

A Battle of the Chatbots: An Exploratory Comparison of ChatGPT, Gemini, Copilot, Claude, Perplexity, and HuggingChat

Karen Okamoto

Librarian and Associate Professor

John Jay College of Criminal Justice, City University of New York (CUNY)

Abstract

This exploratory study compares the output of the free versions of six generative artificial intelligence (AI) chatbots – ChatGPT, Gemini, Copilot, Claude, Perplexity and HuggingChat. Five questions were submitted to all of the chatbots and examined according to the length of the response, tone of the response, and any additional features provided by the chatbots such as citations, web links, and helpful follow-up questions. This comparison found that Perplexity excelled at responding to all of the questions though others provided strong outputs as well, pointing to the possible importance of using multiple chatbot for answers and for continually examining different chatbots given that the field of generative AI is developing and evolving rapidly.

Keywords: generative artificial intelligence (AI), chatbots, comparative study

Karen Okamoto is Reference, Instruction, Interlibrary Loan, and Institutional Repository Librarian at Lloyd Sealy Library, John Jay College of Criminal Justice, City University of New York (CUNY).

Introduction

Since the launch of ChatGPT in November 2022, several generative artificial intelligence (AI) chatbots have been challenging ChatGPT's apparent dominance (Weise et al., 2023). This exploratory study compares the output from the free versions of six generative AI (GenAI) chatbots – ChatGPT, Google's Gemini, Microsoft's Copilot, Claude, Perplexity and HuggingChat – in an attempt to look beyond ChatGPT and examine the current strengths and weaknesses of each chatbot from a library worker's perspective. Five questions were submitted to the six chatbots and analyzed based on the quality of the response and any additional features provided by the chatbots such as citations and helpful follow-up questions. To date, unlike in other disciplines, there have been no comparisons between the free versions of several GenAI chatbots in the library literature in terms of their output. This exploratory study aims to fill this gap and help inform further comparisons of these tools by addressing a need that Atwood (2024) writes about concerning libraries and AI. This exploratory study outlines some of the ways users can compare these different GenAI chatbots and includes an overview of each chatbot and its features.

Before launching into this study and the literature, an explanation of the terminology used throughout this paper follows. Toner's (2023) definition of generative AI as "a broad term that can be used for any AI system whose primary function is to generate content" is used here when referring to chatbots such as ChatGPT (para. 4). Toner (2023) writes that generative AI is different from other AI systems "that perform other functions such as classifying data (e.g., assigning labels to images), grouping data (e.g., identifying customer segments with similar purchasing behavior), or choosing actions (e.g., steering an autonomous vehicle)" (para. 4). Generative AI systems include image generators such as Midjourney and large language models that power ChatGPT and Claude. This study will refer to the chatbots as GenAI chatbots and not large language models (LLMs) given that Microsoft's Copilot, for example, is powered not only by an LLM – that is, the paid version of ChatGPT (4.0) – but it also uses the image generator DALL-E 3, web scraping via Bing, Microsoft's Natural Language Processing technology, and other technology (Muchmore, 2024). Similarly, Perplexity uses a combination of LLMs which connect to online sources (Perplexity Team, 2023). Toner (2023) defines LLMs as AI systems that work with language and are trained on large quantities of data using significant computational power. LLMs "dynamically create text on demand in response to a prompt" (Fernandez, 2023, p. 1). LLMs accomplish this by predicting what words to combine in a probable pattern based on its training data (Houston & Corrado, 2023). However, as Toner (2023) points out, "[t]here is some ambiguity about whether to refer to specific products (such as OpenAI's ChatGPT or

Google's Bard) as LLMs themselves, or to say that they are powered by underlying LLMs" (para. 8). (Note: Google's Bard is now called Gemini.) This paper will use the broader term GenAI to refer to the chatbots included in this comparison.

Literature Review

While at present the library literature does not include extensive comparisons of the free versions of different GenAI chatbots, the literature importantly highlights some of the concerns raised around GenAI. Pfeiffer (2024), for example, succinctly summarizes five growing concerns around GenAI for librarians, with specific references to ChatGPT: GenAI has intellectual limits, particularly free versions which the public at large is likely to use instead of subscription-based platforms; they hallucinate and generate misinformation; they rely on questionable information; they raise privacy issues concerning the information and questions collected from users; and lastly, they can generate harmful stereotypes despite purported guardrails. Others have pointed out, usually with reference to ChatGPT, that GenAI tools do not cite their sources (Hall & McKee, 2024; Houston & Corrado, 2023) or they create fake citations (Pun, 2023). The question of who owns the copyright for generated content has also been raised (Cox & Tzoc, 2023; Frederick, 2023; Houston & Corrado, 2023). Lastly, Hall and McKee (2024) recommend using ChatGPT cautiously, particularly for fact-checking, quantitative analyses of market trends, analyzing and visualizing complex data, and evaluating or analyzing current events.

The Possibilities of and Use Cases for GenAI Chatbots

Despite these concerns, the library literature suggests a number of possibilities and uses for GenAI chatbots, though mostly concerning ChatGPT. Houston and Corrado (2023) outline how academia and libraries can embrace ChatGPT by using it to brainstorm and explore subtopics, refine a research question, create an essay outline, and proofread. They also suggest using ChatGPT to create descriptive metadata for books and other items and to help library staff code. Hall and McKee (2024) see ChatGPT's usefulness in extracting information from documents such as contracts, writing social media posts and marketing copy, conducting reader's advisory, and providing management advice and data analysis. Cox and Tzoc (2023) suggest a range of applications for ChatGPT including being an alternative to search engines such as Google, answering basic reference questions via a 24/7 chat service, assisting with writing instructions by creating syllabi and textbooks, and helping with productivity tasks such as writing emails. Hendrick (2024b) points out the value of GenAI tools that

specifically use the Retrieval Augmented Generation (RAG) framework: RAG tools can connect to specific datasets such as Semantic Scholar or the full web and provide citations to real sources. She briefly highlights the differences between output produced by a non-RAG chatbot such as Claude and one that does such as Perplexity, a difference that will be demonstrated further by this exploratory study. Lastly, prompt engineering has been written about in the library literature as an area in which librarians can further develop their skills and assist users. Parallels have been drawn between the process of crafting an effective prompt and identifying keywords for a topic (Craddock & Wilson, 2023; Frederick, 2023; Lo, 2023; Zhang, 2023). Prompt engineering has been defined as “the process of optimizing textual inputs to AI models” (Lund, 2023, p.6). This includes providing additional context for a question like the intended audience for the information, to generate more relevant output.

