Journal of Academic Writing
The Journal of the European Association for the Teaching of Academic Writing

Journal of Academic Writing
Vol. 14 No. S2 2025, pages 1-12
https://doi.org/10.18552/joaw.v15iS2.1224

AI literacy in the context of working with sources: Pitfalls and possibilities of generative AI models in academic writing

Tine Wirenfeldt Jensen
Aarhus University/METoDo

Søren Wirenfeldt Jensen
Ittybits

Abstract

This study examines the integration of generative AI (GenAI), such as ChatGPT, into students' academic writing practices, focusing on its use for finding and working with sources. Using the concept of 'imagined affordances' we explore how students perceive and interact with this technology in academic contexts. We tested six student-centric prompting strategies across three fields using ChatGPT 3.5 and 4o, simulating realistic academic writing scenarios. Results show significant variations in the accuracy and usability of generated references across fields, strategies, and model versions. Notably, some strategies based on students' imagined affordances, though technically unsound, produced useful outputs for academic writing tasks. ChatGPT 4o generally outperformed 3.5, highlighting rapid advancements in GenAI's potential role in academic writing. These findings reveal a growing gap between institutional guidance on GenAI use in academic writing and students' potential experiences. We advocate for a nuanced approach to AI literacy in higher education that acknowledges students' perspectives, fosters open dialogue, destigmatizes experimentation while emphasizing critical evaluation, and raises awareness of how imagined affordances shape GenAI interactions during the writing process. This study contributes to discussions on AI integration in academic writing, offering insights for writing instructors, librarians, and policymakers.

Introduction

Interactive tools based on large language models, such as the 'chatbot' ChatGPT, have rapidly been integrated into students' academic writing practices (Malmström et al. 2023; Baek et al. 2023). This new technology, broadly known as generative artificial intelligence (GenAI) has greatly impacted education and writing practices. Much has been written about GenAI's tendency to hallucinate in general (Bang et al., 2023) and specifically about ChatGPT's tendency to generate "plausible-looking but fake references" (Sharples, 2022; see also Agrawal, et al., 2023 and Walters & Wilder, 2023). A plausible-looking reference is a completely new type of search outcome that neither students, academic writing teachers, researchers or librarians has had any previous experience with. As accurate references to literature are a cornerstone in academic work, ChatGPT's fake references have naturally been viewed as very problematic (see e.g. Day, 2023 and Gravel et al., 2023 for discussion within disciplinary fields). As a result, students are faced with warnings not to use ChatGPT when working with sources (see e.g. Hicks, 2023).

In this study, we take a student-centric approach to understanding chatbot use when working with sources as part of the academic writing process. This means that our point of departure is neither the technological possibilities of GenAI in general nor ChatGPT specifically. Instead, we situate our approach within the concept of “imagined affordances” of a technology (Nagy & Neff, 2015). A student’s point of departure when interacting with a chatbot is not a deep understanding of GenAI’s technological capabilities, rather it is the imagined affordances of chatbots understood as perceptions of possibilities for use. According to Nagy & Neff, “perceptions of affordances are as much socially constructed for users as they are technologically configured” (2015, p. 6). To understand how students might approach and use chatbots in the context of working with sources, it is therefore crucial to “better understand what people imagine a tool can do” (Nagy & Neff, 2023, p. 278). These imagined affordances of chatbots are constructed within a specific domain (in this case working with sources as part of the academic writing process) and are closely related to students’ perceptions of what they need and want from a tool in this specific context. Using the concept of imagined affordances as a lens, it becomes visible how students’ imagination shapes how they approach and interact with chatbots.

The social realm of academic writing and knowledge production

The imagined affordances of chatbots, understood as perceptions of possibilities for use, are as mentioned above partly socially constructed. This means that an understanding of the specific social realm of use is needed to understand how interactions with chatbots might unfold. As Bhatt et al. (2019) has shown, students need to navigate an increasing number of digital platforms in their working process. Digital platforms and interfaces impact students “writing and knowledge production” in different and often very non-transparent ways, creating a pressing need to educate students to be critical and reflective in what Bhatt et al. (2019) calls “their ritualized practice with digital technology”. With the advent of GenAI chatbots, students’ academic writing and knowledge production are impacted with a new form of technology that functions in a completely different way than any technology they are familiar with. When interacting with chatbots, students are likely to initially map their strategies from interacting with other tools and interfaces for working with sources onto the unknown domain of chatbots (Schön, 1979). What they understand as their needs and wants when working with sources in an academic context also plays a part in how they interact with chatbots.

