https://doi.org/10.37708/ezs.swu.bg.v24i3.5

“YOUR TEXT LOOKS LIKE FROM GPT”: AN ETHICAL STUDY OF AUTHORSHIP IN THE AGE OF LARGE LANGUAGE MODELS

Anastasiya Karnaukh

University of Prešov, Slovakia
Email: karnaukhanastasiya@gmail.com

ORCID ID: 0009-0007-3287-2718

ABSTRACT: The proliferation of large language models (LLMs) has significantly changed academic writing practices and ways of assessing textual authorship. Unlike early instruments of language support, modern generative models demonstrate a high ability to reproduce texts according to the norms of academic discourse. Against this background, the problem of automatic detection of AI-generated content and the associated risks of erroneous labelling are widely discussed in the academic environment. At the same time, existing research mainly focuses on the misuse of AI, while situations in which texts created by a human are perceived as “too similar to AI” remain insufficiently studied. However, considerably less attention has been paid to the opposite, but no less problematic, phenomenon: the misidentification of human authorship, which arises from the stylistic convergence of normative academic writing and LLM output texts. This study assumes that such cases cannot be adequately explained solely by the technical limitations of detection tools, and require consideration within a broader epistemic and ethical context. This qualitative study aims to analyze online narratives of academic authors who report instances in which their texts were perceived as “written by AI” despite the absence of generative tools during their production. The analysis zooms in on how the authors themselves interpret such situations, what arguments they use in defense of human authorship, and how stylistic similarity to LLMs influences trust, academic identity, and a sense of professional legitimacy. The obtained results contribute to a deeper understanding of the limits of algorithmic evaluation and highlight the need for more careful use of detection practices in the academic environment.

KEYWORDS: academic authorship, generative AI, AI detection, epistemic injustice, academic writing, discourse analysis, ethical misrecognition, large language models

1. Introduction

The advent of large language models (LLMs) such as GPT (Generative Pre-trained Transformer) in all and its successors has transformed academic writing practices across disciplines. Unlike earlier tools that offered basic grammar and formatting assistance, contemporary generative models produce a text that increasingly aligns with disciplinary norms, rhetorical structures, and stylistic expectations through large-scale pre-training on academic and professional corpora and deep learning reinforced by the human feedback (Brown et al., 2020; OpenAI, 2023). In academic contexts, this alignment has led to widespread adoption of such tools for drafting literature reviews, refining arguments, and structuring manuscripts across various disciplines in the humanities (van Dis et al., 2023).

Yet this integration has also generated a paradox: a high-quality human writing, especially when composed in accordance with conventional academic norms, can be erroneously classified as AI-generated, thus creating a form of authorship misrecognition that challenges traditional conceptions of authorship and academic integrity. Empirical evaluations of AI detection systems consistently demonstrate that texts adhering closely to academic genre conventions are disproportionately vulnerable to false positive classification (Giray, 2024). For example, detection tools and scholarly perceptions sometimes flag rigorously written, edited human texts as "AI-like," thereby calling into question the reliability of algorithmic classification and the epistemic frameworks that drive judgements about textual authenticity within institutional decision-making contexts such as assessments, peer reviews, or misconduct investigations.

However, this paradox reflects deeper ethical and epistemological concerns. Generative AI models, trained on massive corpora of human academic texts, encapsulate norms of structure, clarity, and argumentation, which are the very qualities esteemed in scholarly writing. As a result, texts that adhere closely to those norms may resemble generative output in the eyes of detection tools or even scholarly peers, leading to false positives that may undermine confidence in human authors' contributions.

Research on AI detectors demonstrates serious limitations in current technology: tools often misclassify even minimally AI-polished or professionally edited human writing as AI-generated, and struggle to differentiate between nuanced degrees of AI involvement or between native and non-native academic language (Almohaimeed et al, 2025; Hastewire Team, 2025).

