

# Has AI written this paper? – A case study in „plagiarism detection”

Anna Małgorzata Kamińska

ORCID: 0000-0002-8935-9735

*Institute of Culture Studies*

*Faculty of Humanities*

*University of Silesia in Katowice*

---

## Abstract

**Purpose/Thesis:** The paper evaluates the use of GenAI content detectors in the context of verifying academic integrity. The main thesis is that content detectors cannot distinguish AI-generated text from human-authored content that has been stylistically edited only by large language models. Consequently, high detection scores do not constitute reliable evidence of plagiarism.

**Approach/Methods:** A case study method was employed. A biography of a fictional character was created by a human and subsequently edited using ChatGPT and Claude, with precise instructions that limited the AI to a purely editorial role. All versions were analysed using the ZeroGPT detector.

**Results and conclusion:** The original text scored a 0% probability of being AI-generated, while the versions edited by ChatGPT and Claude achieved 96.7% and 100%, respectively. Detectors identify stylistic features characteristic of well-edited texts, rather than the actual origin of the intellectual contribution.

**Practical implications:** The findings indicate the need to revise academic policies in light of AI detectors. Institutions should promote transparent rules for using AI as a supporting tool rather than relying on unreliable technologies that can lead to unfair assessments.

**Originality/Value:** The work provides empirical evidence of the unreliability of AI detectors in the context of plagiarism detection, establishes a clear ethical boundary between using AI as a tool and academic dishonesty, and promotes a conscious approach to new technologies in academia.

## Keywords

AI content detection. Case study. Editorial tools. Ethics in science. Generative artificial intelligence (GenAI). Large language models (LLM). Plagiarism. Scholarly communication.

*The text received on the 28th of September 2025.*

---

## 1. Introduction

“Has AI written this paper?” This question and similar ones, while reading academic papers, have currently been asked by more and more people: reviewers and editors of scientific magazines, as well as lecturers who examine the works of their students. It results from the fact that the emergence and rapid popularisation of generative artificial intelligence (GenAI), especially large language models (LLMs), are among the most significant moments in the modern history of technological development. Their influence on the creation, processing and consumption of content is undisputed, and affects nearly every field of life, including, in a particular and multidimensional way, the academic community (Tillmanns *et al.*, 2025; Yusuf *et al.*, 2024). LLMs, such as the GPT family from OpenAI or Claude from Anthropic, ceased to be a technological curiosity and became advanced tools that can serve as assistants in research, programming, data analysis, and, above all, working with text.

However, such a technological transformation became the source of a fundamental conflict. On the one hand, scientists and authors received unprecedented support during the writing process, enabling them to improve the clarity of their works and, consequently, to make the content easier to understand and to reduce the likelihood that they will be misinterpreted by potential readers. It has a positive influence on the process of popularising scientific knowledge and thereby serves the goodwill of the academic community. On the other hand, the ease with which LLMs can generate broad, linguistically consistent texts on any topic has raised justified concerns about academic dishonesty, the integrity of scientific research, and even a new form of plagiarism (Francis *et al.*, 2025). To address such threats, numerous programmes have emerged to detect content allegedly generated by GenAI. It led to a phenomenon that may be called “moral panic”, where the usage of a tool that detects if the text was edited by an LLM seems to be a simple and effective remedy to the complex ethical problem.

However, this paper puts forth a thesis that relying uncritically on such assumptions and solutions is not only ineffective but even harmful to the process of scholarly communication. The following study questions have been asked in this work:

- (1) Can a text which has essentially been written fully by a human being be incorrectly qualified as a work of AI only after correcting it linguistically via an LLM?
- (2) Where is the boundary between using AI ethically as an advanced editorial tool and academic dishonesty?
- (3) What are the potential consequences of relying on AI detectors for students, teachers, researchers, reviewers, editors of magazines and the entire scientific ecosystem?

The main argument of this paper is a statement that the usage of an LLM as a “digital editor” is a natural and logical evolution of many tools that have assisted

us in writing for a long time, and is not plagiarism itself. At the same time, the current detection systems are not naturally able to distinguish generated content (where AI is the essential author) and edited one (where AI is only used as an editorial assistant), which poses a significant risk of false accusations and may restrain the development of good practices in the scholarly communication. To prove this claim, the follow-up to this paper presents a case study that illustrates the problem under controlled conditions.

## 2. Theoretical background and literature review

Technological assistance in the writing process is not a new phenomenon. Its history dates back to the first text editors equipped with simple spell-checking mechanisms. The next step was the implementation of grammar correctors, which could analyse basic sentence structures (Oakman, 1994). A breakthrough came with tools such as Grammarly, which, thanks to analysing large-scale data collections, offered stylistic suggestions, synonyms, and proposals to improve the text's fluency (Barrot, 2022). From this point of view, LLMs are the next, though much more advanced, link in the same evolutionary chain, if we look at it from both functional and technological perspectives. Their ability to identify contexts, linguistic nuances, and complex narrative structures makes them well-suited to the role of an advanced editorial assistant, not just a passive proofreader. Treating them as a completely new, unacceptable category of tools ignores this historical context of the technological development. Voices opting for the complete exclusion of adopting this technology, though justified, especially among people who fail to understand it and are afraid of it as a result, are in contradiction with the general idea of progress. This phenomenon is repeatable in general, and the humanity has experienced it with the appearance of every landmark innovation. However, after the stage of scepticism and fear, questions are usually asked not whether but how to use such a technology for the good of the civilisation and the balanced development of the humanity.

