Edukacja Filozoficzna 79/2025
ISSN 0860-3839, eISSN 2956-8269
DOI: 10.14394/edufil.2025.0006
ORCID: 0000-0002-5960-2909

In Defence of LLM-Based Tools in Scientific Writing: Epistemic and Ethical Considerations of LLM-Restrictive Publishing Policies

Aleksandra Vučković
(Institute for Philosophy, Faculty of Philosophy, University of Belgrade)

Abstract: The growing concerns about using tools based on large language models (LLMs) have caused academic institutions and scientific publishers to adopt rigid policies with little to zero tolerance for LLMs in academic writing. Moreover, some may employ artificial intelligence (AI) tools to differentiate LLM-generated and human essays. We argue that such an approach is inherently limited, as it leaves room for false detection. After analysing recent studies on the effectiveness of AI detection tools and human ability to recognize AI-generated text, we explore epistemic conclusions and the black box problem. Turning to ethical aspects, we argue that non-native English speakers are particularly at risk of false-positive AI detection. We propose the potential benefits of moderate tolerance for LLM-based applications in scientific publishing.

Key words: LLM-based tools, scientific writing, publishing policies, AI tools for LLM detection, linguistic privilege

1. Introduction

This article explores the tension between the growing number of uses of large language models (LLMs) in scientific studies and the policies that universities, research facilities, and academic publishers introduce to avoid the dissemination of papers, in whole or in part, produced by artificial intelligence (AI). The debate on the ethical use of LLMs is multifaceted, with some arguing that the new technologies could improve scientific research and others focusing on data falsification and misrepresentation risks. To ensure that researchers benefit from LLMs while maintaining academic integrity, the scientific community should agree on what classifies as the abuse of this technology and how to prevent it.

This task is more demanding than it appears, as both humans and AI tools have encountered challenges recognizing AI-generated text. The two-way inaccuracy – false positives and false negatives – raises concerns regarding the reliability of AI tools for LLM detection. Additionally, flagging human-written papers as LLM-generated may be more harmful than overlooking the actual use of LLMs, as false accusations may impair researchers' careers and reputations.

Non-native English speakers are especially vulnerable to false positives since AI tools for LLM detection may misinterpret the lack of language fluency as an indication that the paper is AI-generated. Even editors and reviewers can get suspicious when AI tools report possible LLM use. As international researchers are already more disadvantaged in publishing than native English-speaking peers, labelling their manuscripts as AI-generated could widen that gap and further harm their prospects. Moreover, a complete veto on LLMs might deny foreign speakers legitimate assistance, as these tools can improve their writing style and grammar.

In the following section, we explore why recent developments in LLM-powered chatbots have prompted a reaction from academic institutions and publishers. The examples of hallucinations and misrepresentations in AI-generated text provide insight into why many adopted policies that fully ban LLMs. However, there are challenges to this restriction. Section 3 explores the *epistemic* challenge – the difficulty of differentiating between human and AI-generated content. First, we reflect on the studies that reveal how humans struggle to establish whether text was produced by another human or LLM application. Second, we show that even AI tools designed to detect LLM-generated content have made mistakes of false recognition. We argue that this uncertainty, combined with the black box problem, warrants caution before labelling someone's work as AI-generated. In Section 4, we proceed to the *ethical* challenge: the problem of non-native English-speaking researchers being at higher risk of false positives. After introducing the concept of linguistic epistemic injustice and, conversely, linguistic privilege, we turn to studies suggesting that AI tools for LLM detection may disproportionately harm international researchers. Section 5 explores the possible benefits of LLM-based tools, as non-native English speakers can use them to overcome the linguistic gap. After analysing the arguments in favour of LLMs, we highlight some limitations to reliance on them, underscoring the need for responsible use.

2. Academic Response to the Problems of AI-Generated Content

Academic institutions and scientific publishers have changed their policies to prevent the production of AI-generated papers. Harvard guidance for students currently states that while some courses allow moderate exploration of generative AI tools, others classify their use as academic misconduct.[1] Oxford and Cambridge – among other universities in the UK – in 2023 prohibited LLMs, fearing plagiarism.[2] Similarly to academia, scientific publishers adopted new policies. Journals published by Science banned LLMs, while Taylor & Francis and Springer-Nature policies state that these tools do not qualify for authorship. On the other hand, Elsevier adopted a more LLM-friendly policy that limits AI use to language perfection, while the authors are responsible for manuscript content.[3]

To comprehend the unease that recent developments in the AI industry have caused within the academic and publishing community, we need to understand AI-generated content as *any* form of media created as a response to prompts submitted to AI applications. Generative AI is the broad term for various algorithmic procedures based on deep learning and neural networks – such as transformers for language processing or convolutional neural networks for image processing – that assemble seemingly novel content: texts, pictures, music, speech, and videos.[4] Since LLMs generate text and the communication of scientific findings primarily relies on written materials, LLM-based tools are in the middle of the debate on AI abuse within academia and publishing.

The rise of LLM-powered chatbots – such as OpenAI's ChatGPT, Google's Bard (now known as Gemini), Microsoft's Bing AI Chat (now known as Copilot), Anthropic's Claude, Perplexity AI Inc.'s Perplexity – has gained media atten-

---

1 More information is available at their official website: Harvard University, *Generative AI Guidance*, URL: https://oue.fas.harvard.edu/faculty-resources/generative-ai-guidance/.

2 In total, 28 universities across the UK have updated policies to classify the abuse of ChatGPT as plagiarism. For more information, see P. Wood, *Oxford and Cambridge Ban ChatGPT over Plagiarism Fears but Other Universities Embrace AI Bot*, "The iPaper," 23.02.2023, URL: https://inews.co.uk/news/oxford-cambridge-ban-chatgpt-plagiarism-universities-2178391.

3 Y.K. Dwivedi et al., *"So What if ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy*, "International Journal of Information Management" 2023, Vol. 71, 102642, p. 34, https://doi.org/10.1016/j.ijinfomgt.2023.102642.

4 S. Feuerriegel et al., *Generative AI*, "Business & Information Systems Engineering" 2024, Vol. 66, No. 1, p. 111, https://doi.org/10.1007/s12599-023-00834-7.

tion but also raised authorship concerns due to their high accessibility and user-friendliness. These tools are trained on massive data sets, which allows them to mimic human writing and conversations with remarkable fluency.[5] Unlike previous rule-based systems or systems relying on smaller datasets, LLMs possess developed context understanding, reduced biases, and fine-tuning capabilities,[6] which advances their natural-language processing capacity.[7] However, they are not subtle enough not to misrepresent the content. For example, a comparison between different studies on ChatGPT accuracy has shown that it gave correct answers between 60 and 90 percent of the time[8] – a score impressive for casual users but unreliable for scientific purposes.

