Generative AI in Technical Communication: A Review of Research from 2023 to 2024

By Carol Reeves and J.J. Sylvia IV

Abstract

Since its release in late 2022, ChatGPT and subsequent generative artificial intelligence (GAI) tools have raised a wide variety of questions and concerns for the field of technical communication: How will these tools be incorporated into professional settings? How might we appropriately integrate these tools into our research and teaching? In this review, we examine research published in 2023–2024 addressing these questions (N = 28). Overall, we find preliminary evidence that GAI tools can positively impact student writing and assessment; they also have the potential to assist with some aspects of academic and medical research and writing. However, there are concerns about their reliability and the ethical conundrums raised when they are used inappropriately or when their outputs cannot be distinguished from humans. More research is needed for evidence-based teaching and research strategies as well as policies guiding ethical use. We offer suggestions for new research avenues and methods.

Introduction

The field of technical writing and communication is one of several disciplines facing the possibility that GAI tools will alter how we teach writing, conduct research, and prepare our students for their careers. Integrating these tools into these aspects of our work raises significant practical and ethical challenges. What is appropriate and inappropriate use of GAI in student writing, in teaching, and in professional communication? How can our students use GAI tools to master writing skills without being mastered by the tools? How much can we trust these tools to produce valid and truthful results? As GAI tools are fed more data and as our prompting strategies improve, will they change our conceptions of the nature of writing and authorship?

1

What if we cannot distinguish human from GAI written texts? And what if we forget how to write, just as many of us have forgotten how to spell due to spell checker bots? We cannot answer all of these questions in this review; they are the questions for future research.

Considering these challenges, this article provides an overview of what we deem to be the most relevant technical communication-related journal articles about GAI that have been published between January 2023 and February 2024. Table 1, below, includes the criteria used to include and exclude articles. In short, articles needed to address the GAI chat tools that emerged beginning with ChatGPT-3.5 in November of 2022, focus on technical communication or writing, be peer-reviewed (no pre-prints or conference papers) and have clear, reproducible research methods. We analyze the extant literature both to situate what is currently known about GAI as it relates to this discipline, but also to highlight underexplored research areas. Additionally, by examining the results and limitations of current research, it is possible to make some overtures toward emerging best practices in this nascent field. These findings will be valuable for both practitioners and educators in technical communication, as we, collectively, all struggle to make sense of the vast changes that are reshaping the educational landscape.

Methods

We wished to identify studies of the use of GAI tools for writing and communication in technical and professional settings and in the classroom. Because these new tools bring capabilities that exceed those of precursor AI tools and because they are being adopted in workplaces and by students before we can test their reliability or establish policies guiding their use, we decided to narrow our scope to studies of chat-based GAI tools that were published between January 2023 and February 2024. The focus on chat-based GAI tools aligns with the theme of the special issue.

This approach seeks to understand a broad picture of how research is addressing GAI in the field, so we used a non-systematic approach that did not require a particular research question, in part because of the narrow focus on an emerging tool. Table 1 includes our criteria for inclusion and exclusion, as discussed above.

- - - insert Table 1. here - - -

We began by conducting keyword searching (using the terms, "ChatGPT" and "artificial intelligence") of journals related to the fields of technical and professional writing, communication, and pedagogy (see table 2). Search terms were selected that we believed would best surface articles that addressed generative artificial intelligence and topics related broadly to the fields of technical communication/writing. We searched journals in the field first to ensure that articles from this field were explicitly included in the review.

- - - insert Table 2. here - - -

After searching the major technical communication journals, the search (see table 3 for search terms) was expanded to include full databases (listed in table 4). This was done to ensure inclusion of articles that address technical writing and communication but were not published in technical communication journals. The search terms were modified in this case to ensure that articles were related in some way to the field of technical communication/writing.

- - - insert Table 3. here - - -

- - - insert Table 4. here - - -

Search Results

The technical writing and communication journal search results uncovered only two

studies, both published in the *Business and Professional Communication Quarterly* (table 3). Our database searches produced 26 studies meeting our criteria from a variety of fields (table 5). We found significant attention to GAI tools in the sciences and medicine. In the sciences, GAI tools might easily generate research protocols and/or forms of writing that follow generic conventions, such as abstracts or other sections of research reports. In medicine, GAI tools could effectively and efficiently generate technical reports and patient communication. Our search terms (table 2) and a manual review of the articles ensured that those included in these fields addressed technical writing or pedagogy approaches linked to the use of GAI tools. Table 5 indicates the fields and journals in which we found technical writing/communication and GAI related articles.

- - - insert Table 5. here - - -

After narrowing our results down to 28 studies, we completed a narrative style literature review to summarize and interpret the findings. While we believe that ethics and social justice lenses on GAI are of utmost importance, we note that none of the articles we found through this method met the criteria for inclusion. The articles addressing social justice were either not focused on the field of technical communication within the time frame specified or were editorial submissions that did not undergo peer review.[1] This indicates a potential research gap that needs to be filled. Much of the work on social justice related to technical communication and AI predates the release of chat-based GAI systems and may need to be updated or revised in light of the different affordances of these new systems.[2]

---

[1] Although our review focused on journal articles, one book related book of interest is *Augmentation Technologies and Artificial Intelligence in Technical Communication: Designing Ethical Futures* by Ann Hill Duin and Isabel Pedersen.

[2] For a social justice perspective published in 2022, see, for example, "AI for Social Justice: New Methodological Horizons in Technical Communication" by S. Scott Graham and Hannah R. Hopkins.

ChatGPT in Technical Writing and Communication Education

Teaching strategies

Research about teaching with GAI generally indicates that instructors believe it is important for students to learn about AI as part of their curriculum, but integrating material into the curriculum is proving challenging both because of the pace at which AI tools are being developed and improved and because the use of GAI by students challenges existing assessment practices. The studies reviewed below explore perceptions about and strategies for using GAI to teach writing. While GAI integrations were generally found to be effective, important nuances must be noted and investigated further.

Cardon et al. (2023a) completed a survey of 343 communication instructors to learn their opinions about the use of GAI and ChatGPT in businesses and about how they might prepare students for the workforce. The survey contained Likert-style as well as open-ended questions that allowed write-in answers. Participants acknowledged that as businesses integrate GAI, their writing pedagogy will need to adjust to prepare students, but they fear they will not be able to keep up with advancing technology. They also expressed concerns about assessment and academic integrity, which have been discussed widely in education literature, particularly in opinion pieces. Of particular interest was a concern that GAI would reduce students' ability to think critically and creatively. Some responses took this one step further, noting that relying on GAI for writing could reduce writers' agency and ability to find their authentic voices. Participants were asked to explain GAI-related strategies they have used in their classrooms. Many participants reported that they find that asking students to evaluate ChatGPT output and

compare ChatGPT generated text with human written texts is an effective exercise for teaching rhetorical analysis. Other participants found ChatGPT helpful for brainstorming, though this approach is certainly not without controversy, as brainstorming is an important first step in developing one's own ideas and sense of agency in writing. Some participants allow students to use ChatGPT to generate a first draft to revise. The authors emphasize that integrating GAI into the classroom will necessitate a focus on higher-order learning skills assessed through project-based learning, which will require smaller class sizes and teaching loads that enable more personalized interaction between instructors and students.

