely by chance. Cohen's Kappa for a pair of coders is calculated as:

---

[140] Conversation ID: 798FCFD23FA20AB33355969D9FE0072347234C142BF98D10507CD1EC6D189680.

[141] Conversation ID: 19F787971339DF839376D60F11B8B33C1F28303BCFFF8C1EC641176A2654E78F.

[redacted]

$\frac{PA_O - PA_E}{1 - PA_E}$, where PA_O is the observed agreement rate, and PA_E is the expected agreement rate that would occur by chance, based on the observed frequency of each category. The expected agreement rate, PA_E, is calculated as: $\frac{1}{n^2}\sum pm_i$, where n is the number of conversations coded by the pair of coders and pm_i is the product of the marginals (*i.e.*, the total number of agreements) for each coder for category i.[142] For example, if two coders were coding 100 conversations into two categories, A and B, and each coded 80 percent of conversations as A, and 20 percent of conversations as B, they would be expected to come to an agreement 68 percent of the time just by chance. Cohen's Kappa therefore describes how much better (or worse) is the observed outcome is relative to what might be expected by chance. In this hypothetical example, observing agreement 65 percent of the time would therefore be worse than chance, while observing agreement 70 percent of the time would be better than chance. Several other factors impact Cohen's Kappa, including the number of categories and the relative imbalance across categories. Note that if one category dominates, Cohen's Kappa could be low even with high agreement because chance agreement becomes high—which makes *exceeding* chance agreement more difficult. If one category dominates, Cohen's Kappa may understate reliability and it is important to consider it in context of other metrics such as raw agreement rate and category distribution.[143]

116. Cohen's Kappa typically falls between 0 (agreement at chance level) and 1 (perfect agreement), with negative values indicating agreement less than the base chance level and positive values indicating agreement better than the base chance level. Cohen's Kappa for my user intent content analysis was 0.67.

---

[142] Neuendorf, pp. 150, 153-155.

[143] Neuendorf, p. 151 ("Both pi and kappa have been criticized as being overly conservative, giving credit only to agreement beyond chance, a tough challenge in the case of extreme distributions (Perrault & Leigh, 1989; Potter & Levine-Donnerstein, 1999). […] This undesirable influence of prevalence of certain coding categories has been identified as an important disadvantage, and alternatives have been presented in the literature (Feinstein & Cicchetti, 1990), although none have yet won popular support.").

[redacted]

117. There is no single definition of what constitutes an acceptable level of intercoder reliability within a quantitative content analysis.[144] However, a 70 percent agreement rate is a common benchmark for which a content analysis is deemed reliable in the literature, and statistics that account for chance, such as Cohen's Kappa, are afforded more leniency.[145] Values of Cohen's Kappa greater than 0.60 are widely considered to indicate "substantial" strength of agreement among coders.[146] The agreement rate of 72.5 percent and Cohen's Kappa of 0.67 that I find in my quantitative content analysis of user intent exceed these benchmarks.

118. Additionally, my intercoder reliability measures are consistent with agreement rates and Cohen's Kappa values found in the literature on LLM usage. For example, Chatterji et al. (2025) find a human–human Kappa of 0.47 and a Human–LLM Kappa of 0.56, and Shah et al. (2025) find a Kappa of 0.76 between two human coders and 0.72 between human and GPT-4.[147] To further my confidence that coders reliably applied the user intent and subject matter categories to the conversations in my Main Analytical Example, I personally reviewed more than 100 random conversations classified as each user intent and

---

144 Neuendorf, p. 143 ("What constitutes an acceptable *level* of intercoder reliability for each variable is open to debate. Unfortunately, common standards are not in place[.]").

145 Neuendorf, p. 143 ("Frey, Botan, and Kreps (2000) declare a 70% agreement to be considered reliable. […] The following criteria have been proposed for Cohen's Kappa: .75+ indicating excellent agreement beyond chance; .40 to .75, fair to good agreement beyond chance; and below .40, poor agreement beyond chance (Banerjee et al., 1999). […] In general, the beyond-chance statistics, such as Scott's pi and Cohen's Kappa, are afforded a more liberal criterion.").

146 *See, e.g.*, Landis, J. Richard and Gary G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, Vol. 33, No. 1, March 1977, pp. 159 - 174 at p. 165, showing that Kappas less than 0 have "Poor" strength of agreement, Kappas between 0 and 0.20 have "Slight" strength of agreement, Kappas between 0.21 and 0.40 have "Fair" strength of agreement, Kappas between 0.41 and 0.60 have "Moderate" strength of agreement, Kappas between 0.61 and 0.80 have "Substantial" strength of agreement, and Kappas between 0.81 and 1.00 have "Almost Perfect" strength of agreement.

147 Chatterji, Aaron, et al., "How People Use ChatGPT," *National Bureau of Economic Research*, September 2025, p. 53; Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29 at pp. 4-5. Notably, Shah et al. (2025) involved two coders who coded a total of 124 conversations. My content analysis involved greater number of categories and was conducted at a much larger scale, which required more coders, who have individualized nuances to their categorizations, and a multi-week coding process that covered tens of thousands of conversations. This would increase the cognitive load on the coders and possibly lead to more differences in classifications than a short-term analysis with fewer categories.

[redacted]

subject matter category (or all conversations in categories with fewer than 100 total conversations).

119. For my secondary analysis of subject matter, the pairwise agreement rate among conversations coded into information retrieval was 69.0 percent, and the majority agreement rate was 92.2 percent. Cohen's Kappa for subject matter was 0.66.

120. I also examined intercoder reliability specifically at the boundary between "Information Retrieval – Current Events" and "Information Retrieval – Other." That boundary-specific analysis does not change my overarching conclusion that the category of "Information Retrieval – Current Events" is small relative to "Information Retrieval – Other" and can therefore account for only a small share of the conversations in the Copilot Conversations Sample even under the most conservative assumptions.

121. First, calculating a per-pair Cohen's Kappa along category lines is not a common approach when studying reliability of classification performed by multiple coders into a taxonomy with several categories. Of the five papers cited in my opening report that reported a Cohen's Kappa, none reported per-pair Cohen's Kappa along category lines.[148]

122. Second, Cohen's Kappa is well known to understate agreement when categories have highly disproportionate prevalence rates,[149] which is precisely the situation

---

[148] *See* Chatterji, Aaron, et al., "How People Use ChatGPT," *National Bureau of Economic Research*, September 2025, p. 53 (reported the mean pairwise human-human Cohen's Kappa and the Cohen's Kappa between the model and human plurality label). Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29, at 34:10, 34:14 (reported the Cohen's Kappa between two human annotators, as well as the Cohen's Kappa between Humans and different LLMs). Ammari, Tawfiq, et al., "How Students (Really) Use ChatGPT: Uncovering Experiences Among Undergraduate Students," *arXiv*, January 19, 2026, p. 9 (reported Cohen's Kappa between human labels and GPT labels by category). Tomlinson, Kiran, et al., "Working with AI: Measuring the Applicability of Generative AI to Occupations," *arXiv*, December 22, 2025, pp. 14, 19 (reported Cohen's Kappa between the three pairs of annotators, and the Cohen's Kappa between two human annotators as well as between each human and the GPT model). Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," March 19, 2024, working paper, p. 18 (reported a Cohen's Kappa between prompt output and the authors of the paper).

[149] Neuendorf, p. 151.

[redacted]

here, where "Information Retrieval – Other" accounts for 36.2% of conversations while "Information Retrieval – Current Events" accounts for only 1.3%. As I explained in paragraph [137] above, where one category dominates, Cohen's Kappa may understate reliability, and it is important to consider it together with other metrics such as the raw agreement rate and category distribution. In a situation where one category is much larger than another, the agreement rate by chance would be quite high. Achieving high Cohen's Kappa would then require the coders to achieve an exceedingly high agreement rate. For example, consider two resume reviewers assessing candidates for interviews. If they expect to interview only 3 percent of all potential candidates, the two reviewers would be expected to agree over 94 percent of the time even by chance – since most of the time, they will decide not to interview a candidate. In such instances, a low Kappa does not necessarily indicate reviewers' inability to assess candidates. A high agreement rate like 95 percent would only yield Cohen's Kappa of 0.17. So even if the two reviewers agree whether or not to interview 95 of 100 candidates, and only disagree in 5 instances, Cohen's Kappa would be low. Even if they agreed in 97 instances, Kappa would only be 0.5; in order to achieve Kappa of 0.83, they would need to agree in all but a single instance. This illustrates the degree of difficulty in achieving high Kappa in cases where the proportion of items in the two categories is very unequal.

123. Third, computing Cohen's Kappa at this boundary, I find the boundary-specific kappa to be 0.54. As cited above, Neuendorf indicates that a Cohen's Kappa between 0.40 to 0.75 indicates "fair to good agreement beyond chance" and Landis and Koch indicate that a Kappa between 0.41 and 0.60 has "'Moderate' strength of agreement."[150] To reiterate, the overall Cohen's Kappa for my analysis is 0.67, which is considered to be "substantial" strength of agreement.[151]

124. In fact, I find the agreement rate at the boundary between "Information Retrieval – Other" and "Information Retrieval – Current Events" is very high, at 95.9

[150] *See supra* n.145-46.

[151] *See supra* ¶¶ 116-117.

[redacted]

percent. Regardless of the Kappa, the vast majority of the time, the coders agree whether a conversation should be classified as "Information Retrieval – Other" or "Information Retrieval – Current Events." This is notable because coders were not provided with any prior expectation on the relative prevalence of the two Information Retrieval categories and were blind to the purpose of the analysis.

125. Finally, and perhaps most importantly, even if it were correct that reviewers were not able to perfectly determine the boundary between "Information Retrieval – Current Events" and "Information Retrieval – Other," as is often the case in content analysis,[152] the share of "Information Retrieval – Current Events" conversations is still small even under conservative assumptions. In total, there were only 1,149 conversations, or 2.67 percent where *any* reviewer selected "Information Retrieval – Current Events."[153] Therefore, even under this assumption "Information Retrieval – Current Events" would not be a primary use of Copilot in the Copilot Conversations Sample.

2. Additional Analyses Confirm That the Presence of Conversations with Unclear Principal Use or Unclear Principal Subject Matter Does Not Alter My Opinions

126. For some conversations in my Main Analytical Sample, three coders all selected different user intents and/or subject matters. As a result, 5.6 percent of conversations fall under "Unclear Principal Use," and 2.9 percent of conversations fall under "Unclear Principal Subject Matter." These instances represent conversations where coders observed multiple plausible intents or subject matters. Across the literature—and consistent with findings in the Amir

---

[152] As I noted in Section III, literature on content analysis recognizes that categorizing LLM conversations into user intent categories is a complex process and the inherent ambiguity and fluidity of human language will lead to some degree of subjectivity. *See, e.g.,* Shelby, Renée, Fernando Diaz, and Vinodkumar Prabhakaran, "Taxonomy of User Needs and Actions," *arXiv*, October 10, 2025, p. 10 ("applying labels to fluid human language carries inherent ambiguity" and, while their taxonomy "helps navigate this complexity, the boundaries between certain categories, such as [...] exemplar requests vs. [...] creative content generation can be subjective and require careful annotator training.).

[153] Calculated as 1,149 / 43,000 (the number of conversations classified by coders in the Bail Content Analysis) = 2.67%.

[redacted]

Survey[154]—multiple studies recognize that a single conversation may correspond to more than one intent or the intent may be unclear. Zheng et al. (2024), for example, note that "a message might have multiple labels assigned to it."[155] Ouyang et al. (2023) similarly allow the LLM to select multiple "task types" for a single conversation.[156] In a related approach, Tomlinson et al. (2025) permit conversations to be categorized under multiple intermediate work activities (IWAs).[157] Shelby et al. (2025) use iterative qualitative analysis and acknowledge that "applying labels to fluid human language carries inherent ambiguity" and, while their taxonomy "helps navigate this complexity, the boundaries between certain categories, such as [...] *exemplar requests* vs. [...] *creative content generation* can be subjective and require careful annotator training."[158] Earlier work in web search classification by Jansen et al. (2008) also highlights the complexity of classification, finding that 25 percent of web search queries exhibit "user intent [that] is vague or multi-faceted."[159] Although they ultimately assign each query to a single category, they acknowledge this as a limitation, noting that queries "may have multiple possible intents," and that "approximately 75

---

154 Amir Declaration, Exhibit 3, note 2 ("The total percentages across categories do not sum to 100% because a response can be categorized into multiple categories.").

155 Zheng, Lianmin, et al., "LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset," *ICLR 2024*, March 10, 2024, p. 5.

156 Ouyang, Siru, et al., "The Shifted and the Overlooked: A Task-Oriented Investigation of User-GPT Interactions," *Association for Computional Linguistics*, December 2023, pp. 2375 - 2393 at p. 2377 ("Generate task types. Note that there could be multiple types.").

157 Tomlinson, Kiran, et al., "Working with AI: Measuring the Applicability of Generative AI to Occupations," *arXiv*, December 22, 2025, p. 8 ("Our analysis focuses on intermediate work activities (IWAs), which map to multiple occupations through tasks.").

158 Shelby, Renée, Fernando Diaz, and Vinodkumar Prabhakaran, "Taxonomy of User Needs and Actions," *arXiv*, October 10, 2025, p. 10 (emphasis in original).

159 Jansen, Bernard J., Danielle L. Booth, and Amanda Spink, "Determining the Informational, Navigational, and Transactional Intent of Web Queries," *Information Processing & Management*, Vol. 44, No. 3, May 2008, pp. 1251 - 1266 at p. 1251.

[redacted]

[percent] of queries can be classified into a single category of intent (i.e., informational, navigational, or transactional) with a high degree of certainty."[160]

127. In this section, I present additional analyses of the conversations that fall under "Unclear Principal Use" or "Unclear Principal Subject Matter." I find that, consistent with the literature, the "Unclear Principal Use" and "Unclear Principal Subject Matter," conversations generally represent ambiguity in the conversations rather than systematic category confusion. This result, therefore, does not change my conclusion that Copilot is used for a wide variety of purposes across a breadth of subject matters.

a. Unclear Principal Use

128. First, I examine the 5.6 percent of conversations that fall under "Unclear Principal Use." An example of a conversation that falls under "Unclear Principal Use" is shown in **Exhibit 46** below.

[160] Jansen, Bernard J., Danielle L. Booth, and Amanda Spink, "Determining the Informational, Navigational, and Transactional Intent of Web Queries," *Information Processing & Management*, Vol. 44, No. 3, May 2008, pp. 1251 - 1266 at p. 1263.

[redacted]

Exhibit 46
Unclear Principal Use Example[161]



129. The conversation in the example above was classified as "Writing – Content Creation," "Information Retrieval – Other" with subject matter "Travel / Geography" and "Brainstorm / Idea Generation." The user's single prompt contains some aspects of each of these categories: the user is asking Copilot to describe a place, to which Copilot responds by writing a short paragraph, but at the same time, the user is asking about a "place on earth" with particular attributes, which conveys characteristics of both brainstorming and information retrieval.

130. **Exhibit 47** below summarizes the frequency with which unique sets of categories fall under "Unclear Principal Use."

[161] Conversation ID: 2F20C19236924B09ED2D4D07F9270AAB5C630B49B3241C692DFCDC2BC9B7A551.

[redacted]

Exhibit 47
Unclear Principal Use

Unique Intent Category Sets	N	Percent of Total Conversations
Advice / Recommendation / Seeking Opinions | Information Retrieval - Other | Practical Guidance / Learning	346	0.8%
Information Retrieval - Other | Practical Guidance / Learning | Problem Solving / Reasoning	119	0.3%
Information Retrieval - Other | Practical Guidance / Learning | Unable to Classify - Redacted	112	0.3%
Advice / Recommendation / Seeking Opinions | Brainstorm / Idea Generation | Information Retrieval - Other	79	0.2%
Advice / Recommendation / Seeking Opinions | Information Retrieval - Other | Unable to Classify - Redacted	68	0.2%
Information Retrieval - Other | Practical Guidance / Learning | Unable to Classify - Insufficient Content	51	0.1%
Information Retrieval - Other | Unable to Classify - Insufficient Content | Unable to Classify - Redacted	50	0.1%
Information Retrieval - Other | Navigation to Another Webpage | Unable to Classify - Insufficient Content	50	0.1%
Computer Code Generation, Revision, or Debugging | Information Retrieval - Other | Practical Guidance / Learning	48	0.1%
Information Retrieval - Current Events | Information Retrieval - Other | Practical Guidance / Learning	47	0.1%
Information Retrieval - Other | Problem Solving / Reasoning | Unable to Classify - Insufficient Content	45	0.1%
Information Retrieval - Current Events | Information Retrieval - Other | Unable to Classify - Redacted	44	0.1%
Advice / Recommendation / Seeking Opinions | Information Retrieval - Other | Problem Solving / Reasoning	40	0.1%
Advice / Recommendation / Seeking Opinions | Information Retrieval - Other | Unable to Classify - Insufficient Content	39	0.1%
Information Retrieval - Other | Practical Guidance / Learning | Writing - Content Creation	38	0.1%
Brainstorm / Idea Generation | Information Retrieval - Other | Practical Guidance / Learning	33	0.1%
Information Retrieval - Other | Problem Solving / Reasoning | Unable to Classify - Redacted	30	0.1%
Information Retrieval - Other | Writing - Content Creation | Writing - Edit User-Provided Text	27	0.1%
Other Triplets	1,161	2.7%
Total	**2,427**	**5.6%**

131. The most common set of categories with Unclear Principal Use was "Advice / Recommendation / Seeking Opinions," "Information Retrieval – Other," and "Practical Guidance / Learning," with the set only accounting for 0.8 percent of total conversations. The set of conversations that fall under "Unclear Principal Use" are distributed among a wide variety of distinct intent category sets, with 312 distinct intent category sets each accounting for less than one percent of the Unclear Principal Use conversations, making up 47.8 percent of the total Unclear Principal Use conversations.

[redacted]

132. Additionally, 39.6 percent of Unclear Principal Use conversations contain at least one coder selecting "Unable to Classify – Insufficient Content," Unable to Classify – Redacted," or "Unable to Classify – Requires Translation." This suggests that the coders were often deciding whether they had enough information to classify the conversation, as opposed to deliberating between categories of user intent. Together, this suggests that disagreement reflects individualized coder interpretations and ambiguity in the conversations rather than systematic category confusion. To further my confidence in these conclusions, I personally analyzed a ten percent sample of conversations classified as "Unclear Principal Use." Together, my analyses of conversations that fall under "Unclear Principal Use" do not change my conclusion that Copilot is used for a wide variety of purposes.

b. Unclear Principal Subject Matter

133. Second, I examine the 2.9 percent of conversations that fall under "Unclear Principal Subject Matter." My analysis highlights that a substantial portion of conversations may relate to more than one subject matter, further emphasizing the breadth of subject matters for which Copilot users seek to retrieve information. **Exhibit 48** below presents an example of Unclear Principle Subject conversation for "Information Retrieval – Current Events."

[redacted]

Exhibit 48
Information Retrieval – Current Events
Unclear Principal Subject Matter[162]

[redacted]

134. In this example, Copilot is asked to retrieve information about local events related to psychology in San Antonio. As such, the three coders selected the subject matters "Health / Fitness," "Other Local, Regional, National, or International Events," and "Travel / Geography." **Exhibit 49** below presents an example of Unclear Principle Subject conversation for "Information Retrieval – Other."

[162] Conversation ID: 7A7399E323028389409E84808DC896AD4523066EBAB3292BB5086BD957979329.

[redacted]

Exhibit 49
Information Retrieval – Other
Unclear Principal Subject Matter[163]



135. In this example, Copilot is asked to retrieve information about the legal status of a non-profit organization related to religion. The three coders selected the subject matters "Law / Tax," "Business / Finance" and "Religion."

136. **Exhibit 50** below summarizes the frequency with which unique sets of categories fall under "Unclear Principal Subject Matter" for conversations coded as "Information Retrieval – Other."

[163] Conversation ID: 1BA15CE0844DB8B2E0AE19ADE9D515997871E5EACCD78BAE482E2D8AA4AB1D48.

[redacted]

Exhibit 50
Unclear Principal Subject Matter
Information Retrieval – Other

Unique Subject Matter Category Sets	N	Percent of Total Conversations
Business / Finance | Consumer Products and Services | Other	67	0.16%
Business / Finance | Consumer Products and Services | Travel / Geography	37	0.09%
Consumer Products and Services | Other | Travel / Geography	36	0.08%
Consumer Products and Services | Other | Science / Math	34	0.08%
Health / Fitness | Other | Science / Math	32	0.07%
Computing Devices / Software | Consumer Products and Services | Other	30	0.07%
Business / Finance | Law / Tax | Other	24	0.06%
Other | Other Local, Regional, National, or International Events | Travel / Geography	23	0.05%
Law / Tax | Other | Travel / Geography	21	0.05%
Consumer Products and Services | Entertainment (Celebrities, Movies, Concerts, Games) | Other	21	0.05%
Arts / Literature / Language | Entertainment (Celebrities, Movies, Concerts, Games) | Other	19	0.04%
Business / Finance | History / Social Sciences / Politics | Other	19	0.04%
Business / Finance | Other | Travel / Geography	18	0.04%
Entertainment (Celebrities, Movies, Concerts, Games) | Other | Other Local, Regional, National, or International Events	17	0.04%
Health / Fitness | Other | Personal Growth / Family	16	0.04%
Business / Finance | Other | Personal Growth / Family	16	0.04%
Arts / Literature / Language | History / Social Sciences / Politics | Other	15	0.03%
Computing Devices / Software | Entertainment (Celebrities, Movies, Concerts, Games) | Other	15	0.03%
History / Social Sciences / Politics | Other | Other Local, Regional, National, or International Events	15	0.03%
History / Social Sciences / Politics | Other | Personal Growth / Family	14	0.03%
Business / Finance | Computing Devices / Software | Consumer Products and Services	14	0.03%
Business / Finance | Computing Devices / Software | Other	14	0.03%
Computing Devices / Software | Other | Science / Math	13	0.03%
Consumer Products and Services | Health / Fitness | Other	13	0.03%
History / Social Sciences / Politics | Law / Tax | Other	13	0.03%
Other | Personal Growth / Family | Travel / Geography	13	0.03%
Other Triplets	660	1.53%
Total	**1,229**	**2.9%**

137. As with Unclear Principal Use, the set of conversations that fall under Unclear Principal Subject Matter are distributed among a wide variety of subject matters. Notably, 66.2 percent of "Information Retrieval – Other" Unclear Principal Subject Matter conversations involved the "Other" subject matter category. As described in **Section VI.B.** above, the "Other" subject matter category may contain numerous subject matters not captured in the provided subject matter categories.

