

Using Classical Rhetoric in the Age of AI: A Study on Waltzer, Cox, and Heyman’s Study “Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students”

By Daniel R. Fredrick*, Thomas P. Corbin Jr.[‡] & Gregory VanderPyl[°]

This paper analyzes essay writing in AI (ChatGPT) and high school students, focusing on their use of specific details. Discussing the writing examples from Waltzer, Cox, and Heyman’s study, we employ Aristotle’s rhetorical theory to explore how clarity is achieved through specificity in writing. The analysis reveals both ChatGPT and students commonly omit detailed content, affecting clarity and argumentative strength, which affects how humans may detect AI-generated text. The paper suggests reinstating elements of classical rhetoric pedagogy in writing courses, particularly the style device of specificity. This will, we believe, not only improve student writing but also help educators to detect AI-generated text.

Introduction

Waltzer, Cox, and Heyman—in their study “Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students”—discovered that students and teachers alike faced difficulty in determining whether essays were written by real students or by ChatGPT, especially student essays which were ‘well-written’ (2023). One of the writing prompts was the following: ‘Why does literature matter?’ This question is open-ended and subjective which gives students an opportunity to wrestle with and respond to a complex and qualitative idea. To answer such a question well requires originality of thought, some sophistication in understanding literature, and most importantly, specific statements that reveal a deep understanding of the topic. If the responses are based on platitudes or generalities, I am sure all educators would agree, then the answers fall short intellectually.

In their essay, Waltzer, Cox, and Heyman show us various passages; some written by ChatGPT and the others written by students. Although the goal of their study is to discover to what extent teachers and students could determine what was written by AI and what was written by a human, the most striking observation one gathers from the sample passages they provide is how *similarly faulty* are the written responses from both students and ChatGPT. As we reviewed the sample writings

*Associate Professor, American University of Sharjah, United Arab Emirates.

[‡]Senior Lecturer, Stephen F. Austin State University-Nacogdoches, USA.

[°]Senior Instructor, American University of Sharjah, United Arab Emirates.

from their research, we immediately identified which samples were written by ChatGPT because the responses were, as expected, very generic, tightly organized, and grammatical--without the slightest hint of sounding colloquial or awkward. We also noted that the ChatGPT excerpts were void of those qualities in the prose that feel so human, such as complexity or burstiness (D'Agostino, 2023). Unfortunately, the student prose was equally generic, suggesting a lack of sophistication in absorbing the material or appreciating the genre. In other words, we analyzed parallel deficiencies in both ChatGPT and student writing, namely the generic responses.

If the student writing samples were quite similar to the writing samples of ChatGPT, then we can conclude that either the students are writing at least as well as a high powered AI text generator (which they are not) or that ChatGPT is unable to produce text beyond the level of a high school student (which it can). Because of its applicability to generate text in a multitude of fields and in various levels of sophistication and depth, we expect ChatGPT to surpass students. If it does not, then we need to know what is lacking in the features of its writing (or prompting)—not only to be able to increase our ability to detect ChatGPT but also to share this insight with students to improve their writing. The key element missing in both ChatGPT as well as the student writing is *specifics*. Specifics is one of the main principles of style in classical rhetorical theory which helps the writer achieve clarity. According to Aristotle in his *Art of Rhetoric*, "...The virtue of style is to be clear...since if a speech does not indicate clearly it will not be performing its function (2004, 3.2).

This fundamental principle underscores why classical rhetoric remains an effective pedagogical tool in improving student writing. The emphasis on clarity through specificity not only sharpens the discernment of AI-created content but also enhances the overall quality of student writing. By focusing on this key feature of clarity via specificity, this study aims to demonstrate how classical rhetoric can provide the necessary tools for both recognizing the limitations of AI while revealing concrete strategies for improving student writing.

