



Vol.7 No.2 (2024)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : <http://journals.sfu.ca/jalt/index.php/jalt/index>

Can instructors detect AI-generated papers? Postsecondary writing instructor knowledge and perceptions of AI

Nathan Murray^A

A

Algoma University, Canada

Elisa Tersigni^B

B

University of Toronto Mississauga, Canada

Keywords

Academic integrity;
AI;
artificial intelligence;
ChatGPT;
essay writing;
Generative AI;
higher education;
writing pedagogy.

Abstract

Our study assesses the knowledge and perceptions that postsecondary writing instructors have of generative AI programs such as ChatGPT and tests instructors' ability to distinguish between essays written entirely by students and essays generated by ChatGPT. We tested and interviewed twenty experienced postsecondary instructors currently teaching writing. Participants graded four essays and attempted to identify which essays were AI-generated. We found that writing instructors have a moderate level of confidence in their ability to distinguish between student and AI-generated writing but a low level of accuracy—only 35% of instructors could correctly identify the authorship of all four essays. AI-generated essays scored higher than essays written by students, especially in spelling, grammar, and organization, while they scored lower in argumentation and evidence. We suggest that instructors will need to find ways to encourage students to work independently while learning to use AI as a writing support. In our conclusion, we discuss pedagogical solutions that allow the use of AI and propose that these solutions can complement administrative ones.

Correspondence

elisa.tersigni@utoronto.ca ^B

Article Info

Received 1 June 2024

Received in revised form 6 July 2024

Accepted 20 July 2024

Available online 22 July 2024

DOI: <https://doi.org/10.37074/jalt.2024.7.2.12>

Introduction

Large language models such as OpenAI's ChatGPT, Google's Gemini (formerly Bard), Bing Chat, Anthropic's Claude, and Microsoft's Copilot have received a great deal of media attention since their release to the consumer market. Although generative AI will have long-term impacts on many industries, it has had the most immediate impact on secondary and postsecondary education. Because these programs can generate grammatically correct and logically cohesive prose on specific topics, they are able to produce work that appears to satisfy the learning objectives of many secondary and postsecondary writing assignments. Furthermore, because the output is produced in seconds, LLMs can produce work for any unsupervised written assignment, regardless of instructor-set time limits.

Recent studies, as well as the professional experience of the investigators in their respective classrooms, have shown that the use of ChatGPT is widespread among university students in diverse countries and that many students have positive impressions of the potential of generative AI software (Hasanein & Sobaih, 2023; Ibrahim et al., 2023; Thi Thuy, 2023; Zou & Liang, 2023). Although OpenAI has not provided confirmation, observers have suggested that the 10% drop in traffic that ChatGPT experienced in the month of June 2023 was mostly due to high school and university students on summer vacation (Vynck, 2023). A recent study showed that 45% of computer science students at a UK university are using ChatGPT and other LLMs in their academic work (Singh et al., 2023), while a recent survey of American college students showed that 48% of respondents admitted using ChatGPT to help complete graded assignments (Impact Research, 2024). In online message boards such as Reddit, users have posted about submitting assignments composed entirely of ChatGPT output for credit in their university classes and receiving top grades for their submissions (DK-Sonic, 2023).

The use of generative AI creates challenges across the education sector, particularly in "measurement, information accuracy, and skill devaluation" (Steele, 2023, para.1). The availability and popularity of these AI text technologies threaten to upend traditional pedagogical methods for teaching in the humanities and social sciences, in which take-home essays and other written assignments have remained one of the primary modes of assessment in North America for decades (Rudolph et al., 2023) especially because the established consensus emerging among instructors across disciplines and countries is that the use of AI amounts to plagiarism (Ibrahim et al., 2023). These issues are heightened when teaching composition classes where the primary pedagogical goal is learning to write.

Students of composition and rhetoric who are still learning to write effectively may find that these tools can provide a shortcut to a finished essay. The power of these programs to write effectively, as well as the presumption that these programs will improve further over time, has led some commentators to declare the essay (as a meaningful method of assessment) "dead" (Kumar, 2023; Marche, 2022). The ubiquity of AI-powered tools potentially undermines the development of students' writing and critical thinking

skills: the Modern Languages Association has warned that LLMs may result in learning loss as "[s]tudents may miss writing, reading, and thinking practice because they submit generative AI outputs as their own work or depend on generative AI summaries of texts rather than reading" (MLA-CCCC Joint Task Force on Writing and AI, 2023, p. 7). A raft of new online tools have emerged that claim to reliably screen student assignments for AI-generated texts, including GPTZero, Copyleaks, and Turnitin AI Writing Detection, but all of these products are vulnerable to producing false positives (Sadasivan et al., 2023). One recent study found that the best-performing AI detection software was only accurate 71.4% of the time (Chaka, 2023). Students falsely accused of using AI to generate their work have no meaningful recourse as there is no way to prove authorship based on criteria such as perplexity and 'burstiness'. Preliminary research has also demonstrated that AI writing detection programs are more likely to produce false positives when assessing work written by English as an Additional Language (EAL) learners (Liang et al., 2023).

In sum, generative AI presents an urgent problem for postsecondary instructors assessing written assignments, as distinguishing between student and AI writing is often difficult. Furthermore, when university instructors have a faulty understanding of generative AI, they can put student success at risk by falsely accusing students of academic dishonesty (Perkins et al., 2023; Verma, 2023).

Over the past year, researchers have begun to answer the question of whether university instructors can determine whether a text is student-written or AI-generated. Hassoulas et al. (2023) examined the work of 34 experienced markers in medical school, finding that 50% of participants correctly identified student submissions and only 19 to 23% of markers correctly identified the ChatGPT script. Furthermore, marker suspicions of ChatGPT script were "not proven valid on most occasions" (Hassoulas et al., 2023, p. 75). Casal and Kessler (2023) examined whether linguists could differentiate between abstracts written by humans and abstracts written by AI in those same articles. Although they considered the work of 72 frequent reviewers, only 38.9% of the abstracts were accurately identified. Dugan et al. (2023) explored whether graduate students and senior undergraduates were able to detect when a text transitioned from human-written to AI-generated, finding a large variance and relatively low accuracy in ability that could be somewhat improved with practice. Finally, Fleckenstein et al. (2024) tested the ability of pre-service and experienced high school teachers to identify the difference between student-written and AI-generated essays, finding approximately 50% accuracy, with higher accuracy in the experienced instructors. Thus far, as far as we know, no study has examined whether experienced postsecondary writing instructors can differentiate between student-written and AI-generated text, nor has any study conducted interviews to identify why and how instructors are identifying texts as they are. As experienced postsecondary writing instructors are some of the most sophisticated readers in the institution, it is especially important to clarify whether they can differentiate between student and AI text. The primary purpose of the discipline of composition is to teach writing as a critical skill; if AI generators are undermining the acquisition and evaluation of that skill, the discipline's

methods and priorities may need to be reconsidered.

Our study thus assesses instructor knowledge and perceptions of the current state of AI text generators and tests instructors' ability to distinguish between essays written entirely by students and essays generated by ChatGPT. Our research questions are as follows:

1. Are writing instructors knowledgeable about the current state of AI text generators?
2. How are writing instructors adapting their instruction and pedagogy in response to the public availability of AI text generators?
3. Independent of their institutions' policies, do writing instructors perceive students' use of AI text generators as a violation of academic integrity?
4. Are writing instructors able to distinguish between essays written entirely by students and essays generated by ChatGPT using the same essay prompt?