Faiz et al. (2023) employed a mixed-methods approach to evaluate the impact of ChatGPT (no version noted) as a writing aid for intermediate-level EFL learners from four government colleges in Sargodha, Pakistan. ChatGPT was integrated into the EFL curriculum through a phased approach, with guidelines provided for optimal use. Fifty participants balanced in gender, completed writing assignments, some with the help of ChatGPT and some without. They conducted quantitative assessment of grammatical accuracy, lexical richness, and syntactic complexity using tools like Grammarly and Syntax Analyzer, as well as qualitative thematic analysis of learner feedback. Pre-study and post-study writing assessments, along with surveys and interviews, were conducted to gauge shifts in writing proficiency and learner perceptions of ChatGPT's effectiveness. While it seems that the post-tests were analyzed for results, this is never explicitly stated. Unsurprisingly, they found that ChatGPT-assisted writing had fewer spelling errors, fewer lexical errors, and fewer syntactic errors than writing produced without the tool. Most learners also reported that using ChatGPT improved their writing.

ChatGPT may help improve surface features of writing, but how well does it perform when asked to assist with higher order writing and creative processes? Harunasari (2023)

conducted a case study in which 16 undergraduate EFL students were introduced to ChatGPT (no version noted, but likely 3.5) as a tool for assisting with their creative writing assignment. The author assessed the participants' creative writing before and after the use of ChatGPT as an assisting tool. Assessment was based on literary and narrative elements such as creative and literary expression, character, plot, and conflict. The results were mixed; although the maximum individual student score was the same for both the pre- and post-test writing samples, integrating ChatGPT led to improvement for those students whose work was rated the lowest in the pre-test, from 55 to 68, and mean scores improved 4 percentage points. Integrating ChatGPT in creative writing did not lead to universal benefits. While eight students, who scored at the lower end, improved their scores, five students' scores stayed the same, and three students performed worse. The authors did not report which areas of the rubric scores increased or decreased the most.

One explanation for variability in the post test results could be variable prompting strategies. Harunasari (2023) had students submit their ChatGPT chat history using the "Share Link to chat" feature, which allowed for a deeper qualitative understanding of how students were using the tool. Analysis of these chat logs revealed that the quality of ChatGTP's output is impacted by the content of the prompts, suggesting that it could be worthwhile to provide detailed instructions to students about best practices for iterative prompting with GAI. He also found that ChatGPT could be used as a distraction by students, with one student engaging in a game of chess against the GAI.

Assessment and feedback on writing

Assessing student writing can be tedious and time consuming. Moreover, as anyone who has taught writing knows, human raters can be inconsistent, biased, and too busy to thoroughly

assess a large stack of assignments. While we don't want GAI tools to write our students' essays, we may want their help with assessment and feedback. The research highlighted in this section attempts to evaluate whether that feedback is accurate and if it helps or hinders student learning outcomes. Overall, these studies indicate that when provided specific criteria or rubrics, GAI tools can accurately assess student writing and provide useful feedback and recommendations. However, while using GAI tools for assessing and providing feedback may save time, instructors cannot trust them completely.

Parker et al. (2023) tested the use of ChatGPT-3.5 for automated assessment of the writing by student nurses. They assembled 42 human rated texts from a corpus of nursing student papers that had all received a rating of 3 for meets or exceeds criteria. Each essay was submitted to ChatGPT-3.5 separately for assessment with an elaborate prompt that specified writing constructs, criteria, and a point system for each element. They also prompted the tool to provide feedback. ChatGTP-3.5 was a stricter grader, having assigned a 3 (A) to only one of the 42 papers and 2s (B) to the rest. Assessing the tool's ability to provide macro-level feedback, they found that ChatGPT-3.5 addressed broader features of the writing rather than rule-based surface grammar and correctness. The tool also provided appropriate suggestions for improvement using example excerpts from the essays. The authors conclude that GAI tools' ability to provide more complex feedback could lead students to make more substantial improvements in their writing.

Jiang et al. (2023) analyzed how accurate four AI systems were in assessing the correctness of Chinese writing when compared to evaluations made by human raters. The writing analyzed was created by 41 online high school students in the U.S. The AI systems tested included GPT-4, GPT-3.5, iFLYTEK and Baidu Cloud. This research analyzed writing at the T-unit level, with a total set of 2,197 T-units collected. Human raters first analyzed these T-units

for word, structure, and context for comparison to AI results. Although some models performed slightly better than others on certain criteria, overall GPT-4 was the most precise in identifying errors.

In their study of using GAI for assessment, Mizumoto and Eguchi (2023) used a GPT-3.5 API model to assess 12,100 essays that are part of the ETS Corpus of Non-Native Written English (TOEFL11) and compared these to the benchmark scores. As part of the assessment prompt, they included the IELTS TASK 2 Writing band descriptors that included a rubric. In addition to the GPT-3.5 scoring, the authors used a multiple regression model to test if the inclusion of other measures along with the GPT-3.5 scores could better match the baseline scores for the essays. The authors found that GPT-3.5 scoring only had exact matches with baseline scores in 54.33% of cases, but matched with adjacent agreement 89.15% of the times, suggesting this tool is feasible for automatic grading. They also found that the inclusion of linguistic features in the regression model performed the best. In terms of pedagogical implications, the authors noted that GPT-3.5 could also be used to generate a revised version of a student's essay that would demonstrate how the score could be improved. They noted that ethical guidelines would need to be emphasized to prevent students from simply writing or generating their own essays in this manner. They conclude that while this technology could be helpful in lightening the load of assessment for instructors, instructors should not be entirely removed from the process of assessment.

Guo and Wang (2023) compared the amount and type of feedback provided by ChatGTP-3.5 and instructors on essays by fifty Chinese undergraduate students. The results demonstrated that ChatGTP-3.5 produced more feedback than did the human instructors and spread it more evenly across content, organization, and language. While ChatGTP-3.5 provided more feedback,

the instructors expressed concern that students might have trouble interpreting and applying the feedback due to its length and because the tool was unable to provide feedback as annotation, tied to specific problems. The authors argue that their results suggest that there is significant potential for a collaborative approach to grading between humans and GAI. However, this research did not address either student opinions on feedback or how the feedback impacted student learning outcomes. The authors suggest that future research might be able to determine whether the additional feedback offered by ChatGPT helped or harmed student learning.