On AI Literacy

One promising area of the library literature relevant to this study is the development of AI literacy concepts and criteria to evaluate the output of GenAI tools. For example, Archambault and Rincón (2024) evaluate the paid versions of six AI research tools – Litmaps, Connected Papers, ResearchRabbit, Elicit, scite, and Consensus – using the REACT Framework. REACT stands for Relevancy, Ease of Use, Assessing DEIA (diversity, equity, inclusion, and accessibility), Currency, and Transparency & Accuracy. In their Choice white paper, Hervieux and Wheatley (2024) argue that a framework for AI literacy, separate from the Association of College and Research Libraries (ACRL) Framework for Information Literacy, is needed to address some of the emerging skills that the existing information literacy framework does not include, such as prompt engineering, citing generated content, and critically evaluating output by considering questions of ethics and bias. While the ACRL Board is currently drafting AI competencies for release at ACRL’s 2025 conference, Hervieux and Wheatley (2024) propose an AI literacy framework that is comprised of six areas: knowing the basic principles of AI; understanding the differences between AI types; experimenting with AI tools; reviewing AI output; evaluating AI’s social impact; and lastly, engaging with AI discourse. Hendrick (2024c), addressing libraries that are considering subscription-based AI tools, writes about four areas to consider when evaluating GenAI applications: transparency of the responses or citations to the original document that produced the output; the efficiency of the tool in terms of reducing the user’s workload; whether the tool eliminates critical thinking or prevents researchers from retaining or enhancing “their critical thinking skills as they work through their research” (para.5); and lastly, privacy, or more specifically, the privacy policy for the

tool. Lo (2024) proposes seven key competencies for AI literacy: understanding AI system capabilities and limitations; identifying and evaluating AI use cases; using AI tools effectively and appropriately; critically assessing AI quality, biases, and ethics; engaging in informed AI discussions and collaborations; recognizing data privacy and security issues; and anticipating AI's impacts on library stakeholders. Lo (2024) firmly believes in the importance of library staff having hands-on familiarity with AI tools. In a similar vein, Oyarzun (2024, p. 39) values training library staff on AI competencies but also importantly adds that users should ask each AI tool about its thought process to gain a better understanding of its logic and assumptions. In their study of ChatGPT's ability to answer reference questions, Yang and Mason (2024, p. 110) write that students "should learn how to evaluate outputs from ChatGPT, Microsoft Bing, and Google Bard." In other words, evaluating output from different GenAI chatbots is an important and useful practice.

The Predominance of ChatGPT in the Library Literature

The research-based library literature on GenAI chatbots is currently rather limited in scope and is mostly concerned with analyses of ChatGPT, though some authors are exploring other options. Chen (2023), for example, compares responses between ChatGPT and the University of California Irvine's library chatbot (ANTswersays), Lehman College Library's Lightning Bot, and San Jose State University Library's KingBot. Chen posed one question to all of these chatbots: "What is the best library database for finding law reviews and legal cases?" (p. 126). Chen found that ChatGPT provided the best response by naming specific legal databases such as Westlaw and LexisNexis. Other library organizations and practitioner news sites such as the LibTech Insights blog (www.choice360.org/libtech-insights) and Computers in Libraries have provided helpful brief overviews and introductions to the latest GenAI tools such as scite.ai, "an academic search engine" (Hendrick, 2024a, scite.ai section), and Consensus, which searches scholarly papers (Ojala, 2023). Ithaka S+R (n.d.) has a regularly updated table called the "Generative AI Product Tracker" which lists and summarizes over 100 AI tools. These brief overviews of various GenAI tools are useful and they encourage library workers to experiment with and proactively think about these tools and their uses (Bridges et al., 2024; Pfeiffer, 2024).

GenAI Chatbots in the Non-Library Literature

Outside of the library profession, other disciplines have examined and evaluated different GenAI chatbots more extensively. In medicine, for example, Smith et al. (2023) found that Microsoft's Bing Chat, Google's Bard, and ChatGPT 3.5 and 4.0 could pass

an emergency medicine exam in Australia. Warren (2024), writing for the AI newsletter *Exponential View*, compared GPT-4 with Google's Gemini Ultra and Anthropic's Claude 3 on daily business tasks. These three chatbots, however, are not free and require a paid subscription. Warren (2024) tested all three models on tasks such as summarizing a meeting based on a transcript, market sizing, and critiquing an article without using prompt engineering techniques, and found that GPT-4 performed the best with the critique task and Claude performed better on the summarization task and market sizing. Overall, GPT-4 was slow and verbose, Gemini was quick, and Claude was better with statistical analysis (Warren, 2024). In the computer science field, Chatbot Arena is an open-source platform for comparing different LLMs (LMSYS Arena Team, 2024). Chatbot Arena randomly selects LLMs and asks users to pose a question and rate the responses. Aggregate ratings are presented on a leaderboard. As of April 2024, Chatbot Arena included 82 different LLMs and collected over 672,000 votes with GPT-4 Turbo in the lead (LMSYS Chatbot Arena Leaderboard, 2024). Chatbot Arena is an entertaining way to try different anonymized LLMs and review rankings, but it does not provide a detailed evaluation of the types of questions each LLM answered well. It also removes the LLM from the web page or context in which it is presented to a general user, which might include helpful prompt tips and other features such as the ability to save responses and tweak the level of creativeness in its output. Other disciplines have conducted studies that examine, compare, and evaluate different GenAI chatbots. These studies can, in turn, help inform the library literature.

Selecting GenAI Chatbots for the Study

With many different chatbots to choose from—for instance, as listed in the Chatbot Area leaderboard (LMSYS Chatbot Arena LLM Leaderboard (2024)—this study focused on the free version of some of the more widely known GenAI chatbots as well as new and promising ones as mentioned by Britt (2024) and Carr (2024): ChatGPT, Google's Gemini, Microsoft's Copilot, Anthropic's Claude, Perplexity, and the open-source chatbot known as HuggingChat. These six GenAI chatbots are briefly introduced below.

OpenAI's ChatGPT (chat.openai.com)

ChatGPT has been widely covered by mainstream media outlets and in the library literature since its free model 3.5 was released in November 2022 (see Cox & Tzoc, 2023; Houston & Corrado, 2023; Lund & Wang, 2023; Hall & McKee, 2024). GPT model 4.0, a subscription-based version, was later released in March 2023. The earlier free version, ChatGPT-3.5, was based on fixed datasets and was unable to search web sources (Smith et al., 2023); however, OpenAI (2024b) released a newer free version of ChatGPT known as ChatGPT-4o at the end of 2024 that searches the web (OpenAI, 2024a). This greatly improved its output and usefulness.

The ChatGPT interface is simple, uncluttered, and provides options to create an image, summarize text, analyze data, and more. Users can choose to create a free account which saves their chat history and provides access to additional features such as the ability to search the web and upload documents for analysis. The simple interface for ChatGPT, marked with the question "What can I help with?", invites users to start typing their queries and engage with the GenAI chatbot. Suggested tasks under the chat window such as summarizing text and analyzing data illustrate what users can accomplish using ChatGPT.

Google's Gemini (gemini.google.com)

Google released its chatbot, known initially as Bard, in the spring of 2023 (Weise et al., 2023). Almost a year later, Google rebranded Bard to Gemini (Carr, 2024). Google offers three versions of its Gemini chatbot, including a subscription-based version known as Gemini Ultra which costs \$20 per month (Wiggers, 2024c). Gemini has generated more interest than its predecessor, Bard. For instance, in February 2024 it attracted 51.7 million unique visitors (Carr, 2024). Gemini has received negative press, however. In February 2024, it upset users "by producing ahistoric images and blocking requests for depictions of white people" (Berber & Kruppa, 2024, para. 6). Given Google's dominance in the search market (Berber & Kruppa, 2024), this study assesses the free version of Gemini to gauge how it is performing in the GenAI field.

In order to access Gemini, users must create or use an existing Google account to sign in – a requirement that might deter potential users from experimenting with it. Once a user is signed in, a greeting with the user's name is displayed. Below this greeting, sample queries are listed. Gemini gives users the option to upload an image. The author uploaded a photo of their library and asked Gemini to identify where the photo was taken. Gemini correctly, but with a doubtful tone, identified the photo. After a response, users can send a "thumbs up" or a "thumbs down" rating for the output.