Students imagined affordances of chatbots are also shaped by discourses on chatbots in higher education. Students have largely been unable to engage in open discussions with teachers or academic writing instructors about their experiences using ChatGPT for their work with sources. This is due to the widespread perception of any ChatGPT use as a form of cheating and academic dishonesty within educational institutions (see e.g. Bin-Nashwan et al., 2023; Cotton et al., 2024), closely tied to students’ moral status (see Mulholland, 2020; Zwagerman, 2008; Anson, 2022). This view can be understood as an extension of the strong emphasis on plagiarism prevention and the sanctions associated with it. This focus has significantly shaped student experiences with academic writing, especially since the advent of the internet and the subsequent development and widespread adoption of automated plagiarism detection tools based on text matching algorithms (Introna, 2016; Jensen & Bay, 2019; Vardi, 2012). As Anson (2011) and others have shown, this strong focus on plagiarism has not been conducive to helping students develop a complex and nuanced understanding of themselves as academic writers engaging in a wider discourse practice. The characterization of ChatGPT primarily as a new tool for cheating has created an educational environment where students find it risky to initiate open discussions about using ChatGPT as part of their academic writing process, including using it for working with sources.

This study takes a student-centric perspective as its point of departure for investigating how using ChatGPT might be perceived when working with sources in the context of academic writing. We introduce six student-centric strategies for using ChatGPT for working with sources, and systematically evaluate ChatGPT’s proficiency in creating usable academic references from a student perspective. The results add to the ongoing discussion on students’ AI literacy (Laupichler et al., 2022; Long & Magerko, 2020; Miao et al., 2024) as well as show the need for more open dialogue between students, academic writing teachers and librarians about experiences with using ChatGPT in the context of working with sources.

Method

This study applies theoretical insights to understand and evaluate ChatGPT's potential role in students' academic writing practices. Drawing on the concept of imagined affordances (Nagy & Neff, 2015) and Schön's (1979) theory of domain mapping, we developed six prompting strategies that reflect how students might approach and use ChatGPT for finding academic sources (see table 1 below). The strategies range from simple reference requests to more complex approaches involving step-by-step interactions, each designed to emulate different ways students might approach ChatGPT based on their prior experiences with digital research tools and the imagined affordances of chatbots. These strategies were systematically tested across three academic fields using both ChatGPT 3.5 and 4o to evaluate their effectiveness in generating usable academic references. Each strategy was tested three times per topic and model to account for output variability.

We use the term "prompt" when describing these strategies, as the term has been widely accepted as a description of the context a human provides through writing or speech when interacting with GenAI tools such as chatbots (see e.g. White et al., 2023; Mollick & Mollick, 2023).

These six prompting strategies were employed on ChatGPT on topics from three academic fields – rhetoric: 'Racial Discourse and Language', biology: 'Cell-free Systems', and law: 'Privacy Rights and Surveillance'. These academic topics and fields were chosen by the first author, who has extensive experience teaching and producing disciplinary-specific material on academic writing in each of the selected fields. Deploying these six strategies across three different academic fields allowed for identifying potential disciplinary differences in their performance. We tested these strategies both on ChatGPT 3.5 (this work was presented at the EATAW conference 2023: Jensen & Jensen, 2023) and on ChatGPT 4o, which at the time of writing was the newest generally available version of ChatGPT. This allows for identifying potential changes over time.

Six strategies for finding sources with ChatGPT using a student perspective

In the following we describe the six different strategies we developed to explore how ChatGPT might perform from a student perspective. We are of course aware that some of these strategies rely on misunderstandings of how the technology works. The strategies are situated within the concept of imagined affordances (Nagy & Neff, 2015; 2023) from a student perspective, not technological capabilities.

In this study, a reference was considered valid if there was a match in both title and author(s) with sources found in Google Scholar, journal websites, or through general web searches. This criterion was chosen to reflect a pragmatic student perspective: having correct title and author information is typically sufficient for a student to locate the actual paper or book. With this information, a student who is in the process of working on an academic paper will have what might be called a usable reference. We thus prioritize students' practical needs in their academic writing process over strict bibliographic accuracy, where a reference is not correct unless it includes additional elements such as publication year, journal name, volume, issue, and page numbers. The title and author(s) are also the most stable elements of a reference. Other elements, particularly the publication year, can often differ due to various factors such as online-first publication, print publication dates, or database inconsistencies. This stability in title and author information further supports our choice in validation criteria, as it provides a reliable foundation for source identification. By using this pragmatic definition of whether a reference is usable to a student or not, our work differs from studies that seek to test ChatGPT's ability to produce correct bibliographic references (e.g. Walters & Wilder, 2023). We do not aim to test whether a GenAI chatbot can provide the same output as a library database, but rather whether it can produce references useful to a student who is in the process of writing an academic paper.

Table 1 present the six strategies as well as the specific prompts used in full.