Song and Song (2023) compared ChatGPT-based and human feedback on the writing skills of 50 EFL students divided evenly into control and experimental groups. The experimental group received extensive training on how to use ChatGPT (no version noted) to improve their writing skills while maintaining their own distinctive writing style. The results indicate that students receiving feedback from ChatGPT had increased scores compared to the control group in all areas assessed, including writing content, writing organization, language use, and writing motivation. Qualitative results of semi-structured interviews indicate that students in the experimental group had a very positive experience, noting for themselves their improved writing and motivation, while appreciating the accessibility, convenience, and personalized feedback of ChatGPT. Two notes of caution emerged from the interviews. Student interviews revealed that at times the tool did not fully understand the context of the writing and therefore gave poor feedback. Second, students worried that they would come to depend on GAI, using it for feedback on every sentence they write to maintain the improvements gained by its initial use. Based on this data, the authors argue that instructors should consider teaching students how to move beyond using GAI tools in their writing process, weaning themselves from it as they improve their writing.

Escalante et al. (2023) also compared feedback provided by ChatGTP-4 with human feedback. The authors first compared the learning outcomes of 48 students, split into two groups, one of which received feedback from ChatGPT-4 and the other from human tutors. Learning outcomes were measured using pre- and post-tests to assess linguistic progress. The results indicate no statistically significant difference in the outcomes of the two groups, suggesting ChatGPT-4's feedback was equally as helpful as that of the human tutors. 43 students who received feedback from both ChatGTP-4 and tutors were surveyed about which they preferred, and the results were nearly evenly distributed. Those who preferred human feedback found it to be more engaging and noted the benefit of being able to ask follow-up questions, while those who preferred GAI feedback referred to its greater clarity and specificity. The authors conclude that GAI could be effectively used to provide feedback on student essays without any negative impact on learning outcomes. They helpfully provide an appendix with their prompts and an example of the GAI feedback, but no example human feedback is provided for comparison.

Attitudes about appropriate uses of GAI in writing

 Instructors and students may have vastly different attitudes about what constitutes appropriate and ethical uses of GAI tools. While there is instructor enthusiasm for students using GAI in some specific ways, there remains an underlying concern that continued use of GAI will increase cases of academic dishonesty and diminish critical thinking skills, especially if it is used before students have learned to write well on their own. Yet as they enthusiastically experiment with these tools, students may have the same concerns.

Barrett and Pack (2023) compared the attitudes toward the appropriateness of using GAI in students' writing process and in teachers' evaluation process. A questionnaire asked 158

11

university students and 68 professors at a university in the U.S. to rate on a 5-point Likert scale the appropriateness of examples of the use of GAI for student writers and teacher evaluation. ChatGPT-3.5 was prompted to brainstorm topics, provide an outline, write, and revise a writing assignment. It was also prompted to provide assessment and feedback on a poorly written paragraph. Participants then rated the appropriateness of using these methods. The results indicate a surprising unanimity among both groups around a key point: students should not employ GAI tools for writing activities if they cannot perform them on their own. Both groups were equally certain that submitting GAI brainstormed ideas without disclosure is inappropriate, but they were uncertain about whether brainstormed ideas should be submitted at all, with half of both groups approving and the rest either uncertain or disapproving. As for outlining, both groups generally agreed that GAI could be used to produce outlines to model the outlining process, but more students than teachers believed that a GAI outline could form the structure of an essay and could be submitted with disclosure. Both groups agreed that all use of GAI should be disclosed. Teachers and students generally agreed that it was wrong to ask GAI to write an entire essay, even if the student had good writing skills, and wrong to submit a GAI essay with or without disclosure. They agreed that using GAI for drafting was acceptable if the student finished the task without further help from the tool. As for revision, both groups agreed that assistance was acceptable if the student knew how to revise, did most of the revision, and disclosed the use of GAI.

With evaluation and feedback on student writing, both groups agreed that it was wrong for teachers to use GAI evaluation and feedback without disclosing its use. More teachers than students felt that it was appropriate to use GAI feedback and evaluation as models while more students than teachers were concerned about teacher use of GAI for evaluation and feedback.

When asked if their institution had clear policies regarding the use of GAI, students had mixed answers while teachers indicated with certainty that their university had no policies in place. 95% of participants reported that they had no training on how to use GAI while 90% of teachers reported that they had not trained their students.

Xiao and Zhi (2023) conducted semi-structured interviews with 5 university students in China who regularly used ChatGPT-3.5, to better understand how these students were using the tool. Students described using the tool as a peer tutor and an idea generator. The students also explained that they used iterative prompting to get the most useful outputs, noting an awareness that the responses may not be entirely accurate. Finally, while the students interviewed expressed awareness of the educational benefits of ChatGPT, they believed that other students would use ChatGPT to complete entire assignments only when they were not interested or invested in assignments. These advanced students were able to learn how to use the tool iteratively and reflect carefully on how they used it. Because of this, the authors recommend that instructors incorporate GAI into their teaching practices rather than attempting to ban it.

Finally, Hang (2023) surveyed 20 ESL instructors who had already incorporated ChatGPT into their courses about their opinions about the use of ChatGPT in writing classes at Van Lang University in Vietnam. Overall, instructors expressed positive views about using ChatGPT to prepare lesson plans and engaging learning materials for class and to assist students with their writing. However, they were more ambivalent about using ChatGPT to assist with grading and assessment. A major area of concern for half of the participants was the fear that students' overuse of GAI tools would hamper the development of critical thinking, problem-solving, and writing skills. Further, fifty percent of interviewees noted that some students were submitting verbatim responses from ChatGPT rather than revising them or using them to

improve their own work. Further research into the impact of GAI use in critical thinking abilities seems warranted based on these reported observations.

Professional Writing

There is limited research on how professionals are using GAI in the workforce, but the research that exists indicates it has been adopted by professionals quickly, suggesting the need to help students learn to use the tools as they themselves become professionals. Additional research into the biases of GAI will also be important in thinking about how it can best be incorporated in professional situations. For example, if, as the research below indicates, GAI is very truth-biased, that may have implications for how it is incorporated into customer service use.

