The Role of Faculty Expertise and Intuition in Distinguishing between AI-Generated Text and Student Writing

Daniel R. Fredrick
American University of Sharjah, UAE
dfredrick@aus.edu

Laurence Craven
American University of Sharjah, UAE
lcraven@aus.edu

Tor Brodtkorb
American University of Sharjah
tbrodtkorb@aus.edu

Maria Eleftheriou
American University of Sharjah
meleftheriou@aus.edu

Abstract

Detecting plagiarism, which is crucial for upholding institutional credibility, is a persistent and growing challenge for university faculty. The recent emergence of AI language models like ChatGPT introduces a new complication for faculty: discerning whether students have used AI in the production of their course work. Until AI detection software becomes effective at revealing the presence and the extent of the contributions made by AI to student submissions, the ability of faculty to detect, by either expertise or intuition, whether text is generated primarily by AI or student effort, is an important topic for exploration. We investigated faculty members' ability to differentiate paragraphs written by students from those generated by ChatGPT. One group of students used ChatGPT, while another wrote essays without AI assistance. We collected student essays and distributed random samples to 27 faculty members of an English department, 19 of whom responded to a questionnaire in which they reported their beliefs about which texts were written by students and which texts written by AI, and indicated the reasons underlying those beliefs. Faculty were generally successful in identifying ChatGPT-generated content using their expertise and intuition. Linguistic cues, such as syntactic structures, language idiosyncrasies, and contextual coherence gaps were commonly cited as explanations for their decisions.

Keywords

ChatGPT detection, AI models, incorporating AI in education, Academic Integrity

Introduction

ChatGPT and other AI language models have become a major topic in public discussion since the introduction of ChatGPT in 2022, and the implications for academic integrity have been widely discussed among university faculty. AI language models are proficient in grammar and structure, are capable of generating coherent and relevant content, and so naturally, concerns have emerged regarding the potential for AI generated content to be misrepresented as competent or even high-quality student writing. Critical thinking and writing skills are learned through practice, and learning outcomes are assessed through submitted work. If the submitted work has been partly or entirely produced through an AI language model, then the student is missing out on the practice that would lead to proficiency, while illegitimately excelling at the assessment meant to test that proficiency. Additionally, since ChatGPT draws on web-based information, ethical issues arise due to its current inability to provide proper citations, potentially leading to plagiarism or copyright violations in AI produced submissions. Institutional prohibitions on the use of AI are unlikely to be effective, as AI language models are so widely available on a variety of platforms. Therefore, it may be preferable to integrate the use of AI language models as part of the pedagogy, allowing or encouraging students to use them as a way to engage students in critical thinking, as a study buddy, or as an invention tool or editor. Even under this strategy, however, educators must emphasize responsible use of AI language models, especially in assignments completed outside the classroom. In order to teach responsible use, reward responsible use with appropriate grading, and maintain standards of academic integrity, faculty members will need to detect improper use and over-reliance on AI language models that undermine student development. As the technology evolves, faculty members and assessment technologies will need to develop and adopt new techniques to detect or deter plagiarism, which will require a better understanding of the distinctions between human writing and the writing of AI language models. In view of this, the present study becomes significant.

Significance of the Study

The present study holds significance for various reasons. It addresses the issue of academic integrity in the AI era by ascertaining whether faculty members are able to identify AI-generated text, allowing them to properly assess student work. Moreover, the findings may influence the structuring and evaluation of writing assignments, as well as inform the integration of AI tools into writing instruction, because knowing the limitations and strengths of faculty in detecting AI-generated text can guide the development of guidelines and training programs that assist in the effective use of AI for teaching writing. Additionally, these insights could contribute to the creation of tailored pedagogical strategies that blend traditional writing instruction with AI-assisted techniques, ensuring that the educational objectives are met without compromising the authenticity of student work. Such strategies could prove useful in shaping a curriculum that is responsive to the evolving role of AI in academic settings, particularly in the realm of writing and

composition. Additionally, research outcomes may guide universities in providing faculty members with the necessary training and resources to proficiently evaluate student work within the context of AI assistance. Therefore, this study has the potential to contribute to the development of best practices for the incorporation of AI technologies into education, while allowing institutions to maintain the integrity of academic assessment.

Literature Review

In his article provocatively titled "The College Essay Is Dead," Marche argued that the entire "humanist pedagogy" of writing, thinking, and researching, "is about to be disrupted from the ground up" (2022, para. 5) due to the development of effective AI language models like ChatGPT. Marche, as is made clear by his argument, is no fan of the humanities the way they are presently run, describing them as "a journey to an irrelevant, self-consuming future" (2022, para. 9). We live in an increasingly STEM-centered educational world, and the humanities may suffer a fatal, and, if Marche is correct, well-deserved blow if students' performance continues to be judged by the essays they submit, since the essay can now be "significantly automated" (2022, para. 12). In early December 2022, philosophy professor Darren Hick, warned that universities should "expect a flood" of cheating for the obvious reason that students can prompt ChatGPT to write their essays in a matter of seconds (Allen, 2023, para. 15; Chia, 2023).

ChatGPT is excellent at grammar and form, lower order skills, but it also creates intelligible prose and can churn out organized, factual information on nearly any subject (Fitria, 2023). Although ChatGPT seems to be producing this information out of thin air, it is actually retrieving and synthesizing information from the web—all of which was written by humans. ChatGPT is thus "indebted" to humans for its skill, so all of the awe we give it should properly be bestowed on human writers, that is, until AI starts producing its own ideas (Anderson, 2023, para. 25).

The literature and research on ChatGPT are rife with warnings regarding its use. One of the most damning criticisms against it is that it is prone to be inaccurate and is itself the greatest plagiarist, unable to cite its sources and thus unable to be ethically accountable as a writer (Deng, 2022). As an amoral entity, ChatGPT cannot, therefore, qualify to be a "co-author" on published materials, a legality we learn from Siobhan O' Connor, who was not allowed to list ChatGPT as the second author in a peer-reviewed article for *Nurse Education and Practice* (Dunne, 2023). Beyond the obvious concern that students will use ChatGPT to write their assessments for them entirely, even the use of ChatGPT for assistance raises concerns that students might become too dependent on it to clean up their messy prose, relying on it for every writing task. For non-native speakers, some scholars argue that this will have negative effects on language overall, where an English-language focus would reduce the linguistic and stylistic diversity of other languages (Sharma, 2023). Because of these numerous problems—plagiarism, inaccurate information, overreliance, and according to ChatGPT's CEO, its "misleading impression of greatness," New York City schools

formally banned its use in 2023 (though they later removed the ban in May 2023) (Yang, 2023, para. 10).