In light of rising controversies in the area of using GenAI tools in educational and research practices, it becomes necessary to precisely determine where the role of GenAI as an admissible tool ends, and where its usage related with academic dishonesty begins, which requires reconsidering the classical definition of plagiarism in the era of LLMs (Luo, 2024).

The classical definition of plagiarism, deeply rooted in academic ethics, refers to the appropriation of the ideas or data of others and presenting them as one's own, without proper reference to the source (Helgesson & Eriksson, 2015). Therefore, the crucial element is the origin of the intellectual contribution: the content of the paper on its merits. When the author uses an LLM to correct punctuation,

reformulate an unclear sentence or find a better synonym, there is no appropriation of the ideas of others. The author remains fully responsible for every fact, argument and claim contained in the paper. Its author is the source of knowledge, and the language model only assists with putting this knowledge into a more accessible language form. Within this context, the role of the LLM is functionally identical to the role of a human editor or a friend asked to read and give their opinion on the legibility and clarity of the text.

Unfortunately, despite the ethical purity of such usage of GenAI, the academic community quickly addressed this phenomenon, focusing not on the intent of the author but on the technological signature of the text, which directly led to the development of online services offering the possibility to detect AI-generated texts, being rapidly developed. However, numerous studies and technical reports show fundamental problems with their effectiveness and reliability. It was found that these tools tended to generate many false positives, marking human-written texts as AI-generated. This problem is particularly severe in case of texts that are structured by nature and use a specialist, repeatable language, which is visible in scientific, legal and medical texts.

Furthermore, these detectors often identify a feature called low perplexity, which indicates a high level of predictability of consecutive words in a sentence, as “artificial”. This is a distinguishing feature for LLM-generated texts, but, at the same time, this is a desired property of a clear and well-written text, as clarity and unambiguity are necessary in science to precisely convey results and replications of studies, as well as to build knowledge. It means that the scientific text should be predictable syntactically (so it should have low perplexity), so that the reader can focus on the merits, not on interpreting linguistic intricacies. As a result, the more effort the author puts into writing fluently, consistently and comprehensibly, the bigger risk is that their text will be marked by a detector. It creates a paradoxical situation in which tools intended to uphold quality may, in fact, punish the author for pursuing it.

From another point of view, treating the results of AI detectors indiscriminately as proof of dishonesty immediately brought a counter-reaction; namely, the development of tools intended to “humanise” a text. The purpose of these tools is to deliberately disrupt the fluency of a text generated by GenAI. They are tasked with implementing elements with high perplexity (simulating mistakes, inconsistencies, or excessively sophisticated vocabulary) to produce consistent, logically predictable phrases, which aims to reduce the indicator of detection by AI detectors. This process, fully automatic and large-scale, lets one instantly “spoil” content generated by artificial intelligence in order to simulate human editorial imperfections. The existence of such a category of tools proves that, under current technological circumstances, it is no longer an issue to have automatically generated content masked from detectors, also automatically and on a large scale, which additionally exposes the uselessness of the latter as an arbitrator of academic ethics.

### 3. Methodology of the case study

To empirically verify the research thesis, an experimental case study has been conducted. Such a choice of research method is justified by the need to achieve reliability and precision through a controlled course of the study, while at the same time ensuring a deeper, qualitative look typical of such a study. The ability to thoroughly analyse the phenomenon under strictly controlled conditions facilitates an ambiguous illustration and precise exemplification of the problems with the unreliability of AI detectors.

For the study, a text was written: a “raw” biography of a fictional character, Polish jazzman, Bronisław “Borsuk” (Badger) Wierzbicki, supposedly active during the post-war thaw. It can be assumed, though, that this profile has been written by a hypothetical historian based on barely accessible, non-digitalised sources, therefore, absent online as well (archives, interviews, press of the period). Thus, undoubtedly, the merits of this biography, being a synthesis of the specialist knowledge, reflect the intellectual contribution of the hypothetical author. This author, while creating a work-in-progress note, treated it as a research draft, not paying due attention to the fluency of the descriptions at this point, and even allowing themselves to make orthographic and stylistic errors absent-mindedly. The choice of such a fictional character as Bronisław “Borsuk” Wierzbicki was a deliberate methodological measure aimed at demonstrating that models can support editing a text on a topic unknown to them. What is important, the first name and surname of the character were deliberately duplicated with the identity of another well-known online fictional character (meaning he is potentially present in the models’ training data): a character from the TV series *Father Matthew* (*Ojciec Mateusz*). Such a measure aims to conduct the most convincing test: to prove that an LLM’s knowledge in a given field can be completely omitted with precise prompts, using the model only as a helpful editorial assistant and placing particular pressure on its linguistic skills. At the same time, it constituted an additional test of the models’ ability to follow prompts accurately and avoid a tendency to hallucinate or include external information in the text.