Businesses are already relying on ChatGPT for marketing, copywriting, and consumer information. Cardon et al. (2023b) surveyed 692 business practitioners about to GAI adoption. Over 80% of respondents reported they were already using GAI in their work, with 41% doing so weekly. Respondents identified using GAI for research, writing, and, to a lesser extent, editing and revising. They reported that GAI increased efficiency (84%) and quality (67%) of the work and helped them generate ideas (74%). When asked about how best to prepare students for using GAI in the workplace, 78% of participants believed that integrity and strategic vision are the most important AI-related skills. At the same time, due to the presence of GAI, the participants expressed that soft skills, primarily oral communication in a face-to-face setting, would be even more important, as would ethical training. The authors conclude that business students should not only be trained to use GAI but also should be taught ethical principles guiding them to use these tools morally and with careful consideration of the consequences of unethical use.

Businesses and consumers will need to know if GAI can distinguish between truthful and deceptive messages. Markowitz and Hancock (2023) first measured whether GAI was able to detect deceptive messages in two sets of data containing statements about friends and one set of hotel reviews. They then evaluated whether GAI had a truth bias similar to humans, in which humans tend to believe a message is true unless there is a notable reason to think otherwise. They tested three AI models, which included ChatGPT-3.5, Bard, and ChatSonic, based on GPT-4. While human and GAI deception detection rates were similar (both near chance of approximately 50%), GAI was notably more truth-biased than humans, with GAI evaluating nearly 100% of all messages as true. This averaged out to being correct approximately half the time. These results suggest that, at least at this point, these GAI tools may not reliably detect false messaging.

Academic research and writing

GAI tools could become assistants or even participants in academic research and writing if they can be trusted to perform reliably and consistently. For less complex tasks, such as summarizing texts, generating references, simple data analysis, and writing tasks requiring less rhetorical decision making, they could save valuable time. If they can match human performance on higher order analysis and interpretation and writing requiring nimble rhetorical decision making, we might see knowledge progress more quickly. If they match our skills, ethical conundrums naturally arise. For now, it appears that we cannot fully trust the GAI tools. The studies reviewed below, all employing ChatGPT-3.5 and/or 4, may leave us both disappointed because we cannot totally rely on ChatGPT to perform tedious and time-consuming low risk

tasks as well as relieved because, so far, ChatGPT cannot match humans on complex analytical and rhetorical tasks.

Athaluri et al. (2023) investigated ChatGPT-3.5's research planning and proposal writing capabilities. They asked ChatGPT-3.5 to generate a list of scientific research topics that could be investigated by undergraduate medical students and that would meet the FINER criteria (Feasible, Interesting, Novel, Ethical, and Relevant) used in science to formulate research questions. Scientists participating in the study were asked to evaluate the topics; they agreed that all 50 topics generated met the criteria, indicating the efficacy of GAI tools in early research processes. The authors also asked the tool to write a research protocol for each of the 50 topics, with an introduction, objectives, methodology, implications, and references. Evaluation centered on completeness, conventional format, and validity of references rather than on correctness and other limitations in the protocols. They discovered some cases of incompleteness and conventional format violations. The major problem was "hallucinated" references. Of the 178 references provided in the proposals, 109 turned up on a Google search with a valid DOI.

GAI tools may be able to perform higher order research tasks, such as advanced data analysis. The field of corpus or critical discourse analysis has been employing AI tools such as ATLAS.ti to automate data coding and topic modeling for many years. However, these tools cannot replace human discernment in contextualized analysis. To determine ChatGTP-4's performance on key a range of discourse analysis tasks, Curry et al. (2024) prompted it to replicate the results of three published discourse studies. In the first case, a published corpus analysis of key words in an online support group for a health problem, ChatGPT-4 was asked to group keywords from decontextualized data. In the second case, a published concordance analysis of the appearance of the word "homosexual" in newspapers stories about Islam,

ChatGPT-4 was prompted to analyze a set of concordance lines; these concordance lines were taken from a 143-million-word database from a published discourse analysis. In the last case, a study of direct and indirect questions in economics research articles, ChatGPT-4 was asked to identify markers of reader engagement by conducting a contrastive analysis of direct and indirect questions in a corpus of economic research articles. In the first case, the authors found that ChatGPT-4 could group key words semantically, but the categories it produced were mostly too generic to be useful for inductive or thematic coding. In the second case study, ChatGTP-4 produced an inaccurate concordance analysis full of errors. In the third case, ChatGPT-4's question analysis was full of problems—inaccuracies, mis-identified questions, fabricated questions, and inconsistent results. The authors conclude that ChatGPT-4 cannot be trusted to perform complex corpus discourse analysis.

To save time, scholars might employ GAI tools for simple, less rhetorically demanding tasks. One example is the conference abstract. Babl & Babl (2023) asked a non-scientist to prompt ChatGPT-3.5 to write a 250-word conference abstract on plantar fasciitis in children based on a fictitious dataset provided. They prompted the tool to generate a title, an introduction, methods, results, conclusion and two references. Despite one fake reference, the tool generated an abstract the authors expected would pass through the abstract review process. The authors are optimistic that the tool can, with supervision, assist scientists with drafting and polishing documents and non-English speaking researchers with translation. They express no reservations about the use of these tools to write methods and results sections and expect them to be widely incorporated as, they claim, these sections follow generic conventions.

At this point, entire manuscripts submitted to journals that were generated by a GAI tool have been easily detected (Stokel-Walker, 2023), likely because their rhetorical and analytical

skills are still primitive. As these tools advance, they might acquire more sophisticated rhetorical decision-making capabilities and be indistinguishable from human written reports. Ariyaratne et al. (2023) wanted to know if ChatGPT-3.5 could produce a research article good enough to pass through the review process. They fed the tool data but no references from 5 published or under review articles on skeletal radiology and asked it to write research reports on the problems or topics in the original studies, following the appropriate generic requirements. Raters from the radiology field assessed each section of the resulting articles on a 1-to-5-point scale with 5 being excellent. All 5 articles were deemed well-structured, but they contained, to different degrees, serious problems, such as inappropriate techniques, fictitious case reports, contradictory or inappropriate solutions. All contained fictitious references.

Similarly, Chamurliyski (2023) prompted ChatGPT-3.5 to write an introduction to a paper on the yields of common wheat cultivars, a research aims statement, and to perform an analysis of provided data and explain the results. The authors found that the tool was able to perform basic statistical analysis and explain the results, but it provided inaccurate interpretations of the results. ChatGPT-3.5 did not produce an effective introduction and could not generate graphs.

.

Writing and Communications in Healthcare

In Healthcare, practitioners face time constraints that impinge on their ability to produce thorough case and operative reports as well as provide clear patient communication. Interest in the capabilities of GAI tools for medical writing and communication has exploded in the medical community with scores of editorials offering optimistic forecasts and/or warnings about errors and ethical issues. GAI tools' efficacy in medical writing and communication has received little

research attention, however. From the studies we found, GAI tools can perform lower-order and low risk tasks but cannot perform higher-order, high risk tasks.