Literature Review

Since its debut to the public in November 2022, ChatGPT has become a topic of great interest to educational researchers, professionals, freelancers, students, and teachers due to its applicability in perhaps every university course (Baig & Yadegaridehkordi, 2024). Because of its ease of use and incredible language and other text capabilities, ChatGPT is especially popular among students worldwide (Vargas-Murillo, de la Asuncion Pari-Bedoya, & de Jesús Guevara-Soto, 2023). ChatGPT as well is not only gaining popularity in numerous fields outside of education, but it is becoming a revolutionary aid in most professions. For example, in a study by Oh et al., ChatGPT was compared to a standard statistical model to predict the risk of death from sepsis for ICU patients in South Korea. Although ChatGPT was not as accurate as the statistical model, it was accurate in some cases and showed strong potential for accuracy in others (Oh et al., 2024). This is only one of many astonishing examples of how professional fields employ ChatGPT in a range of uses.

In the field of education, AI is undoubtedly playing a “burgeoning role” (Duran, 2024), and no teacher can ignore its presence. Research has shown that teachers across universities worldwide are embracing ChatGPT in nearly every discipline. As a writing tool, it is no wonder that it has been integrated into writing courses. For instance, Su, Lin, and Lai (2023) set up ChatGPT as a peer in courses on argumentative writing, while Bergström and Yvdal (2024) found that compared with peer feedback, ChatGPT’s feedback was more detailed and constructive, especially when prompted to act as a ‘thesis supervisor.’ In other research focusing on ChatGPT’s use in writing courses, ChatGPT served as an aid to generate different versions of an argumentative essay with mediocre results (Bádonyi, 2024). The use of ChatGPT in education can now be considered ubiquitous, as evidenced by the extensive research of Baig and Yadegaridehkordi (2024), who investigated 57 research articles that identified widespread use of ChatGPT, trends, and methods in higher education. It is certain that ChatGPT will be at the forefront of future discussions and research in education.

Despite the innovative if not extraordinary pedagogical uses for ChatGPT, there are also many grave limitations. Of the more alarming problems with ChatGPT, producing false data is at the top of the list. If it had intention, we could say it lies. Instead, experts refer to ChatGPT’s production of inaccuracies as ‘hallucinations’ (Sakib, 2024). Working with inaccurate information is a significant problem in any discipline but could lead to great physical harm in the fields of science and medicine (Cong-Lem, Soyoo, & Tsering, 2024). In addition to generating false data, there are also concerns that ChatGPT reduces critical thinking because its users are relying on it too much (Murtiningsih, Sujito, & Soe, 2024). Other studies have noted similar ethical problems, noting that ChatGPT has a tendency to generate biased responses, a result of the human bias used to train the model (AlAli & Wardat, 2024). Numerous other systematic studies further confirm this wide range of serious problems with ChatGPT (Crompton & Burke, 2024; Baig & Yadegaridehkordi, 2024). These limitations are not trivial. No other technological pedagogical tool is so replete with these issues. Thus, Bhaskar, Misra, and Chopra (2024) ask the most urgent question of all in the title of their essay: “*Shall I use ChatGPT?*” In their study they found that when people perceive a higher risk when dealing with ChatGPT, they are less likely to believe it is useful or trust it, which in turn reduces their intention to use it.

But of all these issues, the inability for professionals in academia to detect the use of ChatGPT could be the most severe problem in education. Detection is key if one is to answer the questions, ‘Are students completely misusing ChatGPT? Is ChatGPT writing the assignments for them?’ Since much of the research shows that students tend to over-rely on ChatGPT (Durango, Blanco, Rodríguez, & de las Moras, 2024), remedies for detecting its use are critically important and urgent. Currently, one of the major ways to detect ChatGPT generated text is with AI detection software. Kar, Bansal, Modi, and Singh (2024) show that *Sapling* and other AI detection tools such as *Quillbot* are very effective in detecting AI-generated text. *Ghostbuster* is also effective. Tested across different types of writing, *Ghostbuster* has shown high accuracy and strong adaptability, making it more effective than other detection tools (Verma, Fleisig, Tomlin, & Klein, 2024). What needs more research, however, is

an explanation of what specific features the AI takes into account when evaluating text. Some research has identified that stylometric features are used in AI detection systems which trace aspects of language such as word frequency, sentence length, and vocabulary (Opara, 2024).