Furthermore, this emerging phenomenon has implications that extend beyond technical detection accuracy. Contemporary academic ethics frameworks have primarily focused on issues such as plagiarism, fabrication, and undisclosed use of generative systems. However, the emerging challenge of stylistic misrecognition, where well-edited human writing is perceived as "AI-generated", raises the need to reconsider concepts of authorship, trust, and academic identity in the humanities and adjacent knowledge-producing fields. These concerns intersect with broader debates about the limits of automated evaluation, the sociotechnical boundaries of authorship, and the potential of algorithmic tools to introduce epistemic injustice, whereby authors are wrongly discredited due to reliance on computational classifications rather than substantive human adjudication.

In this qualitative study, we aim to analyze online narratives from researchers, students, and academics who have encountered accusations or perceptions that their human-written texts look like GPT-ese. By analyzing these narratives, we aim to illuminate how academic communities interpret stylistic similarity to AI output, how such interpretations shape ethical judgments, and what this signifies for the future of academic writing and integrity in the age of generative AI.

2. Theoretical framework

In the classical academic tradition, authorship is considered an individual's original creation and intellectual input, involving originality, ownership, and adherence to genre standards, reflecting an individual's or group's intellectual contribution and responsibility (Pimple, 2012; ICMJE, 2025), where authors are expected to uphold the integrity of their work, which is essential for fostering trust within the scientific community.

In the humanities, authorship is additionally associated with the concept of authorial voice: a rhetorical and stylistic way of constructing knowledge through which the subject positions himself or herself in scientific discourse (Łozińska, 2023). However, even this "individuality" is formed within institutionally fixed genre and stylistic frameworks (Jönsson, 2006; Lillis et al., 2015; Chauhan, 2022). Studies of academic writing show that "a high-quality" scientific text is the result of institutional normalization: genre training, editorial practice, peer review, and stylistic standardization. In other words, academic writing is initially focused not on individual expressiveness but on the reproducibility of norms. It is this normativity that can become a key vulnerability in the context of generative AI.

2.1 Generative language models and academic normativity simulation

As previously mentioned, large language models are taught on a large-scale corpus of texts, including scientific articles, dissertations, educational materials, and reviewed publications. As emphasized by Durt and Fuchs (2024), LLMs do not possess intentionality or understanding but demonstrate a high ability to reproduce patterns of language. So, it means that such models should not be considered as "authors" but as infrastructural simulators of normative discourse.

Several studies show that LLM output texts are characterized by high coherence, correct use of grammar and stylistic "neutrality," which can make them practically indistinguishable from edited human texts, especially in disciplines with a high degree of writing standardization (Skowron and Bączkowska, 2023; Muñoz-Ortiz et al., 2023). Thus, the stylistic similarity between human and machine writing can be understood as a structural consequence of the very logic of academic discourse, rather than a side effect of technology.

2.2 Detection and methodological failure of algorithmic suspicion

Attempts to solve the problem of authorship through automatic AI detection encounter several limitations. Modern detectors rely on specific indicators (stylometry, watermarking, and neural classifiers, among others) (Ganie, 2025). Empirical studies demonstrate high variability in detection results and a significant number of false positives, especially when analyzing academic texts that have undergone professional editing or translation (Giray, 2024; Gotoman et al., 2025). In response, AI developers publicly acknowledge that detectors are recognized as insufficient to serve as conclusive evidence of authorship (OpenAI, 2023). Consequently, an accusation of using AI based on a stylistic impression or the results of a detector may be interpreted as a form of algorithmically amplified suspicion, which can have direct ethical implications.

2.3 Misrecognition of authorship and epistemic injustice

The erroneous classification of human text as AI-generated can be productively conceptualized through the lens of epistemic injustice (Fricker, 2007). In this case, we are talking about testimonial injustice: the subject is denied trust as a carrier of knowledge on the basis of biased and technologically mediated criteria. In an academic context, such misrecognition can have certain consequences: it may result in professional and emotional consequences, institutional sanctions, increased anxiety, and self-censorship, especially among early researchers and non-native speakers of the dominant academic language.