Generative AI tools might, with their summarizing ability, provide clear patient-friendly translations of complex medical information. Lyu et al. (2023) analyzed translations by ChatGTP-3.5 of 138 original radiology reports and, when it became available, by ChatGPT-4. They used three levels of specificity in their prompting. They began with a simple prompt to translate the reports into "plain English" and to provide suggestions for both patients and healthcare providers. The 138 ChatGPT-3.5 reports were first evaluated on a 5-point scale by two experienced radiologists based on accessibility, completeness, and accuracy of information. They found ChatGPT-3.5 did not follow a uniform format but did produce accessible translations that replaced medical jargon with understandable terms, integrated explanations and definitions that were not included in the original reports, and rewrote lists of findings, common in technical reports, into understandable sentences. As far as information completeness and correctness, while receiving an overall score of 4.268, ChatGPT-3.5 reports were not uniformly consistent, with some reports lacking information or providing incorrect information or oversimplifying information. Among the MRI translations, for example, only 32% received a 4 or 5 for correct information.

To determine if the tool would generate uniform translations, the authors prompted ChatGPT-3.5 to generate ten translations of the same radiology report. The tool generated ten translations that varied in format and information value with only five providing clear and correct information. The authors then asked for ten more translations of the same report using an optimized prompt that asked for a specific number of paragraphs and information to include in each paragraph. This time, seven of the ten were rated overall as good. Finally, the authors

experimented with even more specific prompts that asked for translations aimed at audience knowledge or educational level. Interestingly, these prompts led to translations that were similar to those produced from the first prompt and far worse that the earlier set of ten produced with the second optimized prompt. ChatGPT-4's arrival afforded the authors an opportunity to compare version 4 translations with the older version. Using the original and optimized prompts and the same analytical methods, they found that ChatGPT-4 generated translations were dramatically better than those generated by ChatGPT 3.5 as far as information completeness and accuracy. Both versions of ChatGPT produced equally accessible and concise translations, yet both showed problems with random formatting. Even ChatGTP-4 did not follow the format provided by the optimized prompt providing specific structural requirements.

Answering patients' questions requires care and attention to patient's level of knowledge. In their investigation of ChatGTP-3.5's capacity for providing patient information, Wu et al. (2024) prompted the tool to provide patient-friendly answers to 378 questions about epilepsy treatment, general information about epilepsy, types of epilepsy, and five questions about emotional support. To test the tool for bias or offensive language, they prompted the tool to play the role of an experienced neurologist responding to an epilepsy patient's question. Two certified epileptologists independently evaluated the accuracy and consistency of each response then awarded them 1 (correct and comprehensive) to 4 (completely incorrect) points. The evaluators also assessed the tool's ability to respond to online questions of a more personal and emotional nature, such as when patients describe a recent experience, express emotion, and ask what they should do. Their evaluation identified correct and comprehensive answers to 68.4% of the questions and correct but incomplete answers to 17.9% of the questions. ChatGTP-3.5's answers to questions about epilepsy, diagnosis, and treatment were relatively comprehensive, but it

provided correct and comprehensive information about complications and prognosis to 55.6% of the questions. As far as emotional sensitivity in answering questions of a personal nature, they found that ChatGTP-3.5 provided a degree of emotional support, responding with empathetic language first, practical advice, and encouragement to have a positive attitude, follow their treatment plan, and seek out support groups. The authors caution that while ChatGTP-3.5 has the potential to provide more consistent, reliable, and patient friendly information than is available through internet searching and social media, it is limited in its capacity to answer all types of information. It performs better when asked direct and structured questions and much worse on ambiguous questions that a human would be able to interpret. Patients would need to know how to ask direct and structured questions. The tool also tends to provide answers that are too inclusive, providing as many answers as possible to avoid leaving out information. Many possible answers were incorrect. A third problem is that as the tool cannot evaluate evidence and data sources, contradictory or conflicting information can lead it to generate ambiguous answers. The authors expect that patients will likely use ChatGPT for encouragement and emotional support as well as basic information, but they need to be warned against using it to make their healthcare decisions.

Along the same lines, Gravel et al (2023). prompted the ChatGTP-3.5 to answer 20 medical questions and provide sources. They based the questions, mostly beginning with "what," on five public research articles from different topics and fields and did not specify a word limit. Raters assessed validity of the references and the quality of the responses on a 0-to-100-point scale. Out of 59 references, 41 were fabricated even though the authors referenced had previously published articles on the topic, and all titles were appropriately related to the question. When prompted to explain the inaccuracies, ChatGPT-3.5 appeared to "lie" by insisting the

references existed in PubMed or "excused" itself by admitting that it makes mistakes. To assess the quality of the tool's answers, 20 of the authors who wrote the original articles identified major factual errors in a quarter of the answers. The authors, describing ChatGPT's responses as confidently wrong, emphasize the need for an improved algorithm that optimizes reinforcement learning through interaction with humans.

Generative AI tools might be able to generate patient case reports more efficiently and accurately than busy human health care professionals. To test ChatGPT-3.5's capabilities, Ho et al. (2023) asked a medical student to write a surgical case report using the tool. Primed with raw data on the condition, ChatGPT-3.5 was asked to answer a medical question under two conditions: 1) with only raw data, 2) with raw data and the patient's specific medical history. It was then asked follow-up questions to determine whether it retained an understanding of the patient's case and could provide contextualized explanations specific to the case and the operative procedure. Based on its responses, the student then wrote a full operative case report. The authors found that ChatGTP-3.5 was able to answer medical questions specific to the health problem, providing thorough background information and simplified explanations of complex information. It also maintained an understanding of the patient's specific history to answer context and case-specific questions. The student found the tool's help valuable in writing the case report even though it provided fake citations.

Detection

As we formulate personal and professional guidelines about appropriate and inappropriate uses of these tools—in our academic professions and in the classroom—we must ask where we draw the line and whether we can determine when a line is crossed. At this point,

incipient research on detection faces the fast-paced evolution of these tools. As the studies below suggest, there are, as yet, no definitive methods to distinguish GAI-generated and human written text. However, detection might be improved if we develop greater sensitivity to linguistic and rhetorical issues rather than rely on routine or conventional assessments.