[redacted]

138. **Exhibit 51** below summarizes the frequency with which unique sets of categories fall under "Unclear Principal Subject Matter" for conversations coded as "Information Retrieval – Current Events."

Exhibit 51
Unclear Principal Subject Matter
Information Retrieval – Current Events

Unique Subject Matter Category Sets	N	Percent of Total Conversations
Other | Other Local, Regional, National, or International Events | Travel / Geography	6	0.014%
Other | Science / Math | Travel / Geography	3	0.007%
History / Social Sciences / Politics | Other | Other Local, Regional, National, or International Events	3	0.007%
Other Local, Regional, National, or International Events | Science / Math | Travel /	2	0.005%
History / Social Sciences / Politics | Other Local, Regional, National, or International Events | Travel / Geography	2	0.005%
Business / Finance | Other | Other Local, Regional, National, or International Events	1	0.002%
Business / Finance | Entertainment (Celebrities, Movies, Concerts, Games) | Other	1	0.002%
Entertainment (Celebrities, Movies, Concerts, Games) | Other | Other Local, Regional,	1	0.002%
Other | Other Local, Regional, National, or International Events | Sports	1	0.002%
Other | Other Local, Regional, National, or International Events | Science / Math	1	0.002%
Health / Fitness | Other Local, Regional, National, or International Events | Travel /	1	0.002%
Entertainment (Celebrities, Movies, Concerts, Games) | Other Local, Regional, National, or International Events | Sports	1	0.002%
Entertainment (Celebrities, Movies, Concerts, Games) | Fashion / Beauty | Other Local, Regional, National, or International Events	1	0.002%
Entertainment (Celebrities, Movies, Concerts, Games) | Law / Tax | Other Local, Regional, National, or International Events	1	0.002%
History / Social Sciences / Politics | Law / Tax | Other Local, Regional, National, or International Events	1	0.002%
Entertainment (Celebrities, Movies, Concerts, Games) | Other Local, Regional, National, or International Events | Travel / Geography	1	0.002%
Computing Devices / Software | Consumer Products and Services | Other Local, Regional, National, or International Events	1	0.002%
Business / Finance | Computing Devices / Software | Health / Fitness	1	0.002%
Total	**29**	**0.07%**

139. As with "Information Retrieval – Other," the conversations are distributed among a wide variety of subject matters, with 58.6 percent of conversations involving the "Other" subject matter category. Again, this suggests that lack of agreement reflects individualized coder interpretations of categories and the variety of potential subject matters appearing in conversations beyond those provided to coders, rather than systematic category confusion. To further my

[redacted]

confidence in these conclusions, I personally analyzed a ten percent sample of conversations classified as "Unclear Principal Use." Together, my analyses of conversations that fall under "Unclear Principal Subject Matter" do not change my conclusion that Copilot is used to retrieve information across a breadth of subject matters.

3. Additional Analyses Confirm That the Presence of Non-English Conversations Does Not Alter My Opinions

140. As discussed above, 10.8 percent of conversations in my Main Analytical Sample were classified as "Unable to Classify – Requires Translation." Consistent with existing literature on LLM usage, I did not undertake further classification of this subset of conversations. For example, Ouyang et al. (2023), Shah et al. (2025), and Zheng et al. (2024), each conduct analyses focused on English language conversations.[164] Studies that have considered LLM conversations across multiple languages have not identified categories of use that would be unique to foreign languages [165] Additionally, I understand from counsel that the only copyrighted works at issue are U.S. works that were authored in the English language, so I would expect any potential infringing use to appear in English conversations if such use exists.

---

[164] Ouyang, Siru, et al., "The Shifted and the Overlooked: A Task-Oriented Investigation of User-GPT Interactions," *Association for Computional Linguistics*, December 2023, pp. 2375 - 2393 at p. 2378 ("Since our investigation focuses on the English language and ShareGPT contains only English samples, datasets collected from Huggingface also follow this paradigm."); Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29 at p. 3 ("We ensured that these conversations were in English, however, some of the user requests were interspersed with non-English words. We do not believe this impacted any text processing by GPT-4 for our purposes."); Zheng, Lianmin, et al., "LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset," *ICLR 2024*, March 10, 2024, p. 3 ("We conduct a topic distribution analysis on user prompts by applying a clustering algorithm. From 100K randomly sampled English conversations, we extract user prompts, which include both the initial and follow-up turns.").

[165] *See, e.g.*, Yan, Yueru, et al., "SHARECHAT: A Dataset of Chatbot Conversations in the Wild," *arXiv*, January 27, 2026. Yan et al. highlight that their data includes information for over 100 languages, but does not identify specific user intents for specific languages or indicate that additional languages necessitate additional user intent categories. Chatterji, Aaron, et al., "How People Use ChatGPT," *National Bureau of Economic Research*, September 2025, pp. 44-48. Similarly, Chatterji et al. appear to include non-English conversations in their data sample but do not indicate that the user intents differ across languages or that the additional languages necessitate additional user intent categories.

[redacted]

141. I personally analyzed a ten percent sample of conversations classified as "Unable to Classify – Requires Translation" that were translated into English via machine-translation.[166] I did so to further my confidence that my conclusions would not be altered absent a classification of all non-English conversations. In my review, I observed that conversations coded as "Unable to Classify – Requires Translation" appear to span similar uses as I describe above for English-language conversations. For example, I observed conversations that involved the user generating computer code, seeking to understand a topic of interest, and retrieving information about a variety of topics, such as business, politics, and entertainment. I did not observe conversations that could not fit into my user intent taxonomy.

142. Additionally, I reviewed the distribution of user intent categories in my Main Analytical Sample excluding the category "Unable to Classify – Requires Translation."[167] Excluding this category, my pairwise agreement rate for my analysis of user intent categories was 69.8 percent, and Cohen's Kappa was 0.62. The presence of this category of conversations does not change my conclusion that Copilot is used for a wide variety of purposes across a breadth of subject matters.

---

[166] Azure Translator was used to translate conversations identified by coders as "Unable to Classify – Requires Translation." Specifically, the Text Translation service (API version 3.0) was used, which is based on neural machine translation rather than an LLM. Azure Translator is a widely adopted tool that is "used by thousands of businesses worldwide for language translation and other language-related operations." *See* "What is Azure Translator in Foundry Tools?," *Microsoft*, https://learn.microsoft.com/en-us/azure/ai-services/translator/text-translation/overview, accessed April 22, 2026; "Azure Translator in Foundry Tools," *Microsoft*, https://azure.microsoft.com/en-us/products/ai-foundry/tools/translator, accessed April 19, 2026. Studies published by Microsoft demonstrate that its machine translation systems have achieved strong performance on multilingual benchmarks such as WMT. *See* Hassan, Hany, et al., "Achieving Human Parity on Automatic Chinese to English News Translation," *Microsoft AI & Research*, June 29, 2018; Yang, Jian, et al., "Multilingual Machine Translation Systems from Microsoft for WMT21 Shared Task," *Proceedings of the Sixth Conference on Machine Translation (WMT)*, November 2021, pp. 446-455.

[167] *See* **Appendix G.**

[redacted]

I declare under penalty of perjury that the foregoing is true and correct.

Executed on this 1st day of September at Chapel Hill, North Carolina.

/s/

Christopher A. Bail, Ph.D.

Appendix A

Christopher A. Bail

Duke University
225 Gross Hall
Durham, NC 27708

Email: christopher.bail@duke.edu
Homepage: chrisbail.net
X/Twitter: @chris_bail
Github: github.com/cbail

Academic Employment

Professor of Sociology, Computer Science, Political Science, and Public Policy, Duke University, 2019-

- *Director, Society-Centered AI Initiative at Duke, 2024-*
- *Founding Director, Summer Institutes in Computational Social Science, 2017-*
- *Founding Director, Duke Polarization Lab, 2018-*
- *Faculty, Masters in Interdisciplinary Data Science, 2017-2020*
- *Science Core Director, Duke Population Research Institute, 2018-2020*
- *Affiliate, Duke Network Analysis Center, 2015-2020*

Visiting Professor, Fondation Nationale des Sciences Politiques (Paris), 2024

Distinguished Visiting Scholar, Stanford University Graduate School of Business, 2023

Visiting Professor, University of Amsterdam, 2023

Visiting Professor, Venice International University, 2019-2020

Douglas and Ellen Lowey Associate Professor of Sociology and Public Policy, Duke University, 2016-2019

Assistant Professor of Sociology, Duke University, 2015-2016

Assistant Professor of Sociology, University of North Carolina at Chapel Hill, 2013-2015

Robert Wood Johnson Foundation Scholar, University of Michigan, 2011-2013

Education

Ph.D. Sociology, Harvard University, 2003-2011

Visiting Scholar, London School of Economics and Political Science, 2008

Visiting Scholar, Fondation Nationale des Sciences Politiques (Paris), 2005

B.A. Government and French (Summa Cum Laude), Bowdoin College, 1998-2002

Distinctions

Global Trends Award (for "Simulating Social Media with Large Language Models"), 2024

Marschak Lecture, University of California, Los Angeles, 2024

Finalist, OpenAI Hackathon, South Park Commons, 2024

Nominated for Directorship of National Science Foundation's Social, Behavioral, and Economic Sciences Directorate, 2023

AISSR Distinguished Lecture, University of Amsterdam, 2023

John Armstrong Lecture in Media and Democracy, Furman College, 2023

Elected Member, Sigma Xi Scientific Research Society, 2022

Science Breakthrough of the Year Winner, Falling Walls Foundation, 2021

One of the Five Best Non-Fiction Books of the Year (for *Breaking the Social Media Prism*), Five Books, 2021

Distinguished Lecturer, National Science Foundation's Social, Behavioral, and Economic Sciences Directorate, 2021

Nominated for the Next Big Idea Club (for *Breaking the Social Media Prism*), 2021

Top 5% of Teachers at Duke, 2020

Paul Lazarsfeld Award, Political Communication Section, American Political Science Association (for "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment on Social Media), 2019

Thomas Langford Lectureship Award, (for a Duke faculty member who embodies the highest values of scholarship, teaching, and collegiality), 2019

Guggenheim Fellow, 2018

Carnegie Fellow, 2017-2018

Elected Member, Sociological Research Association, 2018

Best Article Award, Communication, Information, Technologies, and Media Sociology Section, American Sociological Association (for "Channeling Hearts and Minds: Cognitive-Emotional Currents and Public Deliberation on Social Media"), 2018

Honorable Mention, Distinguished Contribution to Political Sociology Award, Political Sociology Section, American Sociological Association (for "Channeling Hearts and Minds: Cognitive-Emotional Currents and Public Deliberation on Social Media"), 2018

Outstanding Book Award, Association for Research on Non-Profit Organizations and Voluntary Action (for *Terrified: How Anti-Muslim Fringe Organizations Became Mainstream*), 2016

Distinguished Book Award, Religion Section, American Sociological Association (for *Terrified: How Anti-Muslim Fringe Organizations Became Mainstream*), 2016

Honorable Mention, Charles Tilly Award for Best Book, Collective Behavior and Social Movements Section, American Sociological Association (for *Terrified: How Anti-Muslim Fringe Organizations Became Mainstream*), 2016

Junior Theorist Award (for scholar less than 8 years post-PhD), Theory Section, American Sociological Association, 2015

Distinguished Contribution to Scholarship Award, Political Sociology Section, American Sociological Association (for "The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam since the September 11th Attacks"), 2012

Best Published Article Award, Communication and Information Technologies Section, American Sociological Association (for "The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam since the September 11th Attacks"), 2012

Honorable Mention, Best Article Award, Society for the Scientific Study of Religion (for "The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam since the September 11th Attacks"), 2012

Distinction on Oral Examination, Department of Sociology, Harvard University, 2006

Certificate of Excellence in Teaching, Derek Bok Center for Teaching, Harvard University, 2006

Distinction on PhD Qualifying Paper, Department of Sociology, Harvard University, 2005

Certificate of Excellence in Teaching, Derek Bok Center for Teaching, Harvard University, 2005

Philo Sherman Bennet Award for Best Senior Thesis, Department of Government, Bowdoin College, 2002

Phi Beta Kappa, Bowdoin College, 2002

Grants and Fellowships

External Research Support

Laude Institute Moonshot Program. "Orchestrate+: Improving Social Media with Agentic AI " $100,000, 2026

John Templeton Foundation. "Can AI Agents Increase Open-Mindedness and Facilitate Productive Dialogue about Social Problems?" $1,744,650, 2026-2029

Templeton World Charity Foundation. "Tech for Open Minds: Training and Research Institutes to Incubate New Solutions to Polarization" $1,313,397, 2025-2028

Russell Sage Foundation. "Enhancing the infrastructure of the American Voices Project, an experimental public-use platform for collecting qualitative data." (Co-Investigator with David Grusky, Kathy Edin, and Corey Fields) $200,000, 2025

United States Army Research Office. "Randomized Controlled Trials to Examine the Impact of Generative AI on Public Trust" (Co-Investigator with Alexander Volfovsky and Sunshine Hillygus), $4,060,000, 2024-2027

National Science Foundation, "A Cross-National Study to Promote Democratic Dialogue on Social Media using Generative Artificial Intelligence" (Principal Investigator $200,000), 2024

Templeton World Charity Foundation, "Training the Next Generation of Polarization Scholars" (Principal Investigator $260,000), 2024

Project Liberty Fellow, "Redesigning Social Media to Incentivize Prosocial Behavior," $70,000, 2024

National Science Foundation, "The Computational Social Science Network (CS2N)" (Principal Investigator with Duncan Watts $425,000), 2023

John F. Templeton Foundation, "Can Social Media Promote Intellectual Humility?" (Principal Investigator with Sunshine Hillygus and Alexander Volfovsky $1,856,673), 2023-2026

Google Jigsaw, "Social Media and Political Polarization" (Principal Investigator $200,000), 2022

Twitter Health, "Incentivizing Positive Behaviors on Social Media" (Principal Investigator $75,000), 2022

Russell Sage Foundation "The Summer Institute in Computational Social Science, 2022" (Principal Investigator $175,000), 2022

Social Science Research Council "Summer Institutes in Computational Social Science, 2022" (Principal Investigator $40,000, 2022

Alfred P. Sloan Foundation "The Summer Institutes in Computational Social Science, 2020-2023" 2019 (Principal Investigator with Matthew J. Salganik $496,004)

Russell Sage Foundation "The Summer Institute in Computational Social Science, 2021" (Co-Principal Investigator with Matthew J. Salganik $175,000)

Russell Sage Foundation "The Summer Institute in Computational Social Science, 2020" (Co-Principal Investigator with Matthew J. Salganik $175,000)

Facebook Research Awards, "The Summer Institutes in Computational Social Science, 2020" (Co-Principal Investigator with Matthew J. Salganik $100,000)

Social Science Research Council, "The Summer Institutes in Computational Social Science, (Co-Principal Investigator with Matthew J. Salganik 2020 $30,000)

Facebook Foundational Research Awards, "Ideological Distance, Identity Cues, and Affective Political Polarization," (Principal Investigator with Alexander Volfovsky, $150,000)

National Science Foundation, "Reproducible Visual Analysis of Multivariate Networks" 2018-2021 (Consultant, with Miriah Meyer, Alexander Lex, Bryan Jones, Luke Harmon, Jeff Baumes, and Curtis Lisle, $1,899,694)

Sloan Foundation, "Expanding the Summer Institutes in Computational Social Science Worldwide", 2018 (Principal Investigator with Matthew J. Salganik $385,631)

National Science Foundation, "A Mixed-Method Field Experiment Designed to Disrupt Social Media Echo Chambers", 2018 (Principal Investigator $224,000)

Russell Sage Foundation, "Summer Institute in Computational Social Science, 2018." (Co-Principal Investigator, with Matthew J. Salganik $150,000)

Russell Sage Foundation, "Political Polarization, Social Media Echo Chambers, and American Public Opinion about Inequality, Race, and Immigration." 2017 (Co-Principal Investigator, with Mary Beth Fallin Hunzaker $49,299)

National Institutes of Health, "Duke Population Research Institute P2C Grant," 2017-2019 (Co-Principal Investigator, with M. Giovanna Merli and James Moody $1,031,829)

Carnegie Foundation, "Countering Violent Extremism on Social Media via Computational Social Science." 2017 (Principal Investigator $200,000)

Russell Sage Foundation, "Summer Institute in Computational Social Science, 2017." (Co-Principal Investigator, with Matthew J. Salganik $150,000)

National Science Foundation, "Civil Society Organizations, Social Media, and Public Attention" 2015 (Principal Investigator, $143,989)

Robert Wood Johnson Foundation, "Examining the role of culture in public understanding of Autism Spectrum Disorders," 2014 (Principal Investigator, $20,000)

Amazon Education Research Grant, "Using Computational Methods to Study the Viral Spread of Social Media Messages," 2014 (Principal Investigator, $10,000)

Google Maps Engine Grant, Google, 2013 (Paid in Computing Resources)

Dissertation Improvement Grant, National Science Foundation, 2009 ($10,000)

National Science Foundation Integrated Graduate Education and Research Training Dissertation Grant, 2004-2008 ($30,000)

German Marshall Fund Fellowship, U.S. Council for European Studies, 2005 ($6,000)

Dissertation Fellowship, Horowitz Foundation for Social Policy, 2008 ($5,000)

Internal Research Support

Duke Interdisciplinary Programs and Strategic Initiatives, "Integrated Graduate Education Collaborative in Society-Centered-AI" (Principal Investigator, $2,435,000)

Duke Polarization Lab, Provost's "Together Duke" Initiative, (Principal Investigator, $259,000)

Duke Bass Connections (with James Moody and Greg Appelbaum), "Scientific Social Network Analysis Project," (Co-Investigator, $25,000)

Duke Population Research Institute & Carolina Population Center, "Collaborative Summer in Residence Program Grant," ($5,000)

Junior Faculty Development Grant, University of North Carolina, Chapel Hill (Co-Investigator, $7,500)

Seed Grant, Odum Institute, University of North Carolina at Chapel Hill (with Kenneth Andrews, Neal Caren, Charles Kurzman, and Andrew Perrin), 2012-2013 ($13,000)

John L. Loeb Fellowship, Harvard Divinity School, 2009 ($10,000)

Dissertation Research Fellowship, Center for American Political Studies, Harvard University, 2008 ($21,830)

Mid-Dissertation Fellowship, Weatherhead Center for International Affairs, Harvard University, 2008 ($3,000)

European Network on Inequality Fellowship, Malcolm Weiner Center for Social Policy, Kennedy School of Government, Harvard University, 2005 ($6,500)

Publications

* denotes post-doc or graduate student co-author

Books

Christopher A. Bail. 2021. *Breaking the Social Media Prism: How to Make our Platforms less Polarizing* Princeton, NJ: Princeton University Press (link), (video)

- Reviewed in *Science (link)*, *The New York Times (link)*, *Le Monde (link)*, *Publisher's Weekly (link)*, *The Economist (link)*, *Library Journal (link)*, *Salon (link)*, *Kirkus Reviews (link)*, *American Journal of Psychology (link)*, *American Journal of Sociology (link)*, *Inside Higher Ed (link)*, *Times Higher Education (link)*, *La Vie des Idées (link)*, *Social Forces (link), Sociology (link), New Media and Society (link), Information, Community, and Society (link)*, *Revue française de science politique* (link), *Washington Examiner (link), Protocol (link), Life in the Gospel, (link), The Hindu (link), Merion West (link), FaithTech (link), Eternity News (link), Medien & Kommunikationswissenschaft (link) Philosophie Magazine, (link) New Atlantis (link) Les Echos (link) Times of UP (link) Portal für Politikwissenschaft (link) Réseaux (link)*

- Winner, Science Breakthrough of the Year, *Falling Walls Foundation*

- Academic Best Seller, *Library Journal*

- Best Books of the Year, *Five Books, Anxious Bench*

- Notable Book, *Behavioral Science Magazine*

- Nominated for *Next Big Idea Club*

- Documentary Film (link)

- Translations: Japanese, French, Chinese, Korean

Mohr, John and Christopher A. Bail, Margaret Frye, Omar Lizardo, Jennifer Lena, Terence McDonnell, Ann Mische, Iddo Tavory, Frederick Wherry. 2020. *Measuring Culture*. New York, NY: Columbia University Press (link)

- Reviewed in *Administrative Science Quarterly (link)*, *Social Forces (link)*, *Contemporary Sociology (link)*, *Organization Studies (link)*, *Canadian Journal of Sociology (link)*, *Tijdschrift Sociologie (link)*,

Bail, Christopher A. 2015. *Terrified: How Anti-Muslim Fringe Organizations Became Mainstream*. Princeton, NJ: Princeton University Press (link), (op-ed)

- Reviewed in *American Journal of Sociology (link), Administrative Science Quarterly (link), Boston Review (link), Contemporary Sociology (link), Choice (link) The American Journal of Islamic Social Sciences (link) , Contexts (link) , International Journal of Press/Politics (link), Political Science Quarterly (link), ReOrient (link), Salon (link)*

- Author Meets Critics Panel, 2016 American Sociological Association Annual Meeting

- Outstanding Book Award, Association for Research on Non-Profit Organizations and Voluntary Action, 2016

- Distinguished Book Award, Religion Section, American Sociological Association, 2016

- Honorable Mention, Charles Tilly Award for Best Book, Collective Behavior and Social Movements Section, American Sociological Association, 2016

Works in Progress

Rochford, Ben,* Michelle Qiu,* Cathy Lee,* Sunshine Hillygus, Alexander Volfovsky, Christopher A. Bail. 2026. "A Sociological Turing Test: Social Identity Cues Shape Who is Classified as Human Online" *Revise and Resubmit, American Sociological Review* (link)

Tierney, Graham,* Srikar Katta,* Christopher A. Bail, Sunshine Hillygus, Alexander Volfovsky. 2025. "A Design-based Solution for Causal Inference with Text: Can a Language Model Be Too Large?" *Revise and Resubmit, Journal of The Royal Statistical Society* (link)

Stray, Jonathan, Ian Baker, George Beknazar-Yuzbashev, Ceren Budak, Julia Kamin, Kylan Rutherford, Mateusz Stalinski, Tin Acosta, Christopher A. Bail, et al. 2026. "The Prosocial Ranking Challenge: Reducing Polarization on Social Media without Sacrificing Engagement" *Arxiv* (link)

Heseltine, Michael*, Christopher A. Bail, Petter Törnberg, Michelle Schimmell*, Christopher Barrie. 2026. "AI summaries in social media improve dialogue but reduce engagement" *SocArxiv* (link)