The draft prepared by the hypothetical researcher is presented below. As it was mentioned, it includes purposely introduced deficits in terms of grammar, stylistics and orthography, which was to simulate the natural structure of a draft, created by a human being, before its final edition:

**Base (human) text:** Bronisław “Borsuk” Wierzbicki was one of the most enigmatic jazz musicians during the thaw and 60es. He was respected as a tenor saxophonist and as a composer, and gained the status of a legendary titan of improvisation who combined the American hard bop with the subtle Slavic melancholy. It gave him an unofficial name of the “Polish Coltrane”. He had a band “Echa Północy” (Echoes of the North) with which he was giving concerts on Jazz Jamboree Festival and was successful. Besides, he was giving concerts outside the iron curtain with restrictions which was rare. The top of his career ended with a plot twist that was sudden and happened in 1968. While having

a tour in France, unclear incidents took place after which he did not decide, as many his friends, to stay in the west but returned to the country just to completely withdraw from public life within several next months and quit making music, and begin to work in warsaw as an RTV repairment technician. The reasons for such withdrawal are not known, and rumours on the topic concern being blackmailed by the Security Service and hypotheses on a spiritual crisis that was to be deep. They have remained a topic for speculations of musik historians until today. Which casts a shadow on his earlier legacy, which was splendid.

Before the tasks involving editing the text were carried out, the knowledge bases of the language models chosen for the study, i.e., ChatGPT and Claude Opus 4.1, had been verified. The goal of this procedure was to confirm that the models do not possess knowledge of a fictional character, Bronisław “Borsuk” Wierzbicki, which was an important element of the research project. The answers obtained from both models, confirming that they have no knowledge of the alleged jazzman (see Figure 1 and Figure 2), have guaranteed that their role was limited to linguistic and editorial skills, not to generate the merits.

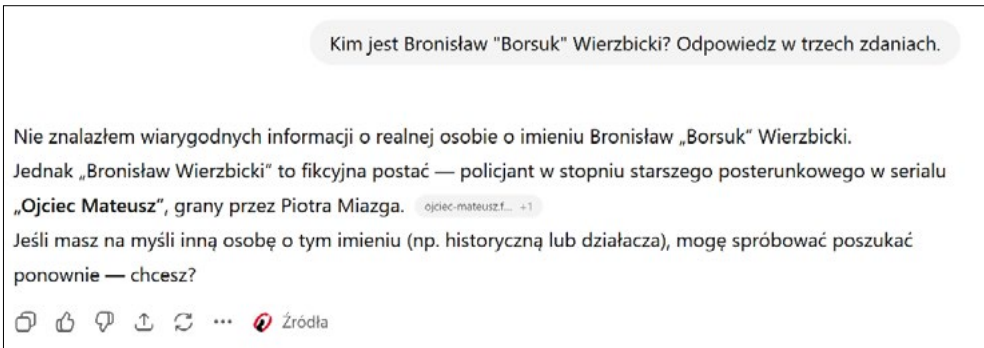


Figure 1. The answer of ChatGPT to the question about Bronisław “Borsuk” Wierzbicki.

Source: own elaboration

(Translation of figure 1):

Who is Bronisław “Borsuk” Wierzbicki? Answer me in three sentences.

I have not found reliable information on a real person called Bronisław “Borsuk” Wierzbicki. However, “Bronisław Wierzbicki” is a fictional character. He is a Senior Police Constable in a TV series *Father Matthew*, played by Piotr Miazga. If you mean a different person having this name (e.g., a historical one or an activist), I can try to search them again. Do you want it?



Figure 2. The answer of Claude Opus 4.1 to the question about Bronisław "Borsuk" Wierzbicki.

Source: own elaboration

(Translation of figure 2):

Who is Bronisław "Borsuk" Wierzbicki? Answer me in three sentences.

Based on available search results, the most probable Bronisław "Borsuk" Wierzbicki is a Senior Police Constable from the TV series Father Matthew, who appears regularly in the third, fourth, fifth, and sixth seasons, as well as sporadically in the seventh. Before he began working at a police station, he had worked in the traffic police. In one of the episodes, he moves to a police station in Kielce because of his mother's illness.

Then, the base text, containing deliberately introduced deficits, served as the input data for ChatGPT and Claude Opus 4.1, with the option to use the formal style. In both cases, an identical prompt was applied, which is presented below.