Can we easily detect AI-generated texts that are well written, follow generic conventions, and contain accurate information and true citations? Casal and Kessler (2023) asked 72 journal editors and reviewers in the field of Linguistics to identify which of eight abstracts, (four generated by ChatGPT-4 and four published abstracts) were written by an AI tool. When asked in a survey to rate their confidence in their ability to distinguish AI-generated from human written texts, most participants (77%) expressed neutrality or low confidence while no one expressed complete confidence. In the identification task, participants performed poorly, with a positive identification rate of 38.9%. Participants were more likely to correctly identify human texts and to misidentify AI texts as human. None of the 72 participants correctly identified all four of the AI abstracts as definitely or likely AI. In interviews with seven participants about what features led them to classify the abstracts, the authors identified four main rationales:

(1) *continuity and coherence,*

(2) *specificity or vagueness of details,*

(3) *familiarity and voice,*

(4) *writing quality at the sentence-level* (p. 6).

Participants applied organizational rationale—coherence, flow, and continuity—most often to distinguish AI-generated from human written abstracts. They identified abstracts they deemed easier to read as human and abstracts containing what they felt were unnatural sentence level relations as AI-generated. However, as Casal and Kessler point out, the continuity and coherence

rationale was unreliable as the participants emphasizing this rationale correctly identified the author only 22.2% of the time. Participants mentioning specificity or vagueness attributed to humans those abstracts containing a high degree of specificity, but they correctly identified the abstract based on this criterion only 28.6% of the time. Some participants felt they could detect a human voice, or familiar tone, or a certain uniqueness in an abstract that would not be possible in an AI-generated text, but this criterion led to correct identification only 20% of the time. As for writing quality, participants mentioning this rationale identified abstracts as human-written if they contained well-constructed sentences. This rationale appears to have been the most reliable, with a 60% correct identification rate.

When asked for their opinions about ethical use of ChatGPT in research and writing, Casal and Kessler's (2023) participants generally agreed that these new tools could not be avoided, but the field needed to proactively determine how they could benefit research ethically. Text editing was the top task identified most often as both beneficial and ethical followed by writing computer code. There was some support for using ChatGPT for writing summaries and abstracts, but clearly, the editors interviewed did not support using the tool for writing sections of academic articles or for analyzing complex data.

Perhaps AI-generated text can be detected if we identified distinguishing syntactic peculiarities. Leong (2023) proposes the use of "clause complexing," essentially a sophisticated stylistic analysis that uncovers the ranking of and relationships among clauses to determine the level of density and complexity. Identifying specific syntactical differences between GAI-generated and human text might improve detection. Loeng assembled a corpus of 200 abstracts, 50 of which were published in *Nature* in 2023 while 150 corresponding to the original *Nature* abstracts were generated by ChatGPT-4, Bard, and Poe. Loeng asked the tools to write an

abstract of 177 words for an article that would be published in a scientific journal. He provided the titles of the original articles and brief topic descriptions. For each set of abstracts, he computed the frequency of clause types (main, subordinate, embedded), clause ranking (main clause to subordinate clauses, and inter-clausal relations (signaling forms of hypotaxis and parataxis) per 100 words. He found that none of the AI tools matched the *Nature* abstracts in all measures, with ChatGPT-4 abstracts coming the closest to matching the *Nature* abstracts but with important differences in subordinate clauses, such as the finding that finite adverbial clauses were much more frequent in *Nature* abstracts. Grammarly, Loeng suggests, can apply the *Nature* abstract clausal complexing results to detect AI-generated text, especially for Poe and Bard. However, since *Nature* is a general science journal, the linguistic profile of the abstracts, while generally following the norms of scientific style, might differ from that of other journals. Still, this study suggests that while Bard and Poe generated abstracts might be accurately detected using clause complexing, ChatGPT-4 generated abstracts may not be as easily detected.

Markowitz et al. (2023) explored whether inherently false AI reviews of hotels could be detected through a linguistic analysis. In terms of business communication work, fake reviews are a widespread problem, whether they occur in the form of intentionally false human generated reviews or inherently false GAI reviews. The authors prompted ChatGPT-3.5 to produce positive reviews for 20 specified Chicago hotels in 120 words (ChatGPT will not generate negative reviews). They then collected 800 human written reviews from Trip Advisor, 400 of which were written by paid workers, thus intentionally false, and 400 written by people who had bought and used the product. The reviews were then analyzed using an analytic writing index, the Flesch Reading Ease metric, and descriptive statistics. They found that GAI reviews were significantly more emotive, with stronger positive language, and more descriptive, with more adjectives, than

were deceptive and truthful human reviews. GAI reviews were also less readable, with longer sentences, than human reviews. The overly positive emotive and descriptive language in ChatGPT-3.5 reviews, the authors conclude, suggests that the tool cannot adjust affective language to suit generic conventions of experience-based writing. While humans are not generally able to differentiate between GAI and human created reviews, the linguistic analysis used here had a high success rate. In terms of implementation, these results could be used to help identify and remove fake reviews, though the authors note that this study will need to be updated as GAI tools continue to improve.

Currently available tools for detecting GAI student essays can produce false positive results, leading teachers to accuse innocent students. Herbold et al. (2023) used human raters and computational linguistic analysis to compare argumentative essays produced by ChatGPT-3.5 and 4 with essays produced by high school students learning English. As we might expect, both ChatGPT-3.5 and 4 produced essays that were more complete, more linguistically complex, and more rigidly structured than student essays. While these differences might not be so evident in the university setting, we might use lexical complexity and rigid structure as possible indicators of AI-produced argumentative essays. Ibrahim (2023) experimented with using an earlier version of GAI, GPT-2, to detect writing generated by ChatGPT-3.5. He employed two RoBERTa classifiers, which are based on a GPT-2 model, to detect ChatGPT generated essays. Among 240 essays, Ibrahim found that the two RoBERTa classifiers could identify GAI writing with 89.2% accuracy, however, it generated inconsistent results because it contained both false negatives and positives. The presence of false positives could certainly present difficulties in its use, as that may lead to making accusations of students who had not used GAI in their work. This research does suggest that further research may be needed with more advanced GAI detection tools. If an

early detection model based on GPT-2 can be mostly correct in detecting GAI generated essays, then more advanced GAI tools may perform even better. More research on the utility of advanced GAI detection capabilities is needed.

Discussion

Our review leads us to conclude that GAI has the potential to assist writing pedagogy and student learning, assist in lower-level academic research and writing tasks, and provide readable translations in healthcare. However, evidence also supports concerns about reliability, misuse and overuse of the technology, as well as ethical violations.

Based on surveys and interviews, instructors are experimenting with GAI in their classrooms rather than banning them and reporting some good results with brainstorming and outlining as well as with assessment and feedback. Instructors and students are concerned about students becoming over-reliant and submitting work entirely generated by GAI, practices that might threaten the development of critical thinking skills. However, fears are not evidence of a negative impact on critical thinking. GAI use in the classroom may actually enhance critical thinking as students learn prompting strategies and apply their rhetorical knowledge to assess the results. Further exploration into the additional affordances of GAI tools is needed. For example, some instructors in Cardon et al.'s (2023a) research expressed that allowing the use of ChatGPT in their classes has allowed them to shift their instructional methods further up Bloom's taxonomy, affording them the opportunity to move more quickly into the levels that focus on analysis, evaluation, and creation.