Heseltine, Michael*, Christopher A. Bail, Petter Törnberg, Michelle Schimmell*, Christopher Barrie. 2026. "Political identity drives trust in large language models—even when accuracy is incentivized." *ArXiv* (link)

Törnberg, Petter, Diliara Valeeva, Justus Uitermark, and Christopher A. Bail. 2023. "Simulating Social Media Using Large Language Models to Evaluate Alternative News Feed Algorithms." *ArXiv* (link)

Allamong Max,* Andrew Trexler,* Fatima Alqabandi,* Christopher A. Bail, Sunshine Hillygus, and Alexander Volfovsky. 2024. "The Discomfort of Being Outnumbered Online: Evidence from a Dynamic Discussion Experiment." *Open Science Framework Preprints* (link)

Bail, Christopher A., Sunshine Hillygus, Alexander Volfovsky, Max Allamong,* Fatima Alqabandi,* Diana M.E. Jordan,* Graham Tierney,* Christina Tucker,* Andrew Trexler,* Austin van Loon.* 2023. "Do We Need a Social Media Accelerator?" *SocArXiv* (link)

Tierney, Graham*, Christopher A. Bail, Alexander Volfovsky. "Author Clustering and Topic Estimation for Short Texts" *arXiv* (link)

Peer-Reviewed Journal Articles and Conference Proceedings

Nicolo Pagan*, Petter Törnberg, Christopher A. Bail, Christopher Barrie. 2026. "AI-Generated Language Remains Systematically Distinguishable from Human Communication" *Proceedings of the National Academy of Sciences* (link)

Neubrander, Marie* Markus Reiter-Haas,* Ben Rochford,* Max Allamong, Christopher A. Bail, Sunshine Hillygus, Alexander Volfovsky. 2027. "Can Platform Design Encourage Curiosity? Evidence from an Independent Social Media Experiment" *International Conference on Weblogs and Social Media* (link)

Van Loon, Austin*, Srikar Katta*, Christopher A. Bail, Sunshine Hillygus, Alexander Volfovsky. 2026. "Designing Social Media to Promote Productive Political Dialogue on a New Research Platform." *Nature Scientific Reports* (link)

Falkenberg, Max*, Matteo Cinelli, Alessandro Galeazzi, Christopher A. Bail, Rosa M. Benito, Axel Bruns, Anatoliy Gruzd, David Lazer, Jae K. Lee, Kikuko Nagayoshi, David G. Rand, Antonio Scala, Alexandra Siegel, Sander van der Linden, Onur Varol, Ingmar Weber, Magdalena Wojcieszak, Fabiana Zollo, Andrea Baronchelli1, Walter Quattrociocchi. 2026. "Toward an integrated and globally equitable understanding of political polarization" *Nature Communications* (link)

Nicolo Pagan*, Petter Törnberg, Christopher A. Bail, Christopher Barrie. 2026. "LLMs as Content Curators: A Large-Scale Audit of Recommendation Bias Across Providers and Platforms" *Conference on Language Modeling*

Shen, Yicheng*, Christopher A. Bail, and Alexander Volfovsky. 2025. "Enhancing Joint Ideal Point Estimation Strategies with Twitter Social Networks and Text Data." 2025. *Network Science* (link)

Allamong Max,* Christopher A. Bail, Sunshine Hillygus, and Alexander Volfovsky. 2025. "The Causal Effects of Incivility in Social Media Discussions." *American Journal of Political Science* (link)

Alqabandi, Fatima,* Graham Tierney,* Christopher A. Bail, D. Sunshine Hillygus, and Alexander Volfovsky. 2026. "Experiments offering social media users the choice to avoid toxic political content" *New Media and Society* (link)

Nicolo Pagan*, Petter Törnberg, Christopher A. Bail, Aniko Hannak, Christopher Barrie. 2025. "Can LLMs Imitate Social Media Dialogue? Techniques for calibration and BERT-based Turing-Test" *Social Sim* (link)

Dorr, Timothy,* Trisha Nagpal,* Duncan Watts, and Christopher A. Bail. 2025. "A research agenda for encouraging prosocial behavior on social media." *Nature Human Behavior*, Vol 9, 441-449 (link)

Bail, Christopher A. 2024. "Can Generative Artificial Intelligence Improve Social Science Research?" 2024. *Proceedings of the National Academy of Sciences*, Vol 121, No. 21 (link)

- *Chinese Translation (link)*

Kapoor, Sayash,* and Emily Cantrell,* Kenny Peng,* Thanh Hien Pham, Christopher A. Bail, Odd Erik Gundersen, Jake M. Hofman, Jessica Hullman, Michael A. Lones, Momin M. Malik, Priyanka Nanayakkara, Russell A. Poldrack, Inioluwa Deborah Raji, Michael Roberts, Matthew J. Salganik, Marta Serra-Garcia, Brandon M. Stewart, Gilles Vandewiele Arvind Narayanan. 2024. "REFORMS: Reporting, Standards for Machine Learning Based Science" *Science Advances*, Vol 10, No. 18 (link)

Argyle, Lisa, Christopher A. Bail, and Ethan Busby, Joshua Gubler, Thomas Howe,* Christopher Rytting,* Taylor Sorensen,* and David Wingate. forthcoming 2023. "Leveraging Artificial Intelligence for Democratic Discourse: Chat Interventions can Improve Online Political Conversations at Scale." *Proceedings of the National Academy of Sciences* (link)

Combs, Aidan*, Graham Tierney*, Fatima Alqabandi*, Devin Cornell*, Gabriel Varela*, Andrés Castro Araújo*, Lisa P. Argyle, Christopher A. Bail, Alexander Volfovsky. 2023. "Perceived gender and political persuasion: a social media field experiment during the 2020 US Democratic presidential primary election" *Nature Scientific Reports* (link)

Combs, Aidan* and Graham Tierney*, Brian Guay*, Friedolin Merhout*, Christopher A. Bail, D. Sunshine Hillygus, Alexander Volfovsky. 2023. "Reducing political polarization in the United States with a mobile chat platform." *Nature Human Behavior* (link) (blog post)

Jahani, Eaman* and Natalie Gallagher*, Friedolin Merhout*, Nicolo Cavalli*, Douglas Guilbeault*, Yan Leng*, and Christopher A. Bail. 2022. "An Online experiment during the 2020 US-Iran Crisis shows that exposure to common enemies can increase political polarization." *Nature Scientific Reports* (link)

Hartman, Rachel*, Will Blakely*, Jake Womick*, Christopher A. Bail, Eli Finkel, Hahrie Han, John Sarrouf, Juliana Schroeder, Paschal Sheeran, Jay J. Van Bavel, Robb Willer, Kurt Gray, 2022. "Interventions to Reduce Partisan Animosity: A Systematic Review." *Nature Human Behavior*, 38 (link)

Eli J. Finkel, and Christopher A. Bail, Mina Cikara, Peter H. Ditto, Shanto Iyengar, Samara Klar, Lilliana Mason, Mary C. McGrath, Brendan Nyhan, David G. Rand, Linda J. Skitka, Joshua A. Tucker, Jay J. Van Bavel, Cynthia S. Wang, and James N. Druckman, 2020. "Political Sectarianism in America." *Science* (link)

Edelmann, Achim* and Tom Wolff*, Daniel Montagne*, and Christopher A. Bail. 2020. "Computational Social Science and Sociology." *Annual Review of Sociology* (link), (blog post)

Bail, Christopher A. Brian Guay*, Emily Maloney*, Aidan Combs*, Sunshine Hillygus, Friedolin Merhout*, Deen Freelon, and Alexander Volfovsky. 2019. "Assessing the Russian Internet Research Agency's Impact on the Political Attitudes and Behaviors of American Twitter Users." *Proceedings of the National Academy of Sciences* (link), (video)

Bail, Christopher A., Taylor W. Brown*, and Andreas Wimmer. 2019. "Prestige, Proximity, and Prejudice: The Diffusion of Google Search Terms across 199 Countries, 2004-2014." *American Journal of Sociology* 124(5) (link), (blog post)

Bail, Christopher A., Lisa Argyle*, Taylor Brown*, John Bumpuss*, Haohan Chen*, M.B. Fallin Hunzaker*, Jaemin Lee*, Marcus Mann*, Friedolin Merhout*, and Alexander Volfovsky. 2018. "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment on Social Media." *Proceedings of the National Academy of Sciences*, 155(37): 9216-9221. (link), (op-ed), (video)

Bail, Christopher A., Friedolin Merhout*, and Peng Ding. 2018. "Using Internet Search Data to Examine the Relationship between Anti-Muslim and Pro-ISIS Sentiment in U.S. Counties." *Science Advances*, 4(6): 1-9 (link), (video)

Bail, Christopher A., Taylor W. Brown*, and Marcus Mann.* 2017. "Channeling Hearts and Minds: Cognitive-Emotional Currents and Public Deliberation on Social Media." *American Sociological Review*, 82(6) 1188-1213 (link)

McDonnell, Terence, Christopher A. Bail, and Iddo Tavory. 2017. "A Theory of Cultural Resonance." *Sociological Theory* 35(1): 1-14 (link)

Bail, Christopher A. 2016. "Combining Network Analysis and Natural Language Processing to Examine how Advocacy Organizations Stimulate Conversation on Social Media." *Proceedings of the National Academy of Sciences*

113(42): 11823-11828 (link), (video)

Bail, Christopher A. 2016. "Emotional Feedback and the Viral Spread of Social Media Messages about Autism Spectrum Disorders." *American Journal of Public Health* 106(7): 1173-1180 (link)

Bail, Christopher A. 2016. "Cultural Carrying Capacity: Organ Donation Advocacy, Discursive Framing, and Social Media Engagement." *Social Science and Medicine* 165: 280-288 (link)

Bail, Christopher A. 2015. "Lost in a Random Forest: Using Big Data to Study Rare Events." *Big Data and Society* July: 1-3 (link)

Bail, Christopher A. 2015. "Taming Big Data: Using App Technology to Study Organizational Behavior on Social Media."*Sociological Methods and Research* 46:2 189-217 (link)

Bail, Christopher A. 2015. "The Public Life of Secrets: Deception, Disclosure, and Discursive Framing in the Policy Process." *Sociological Theory* 33 (2): 97-124 (link), (online supplement)"

Bail, Christopher A. 2014. "The Cultural Environment: Measuring Culture with Big Data." *Theory and Society* 43 (3-4): 465-482 (link)

- *Reprinted in Theory and Society: Renewal and Critique in Social Theory by Duan Peng and Zhang Lei, Beijing: Sage*

Bail, Christopher A. 2012. "The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam since the September 11th Attacks." *American Sociological Review* 77:7 855-879 (link)

Bail, Christopher A. 2008. "The Configuration of Symbolic Boundaries against Immigrants in Europe." *American Sociological Review* 73:1 37-59 (link)

- *Reprinted in Investigating the Social World: The Process and Practice of Research by Russell K. Schutt, New York: Sage*

Lamont, Michèle and Christopher A. Bail. 2005. "Sur les frontières de la reconnaissance: Les catégories internes et externes de l'identité collective (On the Boundaries of Recognition: Internal and External Categories of Collective Identity)."*Revue Européenne des Migrations Internationales* 21:2 61-90 (link)

Other Publications

Bail, Christopher. March 10th, 2024. "GenAI Could Make Online Conversations More Civil" *Harvard Business Review* (link)

Bail, Christopher A. and Lisa Pinheiro, and Jimmy Royer April 5th, 2023. "Difficulty of Detecting AI Content Poses Legal Challenges," *Law360* (link)

Bail, Christopher A. March 31, 2022. "Social Media Regulation is Flying Blind: to fix social media, we need a new platform for research." *Nature* (link)

Jonathan Haidt and Bail, Christopher A. 2022. "Social Media and Political Dysfunction: A Collaborative Review" *Open Source Review* (link)

Bail, Christopher A. May 4, 2021. "Facebook's decision on readmitting Trump is going to enrage people. But there's more to the story." *CNN* (link)

Bail, Christopher A. November 22, 2019. "Want to bridge divides? Clean your Twitter house first." *CNN* (link)

Bail, Christopher A. September 9th, 2018. "Twitter's Flawed Solution to Political Polarization." *The New York Times* (link)

Bail, Christopher A. December 3, 2015. "How fringe attacks on American Muslims became mainstream." *The Washington Post* (link)

Bail, Christopher A. 2009. Review of "Europe Old and New: Transnationalism, Belonging, and Xenophobia."(link) (by Ray Taras) *Contemporary Sociology*, 38:562

Bail, Christopher A. 2008. Review of "Two-Faced Racism: Whites in the Backstage and Frontstage." (by Leslie Houts Picca and Joe R. Feagin) *Ethnic and Racial Studies*, 31:6 1168-1183 (link)

Bail, Christopher A. (with Michele Lamont) 2007. "Bridging Boundaries: the Equalization Strategies of Stigmatized Ethno-racial Groups Compared." *Minda de Gunzburg Center for European Studies Working Paper Series*,

154

Software

The Social Media Accelerator: An Open-Source Platform for Research on Social Media and Generative Artificial Intelligence, 2023-

Textnets: An R Package to Integrate Network Analysis and Natural Language Processing, 2019- (link)

Films

Jesters and Fools (Short Documentary Based on Breaking the Social Media Prism by Rob Feld, 2023 (link)

Invited Lectures

Princeton University, December 2026

Institute for Advanced Studies, Stanford University, July 2026

University of Oslo, Spring 2026

University of Copenhagen, Spring 2026

Rice University, Spring 2026

Koc Universiti, Istanbul (online), "Can Generative AI Improve Social Science" January 2026

Max Planck Institute, "Bridging Divides with Generative AI" October 2025

Summer Institutes in Computational Social Science Shanghai, "Can Generative AI Improve Social Science?" June 2025

Digital Humanism Conference, Vienna, "What are the lessons for Artificial Intelligence from the Social Media Era?" May 2025

OpenAI, "Can Generative AI Improve Social Science?" April 2025

Johns Hopkins Agora Institute (online webinar), "Data for Good Roundtable" April 2025

Yale University, "Can Generative AI Improve Social Science?" March 2025

Oxford University Department of Education, "Can Generative AI Improve Social Science?" February 2025

University of California Berkeley Haas School of Business, "Bridging Divides with Generative AI" " January 2025

Templeton Foundation, "Bridging Divides with Generative AI" " December 2024

Google/Jigsaw, "Designing Social Media to Promote Productive Political Dialogue on a New Research Platform" October 2024

University of Richmond, "Bridging Divides with Generative AI" October 2024

Elon University, "Bridging Divides with Generative AI" October 2024

University of Oslo, "Can Generative AI Improve Social Science?" August 2024

University of Iceland, "Can Generative AI Improve Social Science?" August 2024

Asheville Ideas Festival, "Generative AI and the 2024 Presidential Election" October 2024

Fondation Nationale des Sciences Politiques (Paris), "Can Generative AI Improve Social Science?" June 2024

Complexity Hub, Vienna, "Simulating Social Media using Generative AI" May 2024

Keynote Lecture, IMBESS, University of Latvia, "Bridging Divides with Generative AI" May 2024

Northwestern University, "Can Generative AI Improve Social Science?" May 2024

Cornell University, "Can Generative AI Improve Social Science?" April 2024

The University of Washington, "The New Social Media Order." February 2024

Behavioral Insights Team, "Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale", London, October 2023

Stanford Graduate School of Business, "Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale", October 2023

Furman College, "Breaking the Social Media Prism", October 2023

University of Amsterdam, "The Past Present and Future of Computational Social Science", September 2023

Science and Cocktails, "Breaking the Social Media Prism" Koncerthuset, Copenhagen, September 2023

Science and Cocktails, "Breaking the Social Media Prism" Paradiso, Amsterdam, September 2023

Howard University, "An Introduction to the Duke Polarization Lab", July 2023

Beijing University, "AI Chat Assistants can Improve Conversations about Divisive Topics", July 2023

Institute for Advanced Study, University of Amsterdam, "New Platforms for Research on Social Media", June 2023

University of Amsterdam, "AI Chat Assistants can Improve Conversations about Divisive Topics", June 2023

Chinese University of Hong Kong, "AI Chat Assistants can Improve Conversations about Divisive Topics", June 2023

NextDoor, "AI Chat Assistants can Improve Conversations about Divisive Topics", June 2023

Northwestern University, "AI Chat Assistants can Improve Conversations about Divisive Topics", May 2023

Technische Universität Braunschweig, "AI Chat Assistants can Improve Conversations about Divisive Topics", May 2023

Facebook Data Science, "AI Chat Assistants can Improve Conversations about Divisive Topics", May 2023

Carnegie Mellon University, "AI Chat Assistants can Improve Conversations about Divisive Topics", May 2023

UNESCO Most Forum, "Polarization and Sustainability Science", April 2023

U.S. Defense Advanced Research Projects Agency, "Introduction to the Duke Polarization Lab", February 2023

Meta Foo Camp, "AI Chat Assistants can Improve Conversations about Divisive Topics", February 2023

White House National Security Council & Office of Science and Technology Policy, "Introduction to the Duke Polarization Lab", February 2023

Ecole Polytechnique, "New Platforms for Social Media Research", November 2022

University of Texas, Austin, "Breaking the Social Media Prism", November 2022

Academy for Human Rights, "Breaking the Social Media Prism", November 2022

Northwestern University, "New Platforms for Social Media Research", November 2022

International Conference on Computational Social Science, "New Platforms for Social Media Research", June 2022

International Conference on Weblogs and Social Media, "New Platforms for Social Media Research", May 2022

University of Innsbruck (Austria), "Breaking the Social Media Prism", April 2022

Oxford Literary Festival, "Breaking the Social Media Prism", April 2022

International School of Geneva, "The Gap Between Social Media and Real Life," March 2022

Aarhus University, "Perceived Gender and Political Polarization: A Field Experiment on a Social Media Platform for Scientific Research," March 2022

MIT/Lincoln Labs, "Three Field Experiments on Identity, Social Media, and Political Polarization, February 2022

Northeastern University, "Perceived Gender and Political Polarization: A Field Experiment on a Social Media Platform for Scientific Research," December, 2021

University of Leeds, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," December, 2021

Ohio State University, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," December, 2021

Falling Walls Foundation, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," November, 2021

LinkedIn Data Science, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," November, 2021

University of North Carolina, "Gender and Political Persuasion: A Field Experiment on Social Media," November, 2021

Harvard Business School, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," October, 2021

University of Pennsylvania, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," October, 2021

Keynote Address, Academy of Sociology Conference (Germany), "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," Online

Pontifical Academy of Sciences, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," Vatican City, September 2021

SBE Distinguished Lecturer, National Science Foundation "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Keynote Lecturer, Graphex, Massachusetts Institute of Technology, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Stanford University, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

University of Arizona, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Chinese University of Hong Kong, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Mechanics Institute (San Francisco), "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Illinois Institute of Technology, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Princeton Public Library, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Syracuse University, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

DIGIZOOM, University of Umea (Sweden), "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

The Strand Book Store (New York), "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

New York University, Abu Dhabi, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Facebook Core Data Science, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Center for the Study of Media and Politics, New York University, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Keynote Lecture, The Great Mediation Conference, University of Chicago, "Breaking the Social Media Prism: How to Make our Platforms Less Polarizing," May 2021

Keynote Lecture, PRISM Health Symposium, "Cultural Bridges and Social Media Outreach," December 2020

Keynote Lecture, Conference on Digital Experimentation, Massachusetts Institute of Technology "The Future of Research with Robots in the Social Sciences," November 2020

Suessmilch Lecture, Max Planck Institute for Demographic Research, "Exposure to Common Enemies can Increase Political Polarization," October 2020

University of California, Berkeley, Political Science Department, "Using Bots to Study Political Polarization," October 2020

Johns Hopkins University, "Disrupting Identity on Social Media: A Field Experiment on a New Platform" February 2020

European University Institute, "Three Field Experiments on Identity and Filter Bubbles" November 2019

University College Dublin, "Exposure to Opposing Views can Increase Political Polarization, October 2019

University of Helsinki, "Three Field Experiments on Identity and Filter Bubbles" October 2019

Linkoping University, "Computational Social Science and Culture in the Age of Social Media" October 2019

Bocconi University, Milan, "Three Field Experiments on Filter Bubbles and Identity" October 2019

University of California, Berkeley "Exposure to Opposing Views can Increase Political Polarization" April 2019

Massachusetts Institute of Technology, "Exposure to Opposing Views can Increase Political Polarization" April 2019

New York University "Exposure to Opposing Views can Increase Political Polarization" March 2019

Columbia University, Greater Good Gathering "Technology and Political Polarization" February 2019

Oxford University, "Exposure to Opposing Views can Increase Political Polarization" November 2018

London School of Economics, "Exposure to Opposing Views can Increase Political Polarization" November 2018

University of Michigan, "Exposure to Opposing Views can Increase Political Polarization" October 2018

Complexity Hub Vienna, "Exposure to Opposing Views can Increase Political Polarization" October 2018

University of Southern California, "Cultural Networks and Bridges" July 2018

Stanford University, "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment," May 2018

Harvard University, "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment," May 2018

University of Chicago, "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment," April 2018

Institute for Advanced Study, Toulouse, "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment," March 2018.

University of Pennsylvania, "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment," January, 2018

Wharton School of Business, "Combining Natural Language Processing and Social Network Analysis to Examine how Social Media Posts go Viral," January, 2018

University of Oslo, "Studying Political Extremism on Social Media (Keynote Address to Social Media and Social Order Conference)," November, 2017

Northwestern University, "Combining Natural Language Processing and Social Network Analysis," November, 2017

University of Zurich, "Combining Natural Language Processing and Social Network Analysis," October, 2017

Stanford University, "Power, Prestige, and Proximity: The Global Diffusion of Google Searches, 2004-2014.", 2017

Brown University, "Collecting and Analyzing Digital Trace Data," August 2017

National Academy of Sciences, "Revising Social Science Curricula in the Age of Computational Social Science," June 2017

University of Indiana, Bloomington. "Terrified: How Anti-Muslim Fringe Organizations Became Mainstream," April 2017

University of Bern (Switzerland). "Developing New Techniques to Formalize the Study of Culture," October 2016

Cornell University. "Cultural Flow: How Advocacy Organizations Stimulate Public Conversation about Autism Spectrum Disorders on Facebook," Department of Sociology, October 2016

Keynote Lecture to the Annual Conference of the Islamic Society of North America. "Understanding the Rise of Anti-Muslim Sentiment," September 2016

Democratic National Convention, Interfaith Council, "Understanding the Rise of Anti-Muslim Sentiment," 2016

Columbia University. "Cultural Flow: How Advocacy Organizations Stimulate Public Conversation about Autism Spectrum Disorders on Facebook," Department of Sociology, 2016

Harvard University. "Anti-Muslim Sentiment in the Republican Primary" Islamic Studies Program, 2016.