Table 1. Six Student-centric Strategies and Prompts

No.	Strategy	Prompt used
1	Make reference	List six references on the topic “[Racial Discourse and Language/Cell-free Systems/AI and Intellectual Property]”.
2	Don’t lie – verifiable links	List six references on the topic “[Racial Discourse and Language/Cell-free Systems/AI and Intellectual Property]”. Ensure that all sources included in the review are non-fictitious and have links for verification.
3	Highly cited	List six references on the topic “[Racial Discourse and Language/Cell-free Systems/AI and Intellectual Property]” with a high number of Google Scholar citations.
4	In the tone of a professor	Create a literature review in the tone of a professor for an academic paper on the topic of “[Racial Discourse and Language/Cell-free Systems/AI and Intellectual Property]”. References should be in APA format.
5	Conversational style, using the chat interface	List six influential scholars on the topic of “[Racial Discourse and Language/Cell-free Systems/AI and Intellectual Property]” in the field of [Rhetoric/Biology/Law]. List a highly cited paper by each scholar. Create a literature review using these papers as references. References should be in APA format.
6	Ask ChatGPT to generate the prompt – adjusted version for references	Write a literature review exploring the current research on “[Racial Discourse and Language/Cell-free Systems/AI and Intellectual Property] in the field of [Rhetoric/Biology/Law]”. Your review should analyze and synthesize at least 6 relevant academic sources, highlighting the key themes, controversies, and gaps in knowledge in the field. Be sure to address the following questions: What are the main arguments and evidence presented by researchers in this field? What are the gaps and limitations in current research, and what future directions should researchers take? Finally, how does your review contribute to a deeper understanding of this field and its implications for [social justice, biotechnology, privacy rights].[1]

The first three strategies are examples of different ways of directly asking the chatbot to provide references, using strategies inspired by interacting with other digital interfaces for searching for references (mapping from one domain to another; see Schön, 1979). The last three strategies are examples of providing context when interacting with the chatbot, where some of the affordances of the chatbot are utilized. In the following, the six strategies are briefly described.

The first and simplest strategy is asking the chatbot to list six references. The only context provided is the name of the topic. The second strategy can be viewed as an attempt to weed out any hallucinated references by adding additional demands to the first prompt, asking for non-fictitious sources and links for verification. This strategy does not work on a technical level, as ChatGPT does not have an internal database of non-fictitious sources nor is able to look up links to these. However, the attempt does align with general discourse on prompt engineering, that stresses the importance of being precise when interacting with chatbots. The third strategy asks specifically for references with a high number of Google Scholar citations. This strategy might be viewed as academically more advanced, as identifying highly cited sources helps locate works that are particularly influential in their field. However, ChatGPT is not able to do this on a technical level, as it does not have direct access to Google Scholar. This strategy shows that understanding how academic discourse is organized partly through a web of citations does not necessarily ensure a technically sensitive use of chatbots in the context of finding sources.

The fourth strategy employs a new strategy: the use of a persona (White et al., 2023; Mollick, 2024a). Using this strategy, the student asks ChatGPT to create a literature review in the tone of a professor for an academic paper on the topic, adding that references should be in APA

[1] These descriptions of topics were produced by ChatGPT.

format. Using this strategy, the student tries to utilize an affordance that is unique to the chatbot, namely the ability to produce text in a certain tone, mimicking a specific type of person's voice in writing. Adding to this, the student asks for a specific citation style (APA). The fifth strategy aims to use a conversational style in a step-by-step process (similar to 'chain-of-thought prompting'; see Wei et al., 2022). In the first iteration, the student asks for a list of six influential scholars on the chosen topic, followed by a request for a list of a highly cited paper of each scholar. Then the student asks for a literature review on the basis of these papers, in APA format. The sixth and final strategy is asking ChatGPT to generate a prompt for creating a literature review. Using this strategy also utilizes the technological affordances of the chatbot as it "offloads the job of writing a task-specific prompt to the language model itself" which is referred to as "metaprompt programming" by Reynolds et al. (2021, p. 1).

Method for testing strategies for using ChatGPT for working with sources

In the following, we describe our approach to testing these six strategies' ability to produce valid references. Inspired by Mollick, who has called for improved rigor when researching chatbots performance (Mollick, 2024b), we include the following technical information. The prompts were executed using ChatGPT 3.5 and 4o without additional capabilities, via the official ChatGPT web interface. Both models were accessible in the free version, albeit with a daily usage limit. For convenience, we utilized a personal account with a Plus plan. The model's linguistic variability settings[2] were not adjustable, and specific model release versions were not explicitly provided by the interface. These parameters can typically be adjusted through an application programming interface (API), which allows users to interact with the language model programmatically and enables precise control over text generation settings – a feature particularly useful in academic research on large language models. However, when adopting a student perspective, our research had to rely on the ChatGPT website rather than an API, preventing us from directly modifying these settings and precluding any experimentation with linguistic variability that is possible with API access. Each prompt was executed three times per topic, per strategy, and per model to ensure consistency and account for output variability. Following each chat session, ChatGPT was instructed to reformat the references into BibTeX format. This step was implemented to facilitate structured, programmatic extraction of reference components such as author names and titles.[3] Due to platform limitations, different methods were used to save chat sessions. For ChatGPT 3.5 a Chrome plugin (Fancy GPT) was utilized to export chat history, and for ChatGPT 4o the built-in chat export function was employed. Each generated reference was validated using a combination of Google Scholar, journal websites, and web searches, and the criterion was correct title and author.