Despite critics who want to ban the use of AI language models, or downgrade their abilities, AI language models have become a tempting, powerful and irrepressible tool for anyone who composes text. The majority of academia (Heaven, 2023; Tlili et al., 2023; Villasenor, 2023) has decided to try to incorporate AI language models as tools and control their use, rather than attempt to suppress them or ban them altogether. Of all the departments in the university, it is not surprising that composition teachers are paving the way to develop best practices for AI language models. In many English departments, teachers have already been incorporating ChatGPT in full force, not as a replacement for human writing, but as a sort of automated tutor, a disembodied study buddy, a proficient and talented peer. A framework influencing this integration is Lev Vygotsky's Zone of Proximal Development (ZPD), a concept in educational psychology that refers to a range of tasks a learner can accomplish with the assistance of a more knowledgeable individual, be it a teacher or peer, beyond what they can achieve independently (1978). Central to the ZPD is the concept of scaffolding, where support provided by a knowledgeable person gradually decreases as the learner becomes more competent. AI language models may serve as a form of scaffolding, providing incremental assistance to students across various stages of the writing process, encompassing tasks like brainstorming, outlining, and refining their work. This ongoing support from AI language models could empower students to progressively generate higher-quality writing independently. Vygotsky's emphasis on social interaction in the learning process could also be manifested in student use of AI language models. Despite being non-human entities, AI language models could facilitate interactive learning by allowing students to engage actively in a dynamic exchange. This interaction would allow students to receive valuable feedback, learn from the model's responses, and actively contribute to a collaborative and enriching learning experience. Educators agree that when students use AI language models as a personal tutor for peer-review (Javaid et al., 2023; Ozar et al., 2023), it can contribute to enhancing learning. One constant danger is, of course, plagiarism. It is unusual, but still possible to find educators flippant about the ease at which students can plagiarize, placing their trust in current detection systems such as that provided by Turnitin (Bruder, 2023). This relaxed attitude toward AI language models' potential for encouraging plagiarism is, however, not the norm. Knowing fully well that plagiarism is a constant temptation for students, educators are pointing the way to a proper pedagogical use of AI language models.

As most writing courses emphasize process, AI language models, it is argued, are a good tool for researching or analyzing information, but of course results must be checked for accuracy, because oftentimes they produce incorrect or even invented information. The fact is that, more often than not, AI language models provide correct and relevant results (Firaina & Sulisworo, 2023; Oguz et al., 2023). Due to the increasing acceptance of at least some use of AI language models in writing courses, against the steady background of a prohibition on submitting work that is substantially or

entirely composed by AI, many argue that policies on whether and how students may use AI language models need to be formalized and explained to both students and faculty (Halaweh, 2023).

AI language models may actually have the potential to transform education, by reemphasizing the importance of academic integrity. As a starting point, teachers and university policy-makers will need to develop a general overview of what AI language models are, how they function, what their limitations are, how to write effective prompts for them, how to corroborate the information they provide, and then finally how to incorporate their use of it into pedagogical practices (Tate et al., 2023; Tlili et al., 2023). This process will reform the way plagiarism is detected and addressed (Neumann, 2023). No doubt plagiarism detectors will need to be a crucial resource of academic departments; however, research has shown that presently, it is "impossible to reliably detect content generated with the assistance of AI" (Farrelly & Baker, 2023, para. 8). A study by Elkhatat et al. (2023) revealed the challenges to academic integrity posed by AI-generated content. By focusing on paragraphs from ChatGPT on the topic of cooling towers, and human-written paragraphs on the same topic, the research evaluated various AI content detection tools. They concluded that these tools demonstrated higher accuracy in identifying GPT 3.5 generated content compared to GPT 4. However, the tools also flagged some human writing as AI-produced, raising concerns about false positives. Elkhatat's study indicates the current inadequacy of AI detection tools, and the poorer performance of the tools on GPT 4 generated content raises the possibility that AI will evolve faster than the tools designed to detect it.

Some research suggests that AI detection for plagiarism is going to be problematic, offering no real certainty of detecting AI generated text regardless of the sophistication of the detection software, and thus educators will need to reform assessment strategies (Alexander et al., 2023). Still others believe that teachers need to be aware of the key differences between ChatGPT writing (as it composes currently) to human writing, with one of the key clues being that ChatGPT offers "information" rather than "ideas" (Bishop, 2023). It is on this point that we anchor our research. AI detection software is lagging behind AI development, yet teachers are experts on student writing, so it should follow that their expertise will allow them to detect whether a piece of writing is written by a student, just as their expertise allows them to detect whether a piece of writing is written by a professional.

Research Questions

This study explores the ability of English faculty members at a university in the UAE to accurately differentiate paragraphs generated by ChatGPT from those created solely by students without AI assistance. Students were monitored writing a paragraph on a set topic, which was then checked by Grammarly, for obvious spelling or syntactic errors, and then, separately, asked to write a prompt requesting ChatGPT to write a paragraph on the same topic. The faculty in the department

of English in a UAE university were given a survey asking them to determine which paragraphs were written by students and which by ChatGPT, and to offer reasons for their selections. By analyzing the results, we attempted to answer the following research questions:

R1: Are English department faculty at an institution in the UAE able to detect ChatGPT vs. student generated texts?

R2: What textual clues do faculty rely on in distinguishing between ChatGPT vs. student generated texts?

R3: To what extent does experience and intuition play a role in successful detection?

Methodology

The study involved a selected intact class of 25 junior and sophomore university students enrolled in a highly ranked reputable institution in the United Arab Emirates. We chose this demographic for its diverse linguistic background, including individuals with Arabic and Urdu as their first languages (L1) and English as their second language (L2) (See Table 1). The 19 faculty members who responded to the survey (70% response rate) possessed a wealth of experience and diverse expertise (See Table 2).

Table 1
Demographics of Student Participants

Variable	Percentage
Standing	
--Junior 1	20%
--Junior 2	60%
--Sophomore 1	20%
First Language	
--Arabic	72%
--Urdu	16%
--Other	12%

Table 2
Demographics of English Faculty Participants

Variable	Percentage
Years of University Teaching Experience	

--More than 10 years	89.5%
--5-10 years	10.53%
Primary Field of Expertise	
--Rhetoric	26.32%
--EFL/ESL	21.05%
--Literature	21.05%
--Education	15.79%
--Linguistics	5.26%
--Other	10.53%

This mix of educators from various fields provides a foundation for exploring their ability to discern between text generated with the assistance of ChatGPT and authentic student writing. The study benefits from the depth of experience and varied perspectives these faculty members bring to the evaluation process, potentially offering insights into the detection capabilities within different academic domains.