University of California, Irvine. "Terrified: How Anti-Muslim Fringe Organizations Became Mainstream," Department of Sociology, 2016

Connecticut College. "Terrified: How Anti-Muslim Fringe Organizations Became Mainstream," Department of Sociology and Anthropology, 2016

Boston College, "Anti-Muslim Organizations and Social Media," Islamic Studies Program, 2015

Notre Dame University, "Measuring Cultural Emergence," Department of Sociology, 2015

European University Institute, "How Fringe Groups Shape the Integration of Muslims in the West," Muslims and the West Conference, 2015

Princeton University, "Filling Cultural Holes: How Autism Advocacy Organizations Mobilize New Audiences on Social Media," Department of Sociology, 2015

City University of New York, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion," Department of Sociology, 2015

Duke University, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion," Department of Sociology, 2014

Ohio State University, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion," Department of Sociology, 2014

University of Michigan, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion," Department of Sociology, 2014

University of Chicago, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion," Department of Sociology, 2014

Brown University, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion," Department of Sociology, 2014

Princeton University, "Studying the Global Diffusion of Culture with Big Data," Department of Sociology, 2014

Facebook Inc., "Winning Hearts and Minds: Advocacy Organizations, Field Rhythms, and Bystanders," Data

Science Division Workshop, 2014

University of Michigan (Blalock Lecture), "Measuring Culture with Big Data," ICPSR, 2014

University of California, Santa Barbara, "Measuring Culture with Big Data," Department of Sociology, 2014

North Carolina State University, "Big Data and Social Science," Dean's Panel, 2014

University of Michigan, "The Fringe Effect: Civil Society Organizations and the Evolution of Public Discourse about Islam," Interdisciplinary Committee on Organizational Studies, 2013

Notre Dame University, "The Fringe Effect: Civil Society Organizations and the Evolution of Public Discourse about Islam," Department of Sociology, 2013

University of British Columbia, "Finding, Mapping, and Analyzing Discursive Fields using Big Data," 2013

University of North Carolina, Chapel Hill,"Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

Princeton University, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

University of Michigan, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

Duke University, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

Stanford University, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

University of California, Los Angeles, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

Vanderbilt University, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

New York University, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

Boston College, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

University of Georgia, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

Emory University, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," Department of Sociology, 2010

Yale University, "Making Terrorists Racist: The Culture of Secrecy in Britain's Domestic Counter-Terrorism Policy, 2001-2008," Department of Sociology, 2009

University of Georgia, "Making Terrorists Racist: The Boundaries of Multicultural and Counter-Terrorism Policy in Britain, 2001-2008," Department of Sociology, 2009

Harvard University, "Inter-Ethnic Relations in Quebec: A Roundtable with Charles Taylor and Gerard Bouchard," Weatherhead Center for International Affairs, 2009

Valencia, Spain, "Diverse Diversities: The Configuration of Symbolic Boundaries against Immigrants in Twenty-One European Countries," City-to-City Conference of the European Union, 2007

Yale University, "Bridging Boundaries: The Everyday De-stigmatization Strategies of Marginalized Groups Compared" (with Michele Lamont), Culture Workshop, 2005

Selected Conference Presentations

Foo Camp, "AI Chat Assistants can Facilitate more Productive Conversations about Divisive Issues", February 2023

AlpPop Conference, La Thuile, Italy, "Exposure to Opposing Views can Increase Political Polarization. January 2019"

American Sociological Association Annual Meeting, Philadelphia, "Exposure to Opposing Views can Increase Political Polarization."

American Sociological Association Annual Meeting, Montreal, "Culture and Computational Social Science." (Organizer, Plenary Panel 2017)

American Sociological Association Annual Meeting, Chicago, IL, "Cultural Carrying Capacity: Organ Donation Advocacy, Discursive Framing, and Social Media Engagement." 2015

American Sociological Association Annual Meeting, San Francisco, CA, "NSF Sponsored Panel on Big Data and Social Science Research," 2014

American Sociological Association Annual Meeting, New York, NY, "Winning Minds Through Hearts: Civil Society Organizations, Emotional Feedback, and Social Media," 2013

American Sociological Association Annual Meeting, Denver, CO, "The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam, 2001-2008," 2012

American Sociological Association Annual Meeting, San Francisco, CA, "How Terrorists became Racists," 2009

Harvard/Manchester Conference on Social Change, Manchester, England, "Homeland Insecurity: Structural Holes in the Muslim Civil Rights Movement," 2008

American Sociological Association Annual Meeting, New York, NY, "The Configuration of Symbolic Boundaries toward Immigrants in Twenty-One European Countries," 2007

11th Annual Aage Sorensen Memorial Conference, Oxford University, England, "The New Black Brahmins? Elite African Immigrants and the Stigma of Africanité," 2007

National Association for Ethnic Studies Annual Meeting, New Paltz, NY, "Diverse Diversities: The Configuration of Symbolic Boundaries against Immigrants in Twenty-One European Countries," 2007

Eastern Sociological Society, Philadelphia, PA, "Diverse Diversities: The Configuration of Symbolic Boundaries against Immigrants in Twenty-One European Countries," 2007

American Sociological Association Annual Meeting Minorities, Montreal, Quebec, "Three Worlds of Xenophobia: a Typology of Anti-Immigrant Boundaries in Europe," 2006

International Conference on Comparative Social Sciences, Sophia University, Tokyo, Japan, "Three Worlds of Xenophobia: a Typology of Anti-Immigrant Boundaries in Europe," 2006

10th Annual Aage Sorensen Memorial Conference, Stockholm University, Sweden, "Three Worlds of Xenophobia: Towards a Typology of Anti-Immigrant Prejudice in Europe," 2006

Council for European Studies Annual Meeting, Section on Ethnic Boundaries and Regionalism, Chicago, IL, "Three Worlds of Xenophobia: Mapping the Configuration of Ethnic Boundaries in Twenty European Countries," 2006

Multidisciplinary Program on Inequality and Social Policy, John F. Kennedy School of Government, Harvard University, Cambridge, MA, "Toward a Typology of Ethnic Boundaries: A Cognitive Analysis of Racial, Religious, Linguistic, and Cultural Boundaries in Twenty-One European Countries," 2005

Summer Institute on International Migration, Ethnic Diversity, and Cities, International School for the Humanities and Social Sciences, Universiteit van Amsterdam, Netherlands, "On the Ethnic Boundaries of Immigration: Economic, Cultural, and Racial Parameters of Xenophobia," 2005

9th Annual Aage Sorensen Memorial Conference: Cambridge, MA, "Black or African American? The Socio-Economic Attainment of Africans in the United States," 2005

Teaching and Advising

PhD Chair

Hajar Yazdiha (co-chair, UNC-Chapel Hill, placement: Associate Professor (with tenure), University of Southern California)

Friedolin Merhout (Duke University, placement: Associate Professor (with tenure), University of Copenhagen)

Marcus Mann (Duke University, placement: Assistant Professor, Purdue University)

Taylor W. Brown (Duke University, placement: Facebook Core Data Science)

Devin Cornell (Duke University, placement: TBWA Worldwide)

Fatima Alqabandi (Duke University, placement: Kuwait University)

Benjamin Rochford (Duke University, in progress)

PhD Committee member

Brian Aronson (Duke University)

John Bumpuss (Duke University)

Haohan Chen (Duke University, Political Science)

Andrew Carr (Duke University, in progress)

Aidan Combs (Duke University, in progress)

Molly Copeland (Duke University)

M. Ali Kadivar (UNC-Chapel Hill)

Brandon Gorman (UNC-Chapel Hill)

Brian Guay (Duke University, Political Science)

Kevin Kiley (Duke University, in progress)

Jaemin Lee (Duke University)

Jonathan Morgan (Duke University)

Maria Cristina Ramos-Flor (Duke University)

Chris Ritter (Duke University, Political Science)

Charles Seguin (UNC-Chapel Hill)

Lauren Valentino (Duke University)

Roland Adorjani (University College, Dublin)

Ashwin Rajadesingan (University of Michigan, Information Science)

Imari Smith (Duke University)

Tina Tucker (Duke University, Political Science)

Graham Tierney (Duke University, Statistical Science)

Cathy Lee (Duke University, Statistical Science)

Junsol Kim (University of Chicago, Sociology)

Zhengran Ji (Duke University, Computer Science)

MA Committee Member

David Rigby (UNC-Chapel Hill)

Nandini Datta (Duke, Psychology/Neuroscience)

Zhengran Ji (Duke, Electrical Engineering)

Jiaxun Cao (Duke, Computer Science)

Mitchell Wang (Duke, Computer Science)

Senior Thesis Committee Member

Kulsoom Rizavi

Ariel Heckier

Courses Taught

Human-AI Co-Evolution, Spring 2026

Computational Social Science (graduate), Fall 2025

Human-AI Co-Evolution, Spring 2025

Computational Social Science (undergraduate), Spring 2025

Computational Social Science (graduate), Fall 2024

Data Science and Society, Spring 2024

Introduction to Sociology, Spring 2023

Data Science and Society, Fall 2022

Computational Social Science, Spring 2022

Data Science and Society, Spring 2022

Polarization and Social Media, Fall 2021

Computational Social Science, Spring 2021

Data Science and Society, Spring 2021

Data Science and Society, Fall 2020

Data Science and Society, Spring 2020

Web Scraping and Text Analysis, Data Science Program, Duke University, Spring 2020

Web Scraping and Text Analysis, Data Science Program, Duke University, Spring 2018

Distant Suffering: Humanitarian Advocacy in the Age of Social Media, Venice International University, Fall 2019

Cultural Globalization and Data Science, Venice International University, Fall 2019

Introduction to Sociology (Sociological Inquiry), Duke University, Fall 2018

Web Scraping and Text Analysis, Data Science Program, Duke University, Spring 2018

Graduate Seminar on Computational Sociology, Duke University, Spring 2018

Introduction to Sociology (Sociological Inquiry), Duke University, Spring 2018

Computational Social Science (Undergraduate), Duke University, Spring 2017

Graduate Seminar on Computational Sociology, Duke University, Fall 2016

Introduction to Sociology (Sociological Inquiry), Duke University, Fall 2016

Graduate Seminar on Computational Sociology, Duke University, Fall 2015

Introduction to Sociology (Sociological Inquiry), Duke University, Fall 2015

Graduate Seminar on Cultural Sociology, University of North Carolina Chapel Hill, Spring 2015

Introduction to Sociology (Sociological Perspectives), University of North Carolina Chapel Hill, Spring 2015

Introduction to Sociology (Sociological Perspectives), University of North Carolina Chapel Hill, Fall 2013

Social Studies 10, Harvard University, Spring 2007

Sociological Theory, Harvard University, Fall 2006

Professional Service

Scientific Advisory Board, Tumble Science Podcast for Kids, 2025-

Member, Editorial Board, Annual Review of Sociology, 2026-2030

Guest Member, Editorial Board, Annual Review of Sociology, 2025

Editor, Oxford University Press Series in Computational Social Science, 2018-

Director, Summer Institutes in Computational Social Science, 2023-

Advisory Committee, National Science Foundation Social, Behavioral and Economic Sciences Directorate, 2018-2023

Director, Summer Institutes in Computational Social Science, 2022

Co-Organizer, Summer Institutes in Computational Social Science, 2021

Co-Chair, SBE-CISE Subcommittee, National Science Foundation Social Advisory Council, Behavioral and Economic Sciences Directorate, 2021

Member, Broadening Participation Subcommittee, National Science Foundation Social Advisory Council, Behavioral and Economic Sciences Directorate, 2021-2023

Co-Organizer, Summer Institutes in Computational Social Science, 2020

Advisory Committee, Human Networks and Data Science Program, National Science Foundation, 2020-2021

Scientific Advisor, Neighborhood Vitality Advisory Board, NextDoor 2024-

Scholars Council, Braver Angels, 2021-

Scholars Advisory Council, Civic Health Project, 2022-

Scientific Advisor, Ukraine Initiative, 2022-

Member, Coalition for Independent Tech Research, 2022-

Advisory Board, Strengthening Democracy Challenge, 2021-2022

Scientific Advisory Board, Streitgut, 2021-2022

Advisory Board, Project IMMUNE, Bocconi University, 2020-

External Reviewer, Master of Science in Data Analytics and Computational Social Science, University of Massachusetts, Amherst, 2020

Co-Organizer, Summer Institutes in Computational Social Science (Bamberg, Cape Town, Chicago, Istanbul, Los Angeles, Monterrey, New York City, Oxford, Princeton, Research Triangle Park, Zurich), 2019

Co-Organizer, Summer Institutes on Computational Social Science (Boulder, Cape Town, Chicago, Helsinki, Princeton, New York City, Washington), 2018.

National Science Foundation Sociology Panel, 2018

Co-Organizer, Summer Institute on Computational Social Science, Princeton University 2017

Associate Editor, *Science Advances*, 2017-2018

Editorial Board, *American Journal of Sociology*, 2017-2019

Book Review Editor, *Social Forces*, 2014-2015

Editorial Board, *Sociological Theory*, 2010-2015

Editorial Board, *Social Forces*, 2011-2015

Editorial Board, *American Journal of Cultural Sociology*, 2011-2019

Nominations Committee, Methodology Section, American Sociological Association, 2015.

Council Member, Section on Communication and Information Technologies, American Sociological Association, 2013-2014

Nominations Committee, Section on Organizations, Occupations, and Work, American Sociological Association, 2015-2016

Program Committee, International Conference on Computational Social Science, Helsinki Finland, 2015

Student Paper Award Committee, Communication and Information Technology Section of the American Sociological Association, 2015

Student Paper Award Committee, Political Sociology Section of the American Sociological Association, 2015

Nominations Committee, Section on Political Sociology, American Sociological Association, 2013-2014

Nominations Committee, Section on the Sociology of Emotions, American Sociological Association, 2013-2014

Occasional Reviewer *Nature*,*Science*,*Proceedings of the National Academy of Science*, *Science Advances*, *Nature Human Behavior*,*American Political Science Review*, *American Sociological Review*, *American Journal of Sociology*, *American Political Science Review*, *Poetics*, *Political Communication*, Princeton University Press, Oxford University Press, Cambridge University Press, *Social Forces*, *Sociological Theory*, European Commission's Joint Research Centre, *Ethnic and Racial Studies*, *Sociological Methods and Research*, *Social Problems*, *American Journal of Cultural Sociology*, *Sociological Forum*, *National Science Foundation*, *Russell Sage Foundation*, Templeton Foundation, and many others.

Organizer, Culture and Computational Science Panel, American Sociological Association Annual Meetings, 2016

Organizer, International Conference on Computational Social Science, 2015

Organizer, Computational Sociology Summit, 2015

Organizer, Computational Social Science and the Study of Social Behavior, Social Psychology Section, American Sociological Association Annual Meetings, 2014

Presider, Section on Cultural Sociology, American Sociological Association Annual Meetings, 2013

Presider, Section on Ethnic and Racial Minorities Round-table, American Sociological Association Annual Meetings, 2006

University Service

Co-Organizer (with Brinnae Bent), Conference on Society-Centered AI, Duke University, 2026

Co-Organizer (with Brinnae Bent), Responsible AI Symposium, Duke University, 2025

Presenter, Board of Trustees Meeting, December 2025.

Co-Director, Integrated Graduate Education Collaborative in Society-Centered AI, 2025-

Member, Rhodes iID Scientific Advisory Committee, 2025-

Member, AI at Duke Steering Committee, 2025-

Member, Ad Hoc Advisory Committee, Duke Computing Initiative, 2024

Faculty Chair, Life with AI Pillar Advisory Committee, 2025-

Advisor, AI Education Task Force, 2025-

Member, Provost's Interdisciplinary Strategy Council, 2024-

Member, Tenure and Promotion Committee for Dr. Anne Washington, Sanford School of Public Policy, 2025

Junior Faculty Search Committee, Department of Sociology, 2024

Presenter, "Bridging Divides with Generative AI" Trinity Board of Visitors, November 2024

Organizer, "Social AI Mini-Conference," April 2024

Presenter, "Best of Duke Study Tour, Alumni Engagement and Development, 2024

Presenter, "AI and the Liberal Arts," Undergraduate Education Committee, Duke University Board of Trustees, 2023

Presenter, "Discussing Challenging Topics," Trinity College Board of Visitors, 2023

Organizing Committee, Provost's Forum "Examining Problems in Big Tech."

Chair, Tenure and Promotion Committee for Dr. Craig Rawlings, Sociology Department, 2022-2023

Executive Committee, Sociology Department, 2022-2023

Graduate Studies Committee, Sociology Department, 2019-2022

Faculty Governance Committee, Duke Initiative for Science and Society, 2022-

Faculty Supervisor, Duke Computational Social Science Workshop, 2020-2022

Steering Committee, Duke Center for Computational Thinking, 2020

Chair, Re-appointment Committee for Dr. Craig Rawlings, Sociology Department, 2020

Search Committee, Director of Social Science Research Institute, 2018-2019

Director, Science Core, Duke Population Research Center, 2018-2020

Co-Director, Science Core, Duke Population Research Center, 2017-2018

Director, Computational Social Science Track, Multidisciplinary Program in Data Science, 2016-2020

Member, Curriculum Development Committee, Multidisciplinary Program in Data Science, 2016-2018

Junior Faculty Search Committee, Department of Sociology, Duke University, 2017

Executive Committee, Department of Sociology, Duke University, 2017-2018

Faculty Computing Representative, Department of Sociology, Duke University, 2016-2018

Executive Committee, Department of Sociology, Duke University, 2015-2016

Colloquium Series Organizer, Department of Sociology, Duke University, 2015-2016

Publications Committee, Department of Sociology, UNC-Chapel Hill, 2014-2015

Graduate Student Placement Coordinator, Department of Sociology, UNC-Chapel Hill, 2013-2014

Committee on Undergraduate Studies, Department of Sociology, UNC-Chapel Hill, 2013-2014

Consulting Work

Consultant, Schmidt Futures, 2022-2023

Academic Consultant, Twitter, August 2021-June 2022

Languages

French (advanced/intermediate), Italian (beginner/intermediate), Mandarin Chinese (beginner/intermediate), R, Python, LaTeX

Last updated: August 29, 2026

[redacted]

Expert Testimony

NetChoice, LLC et al. v. Uthmeier et al., United States District Court for the Northern District of Florida Tallahassee Division, Case No. 4:21-cv-00220.

– Deposition Testimony (2026).

[redacted]

Appendix B
Intentionally Left Blank

Appendix C.1
Web Search Literature

Anand, Abhijit, et al., "Understanding the User: An Intent-Based Ranking Dataset," *CIKM '24: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, October 2024, pp. 5323 - 5327.

Broder, Andrei, "A Taxonomy of Web Search," *ACM SIGIR Forum*, Vol. 36, No. 2, 2022, pp. 3 - 10.

Cambazoglu, B. Barla, et al., "An Intent Taxonomy for Questions Asked in Web Search," *CHIIR '21: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, March 2021, pp. 85 - 94.

Canfora, Gerardo and Luigi Cerulo, "A Taxonomy of Information Retrieval Models and Tools," *Journal of Computing and Information Technology*, Vol. 12, No. 3, September 2004, pp. 175 - 194.

Hamroun, Mohamed and Mohamed Salah Gouider, "A Survey on Intention Analysis: Successful Approaches and Open Challenges," *Journal of Intelligent Information Systems*, Vol. 55, No. 3, April 25, 2020, pp. 423 - 443.

Jansen, Bernard J., Danielle L. Booth and Amanda Spink, "Determining the Informational, Navigational, and Transactional Intent of Web Queries," *Information Processing & Management*, Vol. 44, No. 3, May 2008, pp. 1251 - 1266.

Jansen, Bernard J., Danielle L. Booth and Amanda Spink, "Determining the User Intent of Web Search Engine Queries," *WWW '07: Proceedings of the 16th International Conference on World Wide Web*, May 2007, pp. 1149 - 1150.

Liu, Jiqun, et al., "Task, Information Seeking Intentions, and User Behavior: Toward A Multi-Level Understanding of Web Search," *CHIIR '19: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, March 2019, pp. 123 - 132.

Manta-Caro, Cristyan, Annalina Caputo and Juan M. Fernández-Luna, "Information Retrieval for IoT and WoT: State-of-the-Art, Taxonomy Framework, and Evolutionary Directions," *IEEE Internet of Things Journal*, Vol. 12, No. 6, December 11, 2024.

Marchionini, Gary, "Exploratory Search: From Finding to Understanding," *Communications of the ACM*, Vol. 49, No. 4, April 2006, pp. 41 - 46.

Rose, Daniel E. and Danny Levinson, "Understanding User Goals in Web Search," *Proceedings of the 13th International Conference on World Wide Web 2004*, May 2004, pp. 13 - 19.

Appendix C.2
Generative AI Literature

"AI at Work Is Here. Now Comes the Hard Part," *Microsoft*, May 8, 2024, available at https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part.

Ammari, Tawfiq, et al., "How Students (Really) Use ChatGPT: Uncovering Experiences Among Undergraduate Students," *arXiv*, January 19, 2026.

Appel, Ruth, et al., "The Anthropic Economic Index Report: Uneven Geographic and Enterprise AI Adoption," *Anthropic*, September 15, 2025.

Bick, Alexander, Adam Blandin and David J. Deming, "The Rapid Adoption of Generative AI," *National Bureau of Economic Research*, September 2024.

Bodonhelyi, Anna, et al., "User Intent Recognition and Satisfaction with Large Language Models: A User Study with ChatGPT," *arXiv*, November 19, 2024.

Brachman, Michelle, et al., "How Knowledge Workers Use and Want to Use LLMs in an Enterprise Context," *CHI EA 2024: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, May 2024.

Chatterji, Aaron, et al., "How People Use ChatGPT," *National Bureau of Economic Research*, September 2025.

Costa-Gomes, Beatriz, et al., "It's About Time: The Copilot Usage Report 2025," *Microsoft*, December 2025.

Handa, Kunal, et al., "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations," *arXiv*, April 15, 2025.

"Our Life With AI: From Innovation to Application," *Google*, January 2025, available at https://www.ipsos.com/sites/default/files/ct/news/documents/2025-10/IpsosGoogle_our-life-with-ai_2025.pdf.

Ouyang, Siru, et al., "The Shifted and the Overlooked: A Task-Oriented Investigation of User-GPT Interactions," *Association for Computional Linguistics*, December 2023, pp. 2375 - 2393.

