INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2026, VOL. 72, NO. 1, PP. 1-8
Manuscript received January 12, 2026; revised January, 2026. doi: 10.24425/ijet.2026.157881

Artificial Intelligence as a research support tool: reliability, transparency and epistemic risks in academic knowledge production

Miłosz W. Romaniuk, Maciej Gąsienica-Szostak, and Martyna Karwińska

***Abstract*—The rapid integration of generative AI into the research process forces us to look closer at whether these tools can actually be trusted. This generates tension, which becomes particularly visible when AI systems replace transparent analytical procedures with probabilistic outputs that cannot be independently reconstructed or epistemically audited by human researchers. In this paper, we move beyond the excitement over efficiency to examine the accuracy of AI-driven summarization and authorship detection. Our analysis reveals that beneath the speed of these systems lie significant risks, including systematic biases and a tendency toward 'hallucinated' certainty. Rather than rejecting these tools, we propose a new methodological framework that helps scholars use AI while safeguarding the integrity of their results.**

***Keywords*—artificial intelligence; research workflow; text summarization; authorship detection; epistemic reliability**

INTRODUCTION

We have reached a point where artificial intelligence is no longer just an optional add-on in the lab; it is becoming the very backbone of how we search, process, and synthesize academic knowledge. This study builds on a series of earlier investigations into the digital transformation of the social sciences, which highlighted how PhD students utilize ICT in their doctoral theses [1] and examined the broader integration of digital trends in research [2]. While previous work focused on the practical benefits of digital tools for enhancing research workflows and knowledge management [3], the rise of generative AI introduces a new layer of complexity.

Most current literature still treats AI as a neutral instrument for efficiency. We argue, however, that AI is far from neutral. These tools actively reshape our research by framing information in ways that are often hidden behind a "black box" of opacity and bias. This article digs into two specific areas: AI-driven text summarization and the controversial field of automated authorship detection. We aren't just looking at these technologies in isolation. We are asking what happens to scholarly judgment and accountability when we delegate core cognitive tasks to an algorithm. Can AI truly be a reliable partner? By analyzing the "hallucinations" of these systems, we propose a framework for a more disciplined, responsible approach to AI-assisted research that safeguards the methodological rigor identified as crucial in our earlier studies.

I. AI-ASSISTED ACADEMIC TEXTS SUMMARIZATION

A. Introduction

Researchers increase productivity by automating processes through the use of continuously improving AI tools [4]. One of the main fields of interest for social sciences is the use of AI language models for summarizing and analysis of academic texts [5] [4]. AI's abilities for fast and comprehensive reading are also commonly deployed in qualitative research, allowing for automated analysis of interview transcripts or discourse [6]. The capabilities of specialised AI language models are constantly increasing, but the diversity of techniques and the rapid pace of development in this field necessitate an adjustment in the approach of researchers who wish to minimise the risk of misinterpretation, biased results, or unethical practices [7]. Considering the potential of AI tools for critical analysis and summarisation of texts, as well as the "hygiene" of using these tools in a way that limits the above-mentioned risks, is particularly important in the context of doctoral researchers. PhD students are building their academic achievements in an era of rapid development of AI techniques, the use of which has quickly become standard practice. They are usually individuals at an early stage of their professional scientific careers, developing their methodological and ethical skillset. It is therefore important to be aware of the limitations of such technologies and to establish correct patterns of AI support in the scientific process.

The purpose of this study is to assess the reliability of certain commonly available AI tools in summarizing scientific articles, as well as compare their performance across key analytical criteria. Automatic summarization has the potential to increase efficiency when working with large quantities of literature, therefore it could be beneficial for PhD students and early-career researchers. It allows researchers to find and verify the main arguments and key insights of a given article (in more detail than the abstract provides), saving the time it would take to look through the whole text. To effectively support the research process, these tools must accurately represent the source material, identify conceptually relevant information, and produce summaries that are both coherent and practical. This study will have an exploratory character, with one assumption; Automatically produced summaries will exhibit qualitative differences between general-purpose AI and AI specifically tuned for document analysis.