Methodology

We interviewed 20 instructors currently teaching composition at a college and/or university. 17 of these instructors teach in Canada, two teach in the US, and one teaches in New Zealand. We solicited participants using professional listservs and social media and confirmed the participants' institutional affiliations. The interviews took place in March and April 2023 and consisted of two portions: a questionnaire and a semi-structured interview.

All twenty participants were current, highly experienced community-college and/or university-level writing instructors: at the time of the interview, 70% of them had taught for more than five years (65% had taught for 11 or more years), and 60% had taught four or more courses in the past 12 months.

The questionnaire portion involved the instructors reading and assessing four essays using an assigned rubric, and then answering basic questions about their experience teaching writing at the postsecondary level. The four sample essays all responded to the same assignment prompt for a research essay discussing the challenges faced by an industry affected by COVID-19. The assignment prompt and rubric were originally designed by one of the authors, Nathan Murray, for a first-year composition course at Algoma University. They can be found in the Appendix. The assignment required students to provide evidence by reading and citing a minimum of four sources in APA style. Essays 1 and 3 were written and submitted by students for a first-year writing composition class at Algoma University in August 2022, before the public release of ChatGPT. Like the instructor respondents, the students were compensated with an honorarium for their participation in the study. Essays 2 and 4 were constructed by the authors using ChatGPT. Essay 2 was generated using ChatGPT3.5 (January 30, 2023 edition) on Feb 6, 2023. Essay 2 was generated using the identical

essay prompt that the students were given. The text of ChatGPT's output was copied and was not further refined except for the format of the text to match APA standards. Essay 4 was generated using ChatGPT3.5 (February 13, 2023 edition) on March 3, 2023. Essay 4 was generated using the initial prompt, as well as three additional simple prompts to refine the generated text:

1. "Can you rewrite the introduction? It is too short and it doesn't have a specific thesis statement."
2. "Can you rewrite the first two body paragraphs? They are too short and they don't have any evidence or citations."
3. "Can you rewrite the remaining body paragraphs? They are also too short and they don't have any evidence or citations."

These additional prompts are designed to reflect the conversational interactions students may have with the software, encouraged by the conversational user interface (Rospigliosi, 2023). The revised outputs were assembled into a complete paper, but the text shared with the participants consisted entirely of ChatGPT output. The four essays were anonymized and formatted identically; however, titles, references, and spelling and grammar errors were retained.

ChatGPT generated essays on similar topics to the topics that students chose, such as the effects of COVID-19 on the hospitality and travel industries. In both generated essays, the sources were all hallucinations—i.e., none of the provided sources were real. This remains the biggest obstacle for generative AI programs to produce complete assignments independently. While the newest version of ChatGPT (4o) is more likely to produce real sources when asked, the content cited from the real source is still likely to be hallucinated. Nevertheless, none of the claims within the generated papers were false, though this was because the claims were very general.

The participants were asked to assess the essays on four bases—argumentation, evidence, organization, and spelling and grammar, on a spectrum of "needs improvement" to "excellent"—as well as to identify whether the essay was written by a student or generated using AI.

The second portion of the study consisted of a semi-structured interview in which we asked the instructors about their knowledge, perceptions, and uses of AI with respect to their writing composition classes. We also asked how they assessed the four essays and determined which essays were written by students and which essays were generated using AI.

Results

Instructor knowledge of AI

Most instructors (90%) indicated that AI writing technologies were new to them and that they had first learned about them in the media or through discussions with their colleagues shortly after ChatGPT was launched in November 2022. Only two of the 20 instructors indicated that AI writing technologies (such as earlier versions of OpenAI's GPT) were already on their radar prior to the public release of ChatGPT. Of those two, one learned about it through a Twitter thread approximately a year prior and forwarded their concerns and suggestions about it to a professional writing association, of which they were a member, but "then nothing really happened [... until] ChatGPT dropped."

In the questionnaire, the participants were asked to rate their knowledge of AI text generator software on a scale of 1 to 10, where 1 indicated no familiarity at all, and 10 indicated specialist familiarity (for example, they could program software like ChatGPT). The average response indicated a non-specialist familiarity, as might be expected from writing instructors with limited previous experience working with AI: the average was 4.85, and the median was 5.0, with a standard deviation of 1.87. When the participants were asked to rate their confidence in their ability to distinguish between the student-written and the AI-generated essays that they had read on a scale of 1 to 5, they rated their confidence at an average of 3.2 and a median of 3, with a standard deviation of 1.01. This is lower than one might expect from highly experienced writing instructors. In sum, the participants indicated a relatively low confidence in their ability to identify AI-generated essays despite perceived reasonable knowledge about AI writing technologies. A linear regression analysis of these results found no relationship (R^2 value of 0.15%) between their self-reported knowledge of AI writing technologies and their confidence in their identification of the essays.

In the interviews, we tested the instructors' actual knowledge of these technologies by asking them to describe how they worked. 40 percent of the instructors either had an erroneous understanding of how LLMs provide answers to prompts or could not provide an explanation. Common errors included the belief that LLMs could search the internet to find factual answers (which was not true of ChatGPT3.5) or that LLMs search for factual data within their training data and re-present it. Our findings are consistent with other studies' findings on the level of instructor understanding of these new technologies; Zhou et al. (2024) found that even instructors in Computer Science have varying levels of understanding of the mechanisms through which LLMs produce their output.

AI in the classroom

When asked whether they had discussed AI with their students, 70% of the instructors reported that they had. While 60% reported that they had not included rules about the use of AI in their syllabus, there was a great deal of variation in how the instructors had addressed the use of AI

text generators on individual assignments, with only 15% of instructors completely prohibiting the use of them. Of the four instructors who responded to the question of whether they had prohibited the use of AI with "other," two were in the process of determining policies.

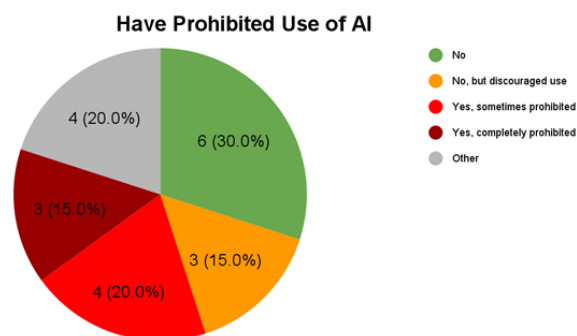


Figure 1. Instructor responses to whether they had prohibited the use of AI in their classes.

While 45% of instructors indicated that they suspected unauthorized use of AI writing technologies in their classes, none had confirmed cases of unauthorized use, and several decided not to confront their students about their suspicions. Participant 10 noted that while in one case they felt "fairly certain" that a student had used AI to write their assignment, "I could not definitively say that this was a generated paper." When asked what made them suspicious, instructors indicated a distinct change in the quality of writing from one assignment to another: "A couple of the students [...] whose English is not good—their English is [that of] second language students—and [...] suddenly their writing is a lot sharper than it was before" (Participant 2). Notably, even when instructors had permitted the transparent use of AI writing technologies, where the use of AI technologies was clearly cited, instructors still had suspicions that students were not declaring their use of them. Despite suspected academic misconduct, many instructors stated that they were not interested in policing the use of these technologies, with one stating, "I don't want to spend more time grading [...] and haranguing this student than they actually spent putting this thing together" (Participant 2).