A case of a retracted article from the scientific journal "Frontiers in Cell and Developmental Biology" with an AI-generated diagram of mouse anatomy became an internet curiosity, as it made little sense even to laypeople, let alone biologists. However, misrepresentations can have vast consequences if inaccurate AI-generated content *appears* authentic. If scientists were to entrust an LLM-based tool with substantial parts of research, its output might seem convincing, but it could also be laden with falsities and inconsistencies. These inaccuracies, known as hallucinations, can vary from statements that contradict the facts (factuality hallucinations) to inconsistencies with the context of the input (faithfulness hallucinations).[9]

Moreover, if LLM applications cannot find the answer to a question, they may invent and cite a non-existent study, thus undermining the research relying on

5 Ö. Aydın, E. Karaarslan, *Is ChatGPT Leading Generative AI? What Is beyond Expectations?*, "Academic Platform Journal of Engineering and Smart Systems" 2023, Vol. 11, No. 3, pp. 118–134, https://doi.org/10.21541/apjess.1293702.

6 P.P. Ray, *ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope*, "Internet of Things and Cyber-Physical Systems" 2023, Vol. 3, p. 122, https://doi.org/10.1016/j.iotcps.2023.04.003.

7 H. Naveed et al., *A Comprehensive Overview of Large Language Models*, arXiv:2307.06435, https://doi.org/10.48550/arXiv.2307.06435; H. Lane, M. Dyshel, *Natural Language Processing in Action*, Simon and Schuster, 2025.

8 K.I. Roumelioti, N.D. Tselikas, *ChatGPT and Open-AI Models: A Preliminary Review*, "Future Internet" 2023, Vol. 15, No. 6, 192, https://doi.org/10.3390/fi15060192.

9 H. Ye et al., *Cognitive Mirage: A Review of Hallucinations in Large Language Models*, arXiv:2309.06794, https://doi.org/10.48550/arXiv.2309.06794; L. Huang et al., *A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions*, "ACM Transactions on Information Systems" 2024, Vol. 43, No. 2, 42, https://doi.org/10.1145/3703155; P.R. Vishwanath et al., *Faithfulness Hallucination Detection in Healthcare AI*, in: *Artificial Intelligence and Data Science for Healthcare: Bridging Data-Centric AI and People-Centric Healthcare*, 2024.

their output.[10] Mosaics of authentic and inaccurate pieces of text are especially dangerous as they, due to illusory credibility, can lead to the dissemination of falsities and fabrications.[11] LLM-based tools may also omit the references. A study has shown that Bard (Gemini) had the lowest score, as it failed to deliver *any* references. Among applications that offered sources, ChatGPT and Bing AI Chat (Copilot) were the least accurate. However, the same study revealed more promising results for Elicit and SciSpace, chatbots designed to explore and analyse scientific literature, as their reference hallucination scores were insignificant.[12]

These findings offer a more optimistic outlook for LLM-based tools in research. Scholars can use them to search for literature and enhance their linguistic competencies, from the proper use of grammar and syntax to the overall writing style and clarity. The latter purpose could contribute to linguistic disparity mitigation – a topic we further explore in section 5. Still, it can be challenging to draw the line between fair usage and misuse of these tools, especially when assessing someone else's work, as we do not know the extent of their reliance on these tools. In the wake of this uncertainty, restrictive publishing policies make sense. However, to justify restrictions, we need to find reliable methods to detect AI-generated text. In the following section, we explore current attempts and challenges in this process.

3. The Epistemic Challenge: (How) Can We Detect AI-Generated Text?

Since the emergence of LLM-based tools among the general public, numerous studies have explored whether their output can be accurately discerned from human-written text. Some studies estimate how well humans can recognize AI-generated content, and others how well AI recognizes AI-generated text. By com-

[10] T. Day, *A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT*, "The Professional Geographer" 2023, Vol. 75, No. 6, pp. 1024–1027, https://doi.org/10.1080/00330124.2023.2190373.

[11] H. Alkaissi, S.I. McFarlane, *Artificial Hallucinations in ChatGPT: Implications in Scientific Writing*, "Cureus" 2023, Vol. 15, No. 2, e35179, p. 4, https://doi.org/10.7759/cureus.35179.

[12] F. Aljamaan et al., *Reference Hallucination Score for Medical Artificial Intelligence Chatbots: Development and Usability Study*, "JMIR Medical Informatics" 2024, Vol. 12, No. 1, e54345, https://doi.org/10.2196/54345.

paring the strengths and weaknesses of human and AI approaches to this issue, we may be able to develop fair future policies for the use of LLMs.

It is troubling that the studies with human participants have shown mixed results – from promising to average. One such study tasked experts in biology with identifying AI-generated abstracts, and their responses were accurate 93 percent of the time,[13] suggesting they did more than just guess. However, a more recent investigation reflected the overall inability of teachers to differentiate between AI-generated and student essays, with 73 percent of correct detection among student articles and only 37.8 percent of correct detection among ChatGPT articles.[14] Another study on university students supports these findings, as out of 376 short essays, teachers correctly classified only 204 as human-written or AI-generated, meaning the accuracy rate was just above 54 percent.[15] Although the contexts of the compared studies differ (experts evaluating experts vs teachers evaluating students), and despite some smaller-scale analyses, where teachers performed better,[16] we are still far from confidently distinguishing AI-generated text. It is also no surprise that expert articles were recognized more accurately than student essays, and this could signify that students lack writing experience and language mastery.

Some studies suggest that humans are intrinsically disadvantaged at recognizing AI-generated text due to our heuristics. For instance, we are inclined to think of first-person texts as human-written. If this is true, we are prone to the manipulations of even more advanced technologies in the future.[17] Therefore, it is unsurprising that we continue to develop AI tools for LLM detection.

---

13 S.L. Cheng et al., *Comparisons of Quality, Correctness, and Similarity between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study*, "Journal of Medical Internet Research" 2023, Vol. 25, e51229, https://doi.org/10.2196/51229.

14 J. Fleckenstein et al., *Do Teachers Spot AI? Evaluating the Detectability of AI-Generated Texts among Student Essays*, "Computers and Education: Artificial Intelligence" 2024, Vol. 6, 100209, https://doi.org/10.1016/j.caeai.2024.100209.

15 C. Saarna, *Identifying Whether a Short Essay Was Written by a University Student or ChatGPT*, "International Journal of Technology in Education" 2024, Vol. 7, No. 3, pp. 618, https://doi.org/10.46328/ijte.773.

16 G. Price, M.D. Sakellarios, *The Effectiveness of Free Software for Detecting AI-Generated Writing*, "International Journal of Teaching, Learning and Education" 2023, Vol. 2, No. 6, pp. 33–34, https://doi.org/10.22161/ijtle.2.6.4.