Authors are with The Maria Grzegorzewska University (APS), Warsaw, Poland (e-mail of corresponding author: mromaniuk@aps.edu.pl). Abstract, introduction, and conclusions are written by M. W. R., as is the overall edition of the paper. Chapter I is written by M. G-S.; chapter II by M. K.

Among the techniques for automatic text summarisation, the extractive technique has emerged earlier. It ‘extracts’ relevant phrases and expressions corresponding to the target query from the database [8]. Strictly extractive summaries produce sets of ‘raw’ text snippets that, out of context, may not be perfectly comprehensible or accessible. Until recently, this was the dominant practice, but with the emergence of sophisticated LLMs, abstract text summarisation methods that rely on semantic understanding of content and produce more complex, paraphrased summaries have gained prominence [9]. The advantage of abstract methods is that they generate original, human-like, and therefore more accessible descriptions of the source content, but they are prone to hallucinations and mixing of threads. Research indicates that approaches combining extractive and abstract functions produce the most reliable summaries of the source text, while maintaining the comprehensibility of the content [10]. There are currently a significant number of LLMs on the market that offer automatic text summarisation capabilities, versatile general-purpose models (e.g., GPT, Copilot, Gemini, Claude) and specialized LLM-based tools optimised for source management, document review, automatic text summarisation (e.g., NotebookLM, AskYourPDF, Scholarcy). Most LLMs use a retrieval-augmented-generation framework (RAG), which combines extractive retrieval mechanisms with abstractive language generation [11]. Recent research indicates that LLM summaries are comparable to human-written ones, aside stylistic differences [12].

B. Method

This article presents an exploratory, comparative case study between several tools: NotebookLM, AskYourPDF, Scholarcy, Claude and ChatGPT. The first three tools are specifically designed for text analysis and source management, whereas the latter two exemplify ‘mainstream’ choice, generalist LLMs. They were selected based on their availability (all available in free use) and ease of use (no complicated setup required), thus representing common tools that are likely to be used by doctoral students. For the purposes of the study, new accounts were created on platforms offering the use of selected AI tools, and tests were conducted in incognito mode browser. One scientific article in the field of social psychology [13] was selected by the author of this study and thoroughly analysed without use of any AI-based tools. The article covers a topic closely related to the author’s own dissertation, making the study simulate an actual doctoral research process. The article is ~37 000 characters in length, uses standard organizational psychology terminology, and quantitative methodology in presenting mediation effects between multiple variables. It’s structure and complexity seem well within standard expectations for an academic reader. Later, the testing procedure began in the same way for each tool. First, a PDF file containing the source article was uploaded. Then, each was given the same prompt „Summarize this article’s main argument and describe how it supports it with evidence”. This prompt was written intentionally concise, domain-neutral and supposed to allow each tool to reveal its natural reasoning behavior. It also reflects a common use case for researchers. Scholarcy did not receive a prompt, as it creates its summary automatically after a file is uploaded. The procedure was repeated three times for each tool in separate browser windows, and these were conducted between October 25th and November 15th, 2025.

C. Results

All comments on the performance of the tools can be found in Tables I, II and III.

TABLE I
COMPARISON OF OPERATIONAL CHARACTERISTICS BETWEEN AI TOOLS IN ACADEMIC TEXT SUMMARIZATION

	Ease of setup	**Transparency of process**	**Citation support**	**Output reproducibility**	**Output formats**
Tool:	How easy it was to start using.	Does it explain which sources were used for each output?	Are citations specific, traceable, and correctly linked to source passages?	Does it produce consistent results when repeating the same query?	Does it provide answers in other forms than notes?
ChatGPT (GPT-5)	Email confirmation.	Doesn’t inherently produce structured citations.	No direct citations Doesn't allow tracking relevant source passages.	Consistent results. The same themes are consistently identified. The focus and specificity of summary format vary with each query.	Charts, images, tables, downloadable files.
NotebookLM	Google account and age confirmation.	Clickable citation markers in chat responses lead to corresponding document.	Citations are specific, traceable, and usually correctly linked. Tendency of highlighting excessive pieces of source material (e.g., end of one paragraph and beginning of another can be treated as one passage) or linking irrelevant passages.	Consistent results. The same themes are consistently identified. The focus and specificity of summary format vary with each query.	NotebookLM can produce audio and video summaries; mind maps, flash cards.
Scholarcy	Email confirmation.	The summary can only be based on one uploaded text.	Citations are specific, traceable and correctly linked. Tool links text passages and highlights exact phrases, which the tool deemed most relevant.	Consistent results. Users can choose between different, tailored summary formats (e.g. "general reader", "high school", "researcher"). Each differs in the number of generated sections, summary depth and specificity.	Scholarcy creates a tailored web page with flashcards.