Professionals are already adopting GAI technologies at a rapid pace, indicating the need to incorporate GAI instruction into the classroom. The pedagogical studies suggest that we need

more research into practices that will enable students to interact with GAI productively and appropriately. Some studies indicate that stronger students are more likely than weaker students to navigate iterative prompting as well reflect critically on issues of academic integrity. Because our students will enter workplaces where they will need these prompting and critical thinking skills, we must also emphasize soft workplace skills to provide them a competitive advantage.

Emerging research on detection reveals many challenges, especially if we want certainty rather than a spectrum of probability. Raters can distinguish between student essays and GAI essays when the students are unskilled, but GAI tools can generate texts following recognized conventions, such as abstracts, that are difficult to detect, even by editors. More research experimenting with linguistic markers, voice, and trained GAI tools trained in detection is needed.

GAI shows promise when it comes to medical writing, with an ability to translate radiology reports into patient-friendly language and even, performing best when given case information. However, although it was able to produce empathetic responses, it struggled in answering questions about complex medical topics and often hallucinated sources, though much of this research was conducted using the less advanced ChatGTP-3.5.

Much of the emerging research around education is focused on whether GAI can be effectively implemented for grading and assessment. While there is also an emerging body of research on how effective ChatGPT is for assisting student learning, there was no research, beyond one survey, related to instructors using ChatGPT to generate materials, lesson plans, or other recommendations for their teaching. Academics wishing to use GAI for research should take a cautious approach, as so far, GAI tools have struggled with more advanced tasks such as thematic coding, concordance analysis, and identifying rhetorical devices.

Our analysis of current research on GAI also suggests several lessons for writing about GAI research in future work. One author specifically noted drastically different results in GAI output depending on the way their prompt was written. One immediate implication is that no one should consider their results authoritatively conclusive about the capabilities of GAI tools, as tweaks to the wording of prompts that are used can potentially change the outcome of the study. We argue that all research that uses GAI prompts should be required to include the exact wording of those prompts in their research as part of the methods, as this is important for being able to reproduce and improve on the methods. We relatedly suggest that it is vital that the version and date of GAI software being used should be clearly stated in the methods. We often saw the use of the phrase "ChatGPT" only, which served to obscure whether ChatGPT-3.5 or ChatGPT-4 was being used. In addition to changes in official numbering, the capabilities of this GAI also change depending on when it is used. For example, ChatGPT-4 has, at various points in time, had the ability to browse the internet added or removed, in addition to the possibility of being supplemented by customized GPTs. These abilities can make very significant differences in the outcomes of research. For example, before ChatGPT-4 had the ability to access the internet, many incorrectly believed that ChatGPT-3.5 had that ability because if given a URL, it often appeared that it was able to access a page based on its extrapolations of what that page might contain based on the URL. Reporting the model and date of use for all GAI tools will help clarify these potential confusions.

This review has several limitations. Excluding opinion and commentary pieces might have narrowed the spectrum of viewpoints about policy and ethics. These types of pieces were much more likely to discuss ethics than the peer reviewed studies that we analyzed for this review. GAI also remains an emerging topic, with new research emerging constantly. This

review can only serve as a snapshot of this moment in time. Further, not only is new research continually emerging, but the GAI tools themselves are in a period of rapid updates, with their abilities steadily improving. There will necessarily be a gap between the latest tools and the research that is being published about them due to the speed of academic publishing. Finally, the narrative style literature review can be subjective, with the potential for bias in which studies are included.

There are many research areas that appear to remain entirely unaddressed. Technical communicators play an important role in designing and evaluating user experiences for both products and services, yet little research has been conducted on how GAI can be used to help create user experiences or be integrated into them. Accessibility will also be an important area of future research, including both how GAI can be leveraged to create accessible materials, such as through generating alternative text for images or closed captions for videos, but also in ensuring that GAI tools are themselves accessible. Finally, while ethical considerations are at least referenced frequently, despite being under-explored, the legal and regulatory frameworks surrounding GAI are almost never discussed. Legal questions around copyright, for example, will be quite important for organizations that want to integrate GAI tools, as current law suggests material generated by AI cannot be copyrighted. Finally, issues of privacy and security are also under-discussed. The ability to run customized, locally hosted GAI tools is a topic that is likely just over the horizon of current research.

Conclusion

GAI tools have rapidly impacted the field of technical writing and communication, as well as pedagogical approaches to the discipline. Current research highlights an overall positive

30

sentiment about the incorporation of the tool into the field, but there are clear concerns regarding how to ensure it is used ethically as well as the long-term impacts on critical thinking. While there is emerging research on how to incorporate GAI into teaching practices, much more experimentation and research needs to be conducted, especially as the technology itself continues to evolve. Future GAI research should report detailed prompts and model information to enable reproducibility and clarify capabilities at the time of the study. Many key issues like user experience, accessibility, legal frameworks and privacy remain under-explored so far.

References

Ariyaratne, S., Iyengar, Karthikeyan. P., Nischal, N., Chitti Babu, N., & Botchu, R. (2023). A comparison of ChatGPT-generated articles with human-written articles. *Skeletal Radiology*, *52*(9), 1755–1758. https://doi.org/10.1007/s00256-023-04340-5

Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*, *15*(4). https://doi.org/10.7759/cureus.37432

Babl, F. E., & Babl, M. P. (2023). Generative artificial intelligence: Can CHATGPT write a quality abstract? *Emergency Medicine Australasia*, *35*(5), 809–811. https://doi.org/10.1111/1742-6723.14233

Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, *20*(1), 59. https://doi.org/10.1186/s41239-023-00427-0

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age. *Business and Professional Communication Quarterly*, *86*(3), 257–295. https://doi.org/10.1177/23294906231176517

Cardon, P., Fleischmann, C., Logemann, M., Heidewald, J., Aritz, J., & Swartz, S. (2023). Competencies Needed by Business Professionals in the AI Age: Character and Communication Lead the Way. *Business and Professional Communication Quarterly*, 23294906231208166. https://doi.org/10.1177/23294906231208166

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, *2*(3). https://doi.org/10.1016/j.rmal.2023.100068

Chamurliyski, P. (2023). Enhancing efficiency in scientific writing in the field of plant growing: Collaborating with AI assistant ChatGPT for enhanced productivity. *Acta Scientifica Naturalis*, *10*(3), 73–81.