We obtained ethical approval from the institutional review board to ensure compliance with ethical standards and distributed informed consent forms to all participants, providing a clear and comprehensive overview of the study's purpose, tasks involved, and a guarantee of confidentiality and anonymity. We also informed participants of their right to withdraw from the study at any point without incurring any negative consequences.

We gave students a well-defined writing task, which involved composing a 300-word opinion paragraph on the topic of whether e-scooters should or should not be permitted on university campuses. To maintain the integrity of the study, we not only explicitly instructed participants not to use ChatGPT for this task, but they were required to use *LockDown Browser,* software that locks down the testing environment within a learning management system*,* and we monitored the complete process of the writing task so that students would be forced to write completely on their own. After collecting the student writings, we used *Grammarly* to fix basic grammatical errors, as minor errors would otherwise easily reveal human writing.

Following the initial writing task, we assigned students a second task. They were instructed to use ChatGPT to generate an additional 300-word opinion paragraph on the same topic. They were permitted to write prompts for ChatGPT in any way, as this freedom replicates how students would use ChatGPT in reality. From the pool of collected writing samples, we selected a subset of eight samples through a randomized process for further scrutiny. We then distributed the chosen samples to 27 English department faculty members at the university, who were challenged with discerning between paragraphs generated by students and those generated by ChatGPT. In conjunction with the distributed writing samples, we provided faculty members with a questionnaire (Appendix A). This questionnaire asked for years of experience and primary field of expertise and was crafted to

evaluate their reasons in distinguishing the source of each paragraph. We asked about their detection process, including their confidence levels in their determinations and any specific linguistic cues they relied upon for their assessments. Upon receipt of the completed questionnaires from the faculty members, we compiled the responses for analysis. The data gleaned from this process enabled us to assess the faculty members' proficiency in accurately identifying the origin of the writing samples.

Results

The survey included questions gauging participants' ability to identify ChatGPT-assisted writing, with each question. In Question 3, 94% of teachers accurately identified the source of the writing samples. For Questions 4 and 5, 74% of teachers correctly identified the source of the writing samples in each case. In Question 6, the accuracy rate rose again, with 94% of teachers correctly identifying the source of the writing samples (See Figure 1). The average success rate across all four paragraphs was 84%.

Figure 1
*Accuracy Rate for Detecting ChatGPT vs. Student Writing*
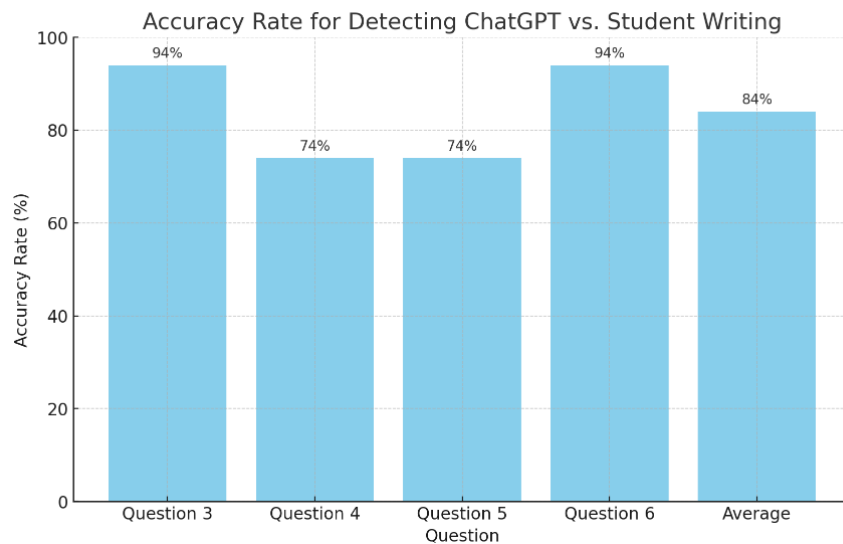
The results demonstrate that faculty members can be highly proficient in identifying text generated by ChatGPT. Notably, a consistently high percentage of 94% accuracy was observed in Question 3 and Question 6. The faculty's ability to identify ChatGPT generated text suggests an awareness and familiarity with the subtleties of student writing, demonstrating the extent of their expertise in the field. This level of accuracy highlights the potential for educators to effectively integrate and evaluate AI-assisted tools within educational contexts. These findings provide insights into the faculty participants' teaching experience, field of expertise, and their ability to discern ChatGPT-assisted writing.

Surprisingly, the survey results reveal that the faculty members' years of experience have a minimal impact on their ability to differentiate between ChatGPT and student writing. Faculty with 5-10 years of experience show an 88% accuracy rate, while those with over 10 years have a slightly higher rate of 89%. (See Figure 2). This mere 1% difference suggests that experience level, within these ranges, does not significantly affect the detection of AI-generated text. The similarity in accuracy rates across these experience levels may imply that discerning between AI-generated and student-written content may not be strongly linked to the duration of teaching experience. It is interesting to note, however, that out of the 27 instructors asked to take the survey, only 19 actually completed it. Most of these instructors had more than ten years of teaching experience. It is possible that the instructors with less experience were reluctant to participate because they were not confident about identifying differences in the text. It may be worth considering when looking at the survey results, given the varying levels of teaching experience among the participating instructors.

Figure 2
Accuracy Rate by Years of Experience for Detecting ChatGPT vs. Student Writing
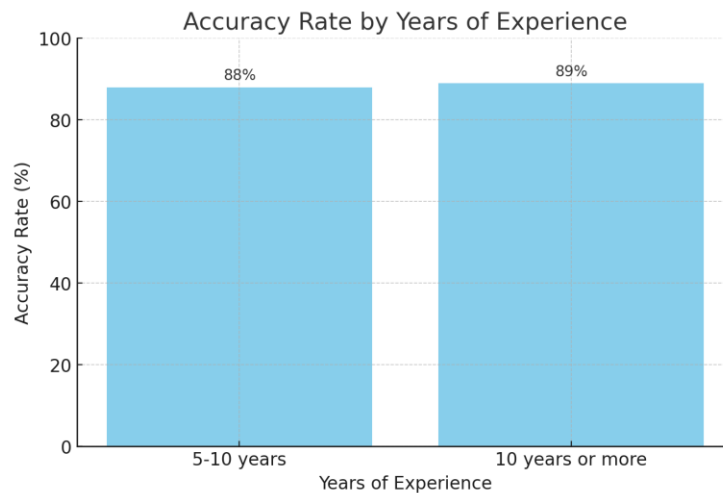

Interestingly, the data does not demonstrate a gradual increase in accuracy with more experience. This observation hints at a potential threshold effect, where faculty members attain a comparable proficiency in identifying AI-generated content after a certain amount of experience, approximately five years or more. However, without more data, especially for faculty with under five years of experience, it is difficult to make conclusive statements about the overall influence of experience on this skill. Overall, this data suggests that faculty training in recognizing AI content might not need to be significantly differentiated based on experience levels.