Although the majority of research focuses on how AI detects AI, there are a growing number of studies that emphasize the importance of using human intuition and experience to detect AI. Human detection may seem less effective since we cannot compute as fast and thoroughly as AI but humans have something AI does not have—thousands of years of using language and the production of billions of human-generated texts, all of which have come from the human heart and brain. Humans have the advantage to detect AI because we know what we sound like. We can respond authentically when a message feels wrong, biased, or uninspiring. To illustrate, a study by Fredrick, Craven, Brodtkorb, and Eleftheriou (2024) at a university in the Middle East shows that faculty members can successfully distinguish between AI-generated text and student-written content, with an average accuracy rate of 84%. Faculty relied on their expertise and intuition, using linguistic cues such as syntax and coherence to make these distinctions. In another study, Markowitz, Hancock, & Bailenson (2024) analyzed hotel reviews, and participants were able to detect human writing 80% of the time. AI hotel reviews were deemed more analytical and emotional but harder to read. AI struggled to express authentic human complaints. Another important study by Waltzer, Cox, and Heyman (2023) shows that while teachers were 70% accurate in their detection of AI, students were lower at 62%, highlighting the importance of experience. The authors did, however, find that a reviewer's confidence, experience with using ChatGPT, and subject expertise did not improve accuracy.

In our initial reading of the Waltzer, Cox, and Heyman study, we were intrigued regarding what the participants were seeing and not seeing from the samples. This led to the idea that perhaps our experience in classical rhetorical theory could shed some unique insights into the study. We believe this insight highlights concrete features that make human writing detectable over AI generated cause. There have been some studies trying to detect explicitly these features. However, the research seems more impressionistic than concrete as to the features of human writing. For example, Gunser, Gottschling, Brucker, Richter, Çakir, and Gerjets (2022) show that participants rated AI-generated texts as lower in stylistic quality, finding them less well-written, inspiring, fascinating, interesting, and aesthetic compared to human-written or original texts. This suggests that while AI can mimic human writing, it may lack the inherent qualities that make human writing more engaging and aesthetically pleasing—what has been called 'intrinsic dimensionality,' writing that is marked by higher complexity in sentence structures, description, and detail (Tulchinskii et al., 2023). For non-linguists, the items for analysis can get less exciting, focusing on differences in n-gram and bigram distributions, where human writing tends to be richer and more varied (André, Eriksen, Jakobsen, Mingolla, & Thomsen, 2023). Our study is unique in that it discusses the exact reasons why AI can be detected. As AI improves, deeper studies obviously will need to follow.

Methods

This research employs a rhetorical analysis and close textual reading of a feature in classical rhetorical theory, namely Aristotle's view of clarity. The analyses and close textual readings stem from three university professors each of whom have taught for over two decades. The study scrutinizes texts produced by ChatGPT and high school students from the article *Testing the Ability of Teachers and Students to Differentiate between Essays Generated by ChatGPT and High School Students* by Waltzer, Cox, and Heyman. The Waltzer, Cox, and Heyman sample consisted of 140 American high school students with an average age of 16.86 years, indicating that they were primarily in the 10th to 12th grades, or approximately 15 to 18 years old. These students completed writing assignments in their English classes, which were graded by teachers, resulting in a mean score of 85.95 out of 100, reflecting their academic performance as average to above average. The student demographic was predominantly white (57%) and native English speakers (96%), and the data collection took place early 2023.

By engaging with and building upon the findings of Waltzer, Cox, and Heyman, this research contributes to the necessary dialogue in the burgeoning field of AI in education, highlighting the importance of classical rhetoric in understanding and improving both AI and student writing. This dialogue is vital for all educators, as it shapes the future direction of teaching writing in an increasingly digital and AI-integrated world.