2.4 Theoretical novelty of the present study

Existing research in AI and academic ethics predominantly focuses on the misuse of AI. This study offers a different perspective: an analysis of possible ethical damage arising not from the use of AI but due to the stylistic convergence of human and machine writing. Thus, the concept of ethical misrecognition of authorship is introduced as an independent category of academic ethics, which requires a rethinking of assessment procedures, trust, and author's identity in the era of generative AI.

Importantly, while the literature outlines the structural and ethical conditions under which misrecognition may arise, it remains largely silent on how such situations are experienced, narrated, and ethically interpreted by academic authors themselves - a gap this study seeks to address.

3. Methodology

This study was carried out within the framework of a qualitative interpretive paradigm focused on the analysis of meanings, narrative structures, and ethical assessments formed by participants in academic discourse. The choice of a qualitative approach is determined by the nature of the phenomenon under study: misrecognition of authorship is not a measurable behavioral deviation but a socially and discursively constructed experience rooted in the interpretations, emotional reactions, and normative expectations of the subjects of the academic field (Denzin and Lincoln, 2018).

Methodologically, the study relies on integrated discursive-thematic analysis, combining elements of reflective thematic analysis, critical discourse analysis (CDA), and analysis of online narratives as a form of "practical knowledge" of the academic community. Thematic analysis is used as a primary analytical tool to identify persistent semantic patterns and repetitive patterns of experience, while the elements of critical discourse analysis are used at the interpretative stage for theoretical understanding of the identified topics in the context of institutional norms, epistemic power, and ethical regimes for assessing authorship. This differentiation of analytical levels avoids methodological eclecticism and ensures the logical coherence of the analysis. Such a combination of methods corresponds to recommendations for studies of digital and institutional discourses, in which individual narratives simultaneously function as subjective evidence and as a reflection of structural asymmetries of power and norms of academic knowledge (Fairclough, 2013; Braun and Clarke, 2021).

The empirical corpus of the study is formed on the basis of publicly available online materials in which members of the academic community describe situations in which their texts claimed to be written without the use of generative AI were perceived as AI-generated. These sources include the following:

- social platforms of professional communication (Reddit: r/academia, AskAcademia, LinkedIn);
- academic forums and personal blogs of researchers;
- public discussions on the X platform (Twitter);
- analytical and reflective publications on Medium;
- comments under the materials on the use of AI detectors in education and scientific publication.

The choice of online narratives as empirical material is due to the fact that digital platforms function as a space for articulating non-institutionalized academic experience, which is rarely recorded in official reports and regulatory documents but plays a key role in shaping collective perceptions, anxieties, and ethical expectations within the academic community.

It is important to note that the study is not aimed at verifying the actual absence of the use of AI but focuses on the analysis of the subjective experience of accusation, its interpretation, and ethical significance. Thus, the epistemological object of the analysis is not the ontological status of the text but the social reaction to the alleged AI-authorship and its consequences for academic identity.

To ensure analytical relevance, the corpus was formed using a target (purposive) sampling strategy recommended for studies of rare, conceptually loaded, and weakly institutionalized phenomena (Tajik et al., 2024).

Inclusion criteria:

1. The author's statement that the text was created without the use of generative AI.
2. Indication that the text was:
   a. accused of AI generation,
   b. characterized as “GPT-like”,
   c. or recognized as AI-generated based on automatic detectors.
3. Description of the institutional or social context (educational, research, publication).
4. The presence of the author's reflection on the consequences of the accusation and its ethical assessment.

Exclusion criteria:

1. texts without indicating the reaction of the institute, teacher, editorial board, or academic community;
2. fragmentary comments that do not have a narrative structure and reflective depth.