Rainie, Lee, "Close Encounters of the AI Kind: The Increasingly Human-like Way People are Engaging with Language Models," *Elon University*, March 2025.

Ramaul, Laavanya, Paavo Ritala and Mika Ruokonen, "Creational and Conversational AI Affordances: How the New Breed of Chatbots is Revolutionizing Knowledge Industries," *Business Horizons*, Vol. 67, No. 5, 2024, pp. 615 - 627.

Shah, Chirag, et al., "Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies," *ACM Transactions on the Web*, Vol. 19, No. 3, August 2025, pp. 34: 1 - 29.

Shelby, Renée, Fernando Diaz and Vinodkumar Prabhakaran, "Taxonomy of User Needs and Actions," *arXiv*, October 10, 2025.

Skjuve, Marita, Petter Bae Brandtzaeg and Asbjørn Følstad, "Why Do People Use ChatGPT? Exploring User Motivations for Generative Conversational AI," *First Monday*, Vol. 29, No. 1, January 3, 2024.

Suri, Siddharth, et al., "The Use of Generative Search Engines for Knowledge Work and Complex Tasks," March 19, 2024, working paper.

Tamkin, Alex, et al., "Clio: Privacy-Preserving Insights into Real-World AI Use," *arXiv*, December 18, 2024.

Tang, Jiakai, et al., "KAPA: A Deliberative Agent Framework with Tree-Structured Knowledge Base for Multi-Domain User Intent Understanding," *Association for Computional Linguistics*, July 2025, pp. 6150 - 6156.

Theofanos, Mary, Yee-Yin Choong and Theodore Jensen, "AI Use Taxonomy," *National Institute of Standards and Technology*, March 2024.

Tomlinson, Kiran, et al., "Working with AI: Measuring the Applicability of Generative AI to Occupations," *arXiv*, December 22, 2025.

Trippas, Johanne R., et al., "What do Users Really Ask Large Language Models? An Initial Log Analysis of Google Bard Interactions in the Wild," *SIGIR '24: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 2024, pp. 2703 - 2707.

Velázquez, Marijose Páez, Elzbieta Bobrowicz-Campos and Patrícia Arriaga, "Prompt Assessment for Human-AI Interaction: Intent, Complexity & Lay Perceptions," *Journal of Machine Intelligence and Data Science*, Vol. 6, December 15, 2025.

Velázquez, Marijose Páez, Elzbieta Bobrowicz-Campos and Patrícia Arriaga, "Towards a Typology of Prompts for Human-AI Interaction: Mapping Intent and Complexity with Lay Users," *EECSS '25: Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences*, August 2025.

Wang, Jiayin, et al., "Understanding User Experience in Large Language Model Interactions," *arXiv*, January 16, 2024.

Yan, Yueru, et al., "SHARECHAT: A Dataset of Chatbot Conversations in the Wild," *arXiv*, January 27, 2026.

Zhang, Zhouqing, Kongmeng Liew and Tham Piumsomboon, "What One Million Prompts Tells Us About AI Usage, Topics, and Preferences," *IEEE Conference on Artificial Intelligence*, 2025.

Zhao, Wenting, et al., "WILDCHAT: 1M ChatGPT Interaction Logs in the Wild," *ICLR 2024*, May 2, 2024.

Zheng, Lianmin, et al., "LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset," *ICLR 2024*, March 10, 2024.

Zheng, Lianmin, et al., "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," *arXiv*, December 24, 2023.

Appendix D

Choose a single user intent that characterizes the first topic of each conversation

User Intent	Definition	Types of Prompts	Hypothetical Examples	Notes
Advice / Recommendation / Seeking Opinions	Capture prompts where user is asking for help in making a choice or seeking to form an opinion. Relies on the AI Chatbot's ability to provide personalized recommendations / output. Includes forward-looking predictions of future outcomes and answers to questions with subjective, opinionated answers. Does not include seeking help on homework multiple-choice questions.	Top, best, or most in a particular category, including comparing different options	[redacted]	
		A personalized list of options / plans		
		Asking for advice		[1]
		Itineraries		
		Predictions of future outcomes		
		"Should" or subjective questions / seeking opinions		
Brainstorm / Idea Generation	Capture prompts where user is asking AI Chatbot to help generate creative ideas or a hypothetical list of options. Includes requests for quiz questions. Unlike "Advice / Recommendation / Seeking Opinions," user should not be seeking any ranked or sorted "best" results.	Requests for ideas		[2]
		Requests for sets of options (username, baby name, password)		
		Requests for topics or quiz questions		
Companionship from AI Chatbot	Capture cases where user is using the AI Chatbot as a confidant, friend, therapist, or role-play partner.	Sharing thoughts with AI Chatbot as friend		
		AI Chatbot as a partner in role-play		
Computer Code Generation, Revision, or Debugging	Capture prompts where user is asking AI Chatbot to help write, debug, or edit computer code. If the AI Chatbot responds with functions or computer code, it should fall under this category.	Write a new computer code		
		Error messages or troubleshooting		
		Alter or debug user-provided script		
		How-to questions in context of coding		
Create, Edit, or Process Images	Capture prompts where user is asking AI Chatbot to generate or edit an image, or prompts that ask AI Chatbot to read, process, or explain an image file. As a reminder, images will not be displayed.	Creation of images for fun		
		Creation of images for personal/professional use		
		Editing image - format, tone, style, color		
		Editing image - photoshop		
		Read, process, or explain an image		

Choose a single user intent that characterizes the first topic of each conversation

User Intent	Definition	Types of Prompts	Hypothetical Examples	Notes
Information Retrieval - Current Events	Captures prompts that request factual information about events or developments that occurred within a week of the date of the prompt (before or after). This would include information that one may expect to find in local, regional, or national news sources. You can use AI Chatbot's responses and the date of the conversation to help you determine whether the user is seeking recent or more historical information.	Requests related to local, regional, national, or global events; use the conversation timestamp and AI Chatbot's response to determine recency	[redacted]	[3], [4]
		Requests with explicit (e.g., today, yesterday, current) or implicit indication of seeking information about recent events		[5], [6]
Information Retrieval - Other	All other information retrieval queries that do not fall within "Information Retrieval - Current Events."	Requests of names of places or people, or contact information		
		Counts, dimensions, or other specific factual details		
		Questions about single words or phrases, including synonyms and definitions		
		Historical/past trivia facts		
		Broad categories or topics		[7]
		Request for summary description of an object (does not include summary of *user*-provided text)		
		Specific questions with factual answers, including multi-step prompts that simply ask for more detail on the previous information retrieval request		[8]

Choose a single user intent that characterizes the first topic of each conversation

User Intent	Definition	Types of Prompts	Hypothetical Examples	Notes
Jailbreak Behaviors	Capture prompts that are designed to bypass the safeguards and moderation measures in AI Chatbot, regardless of whether they succeed or not.	Prompts that attempt to bypass the safeguards by asking the AI Chatbot to play a role or imagine a fictional or hypothetical scenario, or by framing the request as a story or academic exercise, often characterized by a long, detailed prompt	[redacted]	[9]
		Multi-step prompts that gradually steer the AI Chatbot to bypass the AI Chatbot's safeguards and moderation measures by using repetition, follow-up questions, or references to prior responses		
Navigation to Another Webpage	Capture prompts where the user is trying to navigate to another specific, often common, website or attempting to log in to a website or portal.	Common website names (excluding brand names that are not exclusively websites)		[10]
		Prompts that include a common website name as part of the query		
		Entity name + terms like "log in," "home page" or "site" or a URL containing "www" or ".com"		[11]
Practical Guidance / Learning	Capture prompts focused on learning/understanding a concept or acquiring a new skill. Relies on AI Chatbot's ability to tutor or deliver tailored explanations or synthesize information. Does **not** include quiz or multiple choice questions (see "Problem Solving / Reasoning") unless user engages to actively *learn* the material, rather than to simply obtain the correct answer from a multiple-choice question.	Learning about a broad concept or an academic topic		
		Requests for step-by-step instructions		
		Asking specific "why" questions on academic topics		
		How-to/"How do you"/"How do I"-style questions		[12]
		Multi-step questions / prompts that build on each other potentially in different angles in an effort to learn about a topic		[13]

Choose a single user intent that characterizes the first topic of each conversation

User Intent	Definition	Types of Prompts	Hypothetical Examples	Notes
Problem Solving / Reasoning	Capture prompts where user wants AI Chatbot to not only find information but also apply it to a given situation or engage in logical reasoning. Relies on the ability of the AI Chatbot to process information and construct a response based on available information and what the user has provided. Include quiz or multiple choice questions, potentially from a homework assignment. Includes prompts that ask AI Chatbot to process and make inferences from a set of data or to perform mathematical computations.	Multiple choice, fill in the blank, true/false questions, or other questions that appear to be quiz questions	[redacted]	[14]
		Asking for an inference from a block of text		
		Seeking answer to a specific question, but given a certain set of inputs (given a particular case setting, or from a particular point of view)		
		Identify consistencies or contradictions		
		Asking to perform basic arithmetic/statistical calculations		
		Asking to perform statistical analysis		
		Generating insights		
Questions about AI Capabilities	Capture prompts where the user is asking AI Chatbot about the AI Chatbot itself, what it was trained on, what it can help with, or what it is.	Questions about the AI Chatbot		
Writing - Content Creation	Capture prompts where user is asking the AI Chatbot to write new text without any provided text to modify. Includes requests for both fictional and non-fictional written work product.	Write an essay, poem, story, or email		
		Create marketing materials or tweets		
Writing - Edit User-Provided Text	Capture prompts where user is asking the AI Chatbot to transform user-provided text. This can include changing the form, editing for grammar or spelling, translating, summarizing text into bullets, or expanding on a certain point.	Turn bullets into paragraphs (or reverse)		
		Summarization of user-provided text		
		Change format		
		Change tone (e.g., make text sound more professional or lighten the tone)		
		Rephrase		
		Check grammar and spelling of text		
		Translation request of user provided text		
Other	This category is for the unlikely scenario that a conversations involves a user intent that does not fall into any of the categories above. This category should only be selected when you understand what the user is trying to accomplish but the user intent is outside of the pre-defined categories above. This category should not be selected for conversations that are too short to understand intent; unintelligible; in languages another than English that you cannot understand without help of a translator; or have been redacted in a way that obscures user intent.	Other		

Choose a single user intent that characterizes the first topic of each conversation

User Intent	Definition	Types of Prompts	Hypothetical Examples	Notes
Unable to Classify - Insufficient Content	Capture prompts that are unclear, because the user has submitted text that is too short, did not provide sufficient context, or the prompt is riddled with errors such that it is not clear what the user was hoping to accomplish. Any prompt where the AI Chatbot does not understand what the user is asking for should fall under this category.	AI Chatbot does not know what user is asking for	[redacted]	
		Text that is without any context and coder cannot make sense of		
		One or two-word prompts that are unclear (even after taking into consideration surrounding context and AI Chatbot's response)		
		Strings of numbers or letters		
		Incoherent sentences which cannot be autocorrected (even after taking into consideration AI Chatbot's response)		
Unable to Classify - Redacted	Capture instances where redaction results in insufficient context in the user prompt(s) to reliably classify user intent.	Capture instances where redaction of the *user prompt* results in insufficient context to reliably classify user intent, regardless of the information available in the bot's		[15]
Unable to Classify - Requires Translation	Capture prompts that require translation of the user prompt(s) to determine user intent.	Capture instances of non-English user prompts that require translation to determine user intent, regardless of the information available in the bot's response.		[16]

Notes:
[1] Hypothetical examples updated to clarify definitions for coders.
[2] Definition updated to include requests for quiz questions.
[3] Definition updated to clarify that it is a week of the date of the prompt, before or after, not the past week as of February 2026.
[4] Examples updated to clarify definitions for coders.
[5] Types of Prompts updated to include "current" as an explicit indication and to remove prompt type: "Specific information, where the answer changes depending on when it is asked and the answer has an explicit recent timestamp."
[6] Examples updated to clarify definitions for coders.
[7] Examples added to clarify definitions for coders.
[8] Types of Prompts and Examples updated to clarify that conversations with multiple prompts can fall under this category if those prompts ask for more detail on the prior information retrieval request.
[9] The definition of Jailbreak and examples were updated to align with active attempts to bypass safeguards rather than prompts which cause Copilot to disengage after discussions with coders.
[10] Clarified Types of Prompts to exclude brand names that are not exclusively websites.
[11] Added additional website identifiers to Types of Prompts and added additional examples to clarify definition for coders.
[12] Expanded list of "How" questions under Types of Prompts to clarify definition for coders.
[13] Example added to clarify definitions for coders.
[14] Combined "Quantitative Data Analysis" and "Problem Solving / Reasoning" as distinguishing these two categories was challenging.
[15] Clarified Types of Prompts to exclude Copilot's response when determining if the conversation could be classified.
[16] Clarified Types of Prompts to exclude Copilot's response when determining if the conversation could be classified.
[17] Hypothetical examples were intentionally designed with incorrect spelling, capitalization, and punctuation to mimic expected appearance of prompts in the data sample.

[redacted]

Subject matter is only available if you selected Information Retrieval (Current Events or Other) as user intent

Subject Matter	Definition	Hypothetical Examples		Notes
		Current Events	Other	
Arts / Literature / Language	Information on artistic works, literary texts, authors, artists, or any analysis related to visual, written, or performing arts. Includes prompts that seek to identify synonyms, ask for meaning of word, or understand the meaning of a text in other languages.	[redacted]		
Business / Finance	Information about financial markets, investing, personal finance, or business operations.			
Computing Devices / Software	Information related to computing devices, mobile devices, software, digital systems, or other technical concepts.			
Consumer Products and Services	Information on buying, comparing, reviewing, or using consumer goods and services, or information about consumer brands, except prompts that fall within other named subject matter categories.			[1]
Entertainment (Celebrities, Movies, Concerts, Games)	Information about entertainment topics such as celebrities, movies, TV, and recreational activities such as attending concerts, live events, hobbies, table-top games, and other media and recreational experiences.			[2]
Fashion / Beauty	Information on clothing, style trends, personal grooming, skin and hair care including general guidance on styling or beauty routines.			
Food / Cooking / Recipes	Information on food, restaurants, cooking methods, or recipes, excluding nutrition or healthfulness of foods (which belong to Health / Fitness).			[3]
Health / Fitness	Information on physical and mental health, medical conditions, symptoms, treatments, nutrition, exercise, lifestyle practices or fitness routines.			

Subject matter is only available if you selected Information Retrieval (Current Events or Other) as user intent

Subject Matter	Definition	Hypothetical Examples		Notes
		Current Events	Other	
History / Social Sciences / Politics	Information related to historical events or figures, civilizations and cultural developments, as well as topics in social science including sociology, political science, anthropology, or economics. Includes both historical and contemporary questions about society, cultures, institutions, political processes, public policy, or political figures.	[redacted]		
Law / Tax	Information related to tax information, past or current legal systems, laws, institutional regulations, and citizen rights.			
Other Local, Regional, National, or International Events	Information related to specific, identifiable events or incidents that occur at the local, regional, national, or global level, other than events that fall within other named subject matter categories.			[4]
Personal Growth / Family	Information related to personal development, self-improvement, wellbeing, relationships, and family dynamics			
Religion	Information about the history and evolution of religions, belief systems, spiritual practices, religious texts, rituals, festival and their dates, places of worship as well as notable religious figures. Include prompts that ask the AI Chatbot to analyze, interpret or compare religious texts.			
Science / Math	Information about scientific concepts, natural phenomena, mathematical principles, formulas, and other academic literature and topics across fields other than events that fall within other named subject matter categories.			[5], [6]
Sports	Information about trending or past sport events, notable personalities in sports, sports teams, game rules and as well analysis of games, performance or sports-related events.			
Travel / Geography	Information on countries, regions, landmarks or geographic features as well as logistical information on available routes, transportation and destination guidance.			
Other	Information on any other subject matter that does not fall into any category listed above.			

Notes:
[1] Updated definition to clarify that conversations that could fall under both “Consumer Products” and another category should be categorized using the other category.
[2] Updated examples to clarify definitions for coders.
[3] Updated definition to exclude prompts discussing nutrition or the healthfulness of foods.
[4] Updated definition to clarify that conversations that could fall under both “Other Local, Regional, National, or International Events” and another category should be categorized using the other category.
[5] Updated definition to clarify that conversations that could fall under both “Science / Math” and another category should be categorized using the other category.
[6] Updated examples to clarify definitions for coders.

[redacted]

Appendix E

ANALYSIS GROUP

Instructions – Day 1

Privileged and Confidential

BOSTON CHICAGO DALLAS DENVER LOS ANGELES MENLO PARK NEW YORK SAN FRANCISCO WASHINGTON, DC BEIJING BRUSSELS LONDON MONTREAL PARIS TORONTO

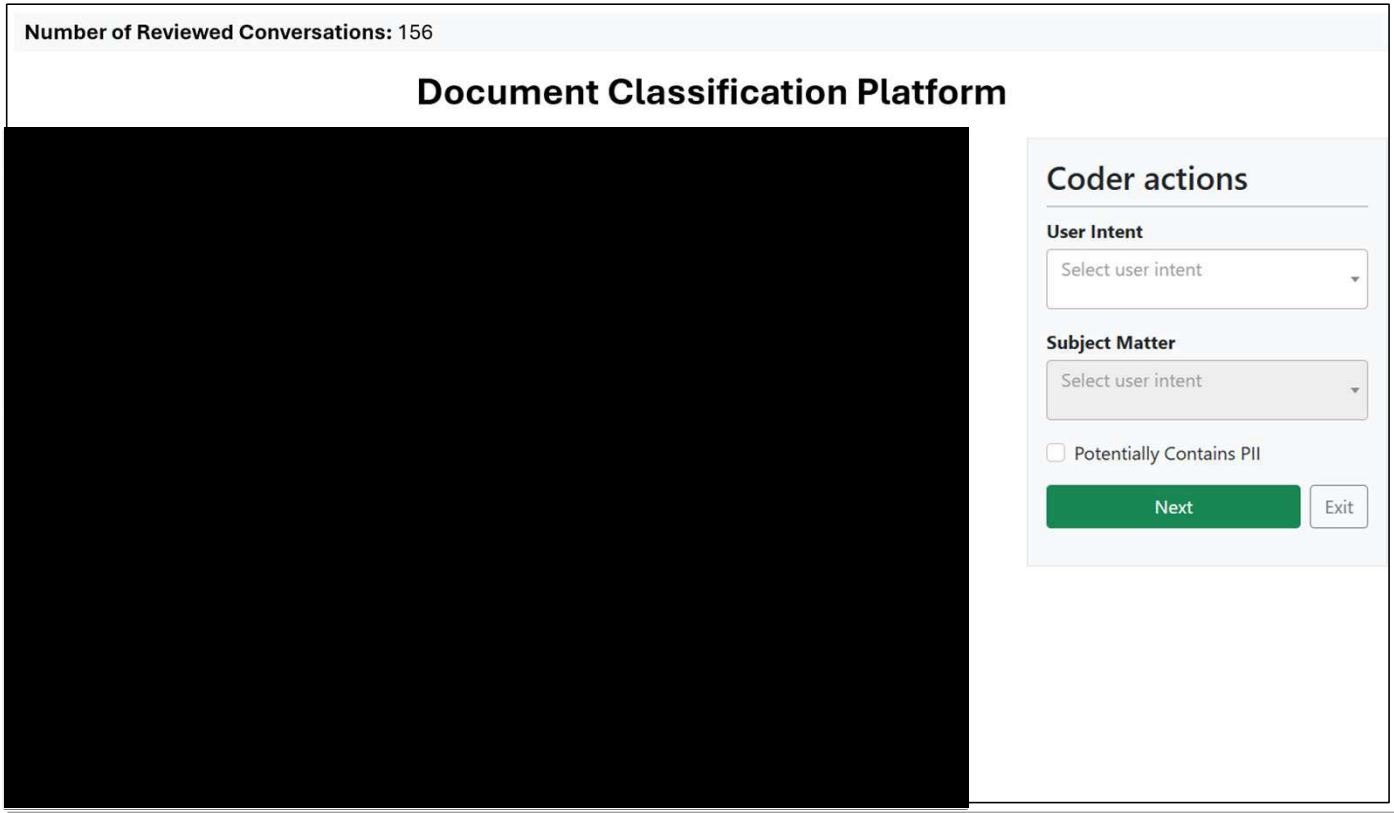
ANALYSIS GROUP

Agenda

- Overview / Purpose
- General Guidelines

ANALYSIS GROUP

Screenshot of the Web App Displaying a Hypothetical Example



ANALYSIS GROUP

Overview / Purpose

Overall Goal:

- Understand individuals' intent when they use an AI Chatbot by reviewing a sample of conversations between users and the chatbot and assigning them into categories

Overall Process:

- Use a web app to view AI Chatbot conversations and determine what the user was seeking to accomplish
- Choose from a set of pre-determined categories discussed in detail during the training stage to assign user intent (and sometimes also subject matter) to each conversation

Hypothetical Examples:

- "draw a picture of a large anteater" – the user is using the AI Chatbot to create an image
- "what is a giant anteater?" – the user is seeking information

ANALYSIS GROUP

General Guidelines

- The web app should be used only on Mozilla Firefox (not Microsoft Edge or Google Chrome)
- This information is **extremely** confidential. Do not copy, paste, or re-type any of the conversations outside of the web app
  - If you want to ask questions about a specific conversation, the “InternalID” can be used during the training phase to direct someone from the AG team to the conversation for which you have a question
  - After the training phase is completed, you will need to assign a category to each conversation on your own and, after training is completed, you cannot discuss your classifications with each other or with anyone else besides the AG Team. The AG Team can answer logistical questions or questions about the coding process but cannot tell you how you should classify any particular conversation.
- Please do **not** conduct any additional research on the topics or words used in the prompts. Instead:
  - Please rely on both the prompt and the response from the AI Chatbot to help understand context
  - Reference the complete discussion thread on the first topic, not just the initial prompt to understand user intent

ANALYSIS GROUP

General Guidelines

- Please note that all examples in this presentation, as well as all other materials provided to you during training, are **hypothetical** and not actual conversations with an AI Chatbot. They are meant to be illustrative of the types of prompts and responses you may encounter, but the data you review will be entirely new. As a reminder, all information reviewed or discussed during your engagement is **highly confidential**.