Table 2 shows number of references tested for each model as well as the time of testing.

Table 1. Six Student-centric Strategies and Prompts

ChatGPT model	**Date**	**References**
ChatGPT 3.5	Early 2023	344
ChatGPT 4o	Mid 2024	386

[2] Linguistic variability settings refer to parameters (such as temperature and top-p) that shape the balance between deterministic and creative outputs. Temperature governs the level of randomness in the model's responses (with lower values yielding more predictable text), while top-p sampling restricts token selection to those that cumulatively account for a specified probability mass, thus influencing the diversity of the generated content.

[3] For each session, the complete text output was captured, and the subsequent analysis was conducted exclusively on the BibTeX-formatted references. For some of the six strategies, reference format was not specified, whereas for others APA style was explicitly requested. APA style conventionally uses initials for given names, but the BibTeX conversion process frequently transformed these initials into full names. This transformation did not affect our findings, as ChatGPT correctly identified the author names for the valid references.

The title provided by ChatGPT was programmatically queried in Google Scholar, and the top result's title and author were compared to the generated reference using fuzzy string matching to account for variations in capitalization, names, or initials. If no match was found, manual verification was performed by checking lower-ranked results in Google Scholar, the journal volume's table of contents, and conducting broader web searches as necessary.[4] See the appendix for a more detailed overview.

Results

This section examines three key aspects of our findings: the variation in reference accuracy across fields and strategies, the comparative performance between ChatGPT 3.5 and 4o, and an analysis of ChatGPT 4o's best and worst outcomes. These results reveal important patterns in how students might experience using ChatGPT for source work across different academic contexts. The findings reveal significant variations in the accuracy and usability of references generated by ChatGPT both across different fields, prompting strategies and models. This means that students in different courses of study might have very different experiences with using identical strategies. For example, a student writing about privacy rights using the 'Don't lie' strategy with ChatGPT 3.5 might experience that this strategy generates less than 6% useful references. However, a student writing about cell-free systems could experience that the "Don't lie" strategy generates 67% useful references (see fig. 1). Again, it is important to note that this strategy does not work on a technical level, but that some students might experience it as useful regardless. Likewise, at student writing about racial discourse and language using ChatGPT 3.5 might experience that the strategy 'Highly cited' generates 100% usable references, even though this strategy also does not work on a technical level. A student using the same strategy writing about privacy rights might experience that it only generates 61% usable references (see fig. 1). While students are not likely to think of the output they get from ChatGPT in terms of percentages, they are likely to form an opinion on whether a strategy 'works' based on the output they receive. It is interesting to note how the imagined affordances of a new technology might play a central role in forming an opinion of what works, and that students might experience getting what they need even though the imagined affordances of the technology are at direct odds with the actual technological affordances.

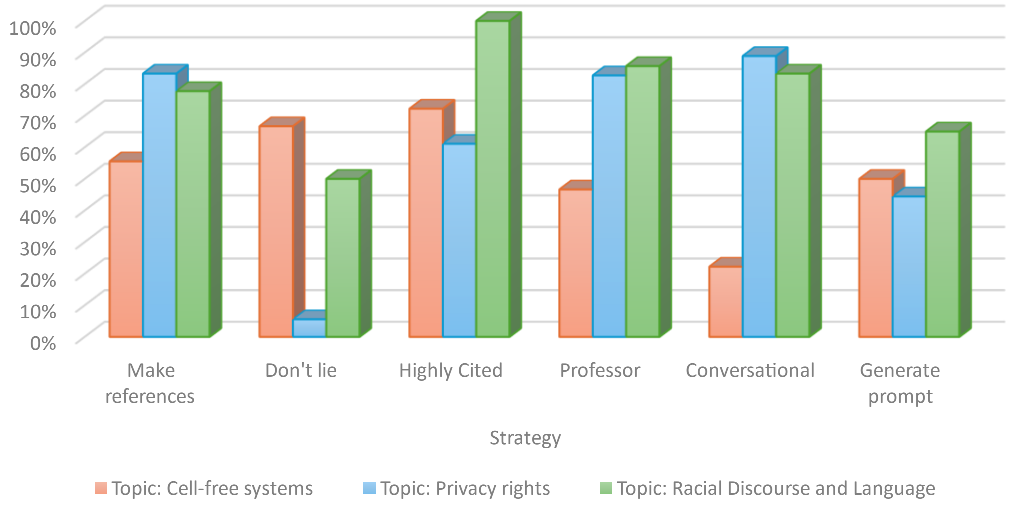

Figure 1. Useable references across six strategies and three topics using ChatGPT 3.5

Overall, both ChatGPT 3.5 and ChatGPT 4o perform better on the topics 'Privacy rights' and 'Racial Discourse and Language' compared to 'Cell-free Systems', indicating that students studying different disciplines might experience the usefulness of ChatGPT very differently when working with sources (see figs. 1 and 2). These differences suggest potential differences in the models' training material across disciplines.