Instructor and student perceptions of AI

When asked their opinion of students who made unauthorized use of AI technologies, 60% of the instructors were sympathetic to their students, citing a culture of stress and overwork. There was widespread acknowledgement that, unless students come from privileged backgrounds, they are juggling significant demands, including full-time work to afford costly school tuition: "I think of them as panicky, rather than jerks. [...] they're not [...] scheming meanies" (Participant 12). Participant 17 was concerned about "going to straight surveillance, and that really bothers me." At the same time, instructors, many of whom were similarly stressed and overworked, expressed resentment at having to assess AI-written material, with one respondent noting that they "would like to outsource grading if [students are] going to outsource writing" (Participant 1). 40% of instructors had a more negative perception of

students who made unauthorized use of AI technologies, with Participant 4 saying “I would think that they were lazy and not interested in learning what we’re supposed to be teaching them in my class.”

When asked about their students’ perceptions of the technologies, the instructors indicated that some students were excited, some were upset at the idea of other students cheating to get their degrees, some were frustrated by the lack of institutional policies around it, and many just didn’t have it on their radar; these findings are consistent with other studies that indicate heterogeneous student responses to AI-powered writing tools (Burkhard, 2022; Firat, 2023; Bedington et al., 2024). According to the instructors, there is a wide range of motivations behind students’ use of these technologies. For some, it is a means to an end: “It’s a tool, especially for assignments that you don’t have any respect for, right? That’s just something you have to do to get done, but you’re not taking it seriously.” (Participant 17). For others, it was a magic wand or simply another tool (not unlike Grammarly) to improve their writing: “Some students think it is a magic bullet [and others] who are a little more on the ball recognize it can be a powerful ally” (Participant 2). One instructor (Participant 17) mentioned that students perceived professors who didn’t speak to them about AI as not understanding the technologies themselves or not wanting the students to know about it. Participant 1 mentioned that in one of their classes, all of their students were excited about ChatGPT, with the exception of one student, “who feels that this is a form of cheating, and it’s kind of like steroids in sports. You know, if everyone’s doing it, you got to.” Participant 17 was similarly concerned that the long-term impact would be unfairly increasing standards of writing, where “graduate-level expectations are going to start being placed on undergraduate student populations.”

When asked about their own perceptions of the technologies, instructors indicated a mix of excitement and anxiety. Some instructors used ChatGPT as a research tool, saying that “It was easier than looking stuff up on the library website or Google Scholar” and that they “wish[ed] this was around when I was a grad student because it would have streamlined a lot of research” (Participant 2). Some instructors hoped that AI could reduce the “busywork” associated with teaching, such as answering emails.

AI and writing pedagogy

When asked whether they were currently using these technologies in their teaching, only 35% of instructors indicated that they were. One instructor mentioned the benefits specifically for supporting students with different learning needs: “I have a son with disabilities, and it’s really helpful for him to use it for reverse outlining or reading, so I can see amazing benefits” (Participant 17). Some instructors explored the technology with students during class time to critique the software’s output, such as asking the software to write the assigned essay in class. Other instructors had used the software to help them generate writing samples. Another 20% indicated that they intended to use these technologies for teaching in the near future and the remaining 45% indicated that they had no immediate plans.

For one instructor (Participant 12) who focused on rhetoric, ChatGPT was an excellent illustration of the “sociality of writing,” as its statistical learning from a corpus of text could demonstrate “how people typically argue” and then allow students to attempt “writing more singularly by contrast.” Ultimately, many instructors acknowledged that students were going to be required to use these technologies in their careers, and so needed to be trained in how to use them effectively.

When asked about how they were changing their teaching and pedagogy to respond to AI writing technologies, most instructors indicated that they had not yet made changes but intended to do so in the future. However, some of their intended changes were based on misunderstandings of the technology: for instance, one instructor understood that any correct statement from ChatGPT resulted from the software searching the internet or its corpus for the information, while another believed that ChatGPT itself could verify whether a text was AI-written. Participant 12 stated that they’ve “considered ditching articles that I like to teach with because they’re in ChatGPT’s database,” although ChatGPT has not revealed their LLM source material (Gillani et al., 2023), nor does ChatGPT search a database when it responds to a prompt.

When asked about the impact of ChatGPT and other LLMs on writing pedagogy, 100% of instructors indicated that this will dramatically change the field. Participants agreed that the use of technologies will be ubiquitous, whether or not they are authorized. For instance, Participant 2 mentioned that a number of their colleagues were “writing into [their] assignments that students cannot use ChatGPT [but] how can you possibly prevent it? There’s no way to prevent it.” While Participant 15 cautioned that “writing pedagogy is [...] prone to apocalyptic thinking,” because of how frequently embattled the field is within universities, they nonetheless conceded that the widespread availability and use of generative AI was likely to deepen inequalities between the students who needed the technology more, such as second-language speakers, and those who needed it less. One participant expressed a common sentiment that the availability of LLMs would lead to a “major rethinking about whether writing is an important thing to instruct students in and what aspects of writing are important” (Participant 12).

Essay assessment

The instructors were asked to assess the four anonymized essays on four bases—argumentation, evidence, organization, and spelling and grammar—and rate each category as “excellent,” “good,” “fair,” or “needs improvement.” Essays 1 and 3 were written by students at Algoma University before the public release of ChatGPT and Essays 2 and 4 were AI generated using ChatGPT; Essay 2 was generated using the same prompt the students were given, and Essay 4 was generated using the prompt and three additional prompts to refine the initial response. As the assessments varied widely across participants, the results are most comprehensible when comparing the four essays.

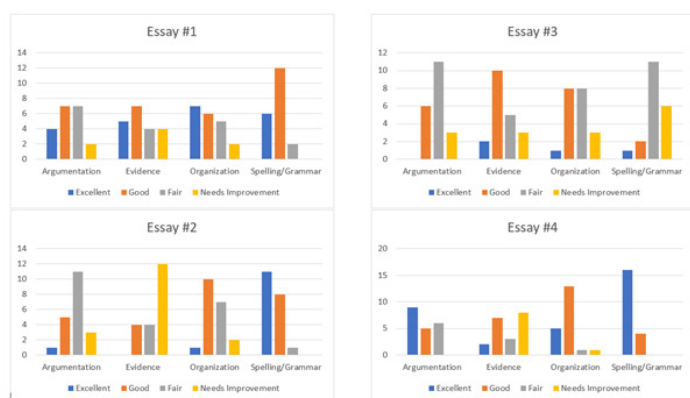


Figure 2. Cumulative essay assessments. Essays 1 and 3 were student-written, and Essays 2 and 4 were AI-generated.

Taken overall, Essays 1 and 4 received the highest overall scores, receiving the most “excellent” and the fewest “needs improvement” scores, at 22 and 8, and 32 and 9, respectively. Essay 1 was assessed as consistently above average across all four categories, whereas Essay 4 was much less consistent, ranked extremely highly on spelling and grammar and relatively low on evidence. Essays 2 and 3 received considerably more “fair” and “needs improvement” scores.

Table 1. Average essay grades.

	Author	Average Grade (%)
Essay 1	Student	73.5 (B)
Essay 2	ChatGPT (unguided)	69.5 (B-)
Essay 3	Student	67.37 (C+)
Essay 4	ChatGPT (guided)	75.5 (B)

Using the rubric and grading system from the original assignment, the average grade earned varied from 67.37% or C+ (Essay 3) to 75.5% or B (Essay 4), which is consistent with average grades in a Canadian context. The essays by ChatGPT were graded slightly higher than those written by students, where the average grade earned by the human students was 70.44% (B-), and the average grade earned by ChatGPT was 72.5% (B).