There is a varied performance across disciplines in the accuracy rates of different academic disciplines in distinguishing ChatGPT from student writing: Education and EFL/ESL faculty have an 88% accuracy rate, Linguistics at 75%, Literature at 94%, Rhetoric at 75%, and Other Disciplines at 75% (See Figure 3). These differences invite a closer look at the potential factors influencing such discrepancies and argue for further research, especially in fields like Rhetoric, which encompass subfields with various theories such as classical, medieval, belletristic, modern, feminist, and environmental rhetoric. Given these findings, it would be beneficial to conduct further research to understand the underlying reasons for these discrepancies within an English department, and even extend the study to faculty who assess writing in disciplines outside of the English department.

Figure 3
Accuracy Rate by Discipline for Detecting ChatGPT vs. Student Writing
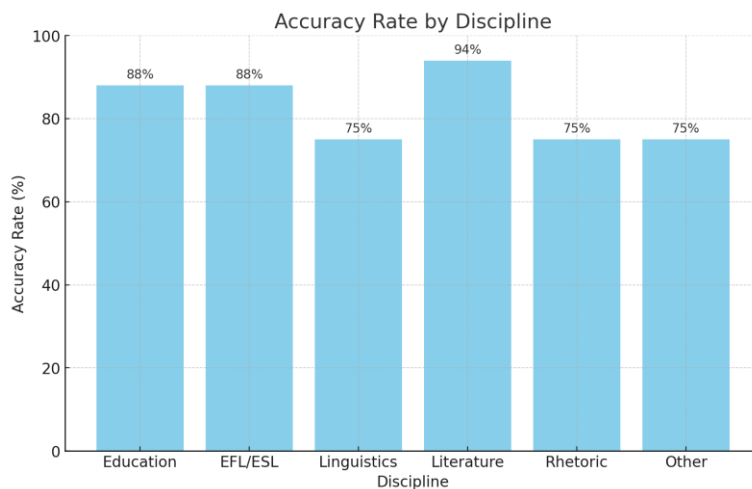
Characteristics Used for Identification

Faculty reported a variety of textual features that helped them to distinguish student writing from ChatGPT-generated content. Faculty detected shifts in perspective, emphasizing students' inclination to share subjective experiences, contrasting with ChatGPT's impersonal tone. Transitions became a focal point, with abrupt shifts and lack of logical organization identified as potential indicators of student writing. Additionally, faculty noted instances of awkward language and coordination issues as common weaknesses in student writing. Beyond expertise, intuition played a pivotal role, as faculty, relying on their gut feelings, successfully discerned ChatGPT's distinct characteristics, demonstrating a blend of analytical and intuitive methods.

Objective Comments

In the analysis of the passages, faculty noted various features which revealed whether the text was assumed to be written by a student versus ChatGPT. Firstly, the use of the passive voice phrase "are known" was deemed unnecessary and therefore 'wordy.' It seemed more characteristic of spoken language than formal writing. Although not grammatically incorrect, this choice was considered stylistically amateurish, reflecting a potential lack of refined writing skills, suggesting the text was written by a student. Faculty also critiqued the casual introduction of a new point about cost, suggesting a lack of formal transitioning in student text. The mention of golf carts in one sample was also deemed unnecessary and therefore suggested that the text was written by a student. Faculty also noted a shift in perspective from direct human experience with e-scooters in contrast to ChatGPT's more impersonal style. Students, faculty believed, were more likely to mention issues such as the heat and the time it takes to go from one building to another. These are 'subjective' experiences as the distance between buildings, or difficulty in walking, would be different at different campuses. Most likely ChatGPT would not have this human, subjective experience.

Faculty members also focused on the use of transitions as a way to discern student writing. They noted the use of the transition "therefore" followed by the introduction of new information about staff needing e-scooters (in addition to students needing them). This transition was considered abrupt, as ChatGPT typically avoids "throwing in" new information. In other words, ChatGPT typically stays on track, whereas student writing is less linearly composed. They emphasized this abrupt use of transitions. One faculty member stated that the transition, "Additionally" indicates a lack of careful organization. Unlike ChatGPT, which typically produces logical transitions, the student's writing was seen as more impulsive.

One faculty member noted the use of an unclear subject in the following sentence, taken from a student excerpt: "In addition, some benefits of the scooters are that they are safe to use as well as eco-friendly: "Some benefits" is functioning as the subject of the sentence but then appears again in the noun clause "are that they are", which creates an awkward repetition. This awkwardness is more likely a feature of student writing than ChatGPT. Other elements of awkward or inelegant writing were also observed: Faculty were attentive to third to second person shifts coupled with coordination issues. The lack of logical coordination with "and" suggested the need for a more fitting conjunction, such as "but." These are mistakes that ChatGPT rarely makes but students often make. There is an ironic possibility that as students increasingly use AI text generators, their own writing might unconsciously start to mirror the AI's style, even some of its better features. Improvements in students' writing proficiency, presumably a desirable outcome, could make it more difficult to distinguish between student writing and AI generated content.

Intuitive Comments

In addition to expertise, intuition played a major rule in faculty detection process, as 100% noted that they did not use any AI detection software in their assessment (**Q8**). Although faculty surveyed are highly experienced, 87% of faculty noted that they relied on intuition to detect the differences (**Q9**) but felt confident in that intuition at a rate of 72%.

Relying on intuition, faculty were highly successful in detecting ChatGPT generated text. Some of their comments regarding intuition were as follows: They noted that students typically do not describe university campuses as "expansive" or use the verb "traverse" to describe movement. The ChatGPT writing was characterized as eerily grammatically correct, and also for graceful use of parallel structures even if clouded by verbose word choices "sustainable, accessible, active." They often identified student-generated writing through subtle inconsistencies in tense, style, and parallelism, viewing these imperfections as indicative of human authorship. The presence of British English expressions, such as using "as" for transitions, also provided a clue for some instructors.

Faculty mentioned that identifying student writing outside the context of coursework, and therefore without familiarity with the individual student, posed a greater challenge than it would in submitted coursework, where they would have some idea of the proficiency of their individual students.