[4] See https://github.com/METoDo-dk/Academic-References for the data sets.

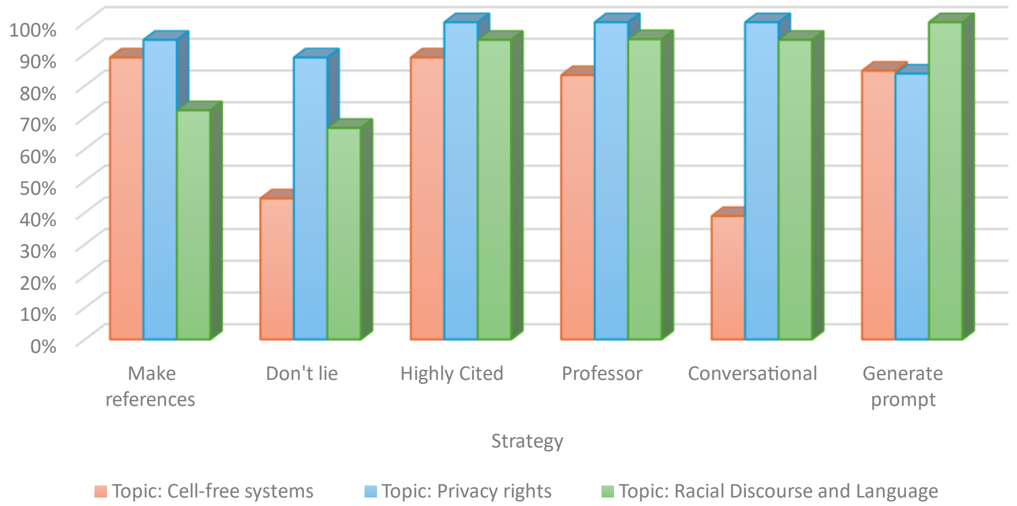

Figure 2. Useable references across six strategies and three topics using ChatGPT 4o

ChatGPT 4o generally outperforms ChatGPT 3.5 in generating usable references across most strategies and topics (see figs. 1 and 2). ChatGPT 4o also shows more consistent performance, with higher percentages of usable references in most categories. Multiple strategies show high effectiveness, with the strategies ‘In the tone of a professor’, ‘Highly cited’ and ‘Conversational style’ often yielding 90–100% usable references. The ‘Conversational style’ strategy shows marked improvement from ChatGPT 3.5 to 4o, suggesting enhanced dialogue capabilities in the newer model. It is important to note that there is no fundamental change to how the models works, and this means that the strategies that do not work on a technical level using ChatGPT 3.5 do not suddenly work using ChatGPT 4o. What has changed is that these three strategies now produce more usable results in the newer model – students attempting these approaches would encounter more reliable outputs compared to working with the earlier version. This would create less friction between imagined and actual affordances, as the model more reliably delivers the expected output.

Best and worst ChatGPT 4o performance

In the above comparison of ChatGPT 3.5 and ChatGPT 4o we use an average of the three times each prompt was run on each model. In the following, we show the best and worst of the three ChatGPT 4o performances. By doing this, we can show how some students might perceive the usefulness of ChatGPT 4o when working with sources.

At its best performance, ChatGPT 4o achieves 100% usable references in many categories (see fig. 3), especially for ‘Privacy Rights’ and ‘Racial Discourse and Language’. This means that an individual student working with privacy rights or racial discourse and language might experience that working with ChatGPT4o in the context of sources is a very good choice, yielding the needed results with a very low risk of generating non-usable references. Overall, ChatGPT 4o shows higher consistency across strategies, with its worst performance (fig. 4) still often outperforming ChatGPT 3.5’s average performance (fig. 1).

ChatGPT 4o represents a significant improvement over 3.5 in generating usable academic references. However, the choice of prompting strategy can significantly impact the number of useful references generated. But even in worst-case scenarios, ChatGPT 4o often produces mainly usable references. The results indicate that students might find ChatGPT4o useful when working with sources in the context of academic writing, even if they are warned against doing so by academic writing teachers and librarians.