Taken overall, Essays 2 and 4 stand out as receiving relatively high assessments on spelling and grammar (11 and 16 “excellent” scores, respectively, and no “needs improvement” scores) and relatively low assessments for evidence, with Essay 2 faring the worst (with 12—or 60% of instructors—rating it as “needs improvement”).

Essay identification

Most instructors correctly identified Essays 1 and 3 as being written by students and Essays 2 and 4 as being generated by AI. However, only a minority of instructors (35%) correctly identified all four essays as written by students or by AI. Of the seven participants who accurately distinguished between all four student-written and AI-generated essays, one participant was able to identify that one essay was

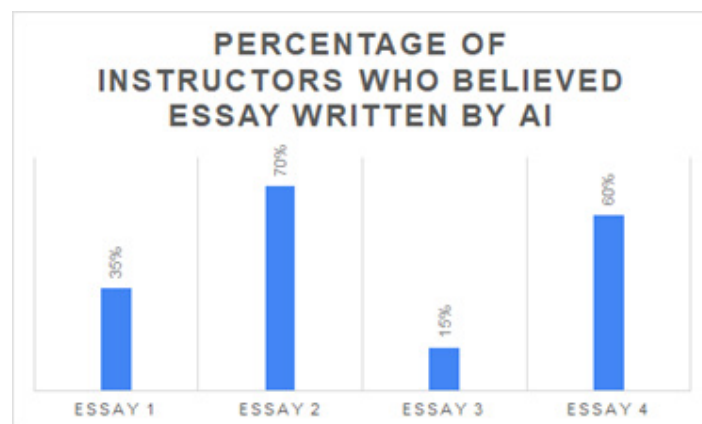


Figure 3. Percentage of instructors who believed essays to be AI-generated.

completely generated by AI and one was generated by AI with further refinement, even though they were not made aware of that distinction by the investigators. Interestingly, this observant participant did not report their confidence in their selections as any higher than any of the participants who made errors in their selections. Furthermore, the participant who reported the highest level of confidence (5 out of 5) correctly identified only one of the essays (Essay 1), incorrectly identifying the three others.

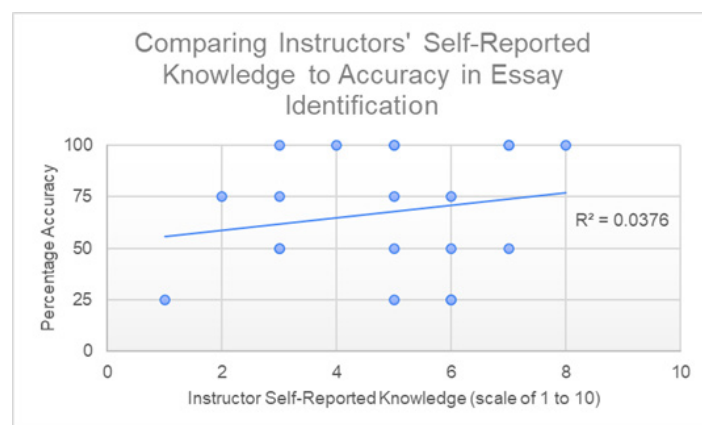


Figure 4. Comparing instructors’ self-reported knowledge of AI writing technologies (expressed on a scale of 1 to 10, where 1 is low and 10 is high) to their percentage accuracy in identifying essays as AI-generated or student-written.

As mentioned above, when we compared their self-reported level of knowledge of AI writing technologies with their self-reported confidence in their identification of the essays using a linear regression model, we found no relationship (R^2 value of 0.15%). When we compared their self-reported confidence in their identification of the essays with their actual accuracy, we also found no relationship (R^2 value of 2.0%). Finally, when we compared their self-reported knowledge with their accuracy in identifying the essays, we found no relationship (R^2 value of 3.8%).

We asked the participants how they identified the essays as student-written or AI-generated. They characterized AI-generated essays as having formulaic writing, few spelling and grammar mistakes, fake references, and surface-level analysis. They noted that these essays lacked evidence in support of their claims and sounded “bland,” “clean,”

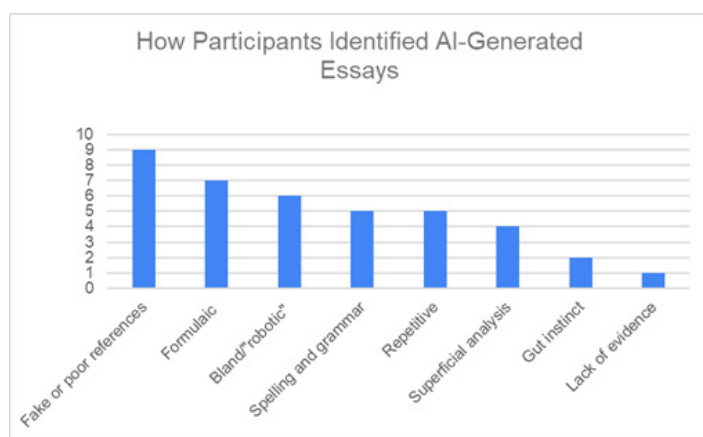


Figure 5. Participant descriptions of AI-generated essays.

"robotic," and "not quite human." The participants also identified a "gut instinct" stemming from years of experience reading student essays. They characterized student-written essays as being "messier," better thought out, and less consistent overall.

Discussion

Instructor knowledge of AI

Given the recent cultural attention to AI writing technologies, it should be no surprise that the survey and interview findings suggest that university and college writing instructors are very much occupied with the impact of these technologies in their writing pedagogy and classes. Although our study only included 20 instructors, if taken to represent writing instructors in Canada and the US as a whole, the average instructor was new to AI-generation technologies until ChatGPT was publicly released in November 2022. Only a small minority of instructors were already considering the impact of these technologies on the field before their release, but even they were not preemptively changing their pedagogy to respond to them.

Although the instructors were new to these technologies, the average instructor felt they had a moderate, non-specialist understanding. Yet, when we tested them on their knowledge, almost half (40%) had an erroneous understanding of how these technologies work. Recent news stories have shown that when faculty have a faulty understanding of these technologies, it can lead to false accusations of academic misconduct (Jimenez, 2023; Klee, 2023; Verma, 2023). Similarly, our findings indicate that a higher degree of knowledge of how these technologies work is the best indicator of accuracy in identifying the presence of generative AI output (albeit a very weak indicator). Taken together, these findings point to the importance of training writing instructors in the design and function of AI writing technologies. Exposure to and understanding of generative AI can help instructors become better able to identify AI output, better equipped to support students in avoiding academic misconduct, and better able to teach students the effective use of these technologies when appropriate.

It is also notable that instructors simultaneously perceived that they had a higher understanding of the AI writing technologies than they actually did and perceived that they had a lower accuracy in the essay identification than they actually did. This might be understood as a manifestation of the Dunning-Kruger effect, a documented cognitive bias: the less one knows about a particular field, the more likely one is to overestimate one's understanding of that field (Kruger & Dunning, 1999). Conversely, experts in a domain, while more self-aware of their limitations, can underestimate their ability (Kruger & Dunning, 1999). The writing instructors in our study, we suggest, are overestimating their abilities outside their knowledge domain (computer science) and underestimating their abilities within their knowledge domain (writing assessment). Training provided to writing instructors might highlight this disconnect between perceived and actual understanding and reassure instructors of their own expertise, specifically in writing composition.