The size of the body was determined not by a predetermined number of analysis units but by the principle of theoretical saturation, in which the further

inclusion of materials did not lead to the emergence of new analytically significant topics (Nascimento et al., 2018).

The primary stage of the analysis was carried out using reflective thematic analysis as interpreted by Braun and Clarke (2006, 2021), which is focused on identifying semantic patterns without data reduction to predetermined categories. The analysis included the following steps:

1. Multiple reading of the corpus to form theoretical sensitivity and immersion in the context.
2. Open coding with fixation:
   a. forms of accusation and suspicion,
   b. arguments used to legitimize suspicion,
   c. emotional reactions of the authors,
   d. strategies to protect authorship and professional identity.
3. Grouping codes into thematic clusters.
4. Iterative validation of topics for internal connectivity, analytical saturation, and conceptual relevance.

The choice of a reflective approach is due to the fact that the purpose of the study is not the quantitative representativeness of the topics, but their interpretive significance in the ethical and epistemic context of academic knowledge.

At the second stage, elements of critical discourse analysis were applied, aimed at identifying regulatory prerequisites for "real" and "permissible" authorship; binary oppositions "human/machine"; and institutional voices (AI detectors, teachers, journal editors) as sources of epistemic power and control. This level of analysis makes it possible to consider individual online narratives not as isolated complaints but as manifestations of broader discursive modes of assessing knowledge and trust that form modern academic ethics (Fairclough, 2013).

The study is based solely on **publicly available data**, which is in line with international guidelines for internet research (AOIR, 2019; British Psychological Society, 2021). However, additional measures have been taken:

- full anonymization of users;
- exclusion of identifying parts.

This approach is consistent with the principle of *ethical reflexivity*, according to which the formal publicity of data does not exempt the researcher from liability for potential harm to participants.

5. Findings

The analysis of the corpus of online narratives made it possible to identify several interrelated thematic patterns that reflect how participants in academic and quasi-academic discourse experience and interpret situations in which their texts, claimed to be written without the use of generative AI, are perceived as "GPT-like" or "AI-generated". Taken together, these patterns indicate not single misunderstandings, but the **formation of new modes of text interpretation**, within which traditional markers of academic competence can be rethought as potential indicators of unreliable authorship.

5.1. Rethinking academic normativity as a source of suspicion

One of the central topics was the participants' feeling of **normative inversion**, in which the qualities previously unambiguously associated with "good academic writing" begin to function as grounds for suspicion in the new conditions. In the narratives analyzed, structure, logical consistency, grammatical correctness, and formal tonality are repeatedly described as factors provoking accusations of the use of AI.

So, the author of the post in Reddit, who has many years of experience in academic writing, describes his style as "the only way he can write," emphasizing his excessive formality, logic, and reliance on sources. Despite this, he regularly encounters suspicions of using AI when publishing. Some more examples of authors complaining about their texts to be accused of looking like AI-generated can be found: "*Clean, concise and correct writing, which is very resemblant to AI and penalising me.", "I had someone ask me why every reply of mine sounded like AI, because they were complete paragraphs!", "What I find funny is how AI Detection will show a paragraph being AI, but I can change a word like "lurch" to "swayed" in a paragraph and it will show 100% human instead of any AI." or "I have taught writing, and have been a journalist and author for decades. I've noticed the same thing. AI detectors assume humans are terrible writers."* These examples illustrate the gap between the author's intentionality and the interpretation of the text, where a certain disciplinary sequence of words ceases to unambiguously indicate human origin.

A similar experience can be traced in the narrative of the authors who received a review with the comments *"Your conclusion reads as though it was written by ChatGPT.", "This is a really solid base, but could you do a rewrite with a more human touch? It sounds a little like it was written by ChatGPT."* In this case, the stylistic assessment is interpreted not as a neutral editorial remark but as a doubt of professional legitimacy; the author emphasizes that the clarity and accuracy of its expression are "stock in trade of the scholar," especially in the humanities. Thus, this indicates a certain **shift in the interpretation framework**, in which disciplinary correctness ceases to be an unambiguous indicator of human authorship.