Discussion

The primary objective of this study was to examine the proficiency of English faculty in distinguishing between paragraphs written by students and those generated by ChatGPT. The methodology, ethical considerations, and analysis of data have provided insights into the dynamics involved in the detection process that could prove useful in designing and assessing composition assignments.

The 84% accuracy rate in identifying Chat GPT-generated writing among the faculty indicates a high level of proficiency in distinguishing student from AI-generated writing. This finding underscores the expertise and strong intuition of English faculty members in deciphering features of language and noticing the structural differences and linguistic features in ChatGPT writing as distinct from those found in essays written by humans (Herbold, 2023). In contrast, studies by Waltzer et al. (2023) and Alexander et al. (2023) reveal varied success in human detection of AI-generated content. Waltzer's study, involving 69 teachers and 140 students, showed teacher accuracy at 70%, with students slightly lower at 62%. Despite their relative success, teachers with high school English teaching experience struggled to distinguish well-written student essays from ChatGPT-generated ones. Alexander et al.'s (2023) qualitative study with six ESL lecturers in Cyprus emphasized a deficit model of assessment, focusing on errors as indicators of second

language (L2) writing output. They identified technical and grammatical accuracy, along with sophisticated language use, as potential AI-generated text indicators. Our study, similar to Alexander et al.'s, acknowledges the growing challenges faced by instructors in AI detection but introduces a nuanced approach. We submitted student-generated essays through Grammarly to eliminate technical and grammatical mistakes, encouraging faculty to focus on more sophisticated stylistic choices rather than basic errors. This strategy aims to enhance the evaluation process and address the complexities associated with distinguishing AI-generated content from human-authored text.

Our study's findings bear implications for the integration of educational technology within academic settings because AI has become a "new form of teacher knowledge" (Dunnigan et al., 2023, para. 32). The key implication is that because faculty can detect ChatGPT writing, they are capable of detecting when students rely excessively on AI language models in their assignments. This awareness allows instructors to adapt their pedagogical practices to teach appropriate and ethical use of AI and to design assignments that will thwart or combat overuse. Moreover, the demonstrated high accuracy in detecting Chat GPT-generated content suggests a heightened awareness among faculty members regarding subtle linguistic distinctions that will help them assure academic integrity. Our study shows that faculty have both concrete knowledge and strong intuition regarding the differences between AI-generated text and student writing. Another crucial implication is that faculty who can detect AI-generated content are better-equipped to guide students in the ethical uses of AI and in integrating technology into best practices for the classroom. Finally, this proficiency enables them to effectively distinguish between human and machine-generated text, underscoring the critical role of human evaluators in maintaining academic integrity.

As educational institutions grapple with ethical considerations of the use of AI technology in education, the discernment exhibited by faculty becomes increasingly important. The study serves as a reminder of the need for cautious and ethically informed use of educational technology, especially powerful AI language models like ChatGPT (Vaccino-Salvadore, 2023). While these tools have the potential to offer support to students across diverse academic tasks, the findings emphasize the importance of maintaining a balance and ethical awareness in their use. This study not only shows that faculty have an ability to detect AI-generated text but also reveals the features that faculty focused on for distinguishing between the AI and student writing. Although students improve as writers and may eventually achieve better organization, language choices, parallelism, and grammar, ChatGPT may also improve its ability to simulate L2 writers, making detection in future versions of ChatGPT more difficult. For this reason, we argue that similar research on human detection of AI-generated text should be conducted frequently and across various disciplines and professions.

To improve and maintain the ability of faculty across different disciplines in identifying AI-generated text versus student writing, educators can adopt a comprehensive approach. This would include organizing cross-disciplinary workshops where faculty members can exchange insights and strategies for recognizing AI-generated texts. Additionally, integrating modules on AI text analysis into existing course curricula could provide faculty with practical experience in identifying AI-generated content. Encouraging interdisciplinary research projects that focus on AI in language and writing would promote a deeper understanding of AI-generated texts. It could also be useful to keep faculty updated on the latest developments in AI writing tools, ensuring they stay informed about the evolving capabilities of these technologies. Facilitating interactions between faculty and AI experts can provide valuable insights into the challenges and strategies for distinguishing AI-generated content. Lastly, establishing peer learning networks would enable faculty to share experiences and best practices in detecting AI-written work, fostering a collaborative learning environment.

Although teachers are proficient at detecting AI-generated text by noticing inconsistencies in writing styles, syntactic peculiarities, and gaps in the logical flow of ideas, grade appeals may arise for any grade reduction based on over-reliance on AI unless the teacher provides concrete evidence. Thus, educators may need more concrete proof to substantiate their claims. This is where *Turnitin*, along with other AI detection tools such as *Copyleaks* and i*Thenticate*, could be useful. These tools can assist a teacher's grading by providing more tangible evidence of plagiarism. Given ChatGPT's reliance on existing web-based information and its inability to cite sources, there's a heightened risk for students to commit plagiarism when they misuse AI. Using *Turnitin*, for example, can support teachers in upholding academic integrity by justifying grade reductions for assignments that reveal signs of AI plagiarism. This strategy can support a teacher's intuition with solid evidence that will be accepted by students and withstand grade appeals.

In short, by addressing these discrepancies through targeted research and implementing these remedies, academic institutions can enhance their faculty's ability to effectively distinguish between AI and student writing, thereby maintaining the integrity and quality of academic work.

The findings resonate with prior research in Applied Linguistics and TESOL, emphasizing the importance of contextual understanding in language assessment. The observation that ChatGPT writing tends to be grammatically correct but lacks the richness of specific examples found in student writing aligns with concerns raised in plagiarism detection literature, where linguistic inconsistencies and generality are often flagged as indicators. Moreover, the instructors' emphasis on the importance of context and specific examples in identifying student writing echoes the principles of Task-Based Language Teaching (TBLT) where language is acquired through contextualized, meaningful tasks (Willis & Willis, 2007).

Because contextualized writing was far richer in student writing, we argue that instructors can design writing assignments to focus on context-specific writing tasks on current issues. This change would force students to engage with real-world challenges, encouraging critical thinking and relevance in their writing—all without the possibility of accessing direct 'answers' from an AI language model. Unlike AI language models, which lack real-time information, context-specific assignments enable students to apply their learning to contemporary issues, staying connected to the world around them. For example, project-based learning (PBL) encourages active engagement and investigation by students, developing critical thinking and independent research skills. In his study, Nunn (2020) described how he modified the traditional individual research essay to incorporate active learning through individual student projects, emphasizing key principles of academic writing in the context of PBL. He argued that freshmen students were able to produce insightful and sophisticated research using this process. The students in his classes effectively conveyed the self-directed nature of their study through language choices and demonstrated a distinctive voice in their writing. He argued that a holistic project-based approach, emphasizing self-directed projects, accommodates various learning methods, including task-based learning activities. His study concluded that project-based learning serves as a holistic framework, not as a standalone method but as a supportive approach that motivates students and encourages autonomous control over their research and writing. By emphasizing real-world and personally meaningful projects, PBL helps students develop a deeper understanding of their subjects. This hands-on approach may reduce the temptation for students to rely solely on AI models for research, as they actively participate in the research and writing process, cultivating a more authentic and independent approach to learning.