AI in the classroom

Most (60%) of the writing instructors had no formal policy for AI use in the classroom because they were not prepared for ChatGPT at the time they had written and distributed their syllabuses. Yet, by the time of the interviews (nearing the end of the first full semester in which ChatGPT was publicly available), many instructors had already observed what they understood as unauthorized use of AI writing technologies in the classroom with no way to confirm that the students had used them. This highlights the clear need for sound policy developed in collaboration between faculty and administration, as the MLA-CCCC Joint Task Force on Writing and AI has explored in their Working Papers (Adisa et al., 2024). Even in cases where the students were authorized to use AI with citations, some students were not citing its use. Ultimately, although the instructors felt confident that there had been inappropriate use of these technologies, they had limited administrative support, limited time, and little desire to police the use of these technologies. Instructors thus should consider redesigning their syllabuses and assessments to put a greater emphasis on in-person assessment, as well as to integrate the use of these technologies since students are using them with or without permission. As instructors are already feeling stretched thin, administrators should offer instructors training in syllabus and assignment redesign to make this transition easier.

Instructor and student perceptions of AI

According to the instructors we interviewed, both instructors and students are experiencing mixed emotions with the introduction of AI writing technologies and could benefit from guidance on best practices around the use of them. To alleviate the anxieties of both instructors and students, university and college administrations should offer support to help both understand the benefits and limitations of using them.

Instructors are fairly evenly split on their perceptions of students making unauthorized use of AI writing technologies, with 60% sympathetic to students and 40% having a more negative view. Clearer guidelines on appropriate use could help improve student-instructor relationships for the latter group of instructors. This mindset would then be aligned with the approach outlined by Sarah Elaine Eaton of transitioning from a punitive to an educative approach to academic integrity, which is made even more urgent by the widespread availability of generative tools (Eaton, 2021; Kumar et al., 2024).

AI and writing pedagogy

The instructors we interviewed described these technologies as paradigm-shifting and demanding significant changes in writing instruction, including changes in pedagogical approaches—such as focusing more on argumentation than grammar—and changes to policies, possibly even to academic integrity policies, which would need to be considered in tandem with administrative positions on the use of these technologies. Yet, both writing instructors and their students (as reported by the instructors) are feeling let down by their administrations, which have not responded quickly enough nor have considered the specific and unique needs around AI technologies and writing instruction.

One participant characterized the situation as an “administrative hot potato that has been very interesting and really disappointing to witness.” At the time of the interview, only 35% of instructors had experimented with LLMs in their teaching. The interviews revealed that while there were many instructors who were interested in integrating these technologies into their pedagogy, they faced major hurdles to actually including them in their classes. The fast-evolving nature of these technologies, instructors’ lack of expertise, and instructors’ workloads were all named factors in their inability to include them. Despite enthusiasm from researchers and educators in other disciplines (Tobler, 2024), no writing instructors in our study expressed serious interest in using generative AI to mark or grade papers.

All the writing instructors acknowledged that the availability of AI writing technologies will have seismic impacts on writing pedagogy, but instructors disagreed on what that impact would or should be. Some instructors were cautiously optimistic that the availability of generative AI could give students the opportunity to improve their writing by streamlining the non-essential elements of the writing process. Many instructors, however, feared that the availability of the technology would result in learning loss. Participant 16 viewed ChatGPT as “technology that has the potential to make people stupider,” while Participant 6 feared that students would lose “critical thinking research skills, even organizing ideas and brainstorming.” These fears are shared with other instructors around the world, many of whom perceive it as “enabling dishonesty and hindering skill development” (Mohammadkarimi, 2023).

However, participants had strongly differing views about what the most formative aspects of essay composition were for students and whether certain processes becoming

automated would have positive or negative effects. Participant 10 hoped that if instructors were able to spend less time teaching “drafting and outlines, perhaps even how to construct a logical paragraph”, they would “get to spend more time on things that I don’t think students get enough exposure to in the current classroom, like editing, revising, [and] developing a mature and unique argumentative voice.” Participant 3, in contrast, called outlining “one of the most integral parts of essay writing,” and suggested that if LLMs were able to obviate this task for most students, this would hinder students’ development.

This heterogeneity of beliefs among study participants mirrors the heterogeneity of opinions on the steps in the writing process in which the use of generative AI is appropriate among researchers worldwide. While Macdonald et al. (2023) argue that “ChatGPT could become a great help to researchers worldwide in designing their studies, conducting analyses and drafting their research articles,” Leon Furze (2024) compares allowing students to edit an AI-generated first draft to encouraging students to have “a middle-aged white American man they’ve never met, write their entire draft, before they tidy it up and hand it in” (para. 16).

Many instructors suggested that if the field of writing instruction did not respond rapidly, it would face an existential threat. Participant 16 feared that instructors in disciplines outside composition might remove writing assignments in response to the availability of generative AI, which in turn would diminish the importance of training in composition. Others suggested that this may also be a space for opportunity; Participant 1 suggested that writing instructors are

[...] going to have to really ramp up our game in terms of proposing that this is a uniquely human form of expression and that it gives meaning to one’s life and allows one an opportunity to reflect on their place in the universe or their purpose. We’re going to have to go to those sort of higher level needs in terms of Maslow—self-actualizing.

Public access to this technology represents a major shift for many industries, but it has already proven to be a sea change for higher education, and especially writing instruction. While, in the future, tech companies such as OpenAI could restrict or charge exorbitant fees to access their software, there are open-source LLMs, such as LLaMA, that can be run from a personal computer. These open-source alternatives are not as effective as the up-to-date models from the largest tech firms, but some are still able to produce prose roughly equivalent to the output of a first-year composition student. There is no possible future state in which these technologies will not be accessible to many students. Thus, this is a transformative moment that will require writing instructors to re-evaluate what exactly is important about writing and what needs to be prioritized in writing instruction, including considering critical AI literacy and ethics as part of a new writing curriculum (Rudolph et al., 2024).

Essay assessment and identification

Our experimental conditions showed that AI-generated output when given simple guiding prompts, can outperform first-year undergraduate writing. The fears of the student (discussed above) who compared the availability of LLMs to the use of steroids in sports and concluded “if everyone’s doing it, you got to” may thus be understandable. These findings are consistent with Ibrahim et al. (2023), who found that “ChatGPT’s performance is comparable, if not superior, to that of students in a multitude of courses.” Many commentators and researchers have observed that ChatGPT and other LLMs are prone to producing prose that features ‘hallucinations’ or false information, poor evidentiary support and faulty logic, and some have pointed to these faults as evidence of why LLMs are ineffective tools for academic and scientific writing (Alkaissi et al., 2023). Indeed, ChatGPT produces ‘bullshit’ more effectively than it does reliable output (Rudolph et al., 2023; Hicks et al., 2024). However, all of these faults are often true of the work of students beginning the learning process, especially students working in a second language. The highest-graded paper out of the four samples, Essay 4, consisted of 100% AI-generated text and was produced with only a handful of guiding prompts in just under seven minutes. 60% of instructors identified this essay as AI-written, but a savvy student could easily avoid detection by moderately editing the output. Students may find it difficult to resist the temptation to ‘juice’ their essay by asking the program to complete some of the work for them, especially if they see their peers’ use going unpunished and even rewarded with higher grades.