Microsoft's Copilot (copilot.microsoft.com)

Microsoft's Copilot has evolved since its initial iteration, Bing Chat, was released in May 2023 (Smith et al., 2023; Muchmore, 2024). Copilot is powered by GPT 4.0 and "has the ability to live search the internet to integrate relevant and up-to-date information when generating responses" (Smith et al., 2023, p. 876). GPT 4.0's standalone version can only be accessed via a paid subscription, but it is available for free via Copilot (Muchmore, 2024). Copilot also creates free images using the DALL-E image generator, which is otherwise a fee-based service (Muchmore, 2024). According to Carr (2024, para. 16), Copilot is part of a "much broader Microsoft strategy to incorporate Co-Pilot-branded AI features into Office, as well as tools for developers and designers." Muchmore (2024) provides a clear explanation of Microsoft's changing and rapidly developing GenAI chat strategy. From the Copilot website and with Copilot in Bing, users can take advantage of plug-ins such as one for OpenTable to make restaurant reservations, for example. Copilot in Windows allows users to change Windows settings to dark mode. Copilot in the Edge web browser allows users to enter text and create images with Microsoft Designer as well as summarize a PDF using the "Ask Copilot" option. The paid subscription allows users to access Copilot in Microsoft 365 apps such as Word and Excel (Muchmore, 2024).

Copilot's interface, unlike that of ChatGPT and Gemini, has large images with suggested topics that users might find helpful or overwhelming. Like Gemini, Copilot allows users to upload an image as part of their query but only after signing in. For example, a user can upload a photograph and ask Copilot where the photo might have been taken, or upload a photo of a plant and ask Copilot to identify it. Users are advised to sign into their Microsoft account to ask more questions, have longer conversations, create images, and have voice interactions with Copilot.

Claude (www.anthropic.com/claude)

Developed by Anthropic, an AI start-up with a \$2 billion investment from Google and a \$4 billion investment from Amazon, Claude offers three LLMs with different capabilities (Irwin, 2024). Anthropic notes that both its paid version ("Opus") and free version ("Sonnet") can analyze visual charts, graphs, and PDF documents (Irwin, 2024). It is important to note that Anthropic was founded by former OpenAI researchers and engineers who left the company with concerns that its chief executive, Sam Altman, did not prioritize safety issues with the rapid growth of AI (Weise et al., 2023). In a test of Claude's ability to answer a set of questions accurately and in a balanced manner, Wiggers (2024b) writes that Claude is slightly disappointing. It is trained on data that can be several months old and lacks a connection to the web. Its

paid plan is similar to OpenAI's and Google's premium chatbot plans (\$20 per month), but it lacks third-party apps and other service integrations which the paid versions of Google's Gemini and ChatGPT provide, such as integration with email and sites like Kayak for travel planning (Wiggers, 2024b). Nevertheless, Claude has been noted as a challenger to ChatGPT (Britt, 2024; Carr, 2024; Warren, 2024) and as such has been included in this study.

Claude requires users to create an account, provide a phone number, and check a box confirming that the user is 18 years of age or older. Alternatively, users can login with their Google account credentials. Claude provides a simple greeting and suggests tasks that the user can complete such as generating interview questions or extracting insights from a report. Unlike ChatGPT and Gemini, Claude allows users to upload not only images but different file types as well. The author uploaded a PDF of an article and requested a summary which Claude quickly and impressively produced. Similarly, users can ask Claude to provide the main argument of an article or summarize a file containing code. Copilot can also summarize documents but only those in PDF format within the Edge browser at this time.

Perplexity (www.perplexity.ai)

Perplexity is an AI start-up "aiming to challenge Google's dominance in web search" by offering a search engine with AI features (Berber & Kruppa, 2024, para. 1). These features give users direct answers to search queries accompanied by links to source material instead of simply providing a list of website links like Google Search (Berber & Kruppa, 2024). This start-up has financial backing from Amazon and is in the process of finalizing a new \$1 billion funding agreement (Berber & Kruppa, 2024). Like ChatGPT, Gemini, and Claude, Perplexity offers a premium version for \$20 per month. As of February 2024, Perplexity's user base has increased by 8.6% to approximately 50 million users (Berber & Kruppa, 2024). Perplexity uses a combination of a homegrown LLM, third-party models including GPT 4 and Anthropic's Claude 2, and web crawlers to retrieve results that are relevant (Sullivan, 2024). Sullivan (2024) notes that Perplexity is effective for answering open-ended questions such as suggestions for vacation planning, but it is less helpful in finding specific information such as local business hours.

Users can access Perplexity without creating an account. With a free account, however, users can try additional features such as a "Pro Search," which claims to search more content, allows file uploads, and saves one's query history. Unlike the other GenAI chatbots included in this study, Perplexity gives users the option to limit

their search to published academic papers and social networks. Users with a free account can upload text or PDFs, and submit five “Pro” searches per day. A “Pro Search”, according to Perplexity’s website, is their most powerful search designed to provide longer answers to complex questions. Unlike Gemini, Copilot, and Claude, Perplexity’s free version does not allow users to upload images when posing a question. Only the paid version of Perplexity allows this.

HuggingChat (huggingface.com/chat)

HuggingChat is an open-source version of ChatGPT developed by Hugging Face, an AI start-up supported by “tens of millions in venture capital” (Wiggers, 2023, para. 1). The model behind HuggingChat was developed by the German nonprofit Open Assistant, which was responsible for creating the dataset on which Stable Diffusion, the text-to-image AI model, was trained (Wiggers, 2023). Open Assistant is mainly comprised of volunteers who seek to not only replicate ChatGPT but to go beyond it, an ambition that Wiggers (2023) writes is still in its infancy, given that HuggingChat’s responses can quickly derail when Wiggers had tested it. Still, it is included in this exploratory study as it is an open-source option that is currently managing to compete with “closed rivals” such as ChatGPT (Franzen, 2024).

HuggingChat gives users the option of working with a specific LLM such as different versions of Meta’s Llama along with other types of GenAI tools such as the image generator DALL-E. HuggingChat maintains a user’s chat history and gives the user the option of searching the web when submitting a query. Like all of the GenAI chatbots included in this study, HuggingChat lists a range of sample search queries such as “Write an email from bullet list” to introduce users to the types of questions that can be asked.

This brief overview of each GenAI chatbot highlighted some of the common and divergent features of each chatbot. Some chatbots require users to create a free account before accessing the chatbot, which raises potential privacy concerns. This can also restrict minors’ access to the chatbots. Others allow users to include results from the web. Perplexity stands out by providing an option to search specific types of sources such as academic publications. Others allow users to analyze PDFs or images. These features are summarized in Table 1. Table 1 attempts to illustrate the types of tasks that each chatbot can complete, thereby potentially helping users to assess each chatbot as part of an AI literacy exercise in determining its usefulness and relevancy to address a specific need.

Table 1
Comparing Selected Features Across the GenAI Chatbots

Feature	ChatGPT	Gemini	Copilot	Claude	Perplexity	HuggingChat
Do users need to create a free account in order to access the chatbot?	x	✓	x	✓ Phone number required to create an account	x Limited "Pro" search requires users to create an account and log in.	✓ Claims to offer a limited number of chats without an account, but this was not the case for the author.
Can it search the web?	✓ Need to login in.	✓	✓	x	✓	✓
Does the chatbot provide web links to sources?	✓ Need to login in.	x Does not always provide links.	✓	x	✓	✓ Sometimes, with web search enabled.
Can you upload and analyze files with the free version?	✓ Need to login in.	x	x	✓	✓ Need to login in.	x
Can you upload and analyze images with the free version?	✓ Need to login in.	✓	✓	✓	x	x

Note. These features are subject to change.