Some other possible assignments to prevent an overreliance on AI could be the following:

1. Reflective Essay: Ask students to reflect on a personal turning point or an ethical dilemma they faced in their lives describing the factors that influenced the decisions they made.

2. Policy Brief: Assign students to research and draft a policy brief addressing a current problem in their university or local community, guiding them to propose practical solutions based on their understanding of the issue and relevant research.

3. Interview-Based Essay: Have students conduct interviews with experts, community members, or stakeholders related to a current event. The assignment could then involve crafting a feature article that integrates these interviews, providing a firsthand perspective on the issue.

These assignments not only cultivate research and analytical skills but also encourage students to apply their writing in a meaningful, real-world context without direct reliance on ChatGPT.

Finally, it is important to stress to students that Human Intelligence surpasses Artificial Intelligence (AI) because humans have the ability to generate new ideas. AI can only work with the data it has been given and lacks the capacity to innovate without ongoing updates from human inputs. The reality is that current AI relies on human intelligence and human input and is incapable of original or critical thinking. Therefore, when students rely on AI tools, they are giving up practicing and using their creative and critical powers of thought, which stunts their intellectual development.

Limitations and Future Directions

We acknowledge the following limitations of this study. We conducted it at a university in the United Arab Emirates (UAE), so its results might not apply to other contexts. It would be useful to repeat this study in various cultural and language settings to make sure the results are valid in different situations. Additionally, focusing on one class and one English department at the university as our samples limits its generalizability. Disciplines may exhibit varying degrees of sensitivity to linguistic cues in writing (Hutchinson & Waters, 1987). Future research could aim for larger and more diverse samples to better understand faculty abilities in discerning AI-assisted and student-generated writing. To obtain a more comprehensive perspective, future research could also broaden the participant pool to include faculty from diverse disciplines. Investigating how faculty from various disciplines approach AI detection could help universities design AI training courses for faculty and develop academic integrity policies. Finally, the use of an early version of ChatGPT, poses a limitation because AI models are evolving rapidly. Future changes and improvements in AI might change the detection clues found in this study, so it will be important to keep updating the research to stay relevant.

In addition, examining the impact of different writing assignments on both AI and student compositions could provide further insight into distinguishing AI-authored essays across varied disciplines from a broader range of writing styles, such as legal, journalistic, scientific, business, and literary. This approach would allow educators to discover additional textual indicators for differentiating between AI-generated and student-written content across disciplines that require writing assignments. Moreover, expanding the study to include graduate students, could offer valuable information on the use of AI-generated content in higher-level academic work and help in understanding the dynamics of plagiarism in more depth. This expanded view could contribute significantly to the development of comprehensive strategies for maintaining academic integrity as AI technologies rapidly progress.

Conclusion

AI language models have the potential to be used by university students to complete many forms of traditional university assignments, creating multiple problems for university pedagogy.

Assignments submitted as graded work are intended to cause the student to learn through practice and mental effort while completing them, and to allow the faculty teaching the course to assess the achievement of course outcomes by the students submitting the work. Both of these goals are frustrated if the student relies excessively or entirely on an AI language model to complete the assignment. Furthermore, the atmosphere of academic integrity at the institution can be compromised by widespread inappropriate student use of AI language models to complete assignments, and the assignments produced by AI language models are themselves likely to contain extensive plagiarism or even copyright infringement, due to the way that AI language models gather information for their responses, and their current inability to properly cite the sources upon which they rely.

While some universities have attempted to prohibit the use of AI language models, there are strong reasons to doubt the efficacy of such prohibitions, and pedagogical reasons for universities to allow or encourage students to use AI language models to assist, rather than to substitute for, their own work and learning. In order to achieve this balance in the use of AI language models, it must be possible for faculty members to discern the difference between student writing and that produced by AI language models. Currently, academic integrity or plagiarism detection software is unable to differentiate between the writing of students and of AI language models reliably. If, however, faculty members can use their own expertise and intuition to detect the use of AI language models, they can encourage responsible and ethical use of AI language models by students while retaining confidence in the integrity of their assessments.

The results showed that faculty members were successful in distinguishing between the paragraphs written by students and those written by ChatGPT. The reasons they cited were a mix of errors or inelegant choices in style found in the student writing, which are not characteristic of ChatGPT, the more impersonal and disembodied tone and examples used in the ChatGPT-written paragraphs, and the faculty members' own intuition.

These results counsel against despair over the current model of assessment used in university English departments, offer potentially fruitful directions for future research, and could provide guidance in developing and improving faculty detection of AI writing. Reliable detection of AI writing would allow universities to encourage responsible and pedagogically useful employment of AI language models, while maintaining reasonable confidence in the integrity of assessments.

References

Al Afnan, M. A., Dishari, S., Jovic, M. & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, *3*(2), 60–68. https://doi.org/10.37965/jait.2023.0184

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in a second-language institution in higher education. *Teaching English with Technology*, *23*(2), 25–43. https://doi.org/10.56297/BUKA4060/XHLD5365

Allen, M. (2022, December 26). Professor warns about chatbot cheating: "Expect a flood." Axios. https://www.axios.com/2022/12/26/chatbot-cheating-university-warning-chatgpt

Bishop, Lea. (January 26, 2023). A computer wrote this paper: What ChatGPT means for education, research, and writing. http://dx.doi.org/10.2139/ssrn.4338981

Bruder, A. (2023, January 4). What ChatGPT means for how we teach writing (Opinion). *Education Week.* https://www.edweek.org/technology/opinion-what-chatgpt-means-for-how-we-teach-writing/2023/01

Chia, O. (2023, January 6). Teachers v ChatGPT: Schools face new challenge in fight against plagiarism. *The Straits Times*. https://www.straitstimes.com/tech/teachers-v-chatgpt-schools-face-new-challenge-in-fight-against-plagiarism

Deng, J., & Lin, Y. (2022). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems*, *2*(2), 81-83.