17 M. Jakesch et al., *Human Heuristics for AI-Generated Language Are Flawed*, "Proceedings of the National Academy of Sciences" 2023, Vol. 120, No. 11, e2208839120, https://doi.org/10.1073/pnas.2208839120.

If we shift our attention to studies that test the effectiveness of these tools, we encounter the epistemic dilemma of whether and to what degree we should trust their results. One study, conducted on 16 different AI detectors, has shown that three of them – Copyleaks, Turnitin, and Originality.ai – had perfect scores in detecting ChatGPT-generated text. The remaining 13 had difficulties distinguishing between LLM-generated and student essays, thus raising concerns about their reliability in the academic context.[18] Furthermore, the available tools for AI-generated text detection recognize earlier versions of ChatGPT (up to GPT 3.5) more successfully than its more recent version – GPT 4.[19] This suggests that the tools we use to identify LLM-generated text tend to fall behind the LLMs they are supposed to detect.

One study tested 14 different tools that scored impressive results of 96 percent accuracy in detecting human-written text and 77 percent in detecting ChatGPT-generated text, with Turnitin, once more, in the lead. However, the initial promising results quickly deteriorated with the introduction of additional parameters. For instance, if a foreign-language article was translated into English using Google Translate, the accuracy of 96 percent dropped to 79 percent, meaning that the non-native authors who use machine translation are about 17 percent more likely to be wrongfully accused of LLM abuse. Additionally, if ChatGPT text was paraphrased via another software, the likelihood of AI tools detecting it dropped from 77 percent to just 31 percent.[20] These findings illustrate a two-fold imprecision. On the one hand, the researchers who use legitimate assistance tools (e.g., machine translation) risk false positives. At the same time, genuine AI abuse can be concealed through just one additional (and AI-generated) step. The significant amount of both false positives and false negatives and the unknown ratio between them raise further concerns regarding how much trust we should put in AI tools for LLM-generated text detection.

---

[18] W.H. Walters, *The Effectiveness of Software Designed to Detect AI-Generated Writing: A Comparison of 16 AI Text Detectors*, "Open Information Science" 2023, Vol. 7, No. 1, 20220158, https://doi.org/10.1515/opis-2022-0158.

[19] A.M. Elkhatat, K. Elsaid, S. Almeer, *Evaluating the Efficacy of AI Content Detection Tools in Differentiating between Human and AI-Generated Text*, "International Journal for Educational Integrity" 2023, Vol. 19, 17, https://doi.org/10.1007/s40979-023-00140-5.

[20] D. Weber-Wulf et al., *Testing of Detection Tools for AI-Generated Text*, "International Journal for Educational Integrity" 2023, Vol. 19, No. 1, pp. 26–65, https://doi.org/10.1007/s40979-023-00146-z.

More recent research,[21] however, revealed improvements in the ability of AI tools to detect text generated through ChatGPT, Perplexity, and Gemini. LLM-generated texts were corrected through Grammarly first, then paraphrased using Quillbot, and finally slightly edited by human experts. Among tested applications, Turnitin had an outstanding 100 percent accuracy in detecting AI-generated content, even with additional paraphrasing. GPTZero and Writer AI had a significant drop in accuracy after Quillbot intervention but still managed to report an AI score of above 50 percent. The only exception was ZeroGPT, which mostly failed to recognize Gemini-generated text.

While these findings suggest that further technological developments could address the risk of LLM abuse, there are epistemic reasons for caution when trusting either LLM-based applications or AI tools for LLM detection. Since the inside of generative AI is a black box, most of the research on the epistemological aspects of these tools is empirical. Contemporary chatbots, unlike their predecessors, do not use traditional models with machine-learning algorithms that create identical outputs for identical inputs (assuming there is no change in training data in between). In modern deep-learning algorithms, the basic idea behind each answer might remain the same. However, the output wording and the choice of relevant information will differ between two identical prompts. The model will change its own classification structure (characterization of learning data) based on the context of the prompt.[22] For this reason, researchers can judge the accuracy of these models solely through their output.

It has been argued that AI ethics is inseparable from the epistemology of AI, with the black box opaqueness as the main problem. To fully assess the moral consequences of the black box applications, we would need to develop *glass-box epistemology*, that is, to understand the processes involved in AI's creation of the output. While glass-box epistemology, in general, may mean any approach that develops procedures that increase the transparency of AI systems, the authors argue for the integration of ethical values throughout the entire development process. At the same time, the evaluation of AI systems should not be limited to experts but include laypeople, which would raise the overall understanding and

[21] M.A. Malik, A.I. Amjad, *AI vs AI: How Effective Are Turnitin, ZeroGPT, GPTZero, and Writer AI in Detecting Text Generated by ChatGPT, Perplexity, and Gemini?*, "Journal of Applied Learning and Teaching" 2024, Vol. 8, No. 1, https://doi.org/10.37074/jalt.2025.8.1.9.

[22] Z. Hao, *Deep Learning Review and Discussion of Its Future Development*, "MATEC Web of Conferences" 2019, Vol. 277, 02035, https://doi.org/10.1051/matecconf/201927702035.

trust in these technologies.[23] Through comprehension of internal processes, we would gain better reasons to trust the output.

At the moment, we cannot prove that AI tools for LLM detection are more efficient than LLMs themselves, and it is a matter of debate whether we can do so even *in principle*. The project of glass-box epistemology (full transparency of all AI systems) may be more of an ideal than a goal attainable in the near future. If LLMs are unreliable, the same applies to AI tools for their detection. Until the latter technologies show a significant amount of transparency compared to the LLMs, they are equally problematic from the epistemological point of view. We argue there is no *epistemic* justification for relying only on AI to detect AI-generated text.

This is not to say that we should abandon our endeavours to identify and sanction the abuse of LLMs. AI tools for LLM detection can be helpful, especially when combined with an independent human evaluation of papers.[24] The takeaway is that we should be cautious of their findings as much as the researchers who use LLMs should be careful about their output. In the following section, we explore *ethical* reasons for this caution and the concerns about false positives disproportionately impacting non-native English speakers.

4. The Ethical Challenge: (How) Do the AI Tools for LLM Detection Maintain Linguistic Privilege?

The question that the discussions on AI tools for LLM detection often overlook is: *What really counts as AI-generated text?* Section 2 defined it as any text created by assigning prompts to the LLM-based application. However, LLM abuse may be more subtle. A typical example would be to skip fact-checking the information we receive from chatbots. Integrating this potentially false information in our (otherwise human-written) article would evade AI tools for LLM detection and pollute our scientific field. As a counter-example, we could collect and check all the research data on our own and use a chatbot as a writing tool afterward. Such a manuscript may get flagged as AI-generated due to suspicious wording, even

[23] F. Russo, E. Schliesser, J. Wagemans, *Connecting Ethics and Epistemology of AI*, "AI & Society" 2023, Vol. 39, pp. 1585–1603, https://doi.org/10.1007/s00146-022-01617-6.