Conducting the Comparison: The Questions Asked and Why

This study asked each GenAI chatbot a series of five questions at the end of 2024. The questions were selected based on some of the types of queries that patrons, mainly undergraduate students, are likely to ask at the author’s institution. (This comparison does not consider academic policies around AI use as established by various academic institutions. These policies alone can constitute separate study.) Librarians at other

institutions may seek to ask questions that are relevant to their contexts. The five questions were as follows:

1. How current is your training data? (Is each chatbot transparent about this?)
2. What are the latest developments in the Russia-Ukraine war? (How does the chatbot handle requests for current information? How current is the response?)
3. Should gender-affirming care be provided to minors? (How does the chatbot respond to potentially polarizing issues?)
4. I need help with an assignment on gun control in the United States. (How does the chatbot respond to a student who is beginning the research process?)
5. I need primary and secondary sources on the Chernobyl nuclear disaster. (How does the chatbot handle a research question about a historical event outside the U.S.?)

These questions were asked to assess how each GenAI chatbot handles different types of questions. The first question about training data was posed to see if the chatbot is transparent about the scope and limitations of its training data. The second question about a current event was used to gauge how a chatbot responds to requests for current information. The third question aims to gauge any biases in the chatbot by posing a potentially polarizing question; this is something that Wiggers (2024a; 2024b) suggests when testing and comparing various chatbots. Wiggers posed controversial questions such as the status of Taiwan, issues with TikTok, Harvard and admissions, the politics around the southern U.S. border, the conflict in Israel and Palestine, the 2020 U.S. presidential election results, and the invasion of Ukraine. Warren (2024) also aptly points out that noting biases in a chatbot is important. Questions four and five are queries that students might ask at the author's institution.

These questions were posed without any "prompt engineering." This is something that Warren (2024, para. 2) discussed when formulating questions for a similar test: he "refrained from any attempt at prompt engineering, keeping the prompts plain and simple, allowing each model's inherent capabilities to shine through" thus mimicking a general user's query, but also assuming that someday prompt engineering will not be necessary as these chatbots and their models continue to improve.

The output from each GenAI chatbot was assessed on a number of dimensions by the author and assigned 1 point if the GenAI chatbot met the criterion. The maximum number of points a GenAI chatbot could receive for a strong response per question was 3 points. The assessment criteria were as follows:

- Did the chatbot answer the question? (1 point)
- Was the length and amount of detail in the response sufficient for the intended audience (e.g., an undergraduate student)? (1 point)
- Was the response easy to understand and helpful (for example, did the chatbot provide real and relevant sources – not fabricated titles – for follow-up and verification, where applicable)? (1 point)

The output was ultimately evaluated by the author who, as a human, also has biases and, unlike Chatbot Arena, did not blindly assess the output. Although the evaluative criteria and assessment here is based on the author's opinion, this exploratory comparison serves to suggest how other library workers can perform similar comparisons for their own libraries and work contexts.

Results

Question One. How Current Is Your Training Data?

Four of the six GenAI chatbots clearly stated the cut-off period for their training data, earning the full 3 points for this question (see Table 2 for scores and Appendix Table A1 for a summary of responses). ChatGPT, Claude, Perplexity, and HuggingChat all provided transparent responses. ChatGPT clearly stated that its training data is current up to October 2023 and that it can use “real-time tools” to access current information. Gemini and Copilot refused to provide this information, which raises questions about the transparency of their output. Gemini refused to provide any information about its training data in order “to protect their competitive advantage and maintain data privacy.” Gemini suggested that users consult “reliable and up-to-date news sources” for more recent events and research results. Copilot also did not disclose the end date of its training data. Claude explained that its training ended in April 2024 and that it does not have knowledge of events and developments after this date. Claude stated that it tries to be transparent about the limits of its knowledge. Perplexity stated that its training data covers up to December 2023. Lastly, HuggingChat responded that its training data ended in October 2023 and, although it is unaware of developments after this time, it will try its best to help. Again, this training data question was asked to compare the potential scope of these GenAI chatbots and gauge how transparent they are about their limitations.

Table 2

Scores for Question One: How Current Is Your Training Data?

	ChatGPT	Gemini	Copilot	Claude	Perplexity	HuggingChat
Did the chatbot answer the question?	1	0	0	1	1	1
Was the length and amount of detail in the response sufficient?	1	1	1	1	1	1
Was the response easy to understand and helpful?	1	0.5	0.5	1	1	1
Total score	3	1.5	1.5	3	3	3

Question Two: What Are the Latest Developments in the Russia-Ukraine War?

This current event question was posed to examine how the chatbots might respond to queries for current information (see Appendix Table A2 for a summary of responses). Would a given chatbot hallucinate a response? How might a chatbot explain its possible limitations and guide users in other ways? Earlier in 2024, ChatGPT clearly stated that it cannot provide “real-time” information and suggested that the user check “reliable news sources.” However, by the end of 2024, ChatGPT provided one of the most helpful responses among the six GenAI chatbots in this study, earning a full 3 points (see Table 3). ChatGPT cited multiple news sources published on the same day and provided subtopics such as “military actions” followed by bullet point information backed by a different source. Gemini provided bullet point information such as North Korea’s involvement in the war but did not cite specific sources for each point. It simply listed three sources at the end of its response and suggested that users consult “reliable” sources but did not explain what makes a source reliable, earning a score of 2 points for this response. Copilot provided four short paragraphs covering different aspects of the war such as the number of casualties and cited sources that were a day to a week old, earning the full 3 points. Claude was clear about its inability to provide current information. Nevertheless, it listed information in bullet points covering events up to the beginning of 2024 and asked if the user wanted to discuss any specific aspects of the war’s impact or earlier developments that it could address with more certainty, earning 2.5 points. Similar to ChatGPT’s response, Perplexity provided subtopics followed by information in bullet points with a link to a source published on the same day or a day earlier. It cited six sources and provided five helpful follow-up questions such as “How are international sanctions affecting Russia’s military efforts?”, earning the full 3 points. HuggingChat summarized information from one site (understandingwar.org) into nine subtopics with a bullet point each, earning 2 points for its response. ChatGPT, Copilot, and Perplexity provided the strongest and most helpful response to this question by

citing multiple current sources. ChatGPT and Perplexity organized information into clearly labelled subtopics. Copilot and Perplexity also provided helpful follow-up questions for users to consider.

Table 3

Scores for Question Two: What Are the Latest Developments in the Russia-Ukraine War?

	ChatGPT	Gemini	Copilot	Claude	Perplexity	HuggingChat
Did the chatbot answer the question?	1	1	1	1	1	1
Was the length and amount of detail in the response sufficient?	1	0.5	1	1	1	0.5
Was the response easy to understand and helpful?	1	0.5	1	0.5	1	0.5
Total score	3	2	3	2.5	3	2

Question Three: Should Gender-Affirming Care Be Provided to Minors?

Overall, there were two types of responses to the question of gender-affirming care for minors (see Appendix Table A3 for a summary of responses). Gemini and Perplexity summarized arguments for and against gender-affirming care, aiming to provide two different perspectives on this question, earning them a full 3 points for their responses and the range of sources they cited, though Perplexity cited the more widely recognized and reliable source, the American Academic of Pediatrics. ChatGPT, Copilot, Claude, and HuggingChat summarized some of the major issues surrounding this question such as the legal, medical, and ethical dimensions. Both Claude and HuggingChat received a score of 2.5 because they did not provide links to sources for users to verify and follow up on. All of the GenAI chatbots, however, did begin their response by acknowledging the complexity and contentiousness of the question. Only ChatGPT, Gemini, Copilot, and Perplexity cited and linked to medical sources such as the American Medical Association and the American Academy of Pediatrics. ChatGPT enhanced its response by also including current news sources and more academic sources such as a research article from PubMed. Gemini and Perplexity provided two opposing views while the remaining chatbots summarized some of the major issues concerning gender-affirming care for minors.