The rhetorical analysis is based on the Aristotelian concept that a key virtue of style is clarity. This analysis seeks to discern whether the writing by ChatGPT and students achieves this virtue. Clarity, in this context, extends beyond mere intelligibility or ease of readability; it encompasses the persuasiveness of the argument, the validity of evidence, and the specificity of the language. The textual analysis involves examining how specific or general the language in the given samples is and what the effects are when the specifics are lacking. This method determines whether the writing uses vague or general terms and whether it fails to provide concrete examples or detailed arguments to support its claims. The objective is to assess whether the writing, while possibly grammatically correct and easy to read (qualities indicated by measures like the Flesch readability index), actually impacts the reader in a rhetorical sense. The study operates under the premise that being specific is not just an ornamental feature of good writing, as suggested by Alsup (2005), but a fundamental element of one's argument.

In conclusion, the research method focuses on how well both ChatGPT-generated texts and student writings adhere to a key component of classical rhetoric, particularly the use of specific, clear, and persuasive language.

Results

If we examine the two samples provided in the Waltzer, Cox, and Heyman essay, we can see that both the AI version and the student version are lacking both

specific words and specific examples (5). First, we will examine the AI. The sentences are numbered.

- (1) Literature is considered important because it offers a means of examining the human experience and exploring various perspectives, cultures, and beliefs. (2) Through reading and studying literature, individuals are able to gain insight into the emotions, motivations, and experiences of others, as well as gain a deeper understanding of themselves. (3) Additionally, literature provides a way to preserve cultural heritage and pass on important values and traditions from one generation to the next. (4) Furthermore, literature can inspire, challenge, and entertain its readers making it a valuable form of art and expression.

This passage is clear in the sense that it is intelligible. But it is not clear in terms of information or argument.

We do not know what the AI means because there is nothing specific. “Literature” could refer to the plays or sonnets of Shakespeare, Martin Luther King Jr’s essays, the letters of Charles Bukowski, or the songs of Bob Dylan. It could also refer to pamphlets at a dry cleaner or onion-skin-thin leaflets in a box of *Panadol* or *Strepsils* etc. We presume that literature means “fiction” as that is a key, functional definition (Todorov et al., 1973, 7). If the AI gave us just one example of an author—Shakespeare, Orwell, Laura Cereta, we could begin to understand its meaning. Or if it gave us one work—*Hamlet, 1984, Letter to Sempronius Bilbous*, we could learn something specific from the passage. Specifics force us to learn the material; specifics invite us to interact cognitively with the message by ‘showing’ us the meaning, by appealing to our senses.

Exploring this further, we read that “Literature is a ‘means... to examine and explore perspectives, cultures, and beliefs.’” But so too does music, painting, technology, food, dance, education, clothing, and even sexual conduct. The AI’s claim about literature tells us nothing about the specific uniqueness of literature. It merely makes an unsubstantiated claim. The problem of course is that anyone (or should we also now say ‘anything’) can make an unsubstantiated claim. What students need to learn is that, to be taken seriously, their claims need evidence. Unsubstantiated claims are the tools of dictators, tyrants, and propagandists. $2+2=5$ is a claim as is $2+2=4$. Without evidence, however, claims are at best intellectually useless; at worst, dangerous. As educators race to deal with ChatGPT, it is urgent that we help students learn the tools of classical rhetoric to heighten not only their critical thinking but to equip them with the ability to compose substantiated claims that can be built upon, challenged, rejected, or praised in a democratic context, be that context in politics, education, or business etc.

In sentence two, AI claims that through literature we ‘gain insight’ into emotions, motivations, and the ability to deeply understand ourselves? For sure, literature offers this, but it is only a claim. How do readers know what the AI means by gaining insight? Would the insights gained from reading Howell’s *The Damnation of Theron Ware* or Steinbeck’s *Of Mice and Men* be different from reading Sophocles’ *Oedipus Rex* or Aristophanes’ *Lysistrata* because the cultures are different?

What if our insight is dark and depressing? Should we read Fitzgerald's *Great Gatsby* to give up on the American dream? Should we read Mark Twain's *Letters from the Earth* to give up on both religion and the human race? If we reject the pessimism of Fitzgerald and Twain—all the while being entertained by them—do we then become more aware of ourselves? We do for sure, but we only can access the meaning of AI's general message if *we* do all the thinking, that is, if we provide all the specifics. The meaning of AI's claims can only become clear through specific language—which it does not offer.