Comparison of essays: AI-generated versus student-written

As Essays 1 (student written) and 4 (AI generated) were consistently ranked more highly and Essays 2 (AI generated) and 3 (student written) were ranked more poorly, there was no clear instructor preference in writing source. However, there did seem to be a bias in perception: that is, when the instructors were asked to explain why they had identified an essay as AI-written or student-written, they consistently described AI using terms like “bland,” “robotic,” and “formulaic,” even when they had misidentified the essays and were actually discussing student-written work. This is aligned with the findings of Alafnan and Mohdzuki (2023), who found that ChatGPT 4.0’s sentence lengths were more consistent than human writers and that the verb tense, mood, and voice varied less than human writers. Similarly, Amirjalili et al. (2024) found that ChatGPT 4.0 hedges less than student writing, potentially revealing less of the uncertainty common to students.

While the essays were sometimes misidentified, there were some patterns that emerged when comparing student-written and AI-generated essays. A common perception is that AI-generated essays have fewer spelling and grammar errors, as articulated in the interviews and confirmed by the blind assessments: overwhelmingly, Essays 2 and 4 were ranked with relatively high scores in that category. Interestingly, while in the interviews the participants characterized student essays as “messier” and less consistent

than AI essays, the student essays were actually ranked more consistently across the four rubric categories. That is, a student who received an “excellent” ranking in one category (such as evidence) was also likely to receive an “excellent” ranking for other categories (such as spelling and grammar), whereas AI was much less even, exceptional in some categories (such as spelling and grammar) and noticeably poor in others (such as evidence). Therefore, unevenness across rubric categories could potentially be used as a flag for instructors looking to distinguish between essays that are student-written and AI-generated.

Essay identification

The results of the experiment confirm the reports of the participants: while, on average, the writing instructors have reasonably high non-specialist familiarity with ChatGPT, they have comparatively low confidence in their ability to distinguish between student-written and AI-generated text. This is lower than other populations studied in other research, such as the pre-service secondary school teachers studied by Fleckenstein et al. (2024), who reported a confidence level of 77.3% in their ability to identify AI-generated texts. This discrepancy may be explained by the skills gap between secondary and post-secondary students, where the latter are more likely to write to the level of ChatGPT and other AI text generators. The instructors had a strong sense of which essays were student-written and which essays were AI-generated because the majority of participants correctly distinguished between the essays most of the time: 70% of instructors correctly identified Essay 2 as AI generated and 60% correctly identified Essay 4 as AI generated. However, there were also many false negatives and false positives: 35% of instructors incorrectly identified Essay 1 as AI generated and 15% incorrectly identified Essay 3 as AI generated. One of the participants who accurately identified all the essays noted that they were uncertain about the more sophisticated of the two AI-generated essays. That essay—Essay 4—was the most prone to identification error; even still, it was accurately identified as AI-generated most of the time.

An important caveat to these findings is that instructors would presumably have greater accuracy in identifying essays written by their own students, with whom they would be familiar, and essays written using prompts they had constructed. In North American universities essays are not graded blind, and composition classes are typically small, allowing for an individual relationship with each student. This is an important difference between the population under consideration and that of Hassoulas et al. (2023), who studied markers at a UK medical school where blind grading is standard. Several of the participants noted that they would feel more confident if they had a sense of a student’s writing voice prior to assessing the essay. Nevertheless, the findings do illustrate the challenge of confidently distinguishing between a student-written and an AI-generated essay, which presents issues in a scenario where the use of AI is strictly prohibited, especially in large classes where one-on-one instructor-student interaction is limited. The results of this study suggest that even experienced instructors are not confident—and cannot be confident—that they are correctly

identifying essays that were generated by AI.

Instructors also noted that, in their personal experience, it was easier to detect the use of AI writing technologies by ELL students whom they personally knew. This raises equity issues, whereby ELL students (who are often marginalized and/or students of color) are more likely to be accused of academic misconduct in situations where AI writing technologies are forbidden. This is one of many reasons that the field of writing instruction might focus more on proactive changes to writing pedagogy that integrate AI writing technologies into their lessons and learning outcomes, rather than focus on policing the use of these technologies. Conversely, as AI writing technologies become more specialized and potentially more costly (as ChatGPT moves out of its demo stage), colleges and universities may want to consider 'leveling the playing field' by paying for institutional subscriptions to an AI-writing technology that makes these technologies equally accessible by all students. Without institutional support, what writing instruction might see is something akin to 'contract writing on steroids,' whereby the students who already had easier access to less-detectable academic misconduct methods will disproportionately benefit from these technologies.

It is important to reiterate that this study did not include student participants and so the assessments of student use of AI writing technologies were strictly perceptions articulated by the instructor participants. Nevertheless, one of the recurring concerns about student use of AI writing technologies was that, even when the use of these technologies was expressly permitted (with citation), instructors suspected students were misusing them by failing to declare their use and/or denying their use of them when their use was suspected. A future study should examine student perceptions and use of AI writing technologies and specifically consider why students do not declare their use of them even when they are expressly permitted with citation.

Conclusion

The recent and quick ubiquity of AI text generators is presenting a significant challenge for instructors of the humanities today. This challenge is all the more heightened for writing instructors, who are tasked with teaching a historically critical skill now in question. As a small study of only 20 writing instructors, the results of this study are preliminary rather than conclusive. Nevertheless, they clearly point to a need for more consideration of the pedagogical implications of AI on writing composition.

Overall, our results indicate that ample experience as a writing instructor is not in and of itself an indication of readiness to grapple with the implications of AI technologies on writing pedagogy. For the average writing instructor, these technologies are new, and they have a limited and/or faulty understanding of them. While writing instructors acknowledge the paradigm-shifting nature of the technologies, they are not yet prepared with new policies or pedagogies to respond to them and are relying on administrative guidance, which has been insufficient to date. As such, we make the following suggestions as strategies to

consider when designing and supporting writing instruction:

1. Instructors should encourage students to work independently while learning to use AI as a writing support.

All evidence thus far suggests that the use of AI writing technologies is widespread among college and university students and that there is no foolproof method of identifying the use of these technologies. As a result, writing instructors must acknowledge that the only student writing that can be said to be produced independently is writing that is produced during a monitored in-class situation. Assessment through in-class writing alone, however, severely limits the metacognitive skills being tested within a writing course, which would typically also focus on planning, outlining, researching, drafting, revising, editing, and proofreading. In order to test this broader range of skills, writing instructors should encourage student independence in their work. Encouraging student independence includes fostering critical thinking, creativity, and originality in student writing, while also teaching them how to ethically and effectively incorporate AI tools. Writing courses often make use of scaffolded assignments in which multiple stages of a larger assignment are assessed by the instructor; we suggest that by using a mix of in-class exercises and take-home work, instructors can guide independent student thought as part of the writing process. It is crucial to guide students in understanding the limitations and biases inherent in AI technologies and to emphasize the importance of their own unique voices and perspectives in their writing. As students work more with generative AI, the probabilistic nature of the software may encourage students to shift their writing style away from human 'burstiness.' Without active intervention from writing instructors, student voices risk being homogenized by the influence of the predictive and predictable style of large language models.

At the same time, instructors should consider integrating interactions with AI tools into their curriculum, offering a structured environment for students to explore these technologies under guidance. This approach not only prepares students for a future where AI is a common tool in many professions but also helps them develop a critical lens towards the content generated by AI.

2. University and college administrators should develop clear yet adaptable policies and invest in instructor and student education rather than funding subscriptions to AI-detection software.

There is currently no foolproof AI-detection software (Sadasivan et al., 2023; Rashidi et al., 2023; Dalalah & Dalalah, 2023), and as AI-generation, AI-detection, and AI-detection evasion software are all developing in response to one another, a foolproof AI-detection software is unlikely to materialize. Thus, funding subscriptions to AI-detection software and using them to police unauthorized use of AI technologies is not likely to solve the many pedagogical issues with the use of AI-generation technologies. Rather, funding access to AI writing technologies may help 'level the playing field' and normalize the use of these technologies.