Table 4

Scores for Question Three: Should Gender-Affirming Care Be Provided to Minors?

	<u>ChatGPT</u>	Gemini	Copilot	Claude	Perplexity	<u>HuggingChat</u>
Did the chatbot answer the question?	1	1	1	1	1	1
Was the length and amount of detail in the response sufficient?	1	1	1	1	1	1
Was the response easy to understand and helpful?	1	1	1	0.5	1	0.5
Total score	3	3	3	2.5	3	2.5

Question Four: I Need Help with an Assignment on Gun Control in the United States

The GenAI chatbots gave varied responses to this question but all provided a helpful tone (see Appendix Table A4). Both ChatGPT and Claude asked the user for more details such as what aspect of gun control the user is interested in, resulting in a score of 2 points. The remaining four chatbots also provided background information on the scope of this topic, earning the full 3 points. Copilot and Perplexity listed links for each bullet point of information. Gemini included a few links to information based on Google searches. HuggingChat went a step further by providing an outline of an essay and writing tips. It listed book and article titles; however, one of the articles could not be verified by research databases, which suggests that it could have been fabricated, resulting in a score of 2 points for its response. Perplexity provided one of the stronger responses by providing links to support each bullet point of information provided and by suggesting related questions such as “How have recent gun control laws impacted crime rates in the U.S.?”. Unlike the other four GenAI chatbots, Gemini and HuggingChat actually advised the user to consider the quality of research sources. For example, Gemini suggested that users consult “[r]eputable news outlets with diverse perspectives” such as the New York Times, Wall Street Journal, and the Washington Post. HuggingChat advised users to consult “reliable sources” but did not explain what these sources are or provide examples. Nevertheless, all of the chatbots had a helpful, friendly, and supportive tone.

Table 5

Scores for Question Four: I Need Help with an Assignment on Gun Control in the United States

	<u>ChatGPT</u>	Gemini	Copilot	Claude	Perplexity	<u>HuggingChat</u>
Did the chatbot answer the question?	1	1	1	1	1	1
Was the length and amount of detail in the response sufficient?	0.5	1	1	0.5	1	0.5
Was the response easy to understand and helpful?	0.5	1	1	0.5	1	0.5
Total score	2	3	3	2	3	2

Question Five: I Need Primary and Secondary Sources on the Chernobyl Nuclear Disaster

The GenAI chatbots provided a range of responses to this question (see Appendix Table A5). HuggingChat’s response was the most problematic. It started with a helpful explanation of primary and secondary sources; however, it then listed fabricated book titles and produced incoherent text toward the latter half of its response, earning a score of only 1.5. Claude produced a rather weak response given that it cannot provide links to sources though it could have, at the very least, explained how one might go about accessing these sources by contacting a local public library (which Gemini did), resulting in a score of 2.5. ChatGPT’s response was also weak as it directed users to sources via a link to a commercial document provider and a commercial essay writing site. ChatGPT also linked to a commercial digital archival site, ignoring some of the freely available online archives such as nsarchive.gwu.edu, which Gemini, Copilot, and Perplexity linked to, resulting in a score of 2.5. For primary sources, all of the GenAI chatbots generally suggested a combination of government reports or reports from energy-related agencies. Some suggested news coverage (ChatGPT, Gemini, HuggingChat) and accounts from individuals (Gemini, Perplexity, HuggingChat). As for secondary sources, a number mentioned books and academic articles and listed a book title or two (ChatGPT, Gemini, Copilot, Claude, Perplexity). Three of the six GenAI chatbots cited the book *Voices from Chernobyl* by Svetlana Alexievich, which is a compilation of oral histories from survivors and witnesses, but the GenAI chatbots diverged on whether it considered this title to be a primary or a second source. Gemini and HuggingChat listed it as a primary source, while Claude considered it a secondary source. Only a few of the GenAI chatbots listed documentary films as a secondary source (ChatGPT and Gemini), while HuggingChat listed footage

from documentary films as a primary source. For an undergraduate student completing this assignment, Perplexity provided the strongest response by including a link to a suggested resource. However, combining responses from Perplexity with Gemini and Copilot, which provided a few different links, would produce an even stronger, though lengthier, response. Perplexity, Gemini, and Copilot all earned the full 3 points for their output. Again, Perplexity provided helpful follow-up questions that users could click on to expand their research. It was also the only chatbot that recommended non-American government sources of information such as Soviet government documents and Soviet accident reports.

Table 6

Scores for Question Five: I Need Primary and Secondary Sources on the Chernobyl Nuclear Disaster

	<u>ChatGPT</u>	Gemini	Copilot	Claude	Perplexity	<u>HuggingChat</u>
Did the chatbot answer the question?	1	1	1	1	1	0.5
Was the length and amount of detail in the response sufficient?	1	1	1	1	1	0.5
Was the response easy to understand and helpful?	0.5	1	1	0.5	1	0.5
Total score	2.5	3	3	2.5	3	1.5

Discussion

This exploratory comparison found that Perplexity consistently provided the most satisfactory responses to all five questions posed, earning a 15 out of 15 score. Perplexity provided links to current information, citations to a range of sources, and useful follow-up questions without overwhelming the user. Perplexity was followed by ChatGPT and Copilot in second place, Gemini and Claude in third place, and HuggingChat in last place (see Table 7 for total scores). A broader discussion of the responses from the chatbots follows.

Table 7

Total Scores for All Five Question Posed to the Six GenAI Chatbots

ChatGPT	Gemini	Copilot	Claude	Perplexity	HuggingChat
13.5	12.5	13.5	12.5	15	11

Note. The maximum score was 15.

On Transparency: Question One and Asking “How Current Is Your Training Data?”

Asking the six GenAI chatbots about the currency of their training data potentially revealed how transparent the chatbots are about their scope and limitations. Archambault and Rincon (2024) emphasized the importance of assessing the transparency of AI tools. Gemini and Copilot refused to provide this information, but ChatGPT, Claude, Perplexity, and HuggingChat clearly stated the month and year their data training ended. Future studies examining the transparency of GenAI chatbots could examine whether the chatbots explain their decision-making processes (Archambault & Rincon, 2024) and the sources of their training data.

On Currency: Question Two and Asking “What Are the Latest Developments in the Russia-Ukraine War?”

This current event question was asked to assess the currency of information and citations the GenAI chatbots could potentially provide. ChatGPT and Perplexity cited sources published the same day, while Copilot cited sources published as recently as one day earlier. Archambault and Rincon (2024) discuss the importance of assessing the currency of information provided by research AI tools in particular. Gemini, Claude, and HuggingChat’s lack of links to a breadth of current sources underscores one of Lo’s (2025) points about developing AI literacy: Being aware of which AI tools are appropriate to use in a specific context is necessary. Knowing these limitations of Gemini, Claude, and HuggingChat, users might decide to not rely solely on these GenAI chatbots for current information.

On Biases and Controversial Topics: Question Three and Asking “Should Gender-Affirming Care Be Provided to Minors?”

All six of the GenAI chatbots acknowledged how complex and controversial this question is. While there were two types of responses to this question – one citing pros and cons and the other summarizing the issues involved – all of the GenAI chatbots

attempted to cover the polarizing aspects of this question. Here, Lo (2025) reminds us about the ethical awareness and critical thinking required to analyze the sources and biases in GenAI chatbot responses: Does the chatbot “uphold fairness, inclusivity, and transparency” (p.121)? While all of the chatbots tried to present at least two different perspectives on this question, ChatGPT enhanced its response by also including current news sources and more academic sources such as a research article in PubMed.