The third sentence as well is crying out for specifics. ChatGPT has an opportunity to discuss how traditions are passed on through literature. A perfect example would be the Homeric epics. ChatGPT, if it knew the subject matter could even reference Eric Havelock's work *A Preface to Plato* which explains how the Homeric verses, the oral formulae, were a kind of technology which aided memory so that an oral society could pass on complex information (such as ship building) and important, moral precepts such as *being hospitable to strangers*, a precept illustrated in the story of the Cyclops Polyphemus who violates the Greek moral code of hospitality by capturing and eating his guests.

The fourth sentence too is a missed opportunity to be specific. How does literature inspire? Again, the foundational principle of style in classical rhetoric—specifics—would help to substantiate the claim. At the end of Steinbeck's novel *The Grapes of Wrath*, for example, Rose of Sharon offers her lactating breast to feed a starving man. This scene inspires us to cheer for the dignity of all human beings regardless of social norms or taboos. Or, as ChatGPT blandly put it, literature challenges us. Again, the claim is true, but only if it is supported by something specific. For example, after he kills a pawnbroker and her sister, should we cheer for Raskolnikov in *Crime and Punishment* to escape the police? Does literature also challenge us to read on if the author's prose style is beautiful yet the topic revolting? If so, ChatGPT could argue that what makes Nabakov's *Lolita* such a masterpiece is that it both challenges and entertains us; challenges us, because the subject of the novel is an illegal love affair between Humbert Humbert, a man about 40 years old, and Lolita, a 12-year old girl; entertains us, because the prose style is exquisite:

“Lolita, light of my life, fire of my loins. My sin, my soul. Lo-lee-ta: the tip of the tongue taking a trip of three steps down the palate to tap, at three, on the teeth.”

We can conclude that one way to detect AI generated prose is that it is fraught with platitudes and generalities—organized fairly well and grammatical. If you want ChatGPT to be specific, you must continually revise your prompts and goad it for more. Basically, if ChatGPT had produced some of the points we made in our analysis above, would it have been harder to detect because it seems more authentic, or would it have been easier to detect because it referenced numerous works?

Student writing, unfortunately, often offers the same platitudes and generalities as ChatGPT, only without good organization or good grammar.

We now turn our analysis to the student sample:

- (1) Literature matters because it shapes people's minds.
- (2) Growing up with a good understanding of different pieces of literature can build intelligence and prepare people for adulthood.
- (3) Literature can expand creativity, help with speaking in conversation, and just overall enlighten the human brain on different ideas in the world.
- (4) Even going into adulthood, it is still important to experience the joy of reading and overall is a key part to a person's ability to be successful in life.

Just as ChatGPT presents general, unsubstantiated claims, so too does the student. The claim "Literature... shapes people's minds" could also apply to any number of school subjects, religious faiths, hobbies, familial relationships etc. Math and biology shape people's minds. Attending church could shape your mind, and so could practicing *jeet kun do*, or obeying your mother or father. Interestingly, the student has made a remarkable point. If literature can shape a mind, then literature can be a rival influence to parental, governmental, or societal conditioning. Literature can therefore be dangerous, subversive, radical. But the general style does not alert the reader to the student's brilliant idea. Thus, in the hands of a tyrant, extremist, or ignoramus, literature can be censored, banned, burned. *Harry Potter*, for showcasing witchcraft; the *Tales of Peter Rabbit*, for only portraying middle-class rabbits; *Huckleberry Finn* because Huck did not wash before church and of course because of the racist wording; Luther's translation of the Bible because it was written in common language, a translation accessible by both the priesthood and the penniless. "Literature matters," the claim is both a reason to read and to celebrate if you are democratic and free-minded, a lover of rights; or it is a reason to confiscate and destroy if you are fascistic, tyrannical, hungry to control the thoughts and behaviors of others. But none of these ideas come through because the style is general; the student has claims but no evidence. If the elements of classical rhetoric would be used, the student writing would be better because it would be clearer, more persuasive, more entertaining.