Dunne, C. (2023). Can ChatGPT be your coauthor? *BCMJ*, *65*(6), 193. https://bcmj.org/editorials/can-chatgpt-be-your-coauthor

Dunnigan, J., Henriksen, D., Mishra, P., et al. (2023). "Can we just please slow it all down?" School leaders take on ChatGPT. *TechTrends*. https://doi.org/10.1007/s11528-023-00914-1

Elkhatat, A.M., Elsaid, K. & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. Int J Educ Integr 19. https://doi.org/10.1007/s40979-023-00140-5

Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Education Sciences*, *13*(11), 1109. MDPI AG. http://dx.doi.org/10.3390/educsci13111109

Firaina, R., & Sulisworo, D. (2023). Exploring the Usage of ChatGPT in higher education: Frequency and impact on productivity. *Buletin Edukasi Indonesia*, *2*(01), 39-46. https://doi.org/10.56741/bei.v2i01.310

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching*, *12*(1), 44-58. https://doi.org/10.15294/elt.v12i1.64069

Heaven, D. (2023). ChatGPT is going to change education, not destroy it. MIT Technology Review: https://www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/

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

Hutchinson, T., & Waters, A. (1987). *English for Specific Purposes: A Learning-Centered Approach*. Cambridge University Press.

Javaid, M., Haleem A., Singh, R., Khan, S. & Khan, I. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. Bench Council Transactions on Benchmarks, Standards and Evaluations, *3*(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115 https://doi.org/10.1016/j.tbench.2023.100115

Marche, S. (2022, December 6). The college essay is dead. Will ChatGPT kill the student essay? *The Atlantic*. https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/

Neumann, M., Rauschenberger, M., & Schön, E.-M. (2023). We need to Talk about ChatGPT: The future of AI and higher education. *IEEE Xplore*. doi:10.25968/opus-2467.

Nunn, R. (2020). Project-based learning: Learning about PBL from successful freshman writing projects ESBB *6*(1), 41-56. https://www.englishscholarsbeyondborders.org/wp-content/uploads/2020/08/Roger-Nunn-2020.pdf

Oguz, F. E., Ekersular, M. N., Sunnetci, K. M., et al. (2023). Can Chat GPT be utilized in scientific and undergraduate Studies? *Annals of Biomedical Engineering*. Advance online publication. https://doi.org/10.1007/s10439-023-03333-8

Ozar, R., Broda, M., & Wereley, M. (2023). A Sheep in wolf's clothing: Thoughtful and practical applications of ChatGPT in teacher education. In T. Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning* (pp. 38-43). Vienna, Austria: Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/222480/.

Sharma, Y. (2023, March 10). What ChatGPT means for linguistic diversity and language learning. *University World News*.
https://www.universityworldnews.com/post.php?story=20230223132734294
Tate, T., Doroudi, S., Ritchie, D., & Xu, Y. (2023). Educational research and AI-generated writing: Confronting the coming tsunami. *Pre-print*. https://edarxiv.org/4mec3/

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT is a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Villasenor, J. (2023). How ChatGPT can improve education, not threaten it. *Scientific American*. https://www.scientificamerican.com/article/how-chatgpt-can-improve-education-not-threaten-it/

Vaccino-Salvadore, S. (2023). Exploring the ethical dimensions of using ChatGPT in language learning and beyond. *Languages*, *8*(3*)*, 191. https://doi.org/10.3390/languages8030191

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Waltzer, T., Cox, R. L., & Heyman, G. D. (2023). Testing the ability of teachers and students to Differentiate between Essays Generated by ChatGPT and High School Students. *Human behavior and emerging Technologies*. https://doi.org/10.1155/2023/1923981

Willis, J., & Willis, D. (2007). *Doing Task-Based Teaching*. Oxford: Oxford University Press.

Yang, M. (2023, January 6). *New York City schools ban AI chatbot that writes essays and answers prompts.* The Guardian. https://www.theguardian.com/us-news/2023/jan/06/new-york-city-schools-ban-ai-chatbot-chatgpt

Appendix A

Faculty Feedback Survey: Detecting ChatGPT-Assisted Writing SurveyMonkey

Q1. How many years of university level teaching do you have?

Q2. What is your primary field of expertise?

Q3. Below are two passages labeled "A" and "B." One passage is generated by ChatGPT, the other by a student. Which passage is written by ChatGPT? (**Answer B)**

PASSAGE A:

There are several reasons why e-scooters should not be banned from university campuses. University campuses are known to have long distances that students find tiring. Walking from one building to another can be exhausting, especially after a long, hectic class. E-scooters may be the easiest transportation for students when moving from one building to another. Therefore, e-scooters can be helpful not only for students but for the staff too. In addition, some benefits of the scooters are that they are safe to use as well as eco-friendly. Now, imagine that you have a class in ten minutes, and you need around fifteen minutes to reach the other class. Using an e-scooter will get you there in around seven minutes, which means it can also help during traffic times. As for the cost, e-scooters are not considered costly compared to golf carts that can be found around the campus sometimes. Additionally, the scooters do not make noise that might disturb the people walking around the university campus. In conclusion, e-scooters benefit the campus rather than harming it. It aids the students and staff in managing their time and arriving at their destination faster. The scooters are also accessible to many and can be found in many places around the campus.

PASSAGE B:

E-scooters should not be banned from university campuses for several compelling reasons. Firstly, e-scooters offer a convenient and eco-friendly mode of transportation for students. They reduce the carbon footprint and help alleviate traffic congestion on campus, contributing to a more sustainable environment. Moreover, e-scooters are an affordable option for many students,

offering a cost-effective alternative to owning a car or relying on expensive rideshare services. This accessibility can be especially important for those who cannot afford a vehicle or do not have access to public transportation. E-scooters also promote physical activity and overall health. Riding them encourages students to spend more time outdoors, which is crucial for their well-being. Additionally, it's a fun and engaging way to stay active and connected with the campus community. Lastly, e-scooters enhance convenience and save time. They allow students to quickly navigate the sprawling university grounds, making it easier to attend classes, access resources, and engage in extracurricular activities. Instead of banning e-scooters, universities should implement sensible regulations, including safety guidelines and designated parking areas, to ensure responsible usage. This approach can promote a harmonious coexistence between e-scooters and the academic community, reaping the benefits of this innovative transportation solution.