The prompt conveyed to the models: As a proficient Polish language proofreader working in a scientific publishing house, edit the source text below, so that 1) it will lose no information in the target text, 2) orthographic, grammar, punctuation and stylistic errors will be corrected, 3) the general flow of the content will be corrected, so that the text will be of a formal nature and easy to read by a potential reader at the same time, 4) adjust the sentence order and move information among sentences to improve the fluency of the utterance, 5) do not use any formatting (bold, italics, numbered/point lists, markdown); it is to be a normal text, separated with paragraphs at the most.

For ChatGPT, the following edited version was received after processing the entry text:

**Text edited by ChatGPT:** Bronisław “Borsuk” Wierzbicki was one of the most enigmatic jazz musicians of the thaw period and the 60s. He was respected both as a tenor saxophonist and a composer. He gained a legendary reputation as a master of improvisation who combined the American hard bop with the subtle Slavic melancholy. This is the reason why he was called the “Polish Coltrane” since then. He was a leader of a band “Echa Północy” (Echoes of the North), with whom he gave performances on Jazz Jamboree Festival, being significantly successful. He also performed, though with restrictions, on both sides of the iron curtain, which was rare those days.

However, the top of his career was ended with a sudden and unexpected twist in 1968. During a tour in France, unclear incidents took place after which Wierzbicki, unlike many his friends who decided to migrate, returned to the country. Within several next months, he completely withdrew from public life, quitted making music, and began to work in Warsaw as an RTV repairment technician. The reasons for such an extreme decision have remained unknown. Diverse speculations have circulated, including a presumed blackmail from the Security Service and hypotheses on his deep spiritual crisis. They have remained a subject matter of discussions among music historians until today, casting a shadow on the legacy of this artist, though it has unquestionably remained outstanding.

For Claude Opus 4.1, the editorial process fruited with the following version of the text:

**Text edited by Claude Opus 4.1:** Bronisław “Borsuk” Wierzbicki belonged to the circle of the most enigmatic jazz musicians of the thaw period and the 60s. As a tenor saxophonist and composer, he gained recognition and the status of a legendary master of improvisation who was able to combine the American hard bop with the subtle Slavic melancholy. Thanks to such a unique style, he obtained an unofficial title of the “Polish Coltrane”.

Being the head of a band “Echa Północy” (Echoes of the North), he gave performances on the prestigious Jazz Jamboree Festival, being significantly successful. Moreover, he performed outside the iron curtain with restrictions, which was indeed rare those days. However, the top of his career was ended with a sudden plot twist in 1968. During a tour in France, unclear incidents took place after which Wierzbicki, unlike many his friends, did not decide to stay in the West but returned to the country. Within several next months, the artist completely withdrew from public life, quitted making music, and began to work in Warsaw as an RTV repairment technician. The reasons for such extreme withdrawal have remained unknown, and rumours circulating include hypotheses on being blackmailed by the Security Service and theses regarding his deep spiritual crisis. These speculations have been a subject of interest of music historians until today, casting a shadow on the earlier, splendid legacy of this artist.

Ultimately, the author’s raw draft and both texts corrected by the LLMs have been assessed by ZeroGPT.

## 4. Results

The edition of the base text carried out by ChatGPT benefited from significant improvements in fluency, consistency, and compliance with the standards of formal language. Those changes were purely stylistic and grammatical, and they fully satisfied the prompt, without introducing new information on the merits.

The model corrected formal and orthographic mistakes, eliminating spelling errors such as “60es” (changed into “60s”), “in warsaw” (into “in Warsaw”) and “musik” (into “music”) as well as improving the incorrect punctuation. One of the crucial interventions was improving consistency and reducing redundancies. The model

eliminated such structures, for example, by changing the sentence regarding the top of his career: “The top of his career ended with a plot twist that was sudden and happened in 1968” into the brief: “However, the top of his career was ended with a sudden and unexpected twist in 1968.” The LLM improved the lexical consistency by replacing repetitions and general statements with a more precise vocabulary. It also changed an informal phrase: “He had a band «Echa Północy»” into a more formal “He was a leader of a band «Echa Północy».” The stylistic and syntactic changes included dividing several long, complex sentences into shorter units, thereby increasing the text’s legibility and transparency. Syntactic inversion was introduced, sentence structures were diversified, and grammar and collocation errors were corrected. To sum up, the version edited by ChatGPT is semantically identical to the base text but shows significantly higher linguistic quality, featuring reduced perplexity (greater predictability and fluency) and higher syntactic consistency, in compliance with the standards of a professional publication.

Claude Opus 4.1 presented as high ability to improve the quality of a text base as ChatGPT, but operated on slightly different, though complementary, editorial strategies. As the previous model, Claude eliminated all the orthographic errors (“60es” into “60s”, “in warsaw” into “in Warsaw”, “musik” into “music”) as well as corrected punctuation and grammar. In the scope of syntax, the model significantly improved the structure of the opening paragraph. For example, the excessively condensed first sentence of the base text has been converted into a more fluent sequence: “Bronisław ‘Borsuk’ Wierzbicki belonged to the circle of the most enigmatic jazz musicians of the thaw period and the 60s” and “As a tenor saxophonist and composer, he gained recognition and the status of a legendary master of improvisation...”