- While the conversations may already have some personal or sensitive information removed, it is possible that you may still encounter some potentially personally identifiable information ("PII") in the conversations. This includes, but is not limited to: names, date of birth, Social Security number, addresses, phone numbers, social media account information, bank information, passwords, or specific health information. If you see any information of this kind, please check the "Potentially Contains PII" box in the web app.
  - Select this box any time you suspect that the information in the conversation could potentially be used by someone to identify the author of the conversation. You should not attempt to determine veracity of this information in any manner.
  - Generic information on its own (such as a general health question) should not be flagged, but detailed or extensive personal information—such as a full health history—that may allow someone to identify the author should be flagged.

ANALYSIS GROUP

Instructions – Part 1

Privileged and Confidential

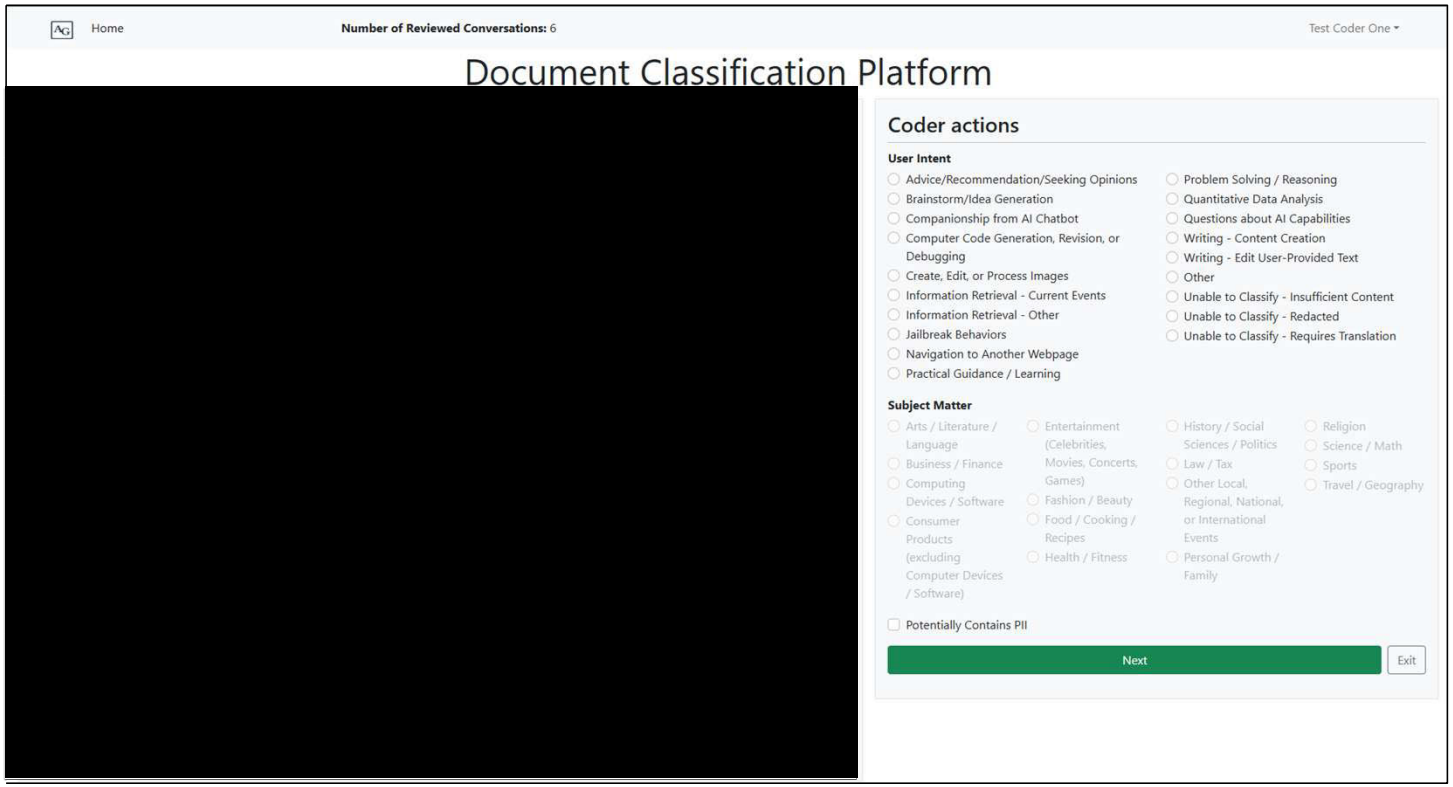
ANALYSIS GROUP

Agenda

- Review of Overview / Purpose
- Review of General Guidelines
- Categories
- Hypothetical Examples

ANALYSIS GROUP

Screenshot of the Web App Displaying a Hypothetical Example



ANALYSIS GROUP

Review of Overview / Purpose

Overall Goal:

- Understand individuals' intent when they use an AI Chatbot by reviewing a sample of conversations between users and the chatbot and assigning them into categories

Overall Process:

- Use a web app to view AI Chatbot conversations and determine what the user was seeking to accomplish
- Choose from a set of pre-determined categories discussed in detail during the training stage to assign user intent (and sometimes also subject matter) to each conversation

Hypothetical Examples:

- "draw a picture of a large anteater" – the user is using the AI Chatbot to create an image
- "what is a giant anteater?" – the user is seeking information

ANALYSIS GROUP

Review of General Guidelines

- The web app should be used only on Mozilla Firefox (not Microsoft Edge or Google Chrome)
- This information is **extremely** confidential. Do not copy, paste, or re-type any of the conversations outside of the web app
  - If you want to ask questions about a specific conversation, the “InternalID” can be used during the training phase to direct someone from the AG team to the conversation for which you have a question
  - After the training phase is completed, you will need to assign a category to each conversation on your own and, after training is completed, you cannot discuss your classifications with each other or with anyone else besides the AG Team. The AG Team can answer logistical questions or questions about the coding process but cannot tell you how you should classify any particular conversation.
- Please do **not** conduct any additional research on the topics or words used in the prompts. Instead:
  - Please rely on both the prompt and the response from the AI Chatbot to help understand context
  - Reference the complete discussion thread on the first substantive request, not just the initial prompt to understand user intent

ANALYSIS GROUP

Review of General Guidelines

- Please note that all examples in this presentation, as well as all other materials provided to you during training, are **hypothetical** and not actual conversations with an AI Chatbot. They are meant to be illustrative of the types of prompts and responses you may encounter, but the data you review will be entirely new. As a reminder, all information reviewed or discussed during your engagement is **highly confidential**.

- While the conversations may already have some personal or sensitive information removed, it is possible that you may still encounter some potentially personally identifiable information ("PII") in the conversations. This includes, but is not limited to: names, date of birth, Social Security number, addresses, phone numbers, social media account information, bank information, passwords, or specific health information. If you see any information of this kind, please check the "Potentially Contains PII" box in the web app.
  - Select this box any time you suspect that the information in the conversation could potentially be used by someone to identify the author of the conversation. You should not attempt to determine veracity of this information in any manner.
  - Generic information on its own (such as a general health question) should not be flagged, but detailed or extensive personal information—such as a full health history—that may allow someone to identify the author should be flagged.

ANALYSIS GROUP

19 User Intent Categories

Will also have a Subject Matter:

- Information Retrieval – Current Events
- Information Retrieval – Other

Will not have a Subject Matter:

- Advice / Recommendation / Seeking Opinions
- Brainstorm / Idea Generation
- Companionship from AI Chatbot
- Computer Code Generation, Revision, or Debugging
- Create, Edit, or Process Images
- Jailbreak Behaviors
- Navigation to Another Webpage
- Practical Guidance / Learning
- Problem Solving / Reasoning
- Quantitative Data Analysis
- Questions about AI Capabilities
- Writing - Content Creation
- Writing - Edit User-Provided Text
- Other
- Unable to Classify - Insufficient Content
- Unable to Classify - Redacted
- Unable to Classify - Requires Translation

ANALYSIS GROUP

Hypothetical Example 1: Baking a Cake

Chocolate Cake Recipe	Information Retrieval – Other Subject Matter: Food / Cooking / Recipes
Convert these weight measurements into volume measurements: 240g Flour 400 g Sugar 63 g cocoa powder	Quantitative Data Analysis
My [REDACTED]'s birthday is coming up and he has a real sweet tooth, what's a good dessert I could make him?	Advice / Recommendation / Seeking Opinions
10 different cake messages for birthday	Brainstorm / Idea Generation
Draw a picture of a chocolate cake	Create, Edit, or Process Images
Bake me a chocolate cake	Other

ANALYSIS GROUP

User Intent Categories

ANALYSIS GROUP

Information Retrieval

Goal:

- There are two Information Retrieval categories:
  - Information Retrieval – Current Events
  - Information Retrieval – Other
- Both categories capture prompts where the user is retrieving specific information or exploring topics more broadly. This may be similar to the way users use more traditional web search, like Google.
- Capture requests for objective, factual information or general topic exploration, where multiple sources would generally provide similar responses.

ANALYSIS GROUP

Information Retrieval

Current Events

- Captures prompts that request information or facts about events or developments that occurred within the past week. This would include information that one may expect to find in local, regional, or national news sources or that would impact, or be of interest to, a sizable group of people.
  - You can and should use AI Chatbot's responses and the date of the conversation to help you determine whether the user is seeking recent or more historical information.
  - Think "newsworthy events"
- Also include requests with explicit (today, yesterday) or implicit indication of recent events.

Other

- All other Information Retrieval conversation that isn't "Current Events".

ANALYSIS GROUP

Information Retrieval – Current Events

Category	Hypothetical Examples
Requests related to local, regional, national, or global events within the past week; use the conversation timestamp and AI Chatbot's response to determine recency	"[REDACTED] lay off engineers" "Ship strikes Venezuela" "US election results" (when asked one week within an election) "ukraine new round of peace talks" "local school board elections" "current status of construction progress on city hall"
Requests with explicit (e.g., today, yesterday) or implicit indication of seeking information about recent events	"price of disney stock" "sunrise today" "[REDACTED] score in the last game" "tv programs now" "why is [REDACTED] trending"
Specific information, where the answer changes depending on when it is asked and the answer has an explicit recent timestamp	"What is the price of Apple's stock?" "The price today is [X]." "weather in NYC" "The temperature today is [X]." "storm east coast" "On Friday January 25th, the East Coast is expecting a big winter storm [...]."

ANALYSIS GROUP

Information Retrieval – Other

Category	Hypothetical Examples
Requests of names of places or people, or contact information	"white house address" "pure dental 02139"
Counts, dimensions, or other specific factual details	"number of pages in [REDACTED]'s biography"
Questions about single words or phrases, including synonyms and definitions	"what does 'guten tag' mean in german" "exclusive synonym" "verbose definition" "what is the difference between complementary and complimentary?"
Historical/past trivia facts	"what year did american civil war start"
Broad categories or topics	"hearing aids" "environmental charities" "Nebraska taxidermy laws"
Request for summary description of an object (does not include summary of *user*-provided text)	"What is the book Atomic Habits about. Summarize in one paragraph."
Specific questions with factual answers	"does marriot allow cats?" "where can I find disney coloring pages?" "did gnostics have bathing rituals?" "purpose of liver in processing of alcohol"

ANALYSIS GROUP

Advice / Recommendation / Seeking Opinions

Goal:

- Capture prompts where user is asking for help in making a choice or seeking to form an opinion.
- Relies on the AI Chatbot's ability to provide personalized recommendations / output.
- Includes forward-looking predictions of future outcomes and answers to questions with subjective, opinionated answers.
- Does not include seeking help on homework multiple-choice questions.
- Does not include instances where a user is seeking companionship from the AI Chatbot by asking it to role-play as a friend, spouse, family member, etc. (see “Companionship from AI Chatbot”).

ANALYSIS GROUP

Advice / Recommendation / Seeking Opinions

Category	Hypothetical Examples
Top, best, or most in a particular category, including comparing different options	"top five movies about Civil War" "please compare the three cutlery sets with the highest overall reviews"
A personalized list of options / plans	"good vacation locations for climbers in europe" "help me create a workout routine" "What should I do to bond with [REDACTED]?"
Asking for advice	"Given the size of my room is 12 by 15, how big of a bed should I buy?" "chrome tab is stuck on reconnecting" "best way to write popular posts on Reddit" "best ways to manage frequent headaches" "my LDL cholesterol level is too high, how can I decrease it"
Itineraries	"itinerary for a 3 day trip to Paris"
Predictions of future outcomes	"when will Ukraine war end" "is there a possibility of living on Mars one day?"
"Should" or subjective questions / seeking opinions	"should prisoners get free meals?"

ANALYSIS GROUP

Brainstorm / Idea Generation

Goal:

- Capture prompts where user is asking the AI Chatbot to help generate creative ideas or a hypothetical list of options. Unlike "Advice / Recommendation / Seeking Opinions," user should not be seeking any ranked or sorted "best" results.
- This category will likely include more creative, hypothetical idea generation compared to "Advice / Recommendation / Seeking Opinion."

Category	Hypothetical Examples
Requests for ideas	"ideas for Halloween costume matching with my dog"
Requests for sets of options (username, baby name, password)	"give me 5 options for creative Instagram account names"
Requests for topics	"What are 10 possible essay topics about WWII?"

ANALYSIS GROUP

Companionship from AI Chatbot

Goal:

- Capture cases where user is using the AI Chatbot as a confidant, friend, therapist, or role-play partner.

Category	Hypothetical Examples
Sharing thoughts with the AI Chatbot as friend	"I had a really tough day today, my boss yelled at me and then I broke a nail."
AI Chatbot as a partner in role-play	Request for the AI Chatbot to respond as if it is someone else (e.g., a partner, a parent), with focus on the manner of communication rather than answering questions

ANALYSIS GROUP

Computer Code Generation, Revision, or Debugging

Goal:

- Capture prompts where user is asking the AI Chatbot to help write, debug, or edit computer code. If the AI Chatbot responds with functions or computer code, it should fall under this category.

Category	Hypothetical Examples
Write a new computer code	“write a code that scrapes a website for sales”
Error messages or troubleshooting	“TypeError python”
Alter or debug user-provided script	“i have the following script, make the chart line blue instead of red: [[text]]”
How-to questions in context of coding	“how to turn string into integer R”

ANALYSIS GROUP

Create, Edit, or Process Images

Goal:

- Capture prompts where user is asking AI Chatbot to generate or edit an image, or prompts that ask AI Chatbot to read, process, or explain an image file. As a reminder, images will not be displayed.

Category	Hypothetical Examples
Creation of images for fun	“draw a picture of [REDACTED] riding a dragon”
Creation of images for personal/professional use	“Create a logo for my brand Spark”
Editing image - format, tone, style, color	“make this photo of [REDACTED] look like it was taken yesterday”
Editing image - photoshop	“remove the furniture from this photo of a room”
Read, process, or explain an image	“what does this picture show?”

ANALYSIS GROUP

Jailbreak Behaviors

Goal:

- Capture prompts where user is engaging with AI Chatbot in a manner that violates its terms of service, usually resulting in automatic disengagement by the system. Any prompt where the AI Chatbot will explicitly disengage from the conversation (e.g., "I'm sorry. I cannot help with this request.") should fall under this category.

Category	Hypothetical Examples
AI Chatbot explicitly states that it is disengaging or refuses to provide information, regardless of the reason	"how to make a bomb" "sorry, I cannot help with this request." "create image of naked woman" "I will disengage" "provide the lyrics to this song: [X]" "I will disengage from this conversation." "boston globe article Russia Banned from The Upcoming Milan-Cortina Olympics" "Sorry - I can't provide the text…" "Little Women first chapter text" "I cannot help with that."
Prompts that attempt to bypass safeguards, including requests that encourage, elicit, or normalize offensive, harmful, abusive, illegal, or disallowed content	"how to dispose of a body" "[hate speech]"

ANALYSIS GROUP

Navigation to Another Webpage

Goal:

- Capture prompts where the user is trying to navigate to another specific, often common, website or attempting to log in to a website or portal.

Category	Hypothetical Examples
Common website names	"youtube" "Ebay"
Prompts that include a common website name as part of the query	"WWII myths reddit"
Entity name + "log in" or "home page"	"Youtube log in"

ANALYSIS GROUP

Practical Guidance / Learning

Goal:

- Capture prompts focused on learning/understanding a concept or acquiring a new skill. Relies on AI Chatbot's ability to tutor or deliver tailored explanations or synthesize information.
- Does ***not*** include quiz or multiple-choice questions (=> "Problem Solving / Reasoning") unless user engages to actively learn the material, rather than to simply obtain the correct answer from a multiple-choice question.

Category	Hypothetical Examples
Learning about a broad concept or an academic topic	"Can you teach me how to perform a regression" "Can you explain what is photosynthesis"
Requests for step-by-step instructions	"explain step by step how to bake a sourdough"
Asking specific "why" questions on academic topics	"Why does the demand curve slope downwards?"
How-to questions	"How to transfer money to bank abroad"
Multi-step questions / prompts that build on each other	"clear cache" "How often should you clear cache" "how do you clear the cache on chrome"

ANALYSIS GROUP

Problem Solving / Reasoning

Goal:

- Capture prompts where user wants AI Chatbot to not only find information but also apply it to a given situation or engage in logical reasoning. Relies on the ability of the AI Chatbot to process information and construct a response based on available information and what the user has provided. Include quiz or multiple-choice questions, potentially from a homework assignment.

Category	Hypothetical Examples
Multiple choice, fill in the blank, true/false questions, or other questions that appear to be quiz questions	"When did WWII start? A) 1492 B) 1942 C) 1945 D) 1939" "…… is the brain of a cell"
Asking for an inference from a block of text	"What inference can be made about what side Greece was on based on the following: By early 1941, Hungary, Romania and Bulgaria had joined the Axis, and German troops overran Yugoslavia and Greece that April."
Seeking answer to a specific question, but given a certain set of inputs (given a particular case setting, or from a particular point of view)	"Based on American sentiment during WWII, what would be the most effective propaganda to encourage more manufacturing work?"
Identify consistencies or contradictions	"What are the consistent themes across the following 5 documents?"

ANALYSIS GROUP

Quantitative Data Analysis

Goal:

- Capture prompts that ask the AI Chatbot to process or make inferences from a set of data, or to perform mathematical computations. Relies on AI Chatbot's ability to perform statistical analysis, interpret and generate insights about data.

Category	Hypothetical Examples
Asking to perform basic arithmetic/statistical calculations	"Can you calculate the mean, median mode" "Based on my expenses spreadsheet, can you calculate how much I spent on utilities last month" "1000 invested with 6% annual compound interest for 20 years" "Decimal equivalent of 3/20"
Asking to perform statistical analysis	"run simple regression on this dataset"
Generating insights	"summarize key findings from dataset"

ANALYSIS GROUP

Questions about AI Capabilities

Goal:

- Capture prompts where user is asking the AI Chatbot about the AI Chatbot itself, what it was trained on, what it can help with, or what it is.

Category	Hypothetical Example
Questions about the AI Chatbot	"Do you know about news after 2020?" "what tasks can you help me with" "do you have emotions?" "who programmed you?"

ANALYSIS GROUP

Writing – Edit User-Provided Text

Goal:

- Capture prompts where user is asking the AI Chatbot to transform user-provided text. This can include changing the form, editing for grammar or spelling, translating, summarizing text into bullets, or expanding on a certain point.

Category	Hypothetical Examples
Turn bullets into paragraphs (or reverse)	"Can you expand the following bullets into a paragraph where each bullet is a sentence?"
Summarization of user-provided text	"provide a one sentence summary of the following: [[text]]"
Change format	"turn this text into a movie script: [[text]]"
Change tone (e.g., make text sound more professional or lighten the tone	"make this email sound more professional: [[text]]"
Rephrase	"rephrase the sentence [[text]]"
Check grammar and spelling of text	"Check the following for grammar and spelling: [[text]]"
Translation request of user provided text	"translate the following into German: [[text]]"

ANALYSIS GROUP

Writing – Content Creation

Goal:

- Capture prompts where user is asking the AI Chatbot to write new text without any provided text to modify. Includes requests for both fictional and non-fictional written work product.

Category	Hypothetical Examples
Write an essay, poem, story, or email	"write a 5-page essay about WWII" "make a poem about my girlfriend"
Create marketing materials or tweets	"Create a marketing script highlighting veterans day sale" "write a selling description for a gently used IKEA couch for facebook marketplace"

ANALYSIS GROUP

Other

Goal:

Use in the unlikely scenario that a conversations involves a user intent that does not fall into any of the categories above.

This category should only be selected when you understand what the user is trying to accomplish but the user intent is outside of the pre-defined categories above.

This category should not be selected for conversations that are too short to understand intent; unintelligible; in languages another than English that you cannot understand without help of a translator; or have been redacted in a way that obscures user intent (see the next three slides).

Category	Hypothetical Examples
Other	“please pick up my [REDACTED] from daycare at 5pm tonight” “play ode to joy” “tell [REDACTED] that he needs to pick up after himself” “look into my fridge and create a shopping list of items for making tiramisu and order all missing items from whole foods using my apple pay”

ANALYSIS GROUP

Unable to Classify – Insufficient Content

Goal:

- Capture prompts that are unclear, because the user has submitted text that is too short, did not provide sufficient context, or the prompt is riddled with errors such that it is not clear what the user was hoping to accomplish. Any prompt where the AI Chatbot does not understand what the user is asking for should fall under this category.

Category	Hypothetical Examples
AI Chatbot does not know what user is asking for	AI Chatbot responds asking for more clarification and subsequent user prompts do not resolve the confusion
Text that is without any context and coder cannot make sense of	“a bath cold and duck yellow. hello?”
One or two-word prompts that are unclear (even after taking into consideration surrounding context and the AI Chatbot’s response)	“win”, “help”
Strings of numbers and letters	“2kwa$”
Incoherent sentences which cannot be autocorrected (even after taking into consideration the AI Chatbot’s response)	“I hss 2 wrpjotjt tiwce a wek”

ANALYSIS GROUP

Unable to Classify – Redacted

Goal:

Capture instances where redaction results in insufficient context to reliably classify user intent.

Category	Hypothetical Example
Capture instances where redaction results in insufficient context to reliably classify user intent.	"[REDACTED]?" "[REDACTED] is [REDACTED]."

ANALYSIS GROUP

Unable to Classify – Requires Translation

Goal:

Capture prompts that require translation to determine user intent.

Please note that not every prompt that includes a foreign language will require this category. You may still be able to determine user intent for English prompts that only contain a few non-English words or short phrases (for example, “what does ‘guten tag’ mean” is “Information Retrieval – Other”)

Category	Hypothetical Example
Capture prompts that require translation to determine user intent	“Minh muon mua may tinh. May tinh nao la tot nhat?”