In sentence two, the student claims that literature can make you smarter and more mature. This sounds like a claim the student believes a teacher wants to hear. It would be clever for the student to explain exactly how literature prepares you for adulthood. The idea seems a bit sweeping as the majority of adults the world over are not well-read in literature. It is easier to see how math or science or computers prepare one for adulthood in terms of a career, but to say that literature prepares one for adulthood is more vague, more inexact. I suppose the student could argue that the study of literature helps with soft skills such as empathy or emotional intelligence, but it all depends on the literature. The masterpieces in Mortimer Adler's Great Books series have the goal of teaching morality and ethics. Take the Greek dramas, for example. They showcase moral conundrums and force us to think critically on gray areas of morality. Orestes must avenge his father's murderer if he (Orestes) is to be a worthy son, but what if the murderer is Clytemnestra, his own mother? Now what should Orestes do? Oedipus faces a similar moral dilemma. In his confusion, he murders his father and marries his own mother. From Greek literature, we also learn that if we offend the gods, we get thrown off course like Odysseus, wandering in exile for ten years, fighting to return home. We also learn that no matter how far

we venture in life, the ultimate goal is returning home, older, wiser, happier. Is not that the same message as Dorothy's in the *Wizard of Oz*? *There's no place like home*. But is this what the student means? We can never know unless the student is specific, unless the student applies the style devices of classical rhetoric. Generalities tell us nothing; they are unclear.

The student's third sentence has too many broad ideas, each worthy of their own focused paragraph. Literature could expand creativity, but how? The Greek and Roman rhetorical schools would have agreed with the student, but they would have shown students a concrete pedagogical tool to bring about that creativity called *imitatio*, or imitation. In the ancient world, students would copy and imitate passages of Homer or Cicero, trying to create similar passages by writing in the same style, using the same tropes and schemes, but writing about their own subject matter. Regardless of the many possible ways literature can enlighten our minds or foster our skills, we need to know exactly how one achieves that skill. Otherwise, it is just an empty claim and can be easily refuted by claiming the opposite: Literature does not matter; literature does not prepare us for adulthood; literature does not enlighten us; or literature does not make us creative. In fact, the data shows that these negative claims about literature might be true, we cringe to say, because we appreciate literature. But in reality, at least outside the university walls, romance/erotic novels matter the most if one judges by sales, then books on crime, religion, science fiction, and horror. Today, more people know who Freddy Kreuger is than Nick Barnes. Literature is nowhere on the list of top sales. *Naked Lunch* is a restaurant in Cap d'Adge, correct? No. It is a novel by William S. Burroughs. Thus, because literature is not popular, both ChatGPT and students need to give us more than a claim regarding how literature prepares us for adulthood.

So, what do we learn in analyzing these passages through the lens of classical rhetoric?

We learn that both ChatGPT and students need to support claims with specifics.

We learn that if either ChatGPT or the student had specifically referenced any piece of literature, any character, any scene, the writing would become clearer, claims enriched with evidence, the writer's voice singing with personality. Without the specifics, the passages are general and ineffective, void of substance, boring and flat, saying nothing.

Conclusions

This study, through a detailed rhetorical analysis of texts generated by ChatGPT and high school students from Waltzer, Cox, and Heyman's study, underscores a significant gap in both the writing generated by AI as well as students. Both lack the classical rhetoric device of specificity. We show that the absence of specificity leads to a deficiency in clarity. This discussion is critical in understanding the limitations of AI-generated text as well as revealing the opportunity to improve student writing by incorporating elements of classical rhetoric into the lesson plans. Our analysis demonstrates that specificity, a foundational principle of classical rhetoric, is not merely an aesthetic choice but an important element of effective communication.

Both ChatGPT and student writing, while varying in grammatical correctness and organization, consistently fall short in providing specific, detailed arguments and examples that are essential for clear and persuasive writing. This issue becomes more pronounced in educational institutions where students are increasingly tempted to use AI technologies like ChatGPT.