[24] M. Melliti, *Using Genre Analysis to Detect AI-Generated Academic Texts*, "Diá-logos" 2024, Vol. 16, No. 29, pp. 9–27, https://doi.org/10.61604/dl.v16i29.377.

though it would not harm the field. One solution would be to prohibit LLMs even as writing tools. However, by doing so, we would be ridding ourselves of an asset for overcoming the linguistic privilege gap in the scientific community.

To understand the concept of linguistic privilege, it is worth looking into linguistic epistemic injustice, particularly Miranda Fricker's distinction between *testimonial* and *hermeneutic* epistemic injustice.[25] Testimonial injustice is the dismissal of someone's findings because they belong to a linguistically marginalized group. An example would be a researcher discredited due to their foreign accent. Hermeneutic injustice occurs due to the novelty of one's findings, that is, in the lack of the conceptual framework to present them. For instance, we could not talk about gender equality before the concept of gender was introduced. The value of one's contribution does not depend on the language one uses to present it, but non-native English speakers are more susceptible to both hermeneutic and testimonial linguistic epistemic injustice.[26] Conversely, being linguistically privileged means a low likelihood of marginalization based on one's native language.

Depending on the circumstances, AI tools can both mitigate and reinforce the disparity between the linguistically privileged and marginalized members of the scientific community. Reliance on LLM-based applications to improve writing style would make the manuscript more approachable and alleviate the linguistic barrier. However, if another AI tool wrongly flagged the paraphrased text as AI-generated, it would harm the international researchers' chance of publishing. In that case, AI tools would widen the gap between native and non-native speakers. Some factors may influence the risk of false positives, although we do not offer an exhaustive list of LLM-detection technologies, nor do we claim that they *will* flag anyone's work as AI-generated. The following examples just illustrate how technological achievements that work for native English speakers could cause damage to international researchers.

A study has shown that reliance on Shannon's equitability – a quantitative measure of diversity – was helpful in differentiating between ChatGPT-generated and human-written texts. The biggest indicator was the use of the article "the," commas, and the connective "and." As humans tend to leave out commas, ar-

[25] M. Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press, Oxford 2007.

[26] A. Vučković, V. Sikimić, *How to Fight Linguistic Injustice in Science: Equity Measures and Mitigating Agents*, "Social Epistemology" 2022, Vol. 37, No. 1, pp. 80–96, https://doi.org/10.1080/02691728.2022.2109531.

ticles, and connectives, ChatGPT is diligent about their correct use in sentences.[27] While these findings offer insights into differences between linguistic structures in human writing and LLM formulations, in the context of our discussion, they also explain some of the false positives. For instance, a cautious researcher who pays attention to the articles could be at greater risk than their more relaxed peer, who occasionally omits them. Perhaps even more concerningly, a non-native English speaker may use Grammarly or a similar digital assistance tool and, as a result, end up with more articles, commas, and connectives than their native English-speaking peers. Their paper would have a higher risk of being flagged as AI-generated.

Detection tools that use *n*-grams – sequences of *n* symbols – to compute the likelihood of the next word based on the occurrence of previous words establish their evaluation using the parameters of predictability, probability, and pattern.[28] Linguistic patterns uncover underlying structures in the data, that is, the parts of the language that often occur together. The probability of the next word is informed by patterns and based on *n*-1 words that precede it. Predictability stems from probability and refers to the algorithm's ability to conjecture the next word based on the previous items in the sequence.[29] The main idea behind this technology is that human writing is more creative and less uniform than the sequences of words and sentence structures in AI-generated text. Such reasoning is acceptable, but its accuracy may depend on the author's English fluency. While native speakers create varied sentence structures and use less-known words, non-native speakers may rely on simplified structures and common words. As a result, their manuscripts may seem robotic, repetitive, and predictable, which puts them at additional risk of "sounding" like a chatbot. For example, more than half of the false positives were discovered among the English essays written by Chinese stu-

[27] D. Ljubisavljević et al., *Homogeneity of Token Probability Distributions in ChatGPT and Human Texts*, "International Association for Development of the Information Society" 2023, pp. 207–213.

[28] P. Picazo-Sanchez, L. Ortiz-Martin, *Analysing the Impact of ChatGPT in Research*, "Applied Intelligence" 2024, Vol. 54, p. 4175, https://doi.org/10.1007/s10489-024-05298-0.

[29] M. Bertin et al., *The Linguistic Patterns and Rhetorical Structure of Citation Context: An Approach Using N-Grams*, "Scientometrics" 2016, Vol. 109, pp. 1417–1434, https://doi.org/10.1007/s11192-016-2134-8; D. Hiemstra, *Language Models*, in: *Encyclopedia of Database Systems*, 2018; A. Tremblay, B.V. Tucker, *The Effects of N-Gram Probabilistic Measures on the Recognition and Production of Four-Word Sequences*, "The Mental Lexicon" 2011, Vol. 6, No. 2, pp. 302–324, https://doi.org/10.1075/ml.6.2.04tre.

dents, as opposed to almost none of the US student essays in the same category.[30] The authors attribute these results to the lack of variability and perplexity in the writing of non-native English speakers. To put it simply, AI detection tools have "deemed" their writing too predictable to be human.

Some models that successfully detect AI-generated content based on writing style were trained on articles from the most prestigious academic journals.[31] This begs the question of what would have happened had they been trained on linguistically inferior examples. As we train AI tools on top-tier papers, they may begin to associate human writing with high linguistic proficiency and AI-generated text with low proficiency. What initially seemed like a double-edged sword of false positives and false negatives is, in reality, a multifaceted dilemma. False positives disproportionately affect non-native English speakers, thus further deepening epistemic injustice and deserving a place in the discussion on linguistic privilege in science.

Finally, the same scepticism should extend to our own ability to differentiate between AI and human text. Wrong accusations are a rising problem even without AI tools for LLM detection. One example concerns an acclaimed biologist whose article has been labelled AI-generated – an unpleasant experience she shared in a "Nature" column.[32] The situation would have been even more alarming if the peer reviewer based their assumptions on the results of a seemingly impartial AI tool. We still do not have reliable methods for LLM recognition, whether due to our heuristics or their remarkable ability to mimic human writing. For these reasons, accusations of AI abuse require caution.

5. Linguistic Benefits of LLM-Based Tools

The academic and publishing communities' overt focus on LLM-related dangers has unfairly shifted our attention from the benefits these tools offer. Apart from

[30] W. Liang et al., *GPT Detectors Are Biased against Non-Native English Writers*, "Patterns" 2023, Vol. 4., No. 7, https://doi.org/10.1016/j.patter.2023.100779.