Our study indicates that writing instructors want to support students in acquiring critical thinking skills and acknowledge that students will need to use AI effectively and responsibly in their future careers. Instructors also expressed that they believe their average student is interested in developing critical skills and preparing for work that will include the use of AI. Furthermore, instructors do not have the capacity, nor are they interested in policing the use of AI in their classrooms. Thus, both students and instructors desire the appropriate and responsible use of these technologies, although they are unclear about what that is in practice.

Instructors also demonstrate overconfidence in their knowledge of these technologies and so are, on average, not currently equipped to deal with them, even when they are highly experienced and competent in traditional writing instruction. Thus, writing instructors need support in acquiring knowledge to develop new writing pedagogy that responds to AI writing technologies. In sum, university and college administrators must invest in instructor and student education specifically around the use of AI writing technologies.

Finally, a large minority of instructors expressed feelings of cynicism around marking AI-generated assignments and negative opinions about students who make unauthorized use of AI writing technologies. In order to preempt poor student-instructor relationships and instructor burnout, we follow the call of other researchers (e.g., Kohnke et al., 2023) for the development of clear, yet adaptable guidelines and policies around the ethical use of AI that permit regular revision in this quickly evolving area. Administrators should outline these policies in addition to offering training for both students and teachers around the responsible use of AI technologies.

Limitations of the present study and future work

As this study was designed and funded prior to the release of ChatGPT-3, when AI text generators were not as well-known and were less accessible, it focused on instructor knowledge, perceptions, and use. As instructors become more familiar with these technologies and with ChatGPT's output, we will continue this study as a longitudinal study to see how instructor perceptions and ability to distinguish AI writing change over time. In addition, now that these technologies are ubiquitous and widely employed by students, our research will inquire into how and why students are using these softwares. While the questions of how students can use them and how they are *actually* using them are topical across fields of study, these questions remain particularly important for the humanities—and especially writing composition—in which essay writing is still the predominant mode of assessment.

Acknowledgements

This work was supported by Algoma University and the Social Sciences and Humanities Research Council (SSHRC) Institutional Grant. We would also like to acknowledge the students and instructors who participated in this study as

well as our research assistant Mary Germaine. Finally, we would like to thank our colleagues, Melissa Auclair and Timothy PJ Perry, as well as the anonymous reviewers, for their comments on earlier drafts. Any remaining errors are our own.

References

- Adisa, T., Ogbonnaya, C., & Adekoya, O. (2021). Remote working and employee engagement: A qualitative study of British workers during the pandemic. *Information Technology & People*. <http://dx.doi.org/10.1108/ITP-12-2020-0850>
- Alafnan, M., & Mohdzuki, S. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial Intelligence and Technology*, 3(3). <http://dx.doi.org/10.37965/jait.2023.0267>
- Alkaiissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, 15(2), e35179. <https://doi.org/10.7759/cureus.35179>
- Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1347421>
- Bedington, A., Halcomb, E. F., McKee, H. A., Sargent, T., & Smith, A. (2024). Writing with generative AI and human-machine teaming: Insights and recommendations from faculty and students. *Computers and Composition*, 71. <https://doi.org/10.1016/j.compcom.2024.102833>
- Burkhard, M. (2022). Student perceptions of AI-powered writing tools: Towards individualized teaching strategies. *19th International conference on cognition and exploratory learning in digital age*. <https://eric.ed.gov/?id=ED626893>
- 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*. <https://doi.org/10.1016/j.rmal.2023.100068>
- Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning & Teaching*, 6(2), 94-104. <https://doi.org/10.37074/jalt.2023.6.2.12>
- Dalalah, D., & Dalalah, O. M. A. (2023). The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. *The International Journal of Management Education*, 21(2). <https://doi.org/10.1016/j.ijme.2023.100822>
- DK-Sonic. (2023, January 11). *I know a few guys fr...* [Reddit Comment]. R/GPT3. www.reddit.com/r/GPT3/comments/108ti98/forget_the_crazy_prompts_forget_the_business/j3wc29g/
- Dugan, L., Ippolito, D., Kirubarajan, A., Shi, S., & Callison-

- Burch, C. (2023). Real or fake text?: Investigating human ability to detect boundaries between human-written and machine-generated text. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(11), 12763–12771. <https://doi.org/10.1609/aaai.v37i11.26501>
- Firat, M. (2023) What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching*, 6(1), 57–63. <https://doi.org/10.37074/jalt.2023.6.1.22>
- Fleckenstein, J., Meyer, J., Jansen, T., Keller, S. D., Köller, O., & Möller, J. (2024). Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays. *Computers and Education: Artificial Intelligence*, 6. <https://doi.org/10.1016/j.caeai.2024.100209>
- Furze, L. (2024, February 21). *The myth of the AI first draft*. Leonfurze.com. <https://leonfurze.com/2024/02/21/the-myth-of-the-ai-first-draft/>
- Gillani, N., Eynon, R., Chiabaut, C., & Finkel, K. (2023). Unpacking the “Black Box” of AI in education. *Educational Technology & Society*, 26(1), 99–111. [https://doi.org/10.30191/ETS.202301_26\(1\).0008](https://doi.org/10.30191/ETS.202301_26(1).0008)
- Hasanein, A. M., Sobaih, A. E. E. (2023). Drivers and consequences of ChatGPT use in higher education: Key stakeholder perspectives. *European Journal of Investigation in Health, Psychology and Education*, 13(11), 2599–2614. <https://doi.org/10.3390/ejihpe13110181>
- Hassoulas, A., Powell, N., Roberts, L., Umla-Runge, K., Gray, L., & Coffey, M. J. (2023). Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT. *Journal of Applied Learning and Teaching*, 6(2), 71–77. <https://doi.org/10.37074/jalt.2023.6.2.13>
- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2), 38. <https://doi.org/10.1007/s10676-024-09775-5>
- Ibrahim, H., Liu, F., Asim, R., Battu, B., Benabderrahmane, S., Alhafi, B., Adnan, W., Alhanai, T., AlShebli, B., Baghdadi, R., Bélanger, J. J., Beretta, E., Celik, K., Chaqfeh, M., Daqaq, M. F., El Bernoussi, Z., Fougny, D., de Soto, B. G., Gandolfi, A., ... & Zaki, Y. (2023). Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Scientific Reports*, 13. <https://doi.org/10.1038/s41598-023-38964-3>
- Impact Research. (2024, May). *AI chatbots in schools findings from a poll of K-12 teachers, students, parents, and college undergraduates*. Walton Family Foundation. <https://8ce82b94a8c4fdc3ea6d-b1d233e3bc3cb10858bea65ff05e18f2.ssl.cf2.rackcdn.com/bf/24/cd3646584af89e7c668c7705a006/deck-impact-analysis-national-schools-tech-tracker-may-2024-1.pdf>
- Jimenez, K. (2023, Apr 13). Professors are using ChatGPT detector tools to accuse students of cheating. But what if the software is wrong? *USA Today*. <https://www.usatoday.com/story/news/education/2023/04/12/how-ai-detection-tool-spawned-false-cheating-case-uc-davis/11600777002/>
- Klee, M. (2023, Jun 6). *She was falsely accused of cheating with AI — and she won't be the last*. Rolling Stone. <https://www.rollingstone.com/culture/culture-features/student-accused-ai-cheating-turnitin-1234747351/>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence*, 5. <https://doi.org/10.1016/j.caeai.2023.100156>
- Kruger, J., & Dunning, D. (1999.) Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <http://dx.doi.org/10.1037//0022-3514.77.6.1121>
- Kumar, R. (2023, Oct 12). *The university essay will die out*. Macleans. <https://macleans.ca/society/technology/ai-university-essay/>
- Kumar, R., Eaton, S. E., Mindzak, M., & Morrison, R. (2024). Academic integrity and artificial intelligence: An overview. In S. E. Eaton (Ed.), *Second handbook of academic integrity* (pp.1583-1596). <https://doi.org/10.1007/978-3-031-54144-5>
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7). <https://doi.org/10.1016/j.patter.2023.100779>
- Macdonald, C., Adeloye, D., Sheikh, A., & Rudan, I. (2023). Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. *Journal of Global Health*, 13. <https://doi.org/10.7189/jogh.13.01003>
- Marche, S. (2022, Dec 6). The college essay is dead. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/>
- MLA-CCCC Joint Task Force on Writing and AI. (2023). *MLA-CCCC joint task force on writing and AI working paper: Overview of the issues, statement of principles, and recommendations*. Humanities Commons. <https://hcommons.org/app/uploads/sites/1003160/2023/07/MLA-CCCC-Joint-Task-Force-on-Writing-and-AI-Working-Paper-1.pdf>
- MLA-CCCC Joint Task Force on Writing and AI. (2024). *Generative AI and policy development: Guidance from the MLA-CCCC task force*. Humanities Commons. <https://hcommons.org/app/uploads/sites/1003160/2024/04/Generative-AI-and-Policy-Development-Guidance-from-the-MLA-CCCC-Task-Force.pdf>
- Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning and Teaching* 6(2), 105–113. <https://doi.org/10.37074/jalt.2023.6.2.10>