In terms of stylistics, Claude Opus 4.1 focused on improving lexical formality and precision. It replaced less formal phrases, such as “He had a band «Echa Północy» with which he was giving concerts...” with “Being the head of a band «Echa Północy», he gave performances...” The model eliminated excessive adjectives, e.g., by converting “with a plot twist that was sudden and happened in 1968” into a brief “with a sudden plot twist in 1968.” Also, the improvement of the ending clause is visible, where “rumours on the topic concern being blackmailed by the Security Service and hypotheses on a spiritual crisis that was to be deep” was changed into a more formal: “rumours circulating include hypotheses on being blackmailed by the Security Service and theses regarding his deep spiritual crisis.” To sum up, the edition made by Claude Opus 4.1 does not change the text on its merits, but achieves a very high level of fluency and consistency, typical for professional editing.

Although both models achieved the same goal, which was eliminating formal errors and improving the fluency of the text to the publishing standard, a detailed analysis lets us observe subtle differences between their editorial strategies. Both ChatGPT and Claude Opus 4.1 successfully corrected orthographic and grammatical errors and removed pleonasms, thereby reducing the text’s perplexity.

In several places, ChatGPT seems to pursue greater brevity and stylistic elegance, introducing, e.g., a phrase about a “sudden and unexpected plot twist” instead of the original, more lengthy description, or changing a “legendary titan of improvisation” into a “legendary reputation as a master of improvisation.” In turn, Claude Opus 4.1 presented a slightly greater tendency to maintain the original semantic structure, making softer changes but introducing more formal and encyclopaedic phrases at the same time, such as “he belonged to the circle of the most enigmatic jazz musicians” and “he obtained an unofficial title.”

Ultimately, while both results are almost perfect, the text generated by Claude Opus 4.1 makes an impression of being slightly more systematised and polished in terms of syntax and vocabulary, achieving the “ideal” level of fluency that is characteristic for machine-assisted generating. Owing to the slight advantage in scope of lexical consistency and formal tone, it can be stated that in this specific editorial assignment, Claude Opus 4.1 is slightly ahead of its competitor, delivering a text of an exceptionally low perplexity.

The subjective opinion, indicating the higher fluency and optimisation of the text by Claude Opus 4.1, has been empirically confirmed by measures carried out with usage of an AI detector, ZeroGPT. The base text, being an author draft, has been correctly identified by the tool, with a 0% probability that it was generated by artificial intelligence (see Figure 3). In this way, it has been verified that the linguistic signature of the entry text was unambiguously human.

On the other hand, although both LLMs, as instructed, fulfilled only the editorial role and made no contribution to the merits, their influence on the text’s syntax and fluency immediately prompted the detector to react. The version edited by ChatGPT received 96.7% of the detection rate (see Figure 4) while the text processed by Claude Opus 4.1 was indicated with the maximum result of 100% as AI-generated (see Figure 5).

These results have unambiguously proven that the detector does not identify the intellectual origin of the content (contribution to the merits) but only focuses on stylistic and syntactic features of the text (low perplexity), which are the effect of linguistic optimisation.

## 5. Discussion

The results of the case study provide an unambiguous answer to the first research question: a text that has been written entirely by a human may be incorrectly classified as a work of AI only after linguistic correction via an LLM. Therefore, the basic conclusion is that the current tools for detecting content generated by artificial intelligence are unable to distinguish between intellectual contribution and stylistic editing. These algorithms, instead of analysing the sources of their

merits, focus on identifying language patterns that correlate with high editing quality: fluency, consistency, and low perplexity. Paradoxically, these features are not only a hallmark of machine-generated texts but also the goal of each author who wants to create a clear and professional scientific text. Consequently, the more effectively the author, assisted with an advanced tool, will correct their text, the more probable is that it will be incorrectly marked as a work of a machine.