5.2. "GPT-like" as a new euphemistic form of delegitimization

The second stable thematic pattern is due to the fact that the formula *"sounds like ChatGPT"* or *"looks AI-generated" in* the narratives of the participants acquires a meaning that goes beyond the scope of stylistic comparison. In some cases, it functions as a **euphemistic form of delegitimization**, allowing you to express distrust of the author without resorting to direct accusations of plagiarism or dishonesty.

This is especially evident in reflective texts, where it is emphasized that such remarks are perceived as a doubt not only in the origin of the text but also in the intellectual viability of the author. One participant notes: *"To say a piece reads like ChatGPT is no longer a neutral comparison; it suggests the work is formulaic, derivative, or suspiciously fluent."* Here, the "like-GPT phrase" acts as a marker

of secondary and insufficient originality, even in the absence of any evidence of the use of AI.

Within the framework of discursive analysis, such statements can be interpreted as **a manifestation of the new hermeneutics of suspicion**, in which the assessment of the text shifts from the content and argumentation to the alleged conditions of its production. At the same time, the participants themselves emphasize that the criteria for this suspicion are still very vague and situational.

5.3. Algorithmic imaginary and pseudo-technical criteria of "humanity"

A separate group consists of narratives in which participants reconstruct and criticize popular ideas about how it is supposedly possible to "recognize" AI text. These statements regularly mention lists of so-called "forbidden" words (*furthermore, moreover, consequently among others*), punctuation elements (em dash, semicolon), as well as symmetrical syntactic constructions.

So, in the LinkedIn-thread, one of the authors describes how his academic work was discredited because of the use of a word that the reviewer considered "too conceptual for a person." Another participant ironically notes that he was forced to advise students to **save grammatical anomalies** when writing academic works in order to demonstrate the "humanity" of the text. These examples indicate that the expectations associated with AI detection are beginning to backfire on writing practices, forming new norms of self-revision that are not always consistent with traditional academic standards.

5.4. Linguistic inequality and a shift in the focus of stigma

The topic related to the experience of multilingual authors and researchers from non-English-speaking contexts was particularly pronounced. In their narratives, erroneous AI attribution is interpreted as a transformation of already existing forms of linguistic inequality. A post called *"I'm Kenyan. I Don't Write Like ChatGPT. ChatGPT Writes Like Me"* suggests how a formal academic style formed within a specific educational tradition can be reinterpreted as "unnatural." The author emphasizes that the characteristics that AI detectors interpret as signs of machineness (predictability, structuredness) were purposefully cultivated in the process of the author taught at higher institutions. The participants themselves emphasize that the need to "prove the humanity" of the text increases their anxiety and self-censorship.

5.5. Emotional and professional dimension of misrecognition

Finally, a significant thematic dimension was the emotional and existential impact of such accusations. Participants describe not only irritation but also deeper consequences: loss of confidence, doubt about the right to their own voice, and a sense of the need to justify the quality of their work. Indicative is the episode with a student concerned about what his essay detector classified as "100% AI-generated" though having been written manually. In this case, we see how the purpose of academic writing shifts from demonstrating individual experience to the need to prove the very fact of the human origin of the text.

6. Discussion

The results of this study allow us to consider the misrecognition of authorship as a complex socio-discursive phenomenon that arises at the intersection of such areas as academic normativity, algorithmic tools and institutional assessment practices. The revealed normative inversion is correlated with the studies of academic writing, emphasizing its institutionally regulated nature (Hyland, 2008), since the academic style is initially focused on the reproducibility of norms, and not on individual expressiveness. Findings show that this reproducibility can make human texts highly vulnerable to erroneous AI attribution. From the point of view of the theory of language models (Bender et al., 2021), this convergence is not an anomaly for the reason that LLMs trained on already existing academic texts reproduce the statistical “average” of academic discourse. Consequently, the stylistic similarity between human and machine writing can be a structural effect of academic communication, not a consequence of its degradation.