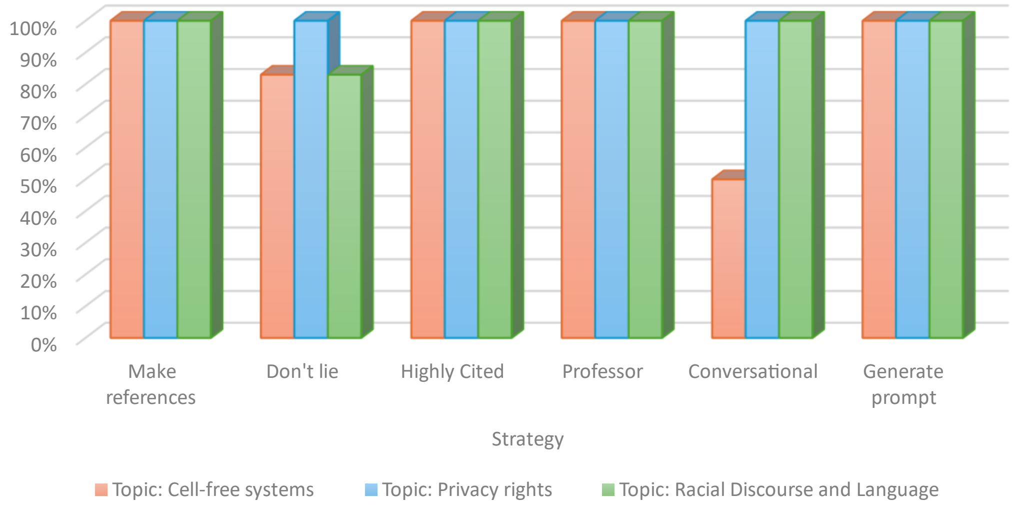

Figure 3. Useable references ChatGPT 4o – best performance

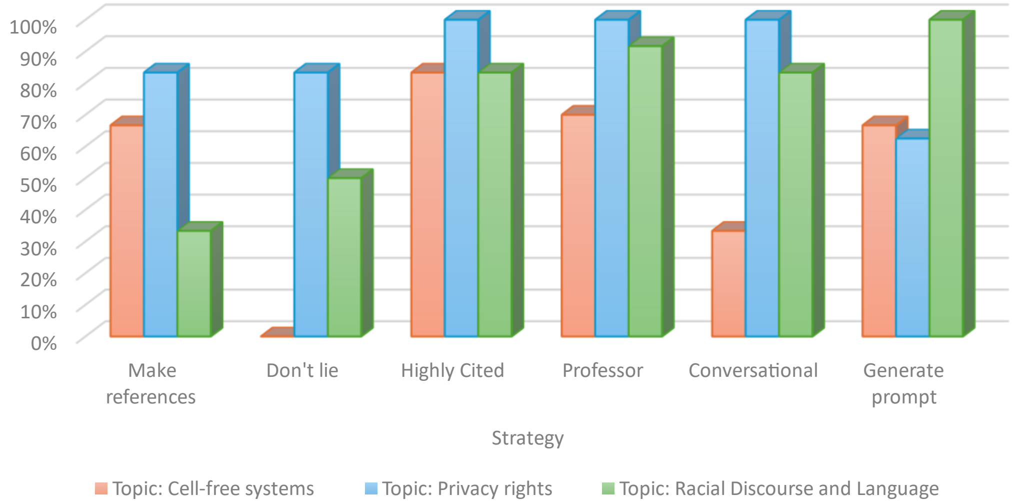

Figure 4. Useable references ChatGPT 4o – worst performance

Discussion

This study was originally undertaken in December 2022 following ChatGPT 3.5's release in late November and submitted to the EATAW conference in early January 2023. The dataset was updated for the conference presentation using ChatGPT 3.5 in early 2023 (Jensen & Jensen, 2023), and for this article we have extended our work with data from ChatGPT 4o in July 2024. With the rapid development of generative AI, it might seem futile to conduct systematic research as presented in this paper. However, it is crucial to record and document these periods in time when radically new technology becomes available, capturing the developments related to these new technologies for future research. Furthermore, it is essential to not only research the development from a purely technological perspective (how the technology functions) but also to consider existing and future human experiences within specific social realms, such as those of students engaged in academic writing. Focusing solely on technological aspects risks overlooking or marginalizing these important human dimensions. But it is difficult to systematically research students' lived experiences with new technology immediately after this becomes accessible with traditional methods such as interviews and focus groups. For this to be feasible, time for adoption must be allowed. We argue that the concept of "imagined affordances" (Nagy & Neff, 2023) has proven useful to imagine how students might interact with ChatGPT in the social realm of academic writing. Using this concept, we have shown that student-centric strategies, often based on “their ritualized practice with digital technology” (Bhatt et al. 2019), may persist even when not technically sound. This underscores the need for a nuanced approach to AI literacy that acknowledges students’ perspectives and experiences.

When describing the method of this study, we have used the term ‘prompt’ (White et al., 2023; Mollick & Mollick, 2023). This terminology, combined with an overly technical focus on providing perfect prompts and an increased market for ‘prompt engineering courses’ has somewhat obscured the fact that these chatbots are inherently natural language models. The notion of ‘prompt’ in the context of GenAI is much more aligned with the tradition of using ‘writing prompts’ in writing courses than any kind of technical or programmatic skill set. In this sense, all writing teachers are, by nature, experienced ‘prompt engineers’.