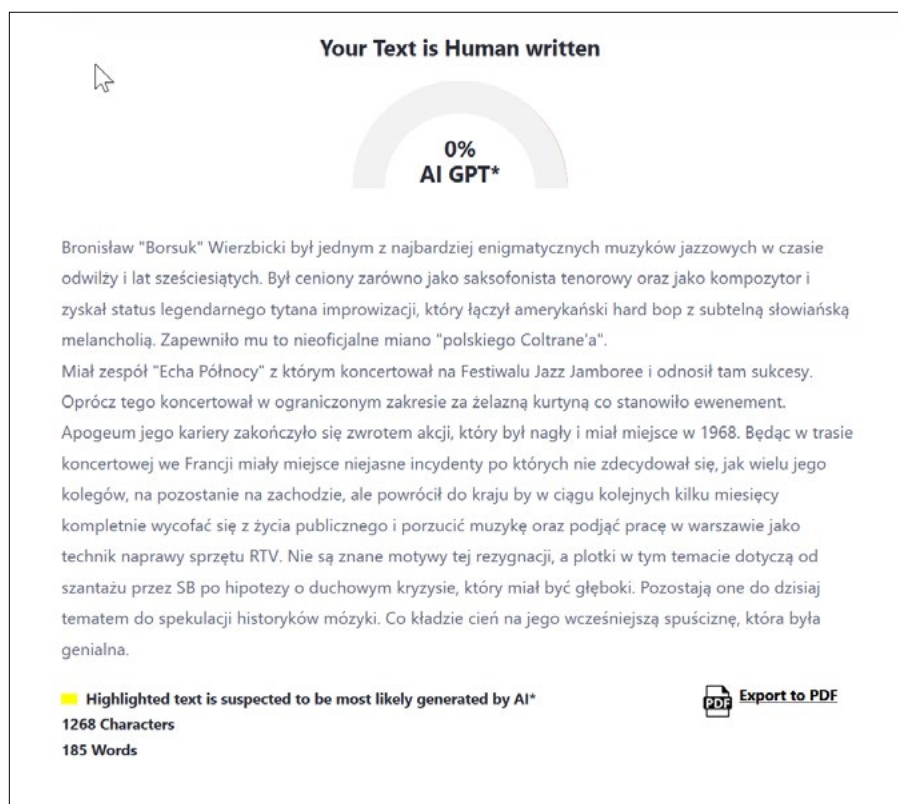



Figure 3. The result of detection of the human text by ZeroGPT.

Source: own elaboration

It directly leads to the second research question, concerning the boundary between the ethical usage of AI and academic dishonesty. Based on the analyses conducted, it should be stated that this boundary lies not in the mere use of the tool itself but in the intent and manner in which it is used. The usage of an LLM as an advanced proofreader whose function is similar to that of a human editor is completely ethical. Academic dishonesty begins when the author assigns responsibility for the intellectual contribution, i.e. generating theses as well as arguments, analysing data and drawing up conclusions, to the machine, and then attributes the

authorship to themselves. Therefore, the origin of the ideas is crucial, not a tool used to improve them in terms of language.

### Your Text is AI/GPT Generated



96.68%  
AI GPT\*

Bronisław „Borsuk” Wierzbicki był jednym z najbardziej enigmatycznych muzyków jazzowych okresu odwilży i lat sześćdziesiątych. Ceniono go zarówno jako saksofonistę tenorowego, jak i kompozytora. Zyskał legendarną opinię mistrza improwizacji, łączącego amerykański hard bop z subtelną, słowiańską melancholią. To właśnie dzięki temu zaczęto określać go mianem „polskiego Coltrane’a”.

Był liderem zespołu „Echa Północy”, z którym występował na Festiwalu Jazz Jamboree, odnosząc tam znaczące sukcesy. Występował również, choć w ograniczonym zakresie, po obu stronach żelaznej kurtyny, co w tamtych czasach stanowiło ewenement.

Apogeum jego kariery zakończyło się jednak nagłym i nieoczekiwanym zwrotem w roku 1968. Podczas trasy koncertowej we Francji doszło do niejasnych incydentów, po których Wierzbicki, w przeciwieństwie do wielu kolegów decydujących się na emigrację, powrócił do kraju. W ciągu kilku następnych miesięcy całkowicie wycofał się z życia publicznego, porzucił muzykę i podjął pracę w Warszawie jako technik naprawy sprzętu RTV.

Motywy tej radykalnej decyzji pozostają nieznane. Krążyły różnorodne spekulacje – od domniemanego szantażu ze strony Służby Bezpieczeństwa po hipotezy o głębokim kryzysie duchowym. Do dziś pozostają one przedmiotem dyskusji historyków muzyki, rzucając cieniem na spuściznę artysty, choć ta bezsprzecznie pozostaje wybitna.

Humanize Text

**Make Your Text Human With Undetectable AI**

Highlighted text is suspected to be most likely generated by AI\*

1370 Characters

181 Words



 **Export to PDF**

Figure 4. The result of detection of the ChatGPT's text by ZeroGPT.

Source: own elaboration

### Your Text is AI/GPT Generated



100%  
AI GPT\*

Bronisław "Borsuk" Wierzbicki należał do grona najbardziej enigmatycznych muzyków jazzowych okresu odwilży i lat sześćdziesiątych. Jako saksofonista tenorowy oraz kompozytor zyskał uznanie i status legendarnego mistrza improwizacji, który potrafił łączyć amerykański hard bop z subtelną słowiańską melancholią. Dzięki temu unikatowemu stylowi otrzymał nieoficjalny tytuł "polskiego Coltrane'a".

Na czele zespołu "Echa Północy" występował na prestiżowym Festiwalu Jazz Jamboree, odnosząc tam znaczące sukcesy. Ponadto koncertował w ograniczonym zakresie za żelazną kurtyną, co stanowiło w tamtych czasach prawdziwy ewenement. Apogeum jego kariery zakończyło się jednak nagłym zwrotem akcji w 1968 roku. Podczas trasy koncertowej we Francji doszło do niejasnych incydentów, po których Wierzbicki, w przeciwieństwie do wielu swoich kolegów, nie zdecydował się na pozostanie na Zachodzie i powrócił do kraju.