The interpretation of the “GPT-like” formula as a socially charged label is consistent with critical discourse analysis (Fairclough, 2013). Findings allow us to suggest that such expressions may function as mechanisms of indirect power operating within academic evaluation regimes. Importantly, this power is not exercised by a single actor, but is distributed across institutional roles and technological artefacts. Peer reviewers, editors, and misconduct committees act as authorized interpreters of textual legitimacy, while AI detection tools and notions of “GPT-like style” function as epistemic supports that legitimize suspicion without requiring evidential justification. In this configuration, authors are seen to be positioned as epistemically subordinate: they are compelled to defend not the quality of their argument, but the presumed conditions of textual production. From a critical discourse analytical perspective, labels such as “sounds like ChatGPT” operate as discursive shortcuts that naturalize distrust. They may result in shifting evaluative authority away from argumentation and toward stylistic surface features, thereby enabling institutional workers to question authorship while remaining insulated from the burden of proof. If institutionalized, such discursive practices may risk consolidating a regime of anticipatory compliance, in which authors may tend to adapt their writing to avoid suspicion rather than to meet disciplinary standards, reinforcing asymmetries of power between evaluators and evaluated subjects.

The interpretation of misrecognition through the concept of testimonial injustice (Fricker, 2007) is particularly evident. In many narratives, the author's credibility is undermined not by the content of the argument, but by the supposed mode of text production. We found out that this injustice can affect multilingual authors. It's important to note that this cannot be considered a radically new phenomenon, but rather a transformation of existing regimes of exclusion: the criteria of “bad English” are being replaced by suspicions of a “too good” writing.

Finally, the study’s results can support arguments about the methodological and ethical limitations of automated AI detection (corresponds with Ibrahim,

2023; Mitchell et al., 2023). Now, the expectation of possible detection may significantly influence writing practices, prompting authors to adapt not to the established writing norms, but to presumed algorithmic criteria.

7. Limitations

This study has a number of limitations that need to be considered when interpreting the results. First of all, the analysis is based on a qualitative corpus of online narratives, which implies self-selection of participants and potential bias towards more emotionally charged or reflective cases. Although such data are methodologically sound for the study of subjective interpretations, moral assessments and discursive practices, they do not allow quantitative generalizations about the prevalence of the phenomenon of erroneous AI attribution in the academic community as a whole.

Secondly, the sources used are predominantly represented by English-language platforms and academic contexts, which limits the cultural and institutional variability of the practices analyzed. Given that the norms of academic writing, evaluation regimes, and attitudes toward generative technologies differ significantly between national and disciplinary traditions, the findings should be viewed as contextually conditioned rather than universal.

The third limitation is the focus of the study on narratives of perception and interpretation, rather than on empirical verification of the technical characteristics of detection systems. The study deliberately did not aim at an experimental assessment of the accuracy or errors of specific AI detectors, but considered their indirect impact through discursive practices of suspicion and legitimation. As a result, the conclusions do not relate to the effectiveness of the algorithms themselves, but to the social and ethical consequences of their use and interpretation in the academic environment.

8. Final takeaways

This study shows that in the context of the widespread use of generative language models, academic writing is involved in the transformation of trust, interpretation and authorial responsibility regimes. Analysis of the phenomenon of erroneous AI attribution of human texts suggests that the emerging tensions cannot be adequately explained within the framework of existing models of academic ethics, focused mainly on the misuse of technologies. On the contrary, the results indicate the need to expand the analytical field in the direction of considering situations in which ethical damage occurs without the use of AI and is due to stylistic convergence between normative academic discourse and statistically reproducible language patterns.