Therefore, the study advocates for a renewed focus on classical rhetoric in writing instruction, emphasizing the need for explicit teaching of specificity as a means to achieve clarity. By doing so, writing teachers can equip students with the necessary tools to not only improve their writing but also critically evaluate and understand AI-generated content. As ChatGPT and similar technologies become more ubiquitous, the role of classical rhetorical principles, particularly in cultivating a specific and clear writing style, becomes increasingly vital. This research contributes to a growing dialogue in the field, highlighting the relevance of classical rhetoric in an AI-integrated educational environment and its importance in preparing students for the challenges of the digital age.

References

- AlAli, R., & Wardat, Y. (2024). How ChatGPT will shape the teaching-learning landscape in the future? *Journal of Educational and Social Research*, 14(2), 336-345. <https://doi.org/10.36941/jesr-2024-0047>
- Alsup, A. (2005). Persuasion, more than argument: Moving toward a literary sensitivity in the classroom. In T. R. Johnson & T. Pace (Eds.), *Refiguring Prose Style: Possibilities for Writing Pedagogy* (pp. 93–106). University Press of Colorado. <https://doi.org/10.2307/j.ctt4cgq34.11>
- André, C. M. J., Eriksen, H. F. L., Jakobsen, E. J., Mingolla, L. C. B., & Thomsen, N. B. (2023). *Detecting AI authorship: Analyzing descriptive features for AI detection*. CEUR Workshop Proceedings, 3551. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85178363236&partnerID=40&md5=8ea03306d9223919e24a29908356c72b>
- Aristotle. (2004). *Rhetoric* (W. Rhys Roberts, Trans.). Dover Publication.
- Bádonyi, R. (2024). *Teaching ChatGPT: How to write an argumentative essay in academic form* (Unpublished manuscript). Kodolányi János University.
- Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, Article 102411. <https://doi.org/10.1016/j.ijer.2024.102411>
- Bergström, O., & Yvdal, A. (2024). *ChatGPT-4's effectiveness in providing feedback on argumentative writing in higher education: A case study*. Degree Project in Technology, First cycle, 15 credits, KTH Royal Institute of Technology, EECS School of Electrical Engineering and Computer Science.
- Bhaskar, P., Misra, P., & Chopra, G. (2024). Shall I use ChatGPT? A study on perceived trust and perceived risk towards ChatGPT usage by teachers at higher education institutions. *International Journal of Information and Learning Technology*. Advance online publication. <https://doi.org/10.1108/IJILT-11-2023-0220>
- Cong-Lem, N., Soyooof, A., & Tsering, D. (2024). A systematic review of the limitations and associated opportunities of ChatGPT. *International Journal of Human-Computer Interaction*. Advance online publication. <https://doi.org/10.1080/10447318.2024.2344142>