Q4. Below are two passages labeled "A" and "B." One passage is generated by ChatGPT, the other by a student. Which passage is written by ChatGPT? (**Answer A**)

Passage A:

E-scooters, although not without their challenges, should not be banned from university campuses. These electric scooters provide numerous benefits to students and the campus community. Firstly, e-scooters offer a convenient and eco-friendly mode of transportation, reducing the need for personal cars and thereby decreasing campus congestion and air pollution. This aligns with the sustainability goals of many universities and fosters a greener campus environment. Secondly, e-scooters are cost-effective for students. They are often more affordable than owning and maintaining a car or even paying for public transportation. Easy access to affordable transportation encourages students to explore their surroundings and engage in off-campus activities, which enhances their overall college experience. Moreover, e-scooters promote physical activity as they require riders to engage their muscles and maintain balance, thus promoting a healthier lifestyle. The health benefits are particularly important for students who spend long hours studying and sitting in classes. Banning e-scooters may also lead to an increase in other forms of personal vehicles, such as bicycles or cars, which could exacerbate traffic and parking issues on campus. In conclusion, e-scooters have a multitude of advantages, from environmental benefits to affordability and improved health for students. Instead of banning them, universities should work with e-scooter providers to establish sensible regulations that ensure safety and proper use, allowing students to continue enjoying these convenient and eco-friendly modes of transportation on-campus.

Passage B:

On many university campuses, moving between buildings and classes can be extremely exhausting and time-consuming for both students and faculty. One of the most common electric transportation modes that is accessible is the electric scooter. Although some safety concerns arose from electric scooters, they should not be banned from university campuses as they provide more benefits than harm. There are several reasons why electric scooters should be allowed on campus. First of all, electric scooters are a far more sustainable, economically friendly transportation mode. When compared to cars, the utilization of electric scooters has far less environmental impact. Another reason why electric scooters should be allowed and even encouraged is that they provide a smoother, easier experience for students and faculty to move between different buildings on campus in a short period of time, especially in hot weather conditions. This can have an impact on their health as well as productivity. With that being said, there are also various ways to minimize the safety issues related to electric scooters. For example, university campuses should dedicate specific e-scooter paths to ensure safe movement for both riders and pedestrians. Another solution is to set a speed limit for scooters on campus to avoid any accidents.

Q5. Below are two passages labeled "A" and "B." One passage is generated by ChatGPT, the other by a student. Which passage is written by ChatGPT? (**Answer A**)

Passage A:

E-scooters offer numerous advantages for university campuses and their prohibition would undermine potential benefits. Firstly, e-scooters provide an eco-friendly alternative to conventional transportation, reducing the carbon footprint of campuses. They promote sustainability and align with universities' environmental initiatives. Moreover, e-scooters enhance campus mobility, making it easier for students to traverse expansive campuses quickly and efficiently. This can lead to improved punctuality and accessibility for classes, fostering a conducive learning environment. Additionally, e-scooters can alleviate parking congestion, reducing the strain on already limited parking spaces. Furthermore, e-scooters encourage physical activity and healthier lifestyles among students, promoting exercise as a part of daily routines. They offer a convenient way for students to stay active while commuting to and from classes. Banning e-scooters would deprive students of these advantages, hindering progress toward a more sustainable, accessible, and active campus.

Passage B:

E-scooters should be allowed on university campuses for several reasons. One reason relates to the design of university campuses. Due to functional reasons, a lot of campuses around the world have different departments sparsely located. This makes navigating between tightly scheduled

classes that belong to two different departments a challenge for a lot of students. Extreme climate might be another justification for using e-scooters on campuses. Because they allow students to travel faster, they end up spending less time exposed to harsh climatic conditions. This is particularly true in hot dry climates where walking under the sun can cause dehydration and fatigue. Finally, prohibiting the use of e-scooters implies that other small-scale transportation equipment like bikes or skateboards should also be banned. This would not be a legitimate option as these are universally used by millions of students around the world.

Q6. Below are two passages labeled "A" and "B." One passage is generated by ChatGPT, the other by a student. Which passage is written by ChatGPT? (**Answer B)**

Passage A:

E-scooters are an effective and sustainable method of transportation. They alleviate the possible hassle of walking to classes while eliminating the carbon footprint of other modes of transportation such as a car or campus shuttle. Furthermore, they may also provide revenue for the university by implementing rentable scooters. E-scooters should not be banned from a university campus. Depending on the size and geographic location of the campus, scooters can provide immense convenience and comfort for students and faculty alike. On a campus where one must walk for 5-10 plus minutes to go from building to building, scooters can help ensure that students utilize the between-class break effectively while not worrying about being late for class. Additionally, in the summer months when the weather is extremely hot, walking can be even more of an inconvenience. Spending ten minutes walking in 40+ degree weather is enough to make anyone start dripping sweat. To get around these inconveniences and discomforts, university members can simply pay a small fee to the university to rent a scooter. In this way, both the university and the community can benefit from the implementation of e-scooters. Some people may be rightfully concerned about the safety implications of allowing e-scooters on campus. These safety concerns can easily be addressed by implementing a dedicated scooter/bike lane and establishing rules for riding a scooter (wear a helmet, one rider per scooter).

Passage B:

E-scooters have swiftly become a popular mode of transportation on university campuses, offering numerous benefits that make them a valuable addition to student life. Banning e-scooters from campuses would be a hasty decision that overlooks their positive impact on both students and the environment. Firstly, e-scooters promote eco-friendly commuting. With campuses striving to reduce their carbon footprint, these electric vehicles offer a green alternative to traditional gas-

powered vehicles. By encouraging students to choose e-scooters over cars, universities contribute significantly to environmental conservation efforts, fostering a culture of sustainability among the younger generation. Secondly, e-scooters enhance campus mobility. Universities are often expansive, and walking from one end to the other can be time-consuming. E-scooters provide a convenient and efficient solution, enabling students to reach their classes, libraries, and other facilities swiftly. This not only saves time but also enhances overall productivity, allowing students to balance their academic responsibilities more effectively. Furthermore, e-scooters support a healthier lifestyle. Riding scooters requires physical activity, contributing to students' daily exercise routines. By promoting active modes of transportation, universities indirectly address the growing concerns about sedentary lifestyles and promote better health among their student populations. In conclusion, banning e-scooters from university campuses would be a step backward in promoting sustainability, mobility, and student well-being. Instead, universities should work collaboratively with e-scooter companies to establish regulations ensuring safe usage. Embracing e-scooters represents a progressive approach that aligns with modern transportation trends and fosters a more environmentally conscious and healthier campus environment

Q7. What specific characteristics helped you identify ChatGPT writing vs. student writing? What specific characteristics made it challenging?

Q8. Did you use any AI detection software or tools to assist in your identification process? If so,

Q9. How confident were you in your ability to accurately identify ChatGPT-assisted writing?

Q10. Did your own intuition/expertise as a writing teacher play a significant role in your ability to detect ChatGPT-assisted writing?