Just months earlier, these GenAI chatbot provided a different answer to this question. ChatGPT, Copilot, and Claude provided arguments for and against gender-affirming care. Claude said it “aimed to objectively outline the major tensions in the multi-faceted debate based on my training.” Perplexity and HuggingChat both argued in favor of gender-affirming care by citing major medical bodies such as the American Academy of Pediatrics, the American Psychological Association, and the Endocrine Society. Asking these GenAI chatbots the same question over a period of time can produce different responses that users should take into consideration. The variability of responses over time and from user-to-user is something to keep in mind when using and evaluating output from these GenAI chatbots.

On Helping a Student Start a Research Project: Question Four and Asking for Help with an Assignment on Gun Control in the United States

This question was posed to gauge what Archambault and Rincon (2024) describe as an issue of “critical thinking and overreliance” on AI tools which “may reduce deep reading and limit serendipitous discoveries” (p. 10). Although Archambault and Rincon (2024) examined the paid versions of AI research tools, their insights and evaluative criteria can be applied here. For example, HuggingChat prescribed a research paper outline which did not allow room for the user to think critically about the topic and formulate their own research paper outline. ChatGPT and Claude took an opposite approach, asking the user important clarifying questions about which aspect of gun control they are interested in researching but did not supply background information to help guide the student in their understanding of the topic. Gemini, Copilot, and Perplexity provided this background information.

Similar to the question about gender-affirming care for minors, a few months earlier when the author posed this same question, all of the GenAI chatbots mentioned the Second Amendment right of individuals to bear arms, as well as the legal and regulatory dimensions of gun control. Also, with the exception of HuggingChat, a few months earlier all of the GenAI chatbots provided brief arguments for and against gun control. Again, posing the same question with the exact wording can result in very

different responses, something that users should keep in mind when using these tools for research.

On Including Non-U.S. Sources: Question Five and Asking for Primary and Secondary Sources on the Chernobyl Nuclear Disaster

This fifth and final question was posed to assess how the GenAI chatbots would respond to a question about a historical event that took place outside of the United States. Gemini, Copilot, and Perplexity earned the full 3 points for this question by providing helpful lists of a broad range of primary and secondary sources. However, Perplexity was the only chatbot that suggested non-American sources such as Soviet government documents and Soviet accident reports. Given all of the valid concerns about chatbots creating hallucinated responses, surprisingly only HuggingChat produced fabricated, incoherent text towards the end of its output. Some of the chatbots diverged in their assessment of whether a specific source is primary or not, highlighting Archambault and Rincon's (2024) and Lo's (2025) point that users need to be critical of and assess the claims made by these chatbots. They are not "infallible or neutral" as Lo (2025, p. 121) reminds us.

This exploratory comparison found that submitting the same query to more than one GenAI chatbot and comparing responses can be instructive and may help users identify the strengths and weaknesses of each chatbot. What was clear through this exploratory study was that, despite all of the attention ChatGPT has received, it has strong competitors. For example, ChatGPT does not provide a helpful list of follow-up questions but Perplexity consistently does this. Claude and other GenAI chatbots that do not connect to the web are probably better suited to help users with tasks such as composing an email, analyzing text, or solving a coding problem. Although, with coding questions, it can be helpful when a GenAI chatbot links to developer sources on the web such as the Mozilla Developer Network (MDN) and Github. Gemini, Copilot, and Perplexity provide this information.

Overall, the insights gained from this exploratory comparative study concur with some of the suggestions and analyses in the library literature on GenAI chatbots. Pfeiffer (2024, Conclusion section), for example, advises librarians and information professionals not to "treat the[se AI] tools as they currently stand as completed": Library workers should continue to be proactive thinkers and testers of GenAI chatbots and any emerging technologies for that matter. While reviewing Copilot, Muchmore (2024) points out that these GenAI chatbots are evolving quickly. As such, Bridges et al. (2024) advise library workers to have hands-on experience with these GenAI chatbots.

This hands-on work is part of “AI literacy” development where users gain an understanding of how chatbots are trained and their possible training biases, as well as an understanding of which tasks they perform well (Bridges et al., 2024). Posing the same questions to these GenAI chatbots after several months can also present illuminating results. Earlier in 2024, the freely available version of ChatGPT (then 3.5, now 4o) did not search the web. As well, Google is now including Gemini responses in Google search results. As mentioned earlier, Carr (2024) writes that Copilot is being integrated in other applications such as Microsoft Office, so in 2025 there may be more GenAI features integrated in other tools. Other features such as video generation from text could become a possibility though limited in the free version of these GenAI chatbots. If anything, this exploratory comparison points out some of the overlapping and diverging responses GenAI chatbots might provide and the possible importance of using more than one chatbot to compare responses and produce more comprehensive outputs.

Conclusion

This exploratory comparison of six generative AI chatbots seeks to be part of what Frederick (2023) currently describes as a period of “exploration and experimentation” (p. 4) with ChatGPT, or in broader terms, GenAI chatbots. Although the future of GenAI, as Fernandez (2023) writes, is unpredictable, “libraries should develop a positive vision for how to use it” (p. 1). In the library literature, several authors have compared the technological changes and disruption to knowledge retrieval and production that ChatGPT and generative AI have introduced to the development of calculators in the 1960s (Frederick, 2023; Houston & Corrado, 2023), Wikipedia (Frederick, 2023), and Google Scholar (Mairn & Rosengarten, 2023). While AI in general may seem new, Houston and Corrado (2023,) remind us that it is in fact not new by pointing out existing tools such as text messaging auto-fill; AI has existed for quite some time now. How then might libraries position themselves at this juncture in generative AI development? Again, Houston and Corrado (2023) offer some insight: “It is imperative for the academic community and libraries to understand and embrace these technologies and their myriad uses, to acknowledge their limitations and ethical issues, and to recognize why these tools, at least for now, cannot replace humans” (p. 77). Library workers should consider experimenting with these tools and help shape and inform how they are used and understood.

References

- Archambault, S. G., & Rincón, J. J. (2024). An evaluation of cutting-edge AI research tools using the REACT Framework. *Computers in Libraries*, 44(8), 4-11. <https://www.infotoday.com/cilmag/oct24/Archambault-Rincon--An-Evaluation-of-Cutting-Edge-AI-Research-Tools-Using-the-REACT-Framework.shtml>
- Atwood, G. S. (2024). Try AI Day: Introducing library faculty and staff to artificial intelligence tools through hands-on experimentation. *Journal of Electronic Resources in Medical Libraries*, 21(2), 108-112. <https://doi.org/10.1080/15424065.2024.2351590>
- Berber, J., & Kruppa, M. (2024, March 5). AI search startup Perplexity set to double valuation to \$1 billion. *The Wall Street Journal*. <https://www.wsj.com/tech/ai/ai-search-startup-perplexity-set-to-double-valuation-to-1-billion-b84be204>
- Bridges, L. M., McElroy, K., & Welhouse, Z. (2024). Generative artificial intelligence: 8 critical questions for libraries. *Journal of Library Administration*, 64(1), 66-79. <https://doi.org/10.1080/01930826.2024.2292484>
- Britt, P. (2024). How to pick the best LLM for your sales activities. *CRM Magazine*, 28(1), 24-27.
- Carr, D. F. (2024, March 8). *ChatGPT, custom GPTs, and AI chat challengers: CoPilot, Gemini, Perplexity, and more*. The Similarweb Blog. <https://www.similarweb.com/blog/insights/ai-news/chatgpt-challengers/>
- Chen, X. (2023). ChatGPT and its possible impact on library reference services. *Internet Reference Services Quarterly*, 27(2), 121-129. <https://doi.org/10.1080/10875301.2023.2181262>
- Cox, C., & Tzoc, E. (2023). ChatGPT: Implications for academic libraries. *College & Research Libraries News*, 84(3), 99-102. <https://doi.org/10.5860/crln.84.3.99>
- Craddock, I., & Wilson, K. (2023). Six prompt tips: Getting the most out of large language model AI. *School Library Journal*, 69(11), 16.
- Fernandez, P. (2023). Some observations on generative text artificial intelligence's impact on libraries Part 1. *Library Hi Tech News*, 40(4), 1-5. <https://doi.org/10.1108/LHTN-05-2023-0076>