- Crompton, H., & Burke, D. (2024). The educational affordances and challenges of ChatGPT: State of the field. *TechTrends*, 68(2), 380-392. <https://doi.org/10.1007/s11528-024-00939-0>
- D'Agostino, S. (2023). Academics work to detect ChatGPT and other AI writing. Inside Higher Ed | Higher Education News, Events and Jobs. <https://www.insidehighered.com/news/2023/01/20/academics-work-detect-chatgpt-and-other-ai-writing>
- Duran, V. (2024). Analyzing teacher candidates' arguments on AI integration in education via different chatbots. *Digital Education Review*, (45), 68-83. <https://doi.org/10.1344/der.2024.45.68-83>
- Fredrick, D. R., Craven, L., Brodtkorb, T., & Eleftheriou, M. (2024). Rethinking Classical Rhetoric: A Comparative Study of Specificity in AI and Student Writing. *English Scholars Beyond Borders*, 10(2).
- Gunser, V. E., Gottschling, S., Brucker, B., Richter, S., Çakir, D. C., & Gerjets, P. (2022). *The pure poet: How good is the subjective credibility and stylistic quality of literary short texts written with an artificial intelligence tool as compared to texts written by human authors?* Proceedings of the 44th Annual Meeting of the Cognitive Science Society: Cognitive Diversity, CogSci 2022, 1744-1750. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85140344936&partnerID=40&md5=9b5cf2bcff5a7c1412231df0dac08974>
- Holcomb, C., & Killingsworth, M. J. (2010). *Performing Prose: The Study and Practice of Style in Composition*. Southern Illinois University Press.
- Johnson, T. R., & Farmer, F. (2020). "Clarity" really means rhythm: Toward a psychoanalytic poetics of prose. In P. Butler, B. Ray, & S. M. Vanguri (Eds.), *Style and the Future of Composition Studies* (pp. 53–65). University Press of Colorado. <http://www.jstor.org/stable/j.ctv194cpgg.8>
- Kar, S. K., Bansal, T., Modi, S., & Singh, A. (2024). How sensitive are the free AI-detector tools in detecting AI-generated texts? A comparison of popular AI-detector tools. *Indian Journal of Psychological Medicine*. Advance online publication. <https://doi.org/10.1177/02537176241247934>
- Markowitz, D. M., Hancock, J. T., & Bailenson, J. N. (2024). Linguistic markers of inherently false AI communication and intentionally false human communication: Evidence from hotel reviews. *Journal of Language and Social Psychology*, 43(1), 63-82. <https://doi.org/10.1177/0261927X23120020>
- Miles, R., Bertonasco, M., & Karns, W. (1991). *Prose Style: A Contemporary Guide* (2nd ed.). Prentice Hall.
- Murtiningsih, S., Sujito, A., & Soe, K. K. (2024). Challenges of using ChatGPT in education: A digital pedagogy analysis. *International Journal of Evaluation and Research in Education*, 13(5), 3466-3473. <https://doi.org/10.11591/ijere.v13i5.29467>
- Oh, N., Cha, W. C., Seo, J. H., Choi, S.-G., Kim, J. M., Chung, C. R., Suh, G. Y., Lee, S. Y., Oh, D. K., Park, M. H., Lim, C.-M., & Ko, R.-E. (2024). ChatGPT predicts in-hospital all-cause mortality for sepsis: In-context learning with the Korean Sepsis Alliance database. *Healthcare Informatics Research*, 30(3), 266-276. <https://doi.org/10.4258/hir.2024.30.3.266>
- Opara, C. (2024). StyloAI: Distinguishing AI-generated content with stylometric analysis. *Communications in Computer and Information Science*, 2151, 105-114. https://doi.org/10.1007/978-3-031-64312-5_13
- Sakib, S. M. N. (2024). *Bane and boon of hallucinations in the context of generative AI*. In *Cases on AI Ethics in Business* (pp. 276-299). IGI Global. <https://doi.org/10.4018/9798369326435.ch016>

- Su, Y., Lin, Y., & Lai, C. (2023). *Collaborating with ChatGPT in argumentative writing classrooms*. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Todorov, T., Moss, L., & Braunrot, B. (1973). The notion of literature. *New Literary History*, 5(1), 5–16. <https://doi.org/10.2307/468404>
- Tulchinskii, E., Kuznetsov, K., Kushnareva, L., Cherniavskii, D., Nikolenko, S., Burnaev, E., Barannikov, S., & Piontkovskaya, I. (2023). *Intrinsic dimension estimation for robust detection of AI-generated texts*. *Advances in Neural Information Processing Systems*, 36. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85191152388&partnerID=40&md5=5efac7da1a951aed0d6e1ce844b2c285>
- Vargas-Murillo, A. R., de la Asuncion Pari-Bedoya, I. N. M., & de Jesús Guevara-Soto, F. (2023). Challenges and opportunities of AI-assisted learning: A systematic literature review on the impact of ChatGPT usage in higher education. *International Journal of Learning, Teaching and Educational Research*, 22(7), 122-135. <https://doi.org/10.26803/ijlter.22.7.7>
- Verma, V., Fleisig, E., Tomlin, N., & Klein, D. (2024). *Ghostbuster: Detecting text ghostwritten by large language models*. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2024*, 1, 1702-1717.
- Waddell, C. (2014). *Basic Prose Style and Mechanics*. Houghton, Michigan.
- 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*, 2023, Article ID 1923981, 9 pages. <https://doi.org/10.1155/2023/1923981>