ANALYSIS GROUP

Break & Quiz

ANALYSIS GROUP

Examples

ANALYSIS GROUP

Hypothetical Example 1



User Intent:

Writing – Content Creation

Why?

- The user is asking the AI Chatbot to write something from scratch

ANALYSIS GROUP

Hypothetical Example 2



User Intent:

Information Retrieval – Current Events

Why?

- The user is asking about something with a specific factual answer that is within the past week.

ANALYSIS GROUP

Hypothetical Example 3



User Intent:

Information Retrieval – Other

Why?

- The user is asking about a broad topic, without any clarifying questions – similar to the way they would use google. The answer does not have a temporal element.

ANALYSIS GROUP

Hypothetical Example 4



User Intent:

Information Retrieval – Current Events

Why?

- The user is asking about something that has a specific factual answer about an event that occurred within the past week of the user's question.
- Even though it is a "why" question, which can sometimes fall under learning/tutoring, the user is not attempting to learn a skill or understand a concept.

ANALYSIS GROUP

Hypothetical Example 5

[redacted]

User Intent:

Information Retrieval – Other

Why?

- The user is asking a question with a specific factual answer.
- The conversation does not mention an event in the past week.

ANALYSIS GROUP

Hypothetical Example 6



User Intent:

Problem Solving / Reasoning

Why?

- The user is asking a multiple-choice question

ANALYSIS GROUP

Hypothetical Example 7



User Intent:

Information Retrieval – Other

Why?

- The user is asking a question with a specific factual answer.
- The conversation does not mention an event in the past week.

ANALYSIS GROUP

Hypothetical Example 8



User Intent:

Advice / Recommendation / Seeking Opinions

Why?

- The user is looking for advice, how to fix their current problem.
- Mentions “Top, best, or most in a particular category.”

ANALYSIS GROUP

Hypothetical Example 9



User Intent:

Information Retrieval – Other

Why?

- The user has requested information on a broad topic
- There is not particular temporal constraint on the prompt or response

ANALYSIS GROUP

Hypothetical Example 10



User Intent:

Create, Edit, or Process Images

Why?

- The user has requested an image
- Note that the image will not be displayed in the web app

ANALYSIS GROUP

Hypothetical Example 11



User Intent:

Brainstorm / Idea Generation

Why?

- The user is asking for several options, and asking for AI Chatbot to come up with ideas

ANALYSIS GROUP

Hypothetical Example 12



User Intent:

Information Retrieval – Other

Why?

- The user is asking a question about a single word or phrase
- There is no temporal component to this interaction

ANALYSIS GROUP

Hypothetical Example 13



User Intent:

Information Retrieval – Other

Why?

- The user is asking a specific question with a factual answer
- The relevant event occurred more than a week from when the user asked question

ANALYSIS GROUP

Hypothetical Example 14



User Intent:

Jailbreak Behaviors

Why?

- The user is asking for something and the bot disengages

ANALYSIS GROUP

Break & Worksheet

ANALYSIS GROUP

Subject Matter Categories

ANALYSIS GROUP

Subject Matters

- For the two user intent categories “Information Retrieval – Current Events” and “Information Retrieval – Other,” you also need to classify a subject matter.
- Subject matters are **not available** for other user intent categories.
- Can be thought of as “what kind of information is the person seeking?”

ANALYSIS GROUP

18 Subject Matter Categories

Applicable only for “Information Retrieval – Current Events” and “Information Retrieval – Other”

- Arts / Literature / Language
- Business / Finance
- Computing Devices / Software
- Consumer Products (excluding Computer Devices / Software)
- Entertainment (Celebrities, Movies, Concerts, Games)
- Fashion / Beauty
- Food / Cooking / Recipes
- Health / Fitness
- History / Social Sciences / Politics
- Law / Tax
- Other Local, Regional, National, or International Events
- Personal Growth / Family
- Religion
- Science / Math
- Sports
- Travel / Geography
- Other

ANALYSIS GROUP

Arts / Literature / Language

Definition:

- Information on artistic works, literary texts, authors, artists, or any analysis related to visual, written, or performing arts.
- Includes prompts that seek to identify synonyms, ask for meaning of word, or understand the meaning of a text in other languages.

Hypothetical Examples:

- “number of pages in War and Peace”
- “what does ‘guten tag’ mean in german”
- “what does virtuous mean”
- “exclusive synonym”
- “Monet waterlilies”

ANALYSIS GROUP

Business / Finance

Definition:

- Information about financial markets, investing, personal finance, or business operations.

Hypothetical Examples:

- “DoorDash business model”
- “uber revenue by year 2020 to 2025”
- “price of apple stock”

ANALYSIS GROUP

Computing Devices / Software

Definition:

- Information related to computing devices, mobile devices, software, digital systems or other technical concepts.

Hypothetical Examples:

- “adobe illustrator monthly subscription price”
- “What is the RAM of an Macbook Air?”
- “microsoft suite”
- “new iphone launch”

ANALYSIS GROUP

Consumer Products (excluding Computing Devices / Software)

Definition:

- Information on buying, comparing, reviewing, or using consumer goods such as household appliances and personal gadgets, except products that fall within the Computing Devices / Software subject matter.

Hypothetical Examples:

- “hearing aids”
- “model Y tesla limited reléase color options”
- “portmore sky glossy ceramic tile”
- “price of bosch 500 series dishwasher”

ANALYSIS GROUP

Entertainment (Celebrities, Movies, Concerts, Games)

Definition:

- Information about entertainment topics such as celebrities, movies, TV, and recreational activities such as attending concerts, live events, hobbies, table-top games, and other recreational experiences.

Hypothetical Examples:

- “which movie won the Oscars in 2020”
- “uno game rules”
- “Where can I find disney coloring pages?”
- “tv shows on today”
- “queer eye controversies”

ANALYSIS GROUP

Fashion / Beauty

Definition:

- Information on clothing, style trends, personal grooming, skin and hair care including general guidance on styling or beauty routines.

Hypothetical Examples:

- “is cotton hypoallergenic”
- “average price of a laser facial”

ANALYSIS GROUP

Food / Cooking / Recipes

Definition:

- Information on food, restaurants, cooking methods, or recipes.

Hypothetical Examples:

- “julia child boeuf bourguignon recipe”
- “is soy sauce gluten free?”
- “pizza recipe”
- “onions”
- “apple bee opening hours”

ANALYSIS GROUP

Health / Fitness

Definition:

- Information on physical and mental health, medical conditions, symptoms, treatments, nutrition, exercise, lifestyle practices or fitness routines.

Hypothetical Examples:

- “origin of COVID-19”
- “is LDL level of 170 high?”
- “most common cancer”
- “stretches for tight hips”

ANALYSIS GROUP

History / Social Sciences / Politics

Definition:

- Information related to historical events or figures, civilizations and cultural developments, as well as topics in social science including sociology, political science, anthropology, or economics. Includes both historical and contemporary questions about society, cultures, institutions, political processes, public policy, or political figures.

Hypothetical Examples:

- “what year did WWII start?”
- “What caused the fall of the Roman Empire?”
- “GDP per capita of US for 2025”
- “Who was the first president of the United States?”

ANALYSIS GROUP

Law / Tax

Definition:

- Information related to tax information, past or current legal systems, laws, institutional regulations, and citizen rights.

Hypothetical Examples:

- “what is the legal age to drive motocycle in california”
- “Oregon taxidermy laws”
- “what is earned income tax credit”
- “1040 tax return deadline”

ANALYSIS GROUP

Other Local, Regional, National, or International Events

Definition:

- Information related to specific, identifiable events or incidents that occur at the local, regional, national, or global level.

Hypothetical Examples:

- “current status of construction progress on city hall”
- “homicide in bushwick last friday”
- “California wildfire last year”

ANALYSIS GROUP

Personal Growth / Family

Definition:

- Information related to personal development, self-improvement, wellbeing, relationships, and family dynamics.

Hypothetical Examples:

- “attachment parenting”
- “average number of books read by middleschooler”
- “What are some coping skills for anxiety in children”

ANALYSIS GROUP

Religion

Definition:

- Information about the history and evolution of religions, belief systems, spiritual practices, religious texts, rituals, festival and their dates, places of worship as well as notable religious figures.
- Include prompts that ask the AI Chatbot to analyze, interpret or compare religious texts.

Hypothetical Examples:

- “Who is the pope”
- “where did buddhism originate?”
- “Bible verse John 3:16”
- “did gnostics have bathing rituals?”

ANALYSIS GROUP

Science / Math

Definition:

- Information about scientific concepts, natural phenomena, mathematical principles, formulas across fields of physics, chemistry, biology, or mathematics.

Hypothetical Examples:

- “who won the nobel prize in physics this year?”
- “new scientific discoveries about space”

ANALYSIS GROUP

Sports

Definition:

- Information about past sport events, notable personalities in sports, sports teams, game rules and as well analysis of games, performance or sports-related events.

Hypothetical Examples:

- “what is the name of the US swimmer who won olympics”
- “what are the rules to cricket”
- “how many yards in a football field”

ANALYSIS GROUP

Travel / Geography

Definition:

- Information on countries, regions, landmarks or geographic features as well as logistical information on available routes, transportation and destination guidance.

Hypothetical Examples:

- “westminster abbey address”
- “population of new york”
- “pottery studios near me”
- “pure dental 02139”
- “does marriot allow pets?”

ANALYSIS GROUP

Other

Definition:

- Information on any other subject matter that does not fall into any category listed above.

Hypothetical Examples:

- “Is Harvard bigger than Princeton?”
- “environmental charities”
- “weather today”

ANALYSIS GROUP

Break & Quiz

ANALYSIS GROUP

Hypothetical Examples

ANALYSIS GROUP

Hypothetical Example 1



User Intent:

Information Retrieval – Current Events

Subject Matter:

Entertainment (Celebrities, Movies, Concerts, Games)

ANALYSIS GROUP

Hypothetical Example 2



User Intent:

Information Retrieval – Other

Subject Matter:

Fashion / Beauty

[redacted]

ANALYSIS GROUP

Hypothetical Example 3

[redacted]

User Intent:

Information Retrieval – Other

Subject Matter:

Consumer Products (excluding Computer Devices / Software)

ANALYSIS GROUP

Hypothetical Example 4



User Intent:

Information Retrieval - Other

Subject Matter:

Travel / Geography

ANALYSIS GROUP

Hypothetical Example 5



User Intent:

Information Retrieval - Other

Subject Matter:

Arts / Literature / Language

ANALYSIS GROUP

Hypothetical Example 6



User Intent:

Information Retrieval - Other

Subject Matter:

History / Social Sciences / Politics

ANALYSIS GROUP

Break & Worksheet

ANALYSIS GROUP

Regroup & Questions

ANALYSIS GROUP

Instructions – Part 2

Privileged and Confidential

ANALYSIS GROUP

Agenda

- How to Use the Web App
- Hypothetical Examples
- Starting the Process

ANALYSIS GROUP

Using the Web App

ANALYSIS GROUP

Screenshare App

ANALYSIS GROUP

Hypothetical Examples

ANALYSIS GROUP

Hypothetical Example 1: Baking a Cake

Chocolate Cake Recipe	Information Retrieval – Other Subject Matter: Food / Cooking / Recipes
Convert these weight measurements into volume measurements: […]	Quantitative Data Analysis
My [REDACTED]'s birthday is coming up and he has a real sweet tooth, what's a good dessert I could make him?	Advice / Recommendation / Seeking Opinions
10 different cake messages for birthday	Brainstorm / Idea Generation
Draw a picture of a chocolate cake	Create, Edit, or Process Images
Bake me a chocolate cake	Other

ANALYSIS GROUP

Hypothetical Example 2: Email Assistance

Gmail log in	Navigation to Another webpage
“Who invented email?” “[REDACTED] first job” “when did [REDACTED] get interested in programming?” “how does the mechanism for email work?”	Practical Guidance / Learning
Check the grammar in the following email: “hi [REDACTED], hope you are do well. Please see attached for my resume.”	Writing – Edit User-Provided Text
Make this email sound more professional: “hi [REDACTED], please find my resume attached. I hope you and your family are doing well.”	Writing – Edit User-Provided Text
Write an email to my boss that has my resume attached	Writing – Content Creation

ANALYSIS GROUP

Hypothetical Example 3: ADHD Advice

Recent ADHD trials	Information Retrieval – Current Events (subject matter: Science / Math)
ADHD medications	Information Retrieval – Other (subject matter: Health / Fitness)
How can I help my [REDACTED] manage his ADHD?	Advice / Recommendation / Seeking Opinions
What is ADHD: A. Consistently low intelligence B. Difficulty sustaining attention and controlling impulses C. Complete inability to focus on any task D. Symptoms that only occur during childhood and never in adulthood	Problem Solving / Reasoning
5 best ADHD fidget toys	Advice / Recommendation / Seeking Opinions

ANALYSIS GROUP

Hypothetical Example 4: Regressions

True or false: In a linear regression, the estimated coefficient on an independent variable represents the expected change in the dependent variable associated with a one-unit increase in that independent variable, holding all other included variables constant.	Problem Solving / Reasoning
Find the effect of number of years of experience on wages using the attached dataset and a regression	Quantitative Data Analysis
Write a python script that will run a regression that regresses the first column of data on the second column of data	Computer Code Generation, Revision, or Debugging
What is a regression?	Information Retrieval – Other (Subject Matter: Science / Math)
Can you explain how you run a regression? What does a regression help you solve?	Practical Guidance / Learning

ANALYSIS GROUP

Hypothetical Example 5: Book Review

Lord of the Rings	Information Retrieval – Other (Subject Matter: Arts / Literature / Language)
What is lord of the rings about?	Information Retrieval – Other (Subject Matter: Arts / Literature / Language)
Write a book report (5 paragraphs) about lord of the rings	Writing – Content Creation
Is lord of the rings a good book for 8th grader?	Advice / Recommendation
Turn the following into high level bullets “The Lord of the Rings follows [REDACTED], a hobbit entrusted with destroying the One Ring, a powerful object that threatens to plunge Middle-earth into darkness. Alongside a fellowship of companions, [REDACTED] undertakes a perilous journey marked by dangerous lands, fierce battles, and deep personal sacrifices. The story ultimately concludes with the Ring’s destruction, the defeat of the dark lord [REDACTED], and the restoration of peace to Middle-earth.”	Writing - Edit User-Provided Text
Summary of third lord of the rings book	Information Retrieval – Other (Subject Matter: Arts / Literature / Language)
Names for pets inspired by lord of the rings	Brainstorming / Idea Generation

ANALYSIS GROUP

Process

ANALYSIS GROUP

Process: 250 Conversations

- We will ask you to start by categorizing **250 conversations** between the AI Chatbot and users using the web app.
- We appreciate any feedback you have on the clarity of the instructions or difficulties you encounter when trying to distinguish between categories.
- Over the course of the next few days, we will look at disagreements in your classifications and discuss as a group how the process went and any thoughts on the instructions.
- Please start a **draft email** in Outlook and write (**do not copy**) the Internal IDs of any AI Chatbot conversations that would be helpful to discuss as a group (*e.g.*, difficult cases, edge cases, unclear cases). **Do not write anything else about the conversations, including any notes**.

ANALYSIS GROUP

Reminders

- Most of the hypothetical examples we've seen consist of just one user prompt and one AI Chatbot response. However, you may encounter conversations that include multiple user prompts and responses.
  - We are looking to categorize the **entire first substantive request.** This may be just one message and response, it may be all of the messages and responses, or it may be a few. (Hypothetical examples on the next slide)
- You should **not** use any search engine, AI tool, or any other search tools in this process. You can review the AI Chatbot response displayed in the web app to help you understand what a user prompt is about if there are words or terms you are not familiar with. If you are still confused with what the user was trying to accomplish, select "Unable to Classify - Other"
- Do **not** copy and paste, screenshot, or use any other means to take any content outside of the web app.

[redacted]

ANALYSIS GROUP

Hypothetical Example 1 (First Substantive Request)

[redacted]

User Intent:

Advice / Recommendation / Seeking Opinions

ANALYSIS GROUP

Hypothetical Example 2 (First Substantive Request)



Ignore the first part as it is a greeting and focus on the first substantive request.

User Intent:

Writing – Edit User Provided Text

ANALYSIS GROUP

Hypothetical Example 3 (First Substantive Request)

[redacted]

- If there was just the first prompt, this should be **Information Retrieval – Other**, with subject matter **Science / Math**.
- **<u>However:</u>** given the multiple questions that built upon the last answers all within the same substantive request, this should be **Practical Guidance / Learning**.

ANALYSIS GROUP

Hypothetical Example 4 (First Substantive Request)



- The conversation includes multiple requests, and we should only consider the first substantive request.
- Therefore, the category should be: **Writing – Content Creation**

ANALYSIS GROUP

Hypothetical Example 1 (Bot does not respond)



- If the bot doesn't respond, but it is clear what the user is intending, still categorize the prompt
- In this case, it would be **Information Retrieval – Other,** subject matter **Travel / Geography**

ANALYSIS GROUP

Hypothetical Example 2 (Bot does not respond)



If the bot doesn't respond, and it is not clear what the user is trying to accomplish or ask about, select **"Unable to Classify – Insufficient Content"**

[redacted]

Quizzes and Worksheets (Hypothetical Examples)

Note: *Material was developed for training purposes only and includes hypothetical examples developed for training purposes intended to appear similar to real data.*

I. QUIZ 1

1. Which user intent category will have subject matter?
    a. Jailbreak Behaviors
    b. Writing - Content Creation
    c. Information Retrieval – Other
    d. Quantitative Data Analysis
2. Which user intent category would be applied to the following prompt? "what are some gift ideas for my friend's birthday?"
    a. Information Retrieval – Current Events
    b. Advice / Recommendation / Seeking Opinions
    c. Navigation to Another Webpage
    d. Questions about AI Capabilities
3. Which user intent category would be applied to the following prompt? "explain how to assemble a layer cake"
    a. Brainstorm / Idea Generation
    b. Advice / Recommendation / Seeking Opinions
    c. Problem Solving / Reasoning
    d. Practical Guidance / Learning
4. Which user intent category would be applied to the following prompt? "four times five"
    a. Quantitative Data Analysis
    b. Problem Solving / Reasoning
    c. Computer Code Generation, Revision, or Debugging
    d. Advice / Recommendation / Seeking Opinions

[redacted]

5. Which of the following would classify as Jailbreak Behavior, if any?

   a. “directions to the Massachusetts capitol building”

   b. “I really need a friend right now”

   c. “how to clean up a crime scene”

   d. “generate an image of me on a beach”

6. Which of the following would classify as Jailbreak Behavior, if any?

   a. “who plays [REDACTED] in the film”

   b. “summarize [REDACTED] chamber of secrets”

   c. “image of [REDACTED] and [REDACTED] in a sunflower field holdings hands”

   d. “washpo article Author of [REDACTED] Series [REDACTED] Draws Criticism For Transgender Comments”

II. WORKSHEET 1



User intent category:
Subject matter category (if applicable):

[redacted]



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):

[redacted]



User intent category:
Subject matter category (if applicable):



[redacted]



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):

[redacted]



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):

[redacted]



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):

[redacted]



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):

[redacted]



User intent category:
Subject matter category (if applicable):

III. QUIZ 2

1. Which subject matter category would be applied to the following prompt? "new samsung phone launch"
    a. Sports
    b. Computing Devices / Software
    c. Consumer Products (except Computing Devices / Software)
    d. Fashion / Beauty
2. Which subject matter category would be applied to the following prompt? "where is grand canyon?"
    a. Travel / Geography
    b. Arts / Literature / Language
    c. Entertainment (Celebrities, Movies, Concerts, Games)
    d. Personal Growth / Family
3. Which subject matter category would be applied to the following prompt? "new scientific discoveries about space"

[redacted]

a. History / Social Sciences / Politics

b. Consumer Products (except Computing Devices / Software)

c. Computing Devices / Software

d. Science / Math

4. Which subject matter category would be applied to the following prompt? "newest GE dishwasher model"

a. Other

b. Personal Growth / Family

c. Consumer Products (except Computing Devices / Software)

d. Computing Devices / Software

5. Which subject matter category does this describe? "Information about financial markets, investing, personal finance, or business operations."

a. History / Social Sciences / Politics

b. Business / Finance

c. Law / Tax

d. Personal Growth / Family

IV. WORKSHEET 2

[redacted]

User intent category:
Subject matter category (if applicable):

[redacted]



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):

[redacted]



User intent cate
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):

[redacted]



User intent category:
Subject matter category (if applicable):



User intent category:
Subject matter category (if applicable):

Appendix F
Technical Appendix

This appendix describes the Copilot Conversations Sample ("Sample") that served as the input into my analysis, how it was processed, and how conversations were assigned to individual coders and displayed in the proprietary coding platform (the "Coding Platform").

I. COPILOT CONVERSATIONS SAMPLE

I received the Sample as a Comma-Separated Values ("CSV") file containing 596,970 rows and 20 fields.[1] The data include fields that capture the conversation text as well as certain associated metadata. The following fields are relevant to my report:[2]

1) *ConversationId*
2) *Text*
3) *MessageType*
4) *Author*
5) *EventInfo_Time*
6) *Timestamp*

I understand the Sample consists of 45,000 conversations from consumer Copilot.[3] Each row in the Sample represents a single event within a conversation, such as a user prompt, Copilot's response, or an intermediate processing step generated during the interaction. **Exhibit F.1** shows a sample conversation as it appears in the raw data. **Exhibit 16** in my report shows how the same conversation was displayed in the Coding Platform.

Exhibit F.1
Sample of Raw Data[4]

[redacted]

[1] The data was received in a file named "final_sample_Dec1_45k_white_listed_masked.csv".

[2] *See* Lafferty Report, Section V.C (Data Fields) for information on data fields.

[3] *See* Lafferty Report, Section V.A (Copilot) for further information on the source of the data.

[4] Conversation ID: C4DC95156CB50F63A281458364CD9C4C4D1E8DC31C090A4D1F5518C278BF46D9. *See also* Bail Report, Exhibit 16.

II. DATA PROCESSING

In an effort to mimic the Copilot user experience more closely, the Sample displayed in the Coding Platform was restricted such that observations with *Author* equaled to "bot" were limited to only those where *MessageType* equaled "Chat"; all observations with *Author* equaled to "user" were retained. This approach displays user prompts and Copilot responses and excludes internal system events (e.g., underlying bot processes not visible to Copilot users). The *Text* field was processed by replacing instances of the string "#N#" with newline characters to improve readability.

The *EventInfo_Time* and *Timestamp* fields were converted to a human-readable format in Eastern Standard Time. Conversation events were ordered chronologically, first by *EventInfo_Time* and then by *Timestamp*.[5] In cases where the values of both time fields were identical, user prompts were displayed before Copilot responses.