W ciągu kilku kolejnych miesięcy artysta całkowicie wycofał się z życia publicznego, porzucił muzykę i podjął pracę w Warszawie jako technik naprawy sprzętu RTV. Motywy tej drastycznej rezygnacji pozostają nieznanne, a krążące plotki obejmują hipotezy od szantażu ze strony Służby Bezpieczeństwa po teorie dotyczące głębokiego kryzysu duchowego. Spekulacje te do dziś stanowią przedmiot zainteresowania historyków muzyki, rzucając cień na wcześniejszą, genialną spuściznę artysty.

Humanize Text

Make Your Text Human With Undetectable AI

Highlighted text is suspected to be most likely generated by AI\*

1386 Characters  
180 Words


 [Export to PDF](#)

Figure 5. The result of detection of the Claude's text by ZeroGPT.

Source: own elaboration

Relying uncritically on AI detectors, which do not recognise this boundary, results in numerous negative consequences for the entire scientific ecosystem, which is the answer to the third research question. For students and researchers, it means making a ridiculous choice between pursuing linguistic perfection and the risk of false accusations. For teachers, reviewers, and editors, the temptation to use a detector as a fast and seemingly objective arbiter may lead to unfair assessments and the rejection of valuable papers. From a broader point of view, scientific institutions,

by creating policies based on unreliable indicators, do not build a culture of trust but promote an atmosphere of suspiciousness, which leads to the pathological “cat-and-mouse game” with tools intended for “humanising” texts.

An additional problematic layer is imposed on this image, and it is connected with the ethical and legal status of the detection services themselves. A lot of those platforms operate under unclear terms of service, not guaranteeing what happens to the content sent to them. Forwarding unpublished scientific papers, student essays, or research data to external commercial entities poses a significant risk of infringing copyright, compromising research secrecy, or simply losing control over the intellectual property. The situation becomes even more complicated when the verifier (e.g., a lecturer or reviewer) is not the author of the text, which may lead to infringement of third parties’ rights.

## 6. Conclusions

This research, despite its pilot nature, provides empirical evidence that the high probability that a text was AI-generated cannot serve as reliable proof of academic dishonesty. The use of a large language model as a digital editor is an authorised and ethically permissible support in the creative process, and the final responsibility for the content, its authenticity, and the originality of the thoughts is always borne by a human being. Therefore, the main message of this paper is the necessity of fundamentally changing the paradigm: from the question “Has artificial intelligence written it?” to “Does the author bear full responsibility for this paper?” Promoting transparency about the tools used, together with education on their capabilities and limitations, is a significantly more constructive approach than creating policies based on unreliable, easily manipulable detection technologies.

It should be emphasised that, although the paper focuses only on a narrow, editorial use of GenAI, the problem should be perceived in a much broader context. Fuelling the phenomenon of “**GENfear**” in academic communities, which is a fear of adopting generative artificial intelligence, poses an actual risk for the competitiveness of the Polish science on the international stage.

While the Polish institutions still debate basic principles, the scientific world does not wait. The admission of GenAI to the broad spectrum of research assignments is becoming the norm. Leading global scientific magazines are withdrawing from prohibitive regulations, focusing on transparency (Wang & Zhao, 2024): they only require that authors declare the scope in which AI has been used, similar to how the contributions of particular authors in papers written by several people are declared. Moreover, leading foreign universities have actively developed comprehensive policies on the ethics of using AI, precisely setting boundaries between the permitted, the support for the application of this technology in both research

and didactics, and the abuse of ethical boundaries. Even the crucial academic infrastructure integrates those changes: the suppliers of scientific databases, such as Scopus and Web of Science, directly implement GenAI technologies into their search interfaces.

Meanwhile, this delay is also a problem in our own backyard: we have observed significant progress even at the national level, in both commercial sectors and public administration. It is visible in the development and ongoing improvement of national large language models (such as Bielik and PLLuM), as well as in the development of policies and pilot integrations of those technologies with governmental infrastructure, at both the central and local levels. Exclusion and stigmatisation of GenAI by the Polish scientific community at the moment may have far-reaching consequences. The science in free-market and democratic systems is to serve society, not to serve itself. If the scholarly communities are going to pretend that the elaborated technological turn does not apply to them, the social justification for their existence and their ability to serve as a catalyst for national progress will be called into question.

However, strictly returning to the editorial use of GenAI, the potentially worst effects of this moral panic should be noted. The reason is that only the suppliers of online services that “humanising” texts, which allow AI detectors to be cheated, are the actual beneficiaries of this situation. We should be aware that the long-term effects of extensive use of such platforms may be highly negative. The texts “spoilt” deliberately to simulate human imperfections will gradually feed into future training data, which may lead to the degradation of large language models and the erosion of the quality of natural language in digital space in the long term. It seems that an effective medicine to the above-mentioned “GENfear” is to make policies of integrating GenAI technologies with didactic and research processes, both in an ethical and active way, by particular academic units.