[31] See, e.g., H. Desaire et al., *Distinguishing Academic Science Writing from Humans or ChatGPT with Over 99% Accuracy Using Off-the-Shelf Machine Learning Tools*, "Cell Reports Physical Science" 2023, Vol. 4, No. 6, pp. 3, https://doi.org/10.1016/j.xcrp.2023.101426.

[32] E.M. Wolkovich, *Obviously ChatGPT: How Reviewers Accused Me of Scientific Fraud*, "Nature," 5.02.2024, https://doi.org/10.1038/d41586-024-00349-5.

enabling us to automate repetitive tasks, LLM-based tools provide learning opportunities, especially for non-native speakers, who can use them to improve their English skills. A study on ChatGPT revealed that it could enhance English for Academic Purposes (EAP) among non-native students by enriching their vocabulary and offering writing examples.[33] LLM applications work for other languages too, as research demonstrated that ChatGPT, Bard (Gemini), Bing AI Chat (Copilot), and Claude all helped non-natives write in Chinese, with some of the tools focusing on grammar and others on the overall style and coherence in writing.[34]

Non-native English speakers are more likely to use LLMs for queries in languages other than English compared to their native peers.[35] However, there are limitations to using LLMs for prompts in less-spoken languages, as studies suggest that the non-English output is less accurate and thorough. Perplexity – a conversational search engine with high accuracy in generating responses in English – struggled to generate output in Russian, as it failed to respond to 86 percent of the tested prompts.[36] Another study revealed a disparity between the accuracy and quality of the LLM output in English and Turkish. The results were attributed to the latter being less present in internet sources and, consequently, in the LLM training data.[37] While LLM tools can help non-natives master high-resource languages (such as English and Chinese), speakers of low-resource languages get limited output if they search in their own language. These findings indicate that linguistic disparity mitigation cannot entirely rely on AI and still requires human involvement.

---

33 W. Tang, *Unlocking Second Language Students' Potential: ChatGPT's Pivotal Role in English for Academic Purposes Writing Success*, in: *Proceedings of the 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*, Atlantis Press, 2023, pp. 694–706, https://doi.org/10.2991/978-2-38476-126-5_79.

34 S. Obaidoon, H. Wei, *ChatGPT, Bard, Bing Chat, and Claude Generate Feedback for Chinese as Foreign Language Writing: A Comparative Case Study*, "Future in Educational Research" 2024, Vol. 2, No. 3, pp. 184–204, https://doi.org/10.1002/fer3.39.

35 I.V. Molina et al., *Leveraging LLM Tutoring Systems for Non-Native English Speakers in Introductory CS Courses*, arXiv:2411.02725, https://doi.org/10.48550/arXiv.2411.02725.

36 M. Makhortykh et al., *LLMs as Information Warriors? Auditing How LLM-Powered Chatbots Tackle Disinformation about Russia's War in Ukraine*, arXiv:2409.10697, https://doi.org/10.48550/arXiv.2409.10697.

37 M.G. Ozsoy, *Multilingual Prompts in LLM-Based Recommenders: Performance across Languages*, arXiv:2409.07604, https://doi.org/10.48550/arXiv.2409.07604.

Varun Grover offers an argument in favour of the use of LLMs by non-native English speakers.[38] He sees chatbots primarily as tools that can help authors linguistically improve and paraphrase manuscripts. We cannot eradicate LLM abuse just by relying on AI tools for LLM detection, as it would entail never-ending competition between these technologies. As LLMs become more developed, so will their detecting counterparts, but a mismatch between them will remain. At times, LLMs will advance so rapidly that the detecting tools will not be able to recognize them, and at other times, AI detectors will be too sensitive and flag human-written text as AI-generated. We should, as Grover argues, focus on the distinction between *communication goals* and *innovation goals*. The innovation goals represent the content of research and are the author's full responsibility. Unlike them, the communication goals are concerned only with *how* the research is linguistically presented. We can assign this task to LLM-based tools, as long as we ensure they do not alter the original meaning of our work. Savvas Papagiannidis agrees with Grover regarding linguistic assistance and suggests that LLMs can improve the communication between the scientific community and the general public through rewriting specialist papers in a more approachable manner.[39] Proper use of LLMs would not only warrant that AI-generated texts are not a source of misinformation but could also lead to better dissemination of the scientific findings.

If we go beyond the advantages of LLMs as language assistants, a study has revealed that the addition of Bing AI Chat to academic libraries improves user experience by personalizing literature research.[40] Similarly, LLM-based tools designed specifically for research purposes – such as Elicit and SciSpace – summarize the scientific literature,[41] which allows researchers to quickly find relevant publications. Finally, LLM-based applications can be a step forward in mitigating the disparity of education quality between the Global South and Global North

[38] V. Grover, *How Does ChatGPT Benefit or Harm Academic Research*, section of Y.K. Dwivedi et al., *"So What if ChatGPT Wrote It?"*, op. cit., pp. 32–33.

[39] S. Papagiannidis, *ChatGPT and Its Potential Impact on Research and Publishing*, section of Y.K. Dwivedi et al., *"So What if ChatGPT Wrote It?"*, op. cit., pp. 34–35.

[40] A.J. Adetayo, *Conversational Assistants in Academic Libraries: Enhancing Reference Services through Bing Chat*, "Library Hi Tech News" 2023, ahead of print, https://doi.org/10.1108/LHTN-08-2023-0142.

[41] H. Berrami et al., *Exploring the Horizon: The Impact of AI Tools on Scientific Research*, "Data and Metadata" 2024, Vol. 3, https://doi.org/10.56294/dm2024289.

as, when properly used, they are highly available and cost-efficient tutoring assistants.[42]

Still, there is room for caution in treating LLMs as handy assistants. One research project revealed that ChatGPT and Bard (Gemini) provided correct feedback for concurrent programming students only 50 percent of the time compared to their teachers.[43] Although this inaccuracy can be attributed to the complex nature of the evaluated assignments, it is clear that the extent of tasks we can entrust to LLMs is still narrow. A part of the problem lies in their limitation in formal reasoning and diminished ability to separate relevant information from irrelevant.[44] LLMs create new text by predicting the words based on their usual occurrence, but do not comprehend the meaning behind them.[45] While they generate human-like writing, they still lag behind in logical thinking and do not understand the words the way we do. For these reasons, authors should be cautious when entrusting them with tasks that require problem-solving skills. The caution should extend to assignments that depend on critical thinking – such as argument structure analysis – as LLMs may misinterpret and twist complex ideas.