- Franzen, C. (2024, February 2). *Hugging Face launches open source AI assistant maker to rival OpenAI's custom GPTs*. VentureBeat. <https://venturebeat.com/ai/hugging-face-launches-open-source-ai-assistant-maker-to-rival-openais-custom-gpts/>
- Frederick, D. E. (2023). ChatGPT: A viral data-driven disruption in the information environment. *Library Hi Tech News*, 40(3), 4-10. <https://doi.org/10.1108/LHTN-04-2023-0063>
- Hall, B., & McKee, J. (2024). An early or somewhat late ChatGPT guide for librarians. *Journal of Business & Finance Librarianship*, 29(1), 58-69. <https://doi.org/10.1080/08963568.2024.2303944>
- Hendrick, R. (2024a, January 3). *Beyond ChatGPT: Gary Price's top AI tools for librarians*. LibTech Insights. <https://www.choice360.org/libtech-insight/beyond-chatgpt-gary-prices-top-ai-tools-for-librarians>
- Hendrick, R. (2024b, May 8). *Considering RAG when evaluating generative AI tools*. LibTech Insights. <https://www.choice360.org/libtech-insight/considering-rag-when-evaluating-generative-ai-tools>
- Hendrick, R. (2024c, November 13). *Evaluating generative AI resources: Separating the tools from the toys*. LibTech Insights. <https://www.choice360.org/libtech-insight/evaluating-generative-ai-resources-separating-the-tools-from-the-toys>
- Hervieux, S., & Wheatley, A. (2023, July 31). *How well can ChatGPT answer library reference questions?* LibTech Insights. <https://www.choice360.org/libtech-insight/how-well-can-chatgpt-answer-library-reference-questions>
- Hervieux, S. & Wheatley, A. (2024). *Building an AI literacy framework: Perspectives from instruction librarians and current information literacy tools*. https://www.choice360.org/wp-content/uploads/2024/08/TaylorFrancis_whitepaper_08.28.24_final.pdf
- Houston, A. B., & Corrado, E. M. (2023). Embracing ChatGPT: Implications of emergent language models for academia and libraries. *Technical Services Quarterly*, 40(2), 76-91. <https://doi.org/10.1080/07317131.2023.2187110>
- Irwin, K. (2024, March 5). *Amazon-backed Anthropic launches Claude 3 AI models to rival ChatGPT*. PC Mag. <https://www.pcmag.com/news/anthropic-launches-claude-3-ai-rival-chatgpt-4>
- Ithaka S+R. (n.d.). *Generative AI product tracker*. Retrieved October 29, 2024, from <https://sr.ithaka.org/our-work/generative-ai-product-tracker>

LMSYS Arena Team. (2024, March 1). *LMSYS chatbot arena: Live and community-driven LLM evaluation*. <https://lmsys.org/blog/2024-03-01-policy/>

LMSYS Chatbot Arena LLM Leaderboard (2024, April 13). Chatbot Arena. <https://chat.lmsys.org/?leaderboard>

Lo, L. S. (2023). The art and science of prompt engineering: A new literacy in the information age. *Internet Reference Services Quarterly*, 27(4), 203-210. <https://doi.org/10.1080/10875301.2023.2227621>

Lo, L. S. (2024). Evaluating AI literacy in academic libraries: A survey study with a focus on U.S. employees. *College & Research Libraries*, 85(5), 635-668. <https://doi.org/10.5860/crl.85.5.635>

Lo, L. S. (2025). AI literacy: A guide for academic libraries. *College & Research Libraries News*, 86(3), 120-122. <https://doi.org/10.5860/crln.86.3.120>

Lund, B. (2023). The prompt engineering librarian. *Library Hi Tech News*, 40(8), 6-8. <https://doi.org/10.1108/LHTN-10-2023-0189>

Lund, B. D. & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*, 40(3), 26-29. <https://doi.org/10.1108/LHTN-01-2023-0009>

Mairn, C., & Rosengarten, S. (2023). Helping students navigate research with AI tools. *Computers in Libraries*, 43(7), 22-26.

Muchmore, M. (2024, February 13). *What is Copilot? Microsoft's AI assistant explained*. PC Mag. <https://www.pcmag.com/explainers/what-is-microsoft-copilot>

Ojala, M. (2023). Search evolves into copilots, chatbots, and research assistants. *Computers in Libraries*, 43(7), 43-44.

OpenAI. (2024a, July 18). *GPT-4o mini: Advancing cost-efficient intelligence*. <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence>

OpenAI. (2024b, October 31). *Introducing ChatGPT search*. <https://openai.com/index/introducing-chatgpt-search>

Oyarzun, J. (2024). Teaching AI in a scholar context: Organizing a workshop on Chatgpt-3.5. *Computers in Libraries*, 44(6), 36-40.

Perplexity Team. (2023, November 29). *Introducing PPLX online LLMs*. www.perplexity.ai/hub/blog/introducing-pplx-online-llms

- Pfeiffer, D. (2024, March 11). *Five growing concerns about generative AI for librarians and information professionals*. LibTech Insights. <https://www.choice360.org/libtech-insight/four-growing-concerns-about-generative-ai-for-librarians-and-information-professionals>
- Pun, R. (2023). Generative AI tools transforming the library? Rethinking possibilities and questions. *Against the Grain*, 35(3), 19-22.
- Smith, J., Choi, P. M., & Buntine, P. (2023). Will code one day run a code? Performance of language models on ACEM primary examinations and implications. *Emergency Medicine Australasia*, 35(5), 876-878. <https://doi.org/10.1111/1742-6723.14280>
- Sullivan, M. (2024, March 19). *How Perplexity aims to reinvent web search*. Fast Company. <https://www.fastcompany.com/91033202/perplexity-most-innovative-companies-2024>
- Toner, H. (2023, May 12). *What are generative AI, large language models, and foundation models?* Center for Security and Emerging Technology, Georgetown University. <https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/>
- Warren, N. (2024, March 13). *Putting GPT-4's new rivals to the test*. Exponential View by Azeem Azhar. <https://www.exponentialview.co/p/putting-gpt-4s-new-rivals-to-the>
- Weise, K., Metz, C., Grant, N., & Isaac, M. (2023, December 5). Inside the A.I. arms race that changed Silicon Valley forever. *The New York Times*. <https://www.nytimes.com/2023/12/05/technology/ai-chatgpt-google-meta.html>
- Wiggers, K. (2023, April 25). *Hugging Face releases its own version of ChatGPT*. TechCrunch. <https://techcrunch.com/2023/04/25/hugging-face-releases-its-own-version-of-chatgpt/>
- Wiggers, K. (2024a, February 15). *We tested Google's Gemini chatbot – here's how it performed*. TechCrunch. <https://techcrunch.com/2024/02/15/we-tested-googles-gemini-chatbot-heres-how-it-performed/>
- Wiggers, K. (2024b, March 7). *We tested Anthropic's new chatbot – and came away a bit disappointed*. TechCrunch. <https://techcrunch.com/2024/03/07/we-tested-anthropics-new-chatbot-and-came-away-a-bit-disappointed/>

Wiggers, K. (2024c, March 18). *Google Gemini: Everything you need to know about the new generative AI platform*. TechCrunch. <https://techcrunch.com/2024/03/18/what-is-google-gemini-ai/>

Yang, S.Q. & Mason, S. (2024). Beyond the algorithm: Understanding how ChatGPT handles complex library queries. *Internet Reference Services Quarterly*, 28(2), 97-151. <https://doi.org/10.1080/10875301.2023.2291441>

Zhang, B. (2023). Prompt engineers or librarians? An exploration. *Medical Reference Services Quarterly*, 42(4), 381-386. <https://doi.org/10.1080/02763869.2023.2250680>

Appendix

Tables Summarizing Responses from the Six GenAI Chatbots to the Five Questions Posed

Table A1

Question One: How Current Is Your Training Data?