In theoretical terms, the article contributes to the rethinking of authorship as an epistemic category in the context of algorithmically mediated communication. It is shown that authorship can no longer be considered solely as a function of intentional control over the text if the criteria for its recognition are shifted towards surface indicators. Thus, the introduced concept of ethical misrecognition of authorship allows us to conceptualize a new type of vulnerability of academic subjects that arises at the intersection of the institutional

normalization of writing and the limitations of the algorithmic imaginary. This contribution expands existing discussions in the field of AI ethics and academic integrity, shifting the focus from the individual responsibility of the author to the conditions of production and assessment of trust in the academic field.

The practical implications of the study are multilevel. For editors and scientific publishers, the results may help to draw attention to the need to revise policies related to the detection and interpretation of the intended use of AI. In particular, the stylistic similarity with generative writing cannot be considered as a reliable basis for ethical conclusions or editorial decisions without contextual and meaningful analysis. This implies the rejection of implicit legitimation of automatic detectors and a clearer distinction between issues of text quality, transparency of the process of its creation and actual violations of academic norms.

For teachers and evaluators in higher education, the study points to the possibility of rethinking academic writing assessment practices in the LLM era. Shifting the focus from argumentation, critical thinking and substantive understanding to attempts to identify the "humanity" of the text carries the risk of undermining educational goals and increasing anxiety among students.

For authors and researchers, the results highlight the structural nature of emerging risks. Erroneous AI attribution is not an individual failure or a consequence of unethical behavior, but rather reflects the systemic tension between the requirements of normative academic clarity and new modes of suspicion. In this context, strategies for the ethical use of AI and authorship self-defense should not be seen as an attempt to "prove the humanity" of a text, but as part of a broader discussion about transparency, contextualization, and institutional accountability.

More broadly, the study indicates that the key challenge posed by generative AI in the humanities is not to replace human writing with machine writing, but to narrow the permissible forms of academic legitimacy. If normative clarity and stylistic discipline begin to be interpreted as signs of inauthenticity, this calls into question the very basis of academic communication as a space for knowledge sharing rather than a demonstration of "recognizable humanity." The answer to this challenge does not require strengthening control, but a more reflective and theoretically grounded approach to assessing authorship, trust and ethical responsibility in the academic environment.

REFERENCES:

Almohaimeed, S., Almohaimeed, S., Jari, M., Alobaid, K.A., & Alotaibi, F. (2025). AI Text Detectors and the Misclassification of Slightly Polished Arabic Text. *Journal of Big Data*. https://doi.org/10.48550/arXiv.2511.16690

AOIR (2019). Internet research: Ethical guidelines 3.0. *Association of Internet Researchers.* Accessible at: https://aoir.org/reports/ethics3.pdf. [Assessed 15.12.2025]

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pp. 610–623.

Braun, V. & Clarke, V. (2006). Using Thematic Analysis in Psychology. Qualitative Research in *Psychology*, 3, 77-101.https://doi.org/10.1191/1478088706qp063oa

Braun, V. & Clarke, V. (2021) *Thematic Analysis: A Practical Guide***.** Sage, London.

British Psychological Society. (2021). *Ethics guidelines for internet-mediated research*. Accessible at: https://cms.bps.org.uk/sites/default/files/2022-06/Ethics%20Guidelines%20for%20Internet-mediated%20Research_0.pdf. [Assessed 15.12.2025]

Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.

Chauhan, P. (2022). Fundamentals of Academic Writing: A Literature Review. *Journal of NELTA*, 27, 161-180. 10.3126/nelta.v27i1-2.53201.

Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE handbook of qualitative research* (5th ed.). SAGE.

Durt, C. & Fuchs, T. (2024). *Large Language Models and the Patterns of Human Language Use*. 10.4324/9781003312284-7.