Conclusion

Our study shows that students might have very different experiences using ChatGPT to generate usable academic references across different fields of study, prompting strategies, and model versions. Students pursuing different disciplines may have vastly different experiences when using ChatGPT for working with sources, supporting different understandings of how chatbots work. Showing the extent of how ChatGPT 4o generally outperforms ChatGPT 3.5 underscores the rapid development of chatbot capabilities. We have shown how some strategies, particularly those leveraging ChatGPT’s unique capabilities (e.g., ‘In the tone of a professor’, ‘Conversational style’), often yield high percentages of usable references, especially with ChatGPT 4o. But interestingly, strategies that do not work on a technical level (e.g., ‘Don’t lie’, ‘Highly cited’) can still produce useful outputs, highlighting the gap between students’ perceptions of ChatGPT’s capabilities and its actual functioning (framed in this paper as imagined vs. actual affordances). And while specialized AI tools for locating sources have become available (such as Elicit, Research Rabbit and Perplexity), students are not necessarily aware of these tools or their technical capabilities compared to chatbots.

These findings have important implications for AI literacy in higher education. While academic institutions, libraries and academic writing teachers have generally advised against using ChatGPT for working with sources, our results suggest that students may find it increasingly useful. Students are likely to perceive at least some ways of working with ChatGPT (especially the 4o version) as very useful when working with sources. Even strategies that do not work on a purely technical level but are employed due to imagined affordances of the technology often generate a useful output. This potentially creates a situation, where students’ lived experiences will not match academic writing teachers’, librarians’ and others’ warnings and expectations, creating a gap between what students gain or perceive to gain and an institutional voice. Consequently, students might overhear and overlook warnings and information in areas where these are fundamental to academic knowledge production, such as heeding content hallucinations and overreliance on RAG (Retrieval–Augment Generation, used e.g. in Perplexity). If academic institutions, academic writing teachers, and university libraries do not take the students’ experiences into account, we risk being viewed as obsolete and irrelevant for future students – especially as there seems to be a lack of space for non-judgmental dialogue about students’ experiences with ChatGPT in their academic writing process. If chatbot use is situated in the realm of cheating and/or students are told that something they successfully do does not actually work, there will be very little incentive to discuss experiences with using ChatGPT for working with sources. There is a pressing need for supporting students’ AI literacy (Laupichler et al., 2022; Long & Magerko, 2020; Miao et al., 2024) and academic writing teachers, tutors and librarians have a key role to play in this. To succeed in this, it is crucial to take the students’ experiences with using ChatGPT when working with sources into account. To bridge the gap between institutional guidance and student practices, we recommend 1) fostering open dialogues about ChatGPT use when working with sources in an academic context, 2) destigmatizing experimentation while emphasizing critical evaluation, and 3) appreciating the role the concept of imagined affordances and domain mapping play in students’ interaction with chatbots.

References

Agrawal, A., Mackey, L., & Kalai, A. T. (2023). *Do language models know when they're hallucinating references?* arXiv. https://doi.org/10.48550/arXiv.2305.18248

Anson, C. M. (2011). Fraudulent practices: Academic misrepresentations of plagiarism in the name of good pedagogy. *Composition Studies, 39*(2), 29–43.

Anson, C. M. (2022). AI-based text generation and the social construction of "fraudulent authorship": A revisitation. *Composition Studies, 50*(1), 37–46.

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu., T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). *A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity*. arXiv. https://doi.org/10.48550/arXiv.2302.04023

Bhatt, I., & MacKenzie, A. (2019). Just Google it! Digital literacy and the epistemology of ignorance. *Teaching in Higher Education, 24*(3), 302–317. https://doi.org/10.1080/13562517.2018.1547276

Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society, 75,* 102370. https://doi.org/10.1016/j.techsoc.2023.102370

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Day, T. (2023). A preliminary investigation of fake peer-reviewed citations and references generated by ChatGPT. *The Professional Geographer*, *75*(6), 1024–1027. https://doi.org/10.1080/00330124.2023.2190373

Gravel, J., D'Amours-Gravel, M., & Osmanlliu, E. (2023). Learning to fake it: Limited responses and fabricated references provided by ChatGPT for medical questions. *Mayo Clinic Proceedings: Digital Health, 1*(3), 226–234. https://doi.org/10.1016/j.mcpdig.2023.05.004

Hicks, M. (2023, August 23). No, ChatGPT can't be your new research assistant. *The Chronicle of Higher Education*. https://www.chronicle.com/article/no-chatgpt-cant-be-your-new-research-assistant

Introna, L. D. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology, & Human Values, 41*(1), 17-49. https://doi.org/10.1177/0162243915587360

Jensen, T. W. & Bay, G. (2019, 1–4 July). *Non-cheater or taking part in the disciplinary dialogue? The impact of plagiarism software on the development of students' authorial identity* [Paper presentation]. 10th Conference of the European Association for the Teaching of Academic Writing, Gøteborg, Sweden. https://easychair.org/smart-program/EATAW2019/2019-07-03.html#talk:106312

Jensen, T. W., & Jensen, S. W. (2023). *AI literacy in the context of working with sources: Pitfalls and possibilities of AI-based natural language production systems in academic writing*. Contribution to EATAW 2023, Winterthur, Switzerland.