Perkins, M., Roe, J., Postma, D., McGaughan, J., & Hickerson, D. (2024). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *Journal of Academic Ethics*, 22(1), 89–113. <https://doi.org/10.1007/s10805-023-09492-6>

Rashidi, H. H., Fennell, B. D., Albahra, S., Hu, B., & Gorbett, T. (2023). The ChatGPT conundrum: Human-generated scientific manuscripts misidentified as AI creations by AI text detection tool. *Journal of Pathology Informatics*, 14. <https://doi.org/10.1016/j.jpi.2023.100342>

Rospigliosi, P. (2023). Artificial intelligence in teaching and learning: What questions should we ask of ChatGPT? *Interactive Learning Environments*, 31(1), 1–3. <https://doi.org/10.1080/10494820.2023.2180191>

Rudolph, J., Ismail, F., & Popenici, S. (2024.) Higher education's generative artificial intelligence paradox: The meaning of chatbot mania. *Journal of University Teaching & Learning Practice*, 21(6). <http://dx.doi.org/10.53761/54fs5e77>

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>

Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). *Can AI-generated text be reliably detected?* arXiv. <https://doi.org/10.48550/arXiv.2303.11156>

Singh, H., Tayarani-Najaran, M.-H., & Yaqoob, M. (2023). Exploring computer science students' perception of ChatGPT in higher education: A descriptive and correlation Study. *Education Sciences*, 13(9). <https://doi.org/10.3390/educsci13090924>

Steele, J. L. (2023). To GPT or not GPT? Empowering our students to learn with AI. *Computers and Education: Artificial Intelligence*, 5. <https://doi.org/10.1016/j.caeai.2023.100160>

Thi Thuy, A. N. The perception by university students of the use of ChatGPT in education. *International Journal of Emerging Technologies in Learning (Online)*, 18(17), 4–19. <https://doi.org/10.3991/ijet.v18i17.39019>

Tobler, S. (2024). Smart grading: A generative AI-based tool for knowledge-grounded answer evaluation in educational assessments. *MethodsX*, 12, 102531. <https://doi.org/10.1016/j.mex.2023.102531>

Verma, P. (2023, May 19). A professor accused his class of using ChatGPT, putting diplomas in jeopardy. *Washington Post*. <https://www.washingtonpost.com/technology/2023/05/18/texas-professor-threatened-fail-class-chatgpt-cheating/>

Vynck, G. D. (2023, July 7). ChatGPT loses users for first time, shaking faith in AI revolution. *Washington Post*. <https://www.washingtonpost.com/technology/2023/07/07/chatgpt-users-decline-future-ai-openai/>

Zhou, K. Z., Kilhoffer, Z., Sanfilippo, M. R., Underwood, T., Gumusel, E., Wei, M., Choudhry, A., & Xiong, J. (2024). *"The teachers are confused as well": A multiple-stakeholder ethics discussion on large language models in computing education*. arXiv preprint. <https://doi.org/10.48550/arXiv.2401.12453>

Zou, M., & Liang, H. (2023). The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-12397-x>

Appendix

Essay Prompt

Write a 1000-word essay on the following topic:

Consider a problem that a particular industry is facing as a result of the COVID-19 pandemic, and make a recommendation for how to solve it. What are the immediate challenges with addressing this problem, and what are some possible solutions you could recommend?

This is an academic research paper, which means that it must cite a minimum of FOUR academic or high-quality journalistic sources within the text of the paper and must include a References page at the end of the paper.

The essay must contain:

1. A strong, clearly articulated and well-reasoned main point that is SPECIFIC, ORIGINAL and CONVINCING.
2. A strong introduction that in some way lays out the stakes of the argument
3. Well-chosen evidence from at least four secondary sources, with APA citations in the text.
5. Logical paragraph and essay structure
6. Clear, grammatical writing

Essay Rubric

	EXCELLENT	GOOD	BELOW AVERAGE	INEFFECTIVE
Argumentation	The essay begins with a clear and specific introduction that states a compelling thesis statement. The essay provides a thoughtful and engaging analysis of the assigned topic.	The essay begins with an introduction that states a thesis statement. The essay analyzes the assigned topic.	The introduction may not state a thesis statement clearly or the main idea may not appear at the beginning of the summary. Demonstrates basic understanding of the assigned topic.	The summary does not state a thesis statement. Demonstrates little to no understanding of the assigned topic.
Evidence	The essay provides more than the minimum amount of evidence and analyzes the evidence thoughtfully. All important details are included.	The essay provides some analysis of evidence. At least the minimum amount of evidence is provided. Most important details are included.	The essay provides less than the minimum amount of evidence. Some critical information is missing. Some important details are left out.	Insufficient evidence. Contains only some details. Substantial copying of key phrases and minimal paraphrasing.
Organization	The sentences follow one another in rigorous logical fashion; connections between ideas are very clear, and transitions between paragraphs are clearly signposted.	The sentences follow one another in logical fashion; connections between ideas are clear, and transitions between paragraphs make sense.	Sentences are disjointed. Ideas are not in a logical order or mostly mimic the original text.	Ideas are in a random order; sentences do not follow each other, or the essay closely mimics the text. The text may be too short or too long.
Spelling and Grammar	No errors exist in spelling, grammar, punctuation, capitalization, and sentence structure.	Minimal errors exist in spelling, grammar, punctuation, capitalization, and sentence structure.	Some errors exist in spelling, grammar, punctuation, capitalization, and sentence structure.	Numerous errors exist in spelling, grammar, punctuation, capitalization, and sentence structure.

Copyright: © 2024. Nathan Murray and Elisa Tersigni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.