ChatGPT	Gemini	Copilot	Claude	Perplexity	HuggingChat
October 2023	Did not disclose. Stated that it is constantly being updated	Did not disclose. Stated that it is constantly being updated.	April 2024	December 2023	October 2023

Note. Currency was evaluated in December 2024.

Table A2
 Question Two: What Are the Latest Developments in the Russia-Ukraine War?

ChatGPT	Gemini	Copilot	Claude	Perplexity	HuggingChat
<ul style="list-style-type: none"> - Cited multiple same-day news sources - Sources included BBC News, Al Jazeera, AP News, Reuters, CNN, The Guardian and more - Provided an extensive summary with headings and bullet points which were linked to a source 	<ul style="list-style-type: none"> - Provided five bullet points without links to sources - Suggested users consult the following 3 “reliable news sources”: The Council on Foreign Relations, Al Jazeera, and The Independent 	<ul style="list-style-type: none"> - Provided four bullet points with links to 3 sources: Al Jazeera, Newsweek and BBC - Sources were a day to seven days old - Posed a follow up question asking the user if there is a specific aspect of the <u>conflict</u> they are interested in 	<ul style="list-style-type: none"> - Listed developments in bullet points up to early 2024 	<ul style="list-style-type: none"> - Provided a summary with headings and bullet points with source links published the same day or 1-2 days earlier - Cited 6 sources including Euractiv, European Council, CBS News, Al Jazeera, and the Institute for the Study of War (ISW) - Suggested follow up questions 	<ul style="list-style-type: none"> - With web search enabled, synthesized information from one source, understandingwar.org, posted the previous day - Provided information with subheadings

Table A3
 Question Three: Should Gender-Affirming Care Be Provided to Minors?

ChatGPT	Gemini	Copilot	Claude	Perplexity	HuggingChat
<ul style="list-style-type: none"> - Summarized some of the major issues involved such as the legal landscape and medical perspective - Provided short paragraphs with citations - Cited sources such as an article in PubMed, the American Medical Association and current news sources such as the Times 	<ul style="list-style-type: none"> - Provided a bulleted list of arguments for and against gender-affirming care - Cited 3 sources: the Trevor Project, Alabama Campaign, AAMC, and LGBTMAP 	<ul style="list-style-type: none"> - Listed 4 major issues involved such as the legal, political, and ethical dimensions - Cited the American Medical Association, opa.hhs.gov, AAMC, the Associated Press, and Yahoo! News 	<ul style="list-style-type: none"> - Provided a two-paragraph response summarizing the issues involved - Mentioned the American Academy of Pediatrics but did not provide a link to the source - Focused on the medical dimensions of the topic 	<ul style="list-style-type: none"> - Provided a bulleted list of arguments for and against gender-affirming care - Cited 7 sources including the American Academy of Pediatrics, and kff.org 	<ul style="list-style-type: none"> - Summarized four subtopics such as the medical, legal, ethical and social dimensions of this question - Mentioned the American Academy of Pediatrics but did not provide links to any sources

Table A4
 Question Four: I Need Help with an Assignment on Gun Control in the United States

<u>ChatGPT</u>	<u>Gemini</u>	<u>Copilot</u>	<u>Claude</u>	<u>Perplexity</u>	<u>HuggingChat</u>
- Asked for more details such as what aspect of gun control the user is focusing on (legislation, public opinion), the format of the assignment (presentation, essay), and any key points or arguments they want to include	- Provided “a breakdown of the key aspects” to explore such as the 2nd Amendment, public safety, mental health, and gun control measures - Listed 3 recommended resources: government websites such as the CDC, research organizations such as Pew, and “Reputable news outlets with diverse perspectives” such as the NY Times, Wall Street Journal, and the Washington Post	- Presented 4 key points on gun control in the U.S.: The 2nd Amendment, legislation, public opinion and recent developments - Cited and recommended 3 sources: Encyclopedia Britannica, Wikipedia and Pew Research Center - Asked if the user would like more detailed information on a specific aspect of or question on gun control	- Asked the user to provide more details on what aspect of gun control they want to explore - Provided a list of gun control topics to consider such as current laws and regulations	- Provided an overview of recent developments and key points to consider such as recent legislation, executive actions, and a 2024 Supreme Court ruling - Cited 7 sources, including everytown.org, Britannica, and WhiteHouse.org	- Produced an outline for a paper on gun control in the United States - Provided 3 sources at the end of response: justice.gov, pewresearch.org and Wikipedia - Provided a list of sources divided by type including 2 book titles, 2 articles (one of which seems to be fabricated), and 4 websites - Provided 4 writing tips such as using reliable sources and citing sources but did not explain in depth what this involves

Table A5
 Question Five: I Need Primary and Secondary Sources on the Chernobyl Nuclear Disaster

<u>ChatGPT</u>	Gemini	Copilot	Claude	Perplexity	<u>HuggingChat</u>
<p>- Primary sources: Listed 3 collections of primary sources (Chernobyl newspapers, CIA reports, US Dept of Energy reports) but provided links to a commercial site (paperlessarchives.com) to obtain these sources.</p> <p>- Secondary sources: Listed 3 general types of sources such as books (cited a real book by Serhii Plokhyy) articles, and films</p>	<p>- Primary sources: Listed 3 types (government documents, personal accounts and news sources) and provided 5 links to sources</p> <p>- Secondary sources: Listed 3 types (academic books, scholarly articles and documentary films) and 2 book titles.</p> <p>- Suggested ways to find resources e.g. using university libraries and online archives</p>	<p>- Primary sources: Listed 3 resources and 2 links, including the Wilson Center’s Digital Archive and the National Security Archive</p> <p>- Secondary sources: Listed 3 sources with links: Encyclopedia Britannica, History.com, and a book title</p> <p>- Asked for more information on a specific aspect of the disaster</p>	<p>- Primary sources: Listed 7 types and titles but did not provide links nor explain how to access them</p> <p>- Secondary sources: Listed a mix of 7 titles and general categories of sources including a book title and World Health Organization health impact studies</p> <p>- Asked if the user needed more details on the sources or on a particular aspect of the disaster</p>	<p>- Primary sources: Listed 3 types (Soviet government documents, eyewitness accounts, Soviet accident reports) and 5 titles with links to these sources</p> <p>- Secondary sources: Listed 3 types (reports, academic analyses and a book) and 6 titles with links to these titles or to further information about them.</p> <p>- Listed useful follow up questions</p>	<p>- Provided a short and helpful explanation of what primary and secondary sources are</p> <p>- Although the “Web Search” tool was selected, links to online sources were not provided</p> <p>- Primary sources: Listed both specific report titles and general sources such as news articles and photographs</p> <p>- Secondary sources: Listed 1 real book title and then produced fake citations and incoherent text</p>