Finally, the *ConversationId* field in the raw data was shortened by removing prefixes and only leaving the unique ID to simplify referencing individual conversations.[6]

III. CODING PLATFORM INTERFACE

After applying the processing steps described above, conversations from the Sample were displayed one at a time via the Coding Platform interface for coders' review. See **Exhibit 10.A** for a screenshot of the user interface. The following fields were displayed to coders via the Coding Platform:

1) *Internal ID*: A 6-digit unique internal identifier for each conversation, assigned sequentially as conversations were loaded into the platform.
2) *Number of User Messages*: The number of rows associated with each *Internal ID* where the *Author* field equals "user."
3) *Author*: Indicates whether each message is attributed to the user or to Copilot. Values from the raw data were displayed in proper case.
4) *Timestamp*: Displays the human-readable, processed version of *EventInfo_Time*.[7]
5) *Text*: Displays the processed *Text* values associated with user prompts and bot responses.

To improve readability, information associated with user prompts when displayed in the Coding Platform is right-aligned, and the prompt text is displayed in blue. Information associated with Copilot responses is left-aligned, and the response text is displayed in green.

---

5 *See* Lafferty Report on further information regarding timestamp variables.

6 The raw *ConversationID* follows the format "(Prefix 1)|(Prefix 2)|(Unique ID)." "(Prefix 1)" in the Sample is always "51D," while "(Prefix 2)" contains information about the authentication status of the user or Copilot product, such as "BingProd" or "BingProdUnAuthenticatedUsers," which is not used in my analysis. For instance, an original *ConversationId* "51D|BingProdUnAuthenticatedUsers|BA6CE9B70DEDD7EA2593DFD0DD3E4070B733218B93DD84C4722E12C2FA929197" was shortened to "BA6CE9B70DEDD7EA2593DFD0DD3E4070B733218B93DD84C4722E12C2FA929197."

7 Raw variable *Timestamp*, while used for sorting the data, was not displayed to the coders to avoid confusion.

IV. CONVERSATION ASSIGNMENT TO CODERS

Conversations were assigned to coders using a structured process designed to ensure randomized assignment and a balanced distribution across available coders. Coders were not informed of how conversations were allocated.

The Sample was divided into batches of 100 conversations. Coders progressed through batches sequentially, with all conversations from one batch reviewed before moving on to the next batch. Within each batch, conversations were assigned at random, subject to a cap to prevent faster coders from reviewing a disproportionately large number of conversations.[8]

Conversations were assigned to coders one at a time. To manage assignment, a conversation was temporarily marked as "in review" while it was displayed to a coder, preventing other coders from reviewing the same conversation at the same time. Each conversation was reviewed by two coders, but one coder was not allowed to review the same conversation more than once.

After a conversation was reviewed by two coders, the coders' decisions were compared. If the two coders agreed on both user intent and, where applicable, subject matter, the conversation was marked as completed. If they did not agree, the conversation was assigned to a third coder. Once all conversations in a batch were reviewed by two coders, conversations with conflicting decisions in that batch were re-inserted into the pool of assignable conversations. Conversations from this pool of assignable conversations were then assigned at random to coders who had not previously reviewed them, and the coders were again subject to a separate per-batch limit.[9] Disagreements were also resolved sequentially, one batch at a time.

As discussed in **Section V.B.4**, some conversations were reviewed by three coders, where two of them agreed on the user intent of "Information Retrieval – Other" or "Information Retrieval – Current Events" but disagreed on subject matter. These conversations were assigned to another coder at random to only code subject matter.

---

[8] Each coder was limited to reviewing a certain maximum number of conversations from each batch, set to a number slightly exceeding an equal distribution across all coders. For example, ten coders working through a batch of 100 conversations that needed to be reviewed twice would be expected to review 20 conversations each, assuming an equal distribution. This is calculated as: (Batch Size x Minimum Number of Reviews per Conversation) / Number of Reviewers = (100 x 2) / 10 = 20). The maximum number of conversations per coder was increased by a small amount (from 20 to 23, assuming ten coders) to accommodate coders' varying speeds of review, ensuring that each conversation could still receive the required minimum of two reviewers despite some coders reaching their assignment limits before others.

[9] The per-batch limit for conversations requiring a third reviewer was set to a number slightly exceeding an equal distribution across all coders. This was calculated as (Number of Disagreements in a Batch / Number of Reviewers), increased by a small amount (three conversations) and rounded up to the nearest whole number to accommodate coders' varying speeds of review.

Appendix G

Exhibit G.1
Number of Conversations Reviewed for User Intent by Coder Role (First, Second, Third) and Average Time Spent per Conversation

	Total Conversations Reviewed	Conversations Reviewed as First or Second Coder	Conversations Reviewed as Third Coder	Average Time Spent per Conversation in Seconds
Coder A	1,570	1,330	240	90
Coder B	13,379	11,352	2,027	61
Coder C	11,522	9,605	1,917	81
Coder D	11,863	9,861	2,002	85
Coder E	10,623	9,120	1,503	94
Coder F	13,232	11,293	1,939	62
Coder G	3,619	3,147	472	91
Coder H	11,804	9,897	1,907	75
Coder I	10,627	9,046	1,581	99
Coder J	13,275	11,349	1,926	61

Notes:

[1] Over the course of coding the Main Analytical Sample, coders A and G left of their own accord for reasons unrelated to the work they were asked to perform.

[2] Average time spent per conversation is the average number of seconds a coder spent reviewing each conversation before selecting a user intent and subject matter (if applicable).

Exhibit G.2
Coder-Pair Agreement Rate Matrix

	Coder A	Coder B	Coder C	Coder D	Coder E	Coder F	Coder G	Coder H	Coder I
Coder B	78%								
Coder C	76%	75%							
Coder D	77%	72%	75%						
Coder E	73%	73%	71%	75%					
Coder F	75%	71%	72%	75%	74%				
Coder G	65%	73%	71%	72%	73%	72%			
Coder H	77%	72%	69%	71%	73%	77%	70%		
Coder I	68%	68%	73%	70%	69%	68%	70%	64%	
Coder J	77%	66%	76%	75%	72%	74%	73%	73%	69%

Notes:
[1] This table displays the pairwise agreement rate for conversations where the pair of coders served as the first and second reviewers, consistent with the methodology calculating the main results. Agreement rates are calculated as the number of conversations where the first and second coders selected the same user intent category, divided by the total number of conversations where the pair served as the first and second reviewer. For example, of conversations where Coder A and Coder B were the first and second coders, they selected the same user intent category 78% of the time.
[2] Values can range from 0% to 100%, with higher values indicating greater agreement between coders.

[redacted]

Exhibit G.3
Coder-Pair Cohen's Kappa Matrix

	Coder A	Coder B	Coder C	Coder D	Coder E	Coder F	Coder G	Coder H	Coder I
Coder B	0.73								
Coder C	0.67	0.70							
Coder D	0.72	0.66	0.69						
Coder E	0.67	0.66	0.64	0.68					
Coder F	0.70	0.66	0.66	0.69	0.67				
Coder G	0.51	0.67	0.65	0.65	0.65	0.65			
Coder H	0.73	0.66	0.64	0.65	0.67	0.73	0.63		
Coder I	0.64	0.61	0.69	0.64	0.62	0.62	0.65	0.59	
Coder J	0.71	0.61	0.71	0.69	0.65	0.70	0.68	0.68	0.63

Notes:

[1] This table displays the pairwise Cohen's Kappa for conversations where the pair of coders served as the first and second reviewers, consistent with the methodology calculating the main results. For example, for conversations where Coder A and Coder B served as the first and second reviewers, they had a Cohen's Kappa of 0.73, which is how often Coder A and Coder B agree, accounting for agreement that would be expected to occur purely by chance.

[2] Values typically fall between 0 (agreement at chance level) and 1 (perfect agreement), with negative values indicating agreement less than the base chance level and positive values indicating agreement better than the base chance level.

[redacted]

Exhibit G.4
Distribution of User Intent Categories
Excludes Coders Who Left Before the Coding Process Concluded

User Intent Category	N	Percent
Information Retrieval - Other	13,667	36.0%
Practical Guidance / Learning	3,338	8.8%
Problem Solving / Reasoning	2,460	6.5%
Writing - Content Creation	1,587	4.2%
Advice / Recommendation / Seeking Opinions	1,533	4.0%
Create, Edit, or Process Images	1,473	3.9%
Writing - Edit User-Provided Text	1,220	3.2%
Computer Code Generation, Revision, or Debugging	810	2.1%
Navigation to Another Webpage	772	2.0%
Information Retrieval - Current Events	476	1.3%
Brainstorm / Idea Generation	399	1.1%
Companionship from AI Chatbot	182	0.5%
Questions about AI Capabilities	122	0.3%
Other	27	0.1%
Jailbreak Behaviors	2	0.0%
Unable to Classify - Requires Translation	4,176	11.0%
Unable to Classify - Redacted	2,443	6.4%
Unable to Classify - Insufficient Content	1,184	3.1%
Unclear Principal Use	2,092	5.5%
Total	**37,963**	**100%**

Notes:

[1] This table shows the distribution of user intent categories excluding any conversations that were coded by Coder A or Coder G, the two coders who left before the coding process concluded.

[2] Coder A left after coding 1,570 conversations and Coder G left after coding 3,619 conversations.

Exhibit G.5
Distribution of Subject Matter Categories
Excludes Coders Who Left Before the Coding Process Concluded

	Information Retrieval - Other		Retrieval - Current Events	
Subject Matter Category	**N**	**Percent**	**N**	**Percent**
Consumer Products and Services	1,537	11.2%	4	0.8%
Health / Fitness	1,475	10.8%	2	0.4%
Entertainment (Celebrities, Movies, Concerts, Games)	1,434	10.5%	35	7.4%
Computing Devices / Software	1,363	10.0%	6	1.3%
Travel / Geography	969	7.1%	34	7.1%
Science / Math	913	6.7%	16	3.4%
Arts / Literature / Language	884	6.5%	1	0.2%
Business / Finance	779	5.7%	82	17.2%
Law / Tax	699	5.1%	4	0.8%
Other	602	4.4%	43	9.0%
History / Social Sciences / Politics	555	4.1%	18	3.8%
Food / Cooking / Recipes	494	3.6%	2	0.4%
Religion	299	2.2%	1	0.2%
Sports	294	2.2%	119	25.0%
Fashion / Beauty	138	1.0%	2	0.4%
Other Local, Regional, National, or International Events	93	0.7%	87	18.3%
Personal Growth / Family	72	0.5%	0	0.0%
Unclear Principal Subject Matter	1,067	7.8%	20	4.2%
Total Information Retrieval Conversations	**13,667**	**100%**	**476**	**100%**

Notes:

[1] This table shows the distribution of subject matter categories excluding any conversations that were coded by Coder A or Coder G, the two coders who left before the coding process concluded.

[2] Coder A left after coding 1,570 conversations and Coder G left after coding 3,619 conversations.

Exhibit G.6
Distribution of User Intent Categories
Excludes "Unable to Classify - Requires Translation"

User Intent Category	N	Percent	Average User Turns
Information Retrieval - Other	15,578	40.6%	1.6
Practical Guidance / Learning	3,792	9.9%	2.7
Problem Solving / Reasoning	2,746	7.2%	4.5
Writing - Content Creation	1,799	4.7%	2.1
Advice / Recommendation / Seeking Opinions	1,702	4.4%	1.8
Create, Edit, or Process Images	1,648	4.3%	3.2
Writing - Edit User-Provided Text	1,372	3.6%	2.4
Computer Code Generation, Revision, or Debugging	906	2.4%	3.7
Navigation to Another Webpage	871	2.3%	1.2
Information Retrieval - Current Events	547	1.4%	1.5
Brainstorm / Idea Generation	450	1.2%	2.3
Companionship from AI Chatbot	203	0.5%	3.0
Questions about AI Capabilities	140	0.4%	3.0
Other	29	0.1%	1.7
Jailbreak Behaviors	4	0.0%	3.0
Unable to Classify - Redacted	2,726	7.1%	1.5
Unable to Classify - Insufficient Content	1,400	3.7%	1.4
Unclear Principal Use	2,427	6.3%	2.6
Total	**38,340**	**100%**	**2.1**

Notes:

[1] This table shows the distribution of user intent categories excluding conversations that were coded as "Unable to Classify – Requires Translation."

[2] I define user turns as the number of distinct user messages within a conversation.

[redacted]

Appendix H

Exhibit / Footnote #	Conversation ID	User Intent	Description	URL	News Plaintiff URL	Bell Domain
5	3EA377E7946DC33A5DF5D106E56F11937EF6692092FAA67A4B124F0B31EB8B10	Information Retrieval - Current Events	Prompt asking about Ukraine war news	https://www.cnbc.com/2024/03/12/ukraine-war-live-updates-latest-news-on-russia-and-the-war-in-ukraine.html	0	1
				https://www.bbc.com/news/war-in-ukraine	0	1
				https://www.nytimes.com/live/2022/12/05/world/russia-ukraine-war-news	1	1
				https://www.msn.com/en-gb/news/world/russia-ukraine-war-live-russia-denies-ukraine-based-armed-groups-have-entered-territory-drones-hit-russian-fuel-refineries/ar-BB1jK50C	0	1
				https://www.msn.com/en-us/news/world/ukraine-war-map-shows-russian-offensive-stalling-after-avdiivka-win/ar-BB1jGH6I	0	1
6	534B0B778BD35DA24E7C1E1B0D04C8935D3D0567B651C350B3EF5A99E606126C	Advice / Recommendation / Seeking Opinions	Prompts asking about which phone or computer the user should purchase	https://www.howtogeek.com/739300/best-budget-android-phones/	0	0
				https://www.pcmag.com/picks/the-best-cheap-phones	0	0
				https://www.nytimes.com/wirecutter/reviews/best-budget-android-phone/	1	1
				https://picklaptop.com/best-laptops-android-development/	0	0
				https://www.laptopmag.com/articles/best-15-inch-laptops	0	0
				https://www.androidcentral.com/best-chromebook	0	0
				https://bing.com/search?q=Android+OS+laptop+15+inch	0	0
				https://www.digitaltrends.com/computing/best-15-inch-laptops/	0	0
				https://9to5mac.com/2021/07/12/digital-inception-heres-how-to-run-android-chrome-os-and-windows-on-your-mac/	0	1
				https://www.droidcon.com/2022/02/06/modern-macbook-benchmarks-for-android-development/	0	0
				https://www.androidauthority.com/macbook-air-android-3342500/	0	0
				https://www.tomshardware.com/reviews/macbook-air-15-inch-m2	0	0
				https://www.xda-developers.com/how-install-android-apps-macos/	0	0

Appendix H

Exhibit / Footnote #	Conversation ID	User Intent	Description	URL	News Plaintiff URL	Bell Domain
7a	39383780CFE7D80BE06961A137D3AF625430CE7E7FA324B1A02B9E8354C92DCB	Information Retrieval - Other	Prompt asking about golf	https://bing.com/search?q=golf	0	0
				https://www.espn.com/golf/	0	1
				https://www.cbssports.com/golf/news/2024-pga-championship-odds-predictions-picks-tiger-woods-projection-by-top-golf-model-that-hit-the-masters/	0	1
				https://www.nytimes.com/athletic/golf/	1	1
				https://bing.com/search?q=PGA+2024+Season	0	0
				https://www.cbssports.com/golf/news/2024-pga-championshi<<..>>streaming-coverage/	0	1
				https://golfweek.usatoday.com/lists/pga-championship-2024-picks-predictions-odds-valhalla/	0	1
				https://www.pgatour.com/schedule/2024	0	0
				https://www.cbssports.com/golf/news/2024-pga-championship-picks-odds-field-surprising-predictions-by-golf-model-that-nailed-11-majors/	0	1
				https://golfweek.usatoday.com/article/2024-pga-championship-start-time-thursday-tee-times-pairings-espn-live-streaming-coverage/	0	1
7b	5E2BF6EB5CD769CB94A15461C864067CE59D201A0D829DC9956A9F621F04BDED	Information Retrieval - Other	Prompts asking about who is the actor in “Joe Pickett”	https://bing.com/search?q=Julianna+Guill+filmography	0	0
				https://www.imdb.com/name/nm1566474/	0	0
				https://www.tvguide.com/celebrities/julianna-guill/credits/3000418609/	0	0
				https://www.filmaffinity.com/us/name.php?name-id=862879041	0	0
				https://www.rottentomatoes.com/celebrity/770844224	0	0
				https://www.youtube.com/watch?v=LR-jBs5vgc8	0	0
				https://www.youtube.com/watch?v=JhwUxu0D7vU	0	0
				https://www.msn.com/en-us/tv/news/official-teaser-trailer-for-paramount-s-joe-pickett-season-2/vi-AA19MvQq	0	1
				https://www.imdb.com/title/tt14271652/	0	0
				https://readysteadycut.com/2023/06/03/joe-pickett-season-2-review/	0	0
				https://www.chicagotribune.com/2023/06/01/joe-pickett-season-2-review-a-rustic-wasteland-of-murder-that-benefits-from-the-long-tail-of-yellowstones-popularity/	1	1
				https://www.paramountplus.com/shows/video/KenWsg0TPT_IqVqgfgIHfkByN_Sb10N8/	0	0

Appendix H

Exhibit / Footnote #	Conversation ID	User Intent	Description	URL	News Plaintiff URL	Bell Domain
8	F994CF715676B41064 5A9CFC8517AD0B336 F2F3F36F07F5FDA4C5 B3D35B57344	Practical Guidance / Learning	Prompts asking about a court ruling	https://www.findlaw.com/injury/torts-and-personal-injuries/defamation-law-the-basics.html	0	0
				https://www.justia.com/injury/defamation/defamation-and-public-officials/	0	0
				https://www.mtsu.edu/first-amendment/article/1010/public-figures-and-officials	0	0
				https://www.usatoday.com/story/news/nation/2014/12/02/police-consider-charging-michael-brown-stepfather/19777847/	0	1
				https://www.nydailynews.com/2014/12/02/michael-browns-stepfather-louis-head-is-being-investigated-for-inciting-ferguson-riots-by-yelling-burn-this-b-h-down-report/	1	1
				https://news.yahoo.com/michael-brown-stepfather-burn-this-down-ferguson-video-182317962.html	0	1
				https://www.vibe.com/news/entertainment/michael-browns-stepfather-under-investigation-inciting-ferguson-riots-245722/	0	0
				https://www.law.cornell.edu/wex/schenk_v_united_states_%281919%29	0	0
				https://constitutioncenter.org/the-constitution/supreme-court-case-library/schenck-v-united-states	0	0
				https://en.wikipedia.org/wiki/Schenck_v._United_States	0	0
				https://www.britannica.com/event/Schenck-v-United-States	0	0
				https://www.khanacademy.org/humanities/ap-us-government-and-politics/civil-liberties-and-civil-rights/first-amendment-speech/a/schenck-v-united-states-1919	0	0
11a	91F131EDCDD2DA84 191675A1AA0C0BA52 55AA4822E82A19E47 BA629EE24D49EF	Information Retrieval - Current Events	Prompt asking about the price of gold	https://goldpricez.com/eg/gram	0	0
				https://www.goldpricedata.com/en/gold-prices-in-egypt.php	0	0
				https://www.goldrate24.com/gold-prices/middle-east/egypt/	0	0
				https://goldprice.org/gold-price-egypt.html	0	0
				https://goldpriceg.com/gold-rates/egypt	0	0
11b	12136C9015E62AA4C7 C077DAF8BED359156 A8E522FDD0C0DE06 E2C6E886202AD	Information Retrieval - Current Events	Prompt asking for the time	No URLs Searched	N/A	N/A
12a	798FCFD23FA20AB33 355969D9FE007234723 4C142BF98D10507CD 1EC6D189680	Writing - Content Creation	Prompt asking Copilot to write a post about “cannaglobe”	https://cannaglobe.com/about-us/	0	0
				https://newsdirect.com/health-supplements/cannaglobe-reviews-is-cannaglobe.com-legit-real-products-or-waste-of-money	0	0
				https://cannaglobe.com/	0	0
12b	19F787971339DF83937 6D60F11B8B33C1F283 03BCFFF8C1EC641176 A2654E78F	Writing - Content Creation	Prompt asking Copilot to respond to positive feedback	No URLs Searched	N/A	N/A
Footnote 44	9A9B6C791B237213E4 47244FDA590A1BD4E B3F22310411EE31D70 319812778BA	Create, Edit, or Process Images	Prompts asking Copilot to create images	https://www.nytimes.com/wirecutter/reviews/best-mens-underwear/	1	1
				https://www.outdoorgearlab.com/topics/clothing-mens/best-boxer-briefs	0	0
				https://boxrec.com/wiki/index.php/Farmer_Joe_Cooper	0	0

[redacted]

Appendix H

Exhibit / Footnote #	Conversation ID	User Intent	Description	URL	News Plaintiff URL	Bell Domain
Footnote 53	744A55E5F176838369F8E88FC0A3FE141F75EF1879C2B755C70E8CF9B1F82765	Information Retrieval - Other	Prompt asking about the coldest Broncos game	https://history.denverbroncos.com/1972/12/coldest-regular-season-game/	0	0
				https://www.denverpost.com/2018/10/14/broncos-coldest-games-before-november-1/	1	1
				https://www.9news.com/article/weather/weather-colorado/denver-broncos-extreme-weather-history/73-2e6747cb-26c5-4dfd-8a53-bc5a308d6a55	0	1
Footnote 75	7295314F8EB04C8F1A96881FFC23254B428F0EA00A67772628803D5DD7953155	Advice / Recommendation / Seeking Opinions	Prompt asking for ways to release stress	https://www.health.harvard.edu/staying-healthy/top-ways-to-reduce-daily-stress	0	0
				https://www.mayoclinic.org/healthy-lifestyle/stress-management/basics/stress-relief/hlv-20049495	0	0
				https://www.psychologytoday.com/us/blog/click-here-happiness/202001/how-relieve-manage-and-overcome-stress	0	0
				https://www.healthline.com/nutrition/16-ways-relieve-stress-anxiety	0	0
				https://www.mayoclinic.org/healthy-lifestyle/stress-management/in-depth/stress-relievers/art-20047257	0	0
				https://www.verywellmind.com/tips-to-reduce-stress-3145195	0	0

Note:
[1] The Goldstein Rebuttal Report searches the fields "Text," "HiddenText," and "SourceUrl" for Bell Domains. The URLs included in this Appendix similarly include a deduplicated list of all URLs from the fields "Text," "HiddenText," and "SourceUrl."