Curry, N., Baker, P., & Brookes, G. (2024). Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT. *Applied Corpus Linguistics*, *4*(1), 100082. https://doi.org/10.1016/j.acorp.2023.100082

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Faiz, R. (2023). Optimizing ChatGPT as a Writing Aid for EFL Learners: Balancing Assistance and Skill Development in Writing Proficiency. *Linguistic Forum*, *5*(3), 24–37. https://doi.org/10.53057/linfo/2023.5.3.3

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: Clinical Health*, 1(3), 226-224.

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12146-0

Harunasari, S. Y. (2023). Examining the Effectiveness of AI-integrated Approach in EFL Writing: A Case of ChatGPT. *International Journal of Progressive Sciences and Technologies*, *39*(2), 357. https://doi.org/10.52155/ijpsat.v39.2.5516

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, *13*(1), 18617. https://doi.org/10.1038/s41598-023-45644-9

Ho, W. L. J., Koussayer, B., & Sujka, J. (2023). ChatGPT: Friend or foe in medical writing? An example of how ChatGPT can be utilized in writing case reports. *Surgery in Practice and Science*, *14*, 100185. https://doi.org/10.1016/j.sipas.2023.100185

Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: "The Terminator Versus the Machines." *Language Testing in Asia*, *13*(1), 46. https://doi.org/10.1186/s40468-023-00260-2

Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning. *Languages*, *8*(4), 247. https://doi.org/10.3390/languages8040247

Leong, A. P. (2023). Clause complexing in research-article abstracts: Comparing human- and AI-generated texts. *Explorations in English Language and Linguistics*, *11*(2), 99–132. https://doi.org/10.2478/exell-2023-0008

Lyu, Q., Tan, J., Zapadka, M. E., Ponnatapura, J., Niu, C., Myers, K. J., Wang, G., & Whitlow, C. T. (2023). Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: Results, limitations, and potential. *Visual Computing for Industry, Biomedicine, and Art*, *6*(1), 9. https://doi.org/10.1186/s42492-023-00136-5

Markowitz, D. M., & Hancock, J. T. (2024). Generative AI Are More Truth-Biased Than Humans: A Replication and Extension of Core Truth-Default Theory Principles. *Journal of Language and Social Psychology*, *43*(2), 261–267. https://doi.org/10.1177/0261927X231220404

Markowitz, D. M., Hancock, J. T., & Bailenson, J. N. (2024). Linguistic Markers of Inherently False AI Communication and Intentionally False Human Communication: Evidence From Hotel Reviews. *Journal of Language and Social Psychology*, *43*(1), 63–82. https://doi.org/10.1177/0261927X231200201

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, *2*(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050

Nguyen, T. T. H. (2023). EFL Teachers' Perspectives toward the Use of ChatGPT in Writing Classes: A Case Study at Van Lang University. *International Journal of Language Instruction*, *2*(3), 1–47. https://doi.org/10.54855/ijli.23231

Parker, J. L., Becker, K., & Carroca, C. (2023). ChatGPT for Automated Writing Evaluation in Scholarly Writing Instruction. *Journal of Nursing Education*, *62*(12), 721–727. https://doi.org/10.3928/01484834-20231006-02

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the

    efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*,

    14, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove.

    Nature, *613*(7945), 620–621. https://doi.org/10.1038/d41586-023-00107-z

Wu, Y., Zhang, Z., Dong, X., Hong, S., Hu, Y., Liang, P., Li, L., Zou, B., Wu, X., Wang, D., Chen,

    H., Qiu, H., Tang, H., Kang, K., Li, Q., & Zhai, X. (2024). Evaluating the performance of the

    language model ChatGPT in responding to common questions of people with epilepsy. *Epilepsy*

    & Behavior, *151*, 109645. https://doi.org/10.1016/j.yebeh.2024.109645

Xiao, Y., & Zhi, Y. (2023). An Exploratory Study of EFL Learners' Use of ChatGPT for Language

    Learning Tasks: Experience and Perceptions. *Languages*, *8*(3), 212.

    https://doi.org/10.3390/languages8030212

Table 1. Criteria for inclusion and exclusion

Criteria	Eligibility	Exclusion
Tools investigated	Generative AI tools	Pre-generative AI tools
Timeline	January 2023- February 2024	Before 2023
Working language	Published in the English language	Not published in the English language
Literature type	Original research articles in print or accepted	Editorials, letters, reviews, conference proceedings, pre-prints
Fields	Technical/Professional Writing and Communication: Writing Pedagogy Technical/Professional Writing and Communication: Research	Medical Diagnostics Scientific research methodology/not writing Online user forums or social media research
Methods	Methods are clearly explained Methodology includes bias controls Reproducible	Methods unclear Methods demonstrate a misunderstanding of how ChatGPT works
Topic focus	Written Communication Writing Pedagogy	Learning subject matter, in general

Table 2. Journals Searched

Fields	No articles on GAI	Contained Articles on GAI
Writing and Communication Pedagogy and Research:	JTWC Language and Communication Journal of Communication JBTC Communication Teacher TCQ BCQ College Composition and Communication Kairos Written Communication Technical Communication IJBC	Business and Professional Communication Quarterly (two studies)

Table 3. Search Terms for Databases

Writing Pedagogy:	Generative artificial intelligence and/or ChatGPT and Writing and pedagogy Artificial Intelligence and Writing Pedagogy ChatGPT and Writing Pedagogy
Professional Writing and Communication	ChatGPT and/or Generative artificial intelligence ChatGPT and/or Generative Artificial intelligence and technical communication ChatGPT and/or Generative artificial intelligence and medical writing

Table 4. Databases Searched

Google Scholar
PubMed
Science Direct
Communications and Mass Media

Table 5. Included studies: Topics and Journals

Topics	Journals
Writing Pedagogy	
Sciences and Medicine	Nature Journal of Nursing Education International Journal of Progressive Sciences and Technologies
Education	International Journal of Educational Journal of Higher Education Thinking Skills and Creativity Education and Information Technologies
Linguistics	Languages Research Methods in Applied Linguistics Linguistic Forum
ELL/ESL/LL	Language Testing In Asia Languages
Psychology	Frontiers in Psychology Journal of Language and Social Psychology
Professional Writing and Research	
Science and Medicine	Skeletal Radiology Emergency Medicine Australasia Acta Scientifica Cureus Surgery in Practice and Science Mayo Clinic Proc Digital health Visual Computing for Industry, Biomedicine, and Art Epilepsy and Behavior
Linguistics	Explorations in English Language and Linguistics Research Methods in Applied Linguistics Applied Corpus Linguistics
Business	Business and Professional Communication Quarterly