## Bibliography

- Barrot, J. S. (2020). Integrating technology into ESL/EFL writing through Grammarly. *RELC Journal*, 53(3), 764–768. doi: 10.1177/0033688220966632
- Francis, N. J., Jones, S., Smith, D. P. (2025). Generative AI in higher education: balancing innovation and integrity. *British Journal of Biomedical Science*, 81(14048), 1–9. doi: 10.3389/bjbs.2024.14048
- Helgesson, G., Eriksson, S. (2015). Plagiarism in research. *Medicine, Health Care and Philosophy*, 18, 91–101. doi: 10.1007/s11019-014-9583-8
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the “originality” of students’ work. *Assessment & Evaluation in Higher Education*, 49(5), 651–664. doi: 10.1080/02602938.2024.2309963
- Oakman, R. L. (1994). The evolution of intelligent writing assistants: trends and future prospects. In: *Proceedings Sixth International Conference on Tools with Artificial*

- Intelligence. TAI 94* (233–234). Institute of Electrical and Electronics Engineer,. doi: 10.1109/TAI.1994.346488
- Tillmanns, T., Filho, A. S., Rudra, S., Weber, P., Dawitz, J., Wiersma, E., Dudenaite, D., Reynolds, S. (2025). Mapping tomorrow's teaching and learning spaces: A systematic review on GenAI in higher education. *Trends in Higher Education*, 4(1), 2, 1–26. doi: 10.3390/higheredu4010002
- Wang, Y., Zhao, L. (2024). Toward the transparent use of Generative Artificial Intelligence in academic articles. *Journal of Scholarly Publishing*, 55(4), 467–484. doi: 10.3138/jsp-2023-0053
- Yusuf, A., Pervin, N., Román-González, M., Noor, N. M. (2024). Generative AI in education and research: A systematic mapping review. *Review of Education*, 12(2), e3489. doi: 10.1002/rev3.3489
- 

## Czy ten artykuł napisała sztuczna inteligencja? – studium przypadku „wykrywania plagiatu”

### Abstrakt

**Cel/Teza:** Artykuł ocenia użycie detektorów treści generowanych przez GenSI w kontekście weryfikacji uczciwości akademickiej. Główna teza głosi, że narzędzia te nie potrafią odróżnić tekstu wygenerowanego przez SI od treści opracowanych przez człowieka poddanych jedynie redakcji stylistycznej przez duże modele językowe, przez co wysokie wyniki detekcji nie stanowią wiarygodnego dowodu plagiatu.

**Koncepcja/Metody badań:** Zastosowano studium przypadku. Stworzono biografię fikcyjnej postaci autorstwa człowieka, którą następnie zredagowano przy użyciu ChatGPT i Claude z precyzyjnymi poleceniami ograniczającymi SI do roli redakcyjnej. Wszystkie wersje przeanalizowano detektorem ZeroGPT.

**Wyniki i wnioski:** Tekst oryginalny uzyskał wynik 0% prawdopodobieństwa bycia wygenerowanym przez SI, podczas gdy wersje zredagowane przez ChatGPT i Claude osiągnęły odpowiednio 96,7% i 100%. Detektory identyfikują cechy stylistyczne charakterystyczne dla dobrze zredagowanych tekstów, a nie rzeczywiste pochodzenie wkładu intelektualnego.

**Zastosowania praktyczne:** Wyniki wskazują na konieczność rewizji polityk akademickich opartych na detektorach SI. Instytucje powinny promować transparentne zasady wykorzystania SI jako narzędzia wspierającego, zamiast polegać na zawodnych technologiach prowadzących do niesprawiedliwych ocen.

**Oryginalność/Wartość poznawcza:** Praca dostarcza empirycznych dowodów na brak wiarygodności detektorów SI w kontekście wykrywania plagiatów, wyznaczając jasną granicę etyczną między wykorzystaniem SI jako narzędzia a nieuczciwością akademicką oraz promując świadome podejście do wdrażania nowych technologii w nauce.

### Słowa kluczowe

Detekcja treści SI. Duże modele językowe (DMJ). Etyka w nauce. Generatywna sztuczna inteligencja. GenSI. Komunikacja naukowa. Narzędzia redakcyjne. Plagiat. Studium przypadku.

---

*ANNA MAŁGORZATA KAMIŃSKA is an assistant professor in the Institute of Culture Studies of the University of Silesia in Katowice and an employee of the Library of the Silesian University of Technology in Gliwice. Her scientific interests include information science, scientometrics, bibliology, scholarly communication and issues concerning processing the natural language. In her research, she focuses on analysing citations, bibliographic bases and sustainable development in the library science.*

*Contact:*

*anna.kaminska@us.edu.pl*

*University of Silesia in Katowice*

*Faculty of Humanities, Institute of Culture Studies*

*Uniwersytecka 4, 40-007 Katowice*