Fairclough, N. (2013). *Critical Discourse Analysis: The Critical Study of Language*. Routledge. https://doi.org/10.4324/9781315834368

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford, online edn. Oxford Academic. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001.

Ganie, A. (2025). Uncertainty in Authorship: Why Perfect AI Detection Is Mathematically Impossible. 10.48550/arXiv.2509.11915.

Giray, L. (2024). The Problem with False Positives: AI Detection Unfairly Accuses Scholars of AI Plagiarism. *The Serials Librarian*, 85(5–6), 181–189. https://doi.org/10.1080/0361526X.2024.2433256

Gotoman, J.E., Luna, H.L., Sangria, J.C., Santiago Jr, C.S., & Barbuco, D.D. (2025). Accuracy and Reliability of AI-Generated Text Detection Tools: A Literature Review. *American Journal of IR 4.0 and Beyond*, 4(1), 1-9.

Hyland, K. (2008). Genre and Academic Writing in the Disciplines. *Language Teaching*, 41, 543-562. https://doi.org/10.1017/S0261444808005235

OpenAI (2023). Jan Hendrik Kirchner, Lama Ahmad, Scott Aaronson, Jan Leike. New AI classifier for indicating AI-written text January 31, 2023. Available at: https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/ [accessed 25.12.2025]

Jönsson, S. (2006). On academic writing. European Business Review. 18. 479-490. 10.1108/09555340610711102.

Hastewire Team (2025). *Study Reveals AI Detectors' False Positives on Non-Native Writers*. Accessible at: https://hastewire.com/blog/study-reveals-ai-detectors-false-positives-on-non-native-writers. [Assessed 20.12.2025]

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

ISMJE (2025). *Defining the Role of Authors and Contributors*. Accessible at: https://www.icmje.org/recommendations/browse/roles-and-

responsibilities/defining-the-role-of-authors-and-contributors.html#four. [Assessed 23.12.2025]

Lillis, T. & Harrington, K. & Mitchell, S. & Lea, M. (2015). *Working with Academic Literacies: case studies towards transformative practice*. Parlor Press

Łozińska, A. (2023). Authorial voice in writing: A literature review. *Social Sciences and Humanities Open*, 8(1), 100550. https://doi.org/10.1016/j.ssaho.2023.100550

Mitchell, E; Lee, Y.; Khazatsky, A.; Manning, Ch. & Finn. Ch. (2023). Detectgpt: Zero-shot machine-generated text detection using probability curvature. *International Conference on Machine Learning (ICML)*, pp. 24950– 24962. PMLR.

Muñoz-Ortiz, A., Gómez-Rodríguez, C., and Vilares, D. (2023). Contrasting Linguistic Patterns in Human and LLM-Generated Text. *arXiv.Org*, *abs/2308.09067*. https://doi.org/10.48550/arxiv.2308.09067

Nascimento, L. & Souza, T. & Oliveira, I. & de Moraes, J. & Aguiar, R. & Silva, L. (2018). Theoretical saturation in qualitative research: an experience report in interview with schoolchildren. *Revista brasileira de enfermagem*, 71. 228-233. 10.1590/0034-7167-2016-0616.

Pimple, K. D. (2012). Authorship in Scientific and Academic Research. *CORE Issues in Professional and Research Ethics*, 1(4). ISSN: 2160-8784

Skowron, D. & Bączkowska, A. (2023). Assessment of various GPT Models versus the human text: a quantitative analysis of lexis and cohesion. Forum Filologiczne Ateneum. https://doi.org/10.36575/2353-2912/1(11)2023-09

Tajik, O. & Golzar, J. & Noor, S. (2024). Purposive Sampling. 2. 1-9. 10.22034/ijels.2025.490681.1029.

van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. H. (2023). ChatGPT: Five priorities for research. *Nature*, 614(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

Copyright © 2026 Karnaukh. This is an open-access publication distributed under the terms of the Creative Commons Attribution 4.0 International Licence CC BY 4.0