This may change with the development of reasoning models that are more efficient at problem-solving tasks, such as DeepSeek's R1-Zero and R1.[46] However, this will open a different set of concerns regarding authorship. Currently, we can rely on LLMs for language perfection and literature navigation but not for solving complex problems. Those who engage in academic misconduct using LLMs are still more likely to be caught now than they will be in the future. However, LLMs will eventually become more efficient in critical thinking. Using them to formulate novel ideas and solutions would tamper with innovation goals, and

[42] A. Vučković, V. Sikimić, *Global Justice and the Use of AI in Education: Ethical and Epistemic Aspects*, "AI & Society", Vol. 40, pp. 3087–3104, https://doi.org/10.1007/s00146-024-02076-x.

[43] I. Estévez-Ayres et al., *Evaluation of LLM Tools for Feedback Generation in a Course on Concurrent Programming*, "International Journal of Artificial Intelligence in Education" 2024, Vol. 35, pp. 774–790, https://doi.org/10.1007/s40593-024-00406-0.

[44] I. Mirzadeh et al., *GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models*, arXiv:2410.05229, https://doi.org/10.48550/arXiv.2410.05229.

[45] J. Grindrod, *Large Language Models and Linguistic Intentionality*, "Synthese" 2024, Vol. 204, 71, https://doi.org/10.1007/s11229-024-04723-8; Hannigan et al., *Beware of Botshit: How to Manage the Epistemic Risks of Generative Chatbots*, "Business Horizons" 2024, Vol. 67, No. 5, pp. 471–486, https://doi.org/10.1016/j.bushor.2024.03.001.

[46] D. Guo et al., *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*, arXiv:2501.12948, https://doi.org/10.48550/arXiv.2501.12948.

this misconduct would be much harder to detect. Hence, careful revisions of the manuscripts will be even more necessary in the future.

For now, if publishers allowed moderate use of chatbots, non-native English-speaking researchers could use them alongside traditional editorial services to refine the language.[47] The number of international researchers publishing in prestigious journals could, in the long run, indicate whether the scientific community has embraced the benefits of LLMs. However, we need to be cautious before drawing any conclusions from the sheer number of published papers. LLMs also create fertile ground for academic misconduct which increases the number of publications, like in the case of paper mills – multiple rewritings of the same paper.[48]

The question of how strict LLM policies should be is a matter of trust – whether we put our confidence in peers or the technology, we do not fully understand it, nor can we vouch for its reliability. The argument for putting more faith in our colleagues than AI tools for LLM detection is as epistemological as it is based on goodwill. From an ethical point of view, detection tools will be unproblematic only after we minimize the risk of false positives and ensure that they work equally well for native and non-native English speakers. From the perspective of epistemology, it is rational to give preference to our peers, as they – unlike AI applications – are not a black box. There are standards and procedures for testing claims and findings of other scholars. Ideally, we will manage to (re)establish epistemic trust in the scientific community[49] and approach our peers in the belief that they seek true answers, not instant gratification through reliance on unverified data. The path towards the fair use of LLMs in research, thus, requires broad discussions on responsibility, intellectual honesty, and the risks of relying on unverified data.

47 S.I. Hwang et al., *Is ChatGPT a "Fire of Prometheus" for Non-Native English-Speaking Researchers in Academic Writing?*, "Korean Journal of Radiology" 2023, Vol. 24, No. 10, 952, https://doi.org/10.3348/kjr.2023.0773.

48 G. Kendall, J.A. Teixeira da Silva, *Risks of Abuse of Large Language Models, Like ChatGPT, in Scientific Publishing: Authorship, Predatory Publishing, and Paper Mills*, "Learned Publishing" 2024, Vol. 37, No. 1, https://doi.org/10.1002/leap.1578.

49 W. Torsten, *Epistemic Trust in Science*, "British Journal for the Philosophy of Science" 2013, Vol. 64, No. 2, pp. 233–253, https://doi.org/10.1093/bjps/axs007.

6. Conclusions

LLM-based tools have changed the academic and scientific landscape. Laborious and time-consuming tasks, such as grammar checking and rare-literature searches, can now be assigned to machines, allowing researchers to focus more on intellectual pursuits. At the same time, the level of trust within the scientific community has decreased, as researchers may include AI-generated content in manuscripts. If unsanctioned, this trend could lead to numerous problems – from false authorship claims to unverified and incorrect data in scientific journals. In response, many academic institutions and publishers have banned LLMs to preserve the quality and integrity of research dissemination.

In this study we investigated whether such measures are justified and how their consequences unravel over time, especially for researchers who write in English but are not native speakers. We argue that the question of LLM restriction belongs in the discussion on linguistic privilege. AI detection tools not only report both false negatives and false positives, but non-native English speakers are more vulnerable to the latter due to their lower language proficiency. Labelling someone's paper as AI-generated warrants caution as it might harm their career and contribute to the linguistic privilege gap.

If academic institutions and scientific publishers continue to ban the use of LLMs, we risk forfeiting the benefits these technologies offer. LLM-based tools can help us mitigate linguistic disparity in the scientific community, as they offer learning opportunities, particularly for international researchers, who can use them for translation, paraphrasing, and grammar checking. However, even simple AI-generated essays require checking, as they may contain inaccuracies in terms of content and references. Additionally, these tools may not work as well in low-resource languages, and their reasoning skills are suboptimal. When LLMs improve in solving problems, a new challenge in verifying authorship will arise, as generated content will be even harder to detect.

From the epistemological point of view, the main concern is whether we can accurately distinguish AI-generated and human-written content. Relying on human judgement alone is insufficient, as we often fail to recognize whether LLMs were involved in manuscript writing. Studies that analyse the efficiency of AI tools for LLM detection reveal mixed results. Some of these tools are highly ac-

curate, but we encounter the black box problem. Both LLMs *and* AI tools we use to detect them need to become more transparent to earn our trust.

From an ethical perspective, the focus is on the impact of false positives, especially among international researchers. Relying on the discourse of linguistic epistemic injustice, we explored the concept of linguistic privilege. After that, we analysed some of the technologies in AI detection that contribute to a disproportionately higher rate of false positives among researchers who write in English as a second language.

Addressing the risks posed by LLMs is a task for the whole scientific community. The first step is to acknowledge the ethical and epistemic risk of putting too much trust in either LLMs or AI tools for their detection. We need more research on the differences in linguistic structures that native and non-native English speakers use. This could lead to further development of AI tools for LLM detection so they no longer target non-native speakers disproportionately. Employing these tools alongside human evaluation will help us avoid academic misconduct and maintain an inclusive approach. Finally, we should encourage a broad discussion on the long-term means of maintaining responsibility in science while enjoying the benefits of these technologies.