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence, 3*, 100101. https://doi.org/10.1016/j.caeai.2022.100101

Long, D. & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3313831.3376727

Miao, F. & Shiohira, K. (2024). *AI competency framework for students*. UNESCO. https://doi.org/10.54675/JKJB9835

Mollick, E. (2024a). *Co-Intelligence*. Random House UK.

Mollick, E. (2024b, March). *Academic journals need to establish AI methods requirements, not just AI disclosure and ethics statements* [Post]. LinkedIn. https://www.linkedin.com/posts/emollick_academic-journals-need-to-establish-ai-methods-activity-7157502786635534336-UQGW

Mollick, E., & Mollick, L. (2023). *Assigning AI: Seven approaches for students, with prompts*. arXiv. https://doi.org/10.48550/arXiv.2306.10052

Mulholland, M.-L. (2020). Honor and shame: Plagiarism and governing student morality. *Journal of College and Character, 21*(2), 104–115. https://doi.org/10.1080/2194587X.2020.1741394

Nagy, P., & Neff, G. (2015). Imagined affordance: Reconstructing a keyword for communication theory. *Social Media + Society, 1*(2). https://doi.org/10.1177/2056305115603385

Nagy, P., & Neff, G. (2023). Rethinking affordances for human–machine communication research. In S. Zuboff (Ed.), *The SAGE handbook of human-machine communication* (pp. 273–279).

Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3411763.3451760

Schön, D. (1979). Generative metaphor: A perspective on problem-setting in social policy. In A. Ortony (Ed.), *Metaphor and thought* (pp. 254–283). Cambridge University Press.

Sharples, M. (2022). Automated essay writing: An AIED opinion. *International Journal of Artificial Intelligence in Education, 32*(4), 1119–1126. https://doi.org/10.1007/s40593-022-00300-7

Vardi, I. (2012). Developing students' referencing skills: A matter of plagiarism, punishment and morality or of learning to write critically? *Higher Education Research & Development, 31*(6), 921–930. https://doi.org/10.1080/07294360.2012.673120

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*, 14045. https://doi.org/10.1038/s41598-023-41032-5

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *NIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems,* 24824–24837. https://dl.acm.org/doi/10.5555/3600270.3602070

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. arXiv. https://doi.org/10.48550/arXiv.2302.11382

Zwagerman, S. (2008). The scarlet P: Plagiarism, panopticism, and the rhetoric of academic integrity. *College Composition and Communication, 59*(4), 676–710. https://doi.org/10.58680/ccc20086674

Appendix. Detailed Overview of Results

ChatGPT 3.5 Early 2023

Topic	Strategy	Number of references	Useable references	Useable %
Cell-free Systems	1	18	10	56%
Cell-free Systems	2	18	12	67%
Cell-free Systems	3	18	13	72%
Cell-free Systems	4	15	7	47%
Cell-free Systems	5	18	4	22%
Cell-free Systems	6	18	9	50%
Privacy Rights	1	18	15	83%
Privacy Rights	2	18	1	6%
Privacy Rights	3	18	11	61%
Privacy Rights	4	29	24	83%
Privacy Rights	5	18	16	89%
Privacy Rights	6	18	8	44%
Racial Discourse	1	18	14	78%
Racial Discourse	2	18	9	50%
Racial Discourse	3	18	18	100%
Racial Discourse	4	28	24	86%
Racial Discourse	5	18	15	83%
Racial Discourse	6	20	13	65%

ChatGPT 4o, Mid 2024

Topic	Strategy	Number of references	Useable references	Useable %
Cell-free Systems	1	18	16	89%
Cell-free Systems	2	18	8	44%
Cell-free Systems	3	18	16	89%
Cell-Free Systems	4	30	25	83%
Cell-Free Systems	5	18	7	39%
Cell-Free Systems	6	20	17	85%
Privacy Rights	1	18	17	94%
Privacy Rights	2	18	16	89%
Privacy Rights	3	18	18	100%
Privacy Rights	4	34	34	100%
Privacy Rights	5	18	18	100%
Privacy Rights	6	29	25	84%
Racial Discourse	1	18	13	72%
Racial Discourse	2	18	12	67%
Racial Discourse	3	18	17	94%
Racial Discourse	4	33	31	95%
Racial Discourse	5	18	17	94%
Racial Discourse	6	24	24	100%