Bibliography

Adetayo A.J., *Conversational Assistants in Academic Libraries: Enhancing Reference Services through Bing Chat*, "Library Hi Tech News" 2023, ahead of print, https://doi.org/10.1108/LHTN-08-2023-0142.

Aljamaan F., Temsah M.H, Altamimi I., Al-Eyadhy A., Jamal A., Alhasan K., Mesallam T.A., Farahat M., Malki K.H., *Reference Hallucination Score for Medical Artificial Intelligence Chatbots: Development and Usability Study*, "JMIR Medical Informatics" 2024, Vol. 12, No. 1, e54345, https://doi.org/10.2196/54345.

Alkaissi H., McFarlane S.I., *Artificial Hallucinations in ChatGPT: Implications in Scientific Writing*, "Cureus" 2023, Vol. 15, No. 2, e35179, pp. 1–4, https://doi.org/10.7759/cureus.35179.

Aydın Ö., Karaarslan E., *Is ChatGPT Leading Generative AI? What Is beyond Expectations?*, "Academic Platform Journal of Engineering and Smart Systems" 2023, Vol. 11, No. 3, pp. 118–134, https://doi.org/10.21541/apjess.1293702.

Berrami H., Jallal M., Serhier Z., Othmani M.B., *Exploring the Horizon: The Impact of AI Tools on Scientific Research*, "Data and Metadata" 2024, Vol. 3, https://doi.org/10.56294/dm2024289.

Bertin M., Atanassova I., Sugimoto C.R., Lariviere V., *The Linguistic Patterns and Rhetorical Structure of Citation Context: An Approach Using N-Grams*, "Scientometrics" 2016, Vol. 109, pp. 1417–1434, https://doi.org/10.1007/s11192-016-2134-8.

Cheng S.L., Tsai S.J., Bai Y.M., Ko C.H., Hsu C.W., Yang F.C., Tsai C.K., Tu Y.K., Yang S.N., Tseng P.T., Hsu T.W., *Comparisons of Quality, Correctness, and Similarity between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study*, "Journal of Medical Internet Research" 2023, Vol. 25, e51229, https://doi.org/10.2196/51229.

Day T., *A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT*, "The Professional Geographer" 2023, Vol. 75, No. 6, pp. 1024–1027, https://doi.org/10.1080/00330124.2023.2190373.

Desaire H., Chua A.E., Isom M., Jarosova R., Hua D., *Distinguishing Academic Science Writing from Humans or ChatGPT with Over 99% Accuracy Using Off-the-Shelf Machine Learning Tools*, "Cell Reports Physical Science" 2023, Vol. 4, No. 6, pp. 1–11, https://doi.org/10.1016/j.xcrp.2023.101426.

Dwivedi Y.K., Kshetri N., Hughes L., Slade E.L., Jeyaraj A., Kar A.K., Baabdullah A.M., Koohang A., Raghavan V., Ahuja M., Albanna H., *"So What if ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy*, "International Journal of Information Management" 2023, Vol. 71, 102642, https://doi.org/10.1016/j.ijinfomgt.2023.102642.

Elkhatat A.M., Elsaid K., Almeer S., *Evaluating the Efficacy of AI Content Detection Tools in Differentiating between Human and AI-Generated Text*, "International Journal for Educational Integrity" 2023, Vol. 19, 17, https://doi.org/10.1007/s40979-023-00140-5.

Estévez-Ayres I., Callejo P., Hombrados-Herrera M.A., Alario-Hoyos C., Delgado Kloos C., *Evaluation of LLM Tools for Feedback Generation in a Course on Concurrent Programming*, "International Journal of Artificial Intelligence in Education" 2024, Vol. 35, pp. 774–790, https://doi.org/10.1007/s40593-024-00406-0.

Feuerriegel S., Hartmann J., Janiesch C., Zschech P., *Generative AI*, "Business & Information Systems Engineering" 2024, Vol. 66, No. 1, pp. 111–126, https://doi.org/10.1007/s12599-023-00834-7.

Fleckenstein J., Meyer J., Jansen T., Keller S.D., Köller O., Möller J., *Do Teachers Spot AI? Evaluating the Detectability of AI-Generated Texts among Student Essays*, "Computers and Education: Artificial Intelligence" 2024, Vol. 6, 100209, https://doi.org/10.1016/j.caeai.2024.100209.

Fricker M., *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press, Oxford 2007.

Grindrod J., *Large Language Models and Linguistic Intentionality*, "Synthese" 2024, Vol. 204, 71, https://doi.org/10.1007/s11229-024-04723-8.

Guo D., et al., *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*, arXiv:2501.12948, https://doi.org/10.48550/arXiv.2501.12948.

Hannigan T.R., McCarthy I.P., Spicer A., *Beware of Botshit: How to Manage the Epistemic Risks of Generative Chatbots*, "Business Horizons" 2024, Vol. 67, No. 5, pp. 471–486, https://doi.org/10.1016/j.bushor.2024.03.001.

Hao Z., *Deep Learning Review and Discussion of Its Future Development*, "MATEC Web of Conferences" 2019, Vol. 277, 02035, https://doi.org/10.1051/matecconf/201927702035.

Harvard University, *Generative AI Guidance*, URL: https://oue.fas.harvard.edu/faculty-resources/generative-ai-guidance/.

Hiemstra D., *Language Models*, in: *Encyclopedia of Database Systems*, eds. L. Liu, M.T. Özsu, Springer, New York 2018, pp. 2061–2065, https://doi.org/10.1007/978-1-4614-8265-9_923.

Huang L., et al., *A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions*, "ACM Transactions on Information Systems" 2024, Vol. 43, No. 2, 42, https://doi.org/10.1145/3703155.

Hwang S.I., Lim J.S., Lee R.W., Matsui Y., Iguchi T., Hiraki T., Ahn H., *Is ChatGPT a "Fire of Prometheus" for Non-Native English-Speaking Researchers in Academic Writing?*, "Korean Journal of Radiology" 2023, Vol. 24, No. 10, 952, https://doi.org/10.3348/kjr.2023.0773.

Jakesch M., Hancock J.T., Naaman M., *Human Heuristics for AI-Generated Language Are Flawed*, "Proceedings of the National Academy of Sciences" 2023, Vol. 120, No. 11, e2208839120, https://doi.org/10.1073/pnas.2208839120.

Kendall G., Teixeira da Silva J.A., *Risks of Abuse of Large Language Models, Like ChatGPT, in Scientific Publishing: Authorship, Predatory Publishing, and Paper Mills*, "Learned Publishing" 2024, Vol. 37, No. 1, https://doi.org/10.1002/leap.1578.

Lane H., Dyshel M., *Natural Language Processing in Action*, Manning Publications, Shelter Island 2025.

Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J., *GPT Detectors Are Biased against Non-Native English Writers*, "Patterns" 2023, Vol. 4., No. 7, https://doi.org/10.1016/j.patter.2023.100779.

Ljubisavljevic D., Koprivica M., Kostic A., Devedžic V., *Homogeneity of Token Probability Distributions in ChatGPT and Human Texts*, "International Association for Development of the Information Society" 2023, pp. 207–213.

Makhortykh M., Baghumyan A., Vziatysheva V., Sydorova M., Kuznetsova E., *LLMs as Information Warriors? Auditing How LLM-Powered Chatbots Tackle Disinformation about Russia's War in Ukraine*, arXiv:2409.10697, https://doi.org/10.48550/arXiv.2409.10697.

Malik M.A., Amjad A.I., *AI vs AI: How Effective Are Turnitin, ZeroGPT, GPTZero, and Writer AI in Detecting Text Generated by ChatGPT, Perplexity, and Gemini?*, "Journal of Applied Learning and Teaching" 2024, Vol. 8, No. 1, https://doi.org/10.37074/jalt.2025.8.1.9.

Melliti M., *Using Genre Analysis to Detect AI-Generated Academic Texts*, "Diálogos" 2024, Vol. 16, No. 29, pp. 9–27, https://doi.org/10.61604/dl.v16i29.377.

Mirzadeh I., Alizadeh K., Shahrokhi H., Tuzel O., Bengio S., Farajtabar M., *GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models*, arXiv:2410.05229, https://doi.org/10.48550/arXiv.2410.05229.

Molina I.V., Montalvo A., Ochoa B., Denny P., Porter L., *Leveraging LLM Tutoring Systems for Non-Native English Speakers in Introductory CS Courses*, arXiv:2411.02725, https://doi.org/10.48550/arXiv.2411.02725.

Naveed H., Khan A.U., Qiu S., Saqib M., Anwar S., Usman M., Akhtar N., Barnes N., Mian A., *A Comprehensive Overview of Large Language Models*, arXiv:2307.06435, https://doi.org/10.48550/arXiv.2307.06435.

Obaidoon S., Wei H., *ChatGPT, Bard, Bing Chat, and Claude Generate Feedback for Chinese as Foreign Language Writing: A Comparative Case Study*, "Future in Educational Research" 2024, Vol. 2, No. 3, pp. 184–204, https://doi.org/10.1002/fer3.39.

Ozsoy M.G., *Multilingual Prompts in LLM-Based Recommenders: Performance across Languages*, arXiv:2409.07604, https://doi.org/10.48550/arXiv.2409.07604.

Picazo-Sanchez P., Ortiz-Martin L., *Analysing the Impact of ChatGPT in Research*, "Applied Intelligence" 2024, Vol. 54, pp. 4172–4188, https://doi.org/10.1007/s10489-024-05298-0.

Price G., Sakellarios M.D., *The Effectiveness of Free Software for Detecting AI-Generated Writing*, "International Journal of Teaching, Learning and Education" 2023, Vol. 2, No. 6, pp. 31–38, https://doi.org/10.22161/ijtle.2.6.4.

Ray P.P., *ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope*, "Internet of Things and Cyber-Physical Systems" 2023, Vol. 3, pp. 121–154, https://doi.org/10.1016/j.iotcps.2023.04.003.

Roumeliotis K.I., Tselikas N.D., *ChatGPT and Open-AI Models: A Preliminary Review*, "Future Internet" 2023, Vol. 15, No. 6, 192, https://doi.org/10.3390/fi15060192.

Russo F., Schliesser E., Wagemans J., *Connecting Ethics and Epistemology of AI*, "AI & Society" 2023, Vol. 39, pp. 1585–1603, https://doi.org/10.1007/s00146-022-01617-6.

Saarna C., *Identifying Whether a Short Essay Was Written by a University Student or ChatGPT*, "International Journal of Technology in Education" 2024, Vol. 7, No. 3, pp. 611–633, https://doi.org/10.46328/ijte.773.

Tang W., *Unlocking Second Language Students' Potential: ChatGPT's Pivotal Role in English for Academic Purposes Writing Success*, in: *Proceedings of the 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*, Atlantis Press, 2023, pp. 694–706, https://doi.org/10.2991/978-2-38476-126-5_79.

Torsten W., *Epistemic Trust in Science*, "British Journal for the Philosophy of Science" 2013, Vol. 64, No. 2, pp. 233–253, https://doi.org/10.1093/bjps/axs007.

Tremblay A., Tucker B.V., *The Effects of N-Gram Probabilistic Measures on the Recognition and Production of Four-Word Sequences*, "The Mental Lexicon" 2011, Vol. 6, No. 2, pp. 302–324, https://doi.org/10.1075/ml.6.2.04tre.

Vishwanath P.R., et al., *Faithfulness Hallucination Detection in Healthcare AI*, in: *Artificial Intelligence and Data Science for Healthcare: Bridging Data-Centric AI and People-Centric Healthcare*, 2024.

Vučković A., Sikimić V., *Global Justice and the Use of AI in Education: Ethical and Epistemic Aspects*, "AI & Society", Vol. 40, pp. 3087–3104, https://doi.org/10.1007/s00146-024-02076-x.

Vučković A., Sikimić V., *How to Fight Linguistic Injustice in Science: Equity Measures and Mitigating Agents*, "Social Epistemology" 2022, Vol. 37, No. 1, pp. 80–96, https://doi.org/10.1080/02691728.2022.2109531.

Walters W.H., *The Effectiveness of Software Designed to Detect AI-Generated Writing: A Comparison of 16 AI Text Detectors*, "Open Information Science" 2023, Vol. 7, No. 1, 20220158, https://doi.org/10.1515/opis-2022-0158.

Weber-Wulff D., Anohina-Naumeca A., Bjelobaba S., Foltýnek T., Guerrero-Dib J., Popoola O., Šigut P., Waddington L., *Testing of Detection Tools for AI-Generated Text*, "International Journal for Educational Integrity" 2023, Vol. 19, No. 1, pp. 26–65, https://doi.org/10.1007/s40979-023-00146-z.

Wolkovich E.M., *Obviously ChatGPT: How Reviewers Accused Me of Scientific Fraud*, "Nature," 5.02.2024, https://doi.org/10.1038/d41586-024-00349-5.

Wood P., *Oxford and Cambridge Ban ChatGPT over Plagiarism Fears but Other Universities Embrace AI Bot*, "The iPaper," 23.02.2023, URL: https://inews.co.uk/news/oxford-cambridge-ban-chatgpt-plagiarism-universities-2178391.

Ye H., Liu T., Zhang A., Hua W., Jia W., *Cognitive Mirage: A Review of Hallucinations in Large Language Models*, arXiv:2309.06794, https://doi.org/10.48550/arXiv.2309.06794.