AskYourPDF	Email confirmation.	Clickable citation markers in chat responses lead to corresponding document.	Citations are traceable. Their relation to linked summary pieces tends to be unclear. Highlights passages and single phrases scattered within one page of the source document, no citation crossed between pages.	Consistent results. The same themes are consistently identified. Highlighted source passages differ with each query. The specificity of summary varies with each query.	AskYourPDF can transform documents into AI generated podcasts.
Claude (Sonnet 4.5)	Email and phone number confirmation.	Doesn't inherently produce structured citations.	No direct citations, the output doesn't provide a way of tracking relevant source passages.	Consistent results. The same themes are consistently identified. The focus and specificity of summary format vary with each query.	Downloadable files (only paid version).

TABLE II
COMPARISON OF ANALYTICAL PERFORMANCE BETWEEN AI TOOLS IN ACADEMIC TEXT SUMMARIZATION

	Fidelity to source	**Comprehensiveness**	**Interpretive depth**	**Coherence & structure**	**Terminological accuracy**	**Critical neutrality**
Tool:	Does it stay within the uploaded material?	Does it capture all main arguments and themes?	Does it go beyond summarizing to explain reasoning or evidence?	Logical flow and clarity of the generated analysis	Proper use of discipline-specific terms and concepts	Does it maintain objectivity, or inject value opinions?
ChatGPT (GPT-5)	No observed deviation from source.	Main argument and relevant themes were properly identified.	Focus on summarizing with simple, digestible explanations.	The tool produces a comprehensible summary, focused on key insights.	The tool is consistent in terminology, uses proper terms and concepts from the source material.	No opinions deviating from the source material were noted.
NotebookLM	No observed deviation from source.	Main argument and relevant themes were properly identified.	The tool can mix excerpts from different parts of source material to explain some of the output sentences (e.g., in summarizing the statistical results, an additional passage from the conclusion was linked for clarity). This varies across repetitions of the query.	The tool followed the source material structure, creating a detailed and coherent summary.	The tool is consistent in terminology, uses proper terms and concepts from the source material.	No opinions deviating from the source material were noted.
Scholarcy	No observed deviation from source.	Main argument and relevant themes were properly identified.	In the "analysis" section, Scholarcy generated insights on research quality and links to Wikipedia articles on certain methodology and statistical results. It sometimes inferred insights based on the source.	The output is divided into main sections: summary, analysis, original text; with subsections. Summary contained logical errors (e.g., correct summary of a certain paragraph ended with a statement that such information was not described in the source text).	The tool is consistent in terminology, uses proper terms and concepts from the source material.	No opinions deviating from the source material were noted.
AskYourPDF	No observed deviation from source.	Main argument and relevant themes were properly identified.	Focus on summarizing.	Produces a comprehensible summary, with a brief introduction and conclusion. Relevant insights were presented in equally sized bullet points.	The tool is consistent in terminology, uses proper terms and concepts from the source material.	No opinions deviating from the source material were noted.
Claude (Sonnet 4.5)	No observed deviation from source.	Main argument and relevant themes were properly identified.	Focus on summarizing.	The tool produces a comprehensible summary, focused on key insights.	The tool is consistent in terminology, uses proper terms and concepts from the source material.	No opinions deviating from the source material were noted.

TABLE III
EPISTEMIC BEHAVIOR OBSERVED IN AI TOOLS IN ACADEMIC TEXT SUMMARIZATION

	Extractive	Abstractive	Interpretive	Speculative	Grounded
Indicators	Focuses on direct quotations and factual recall	Rephrases and synthesizes ideas in its own words	Adds thematic connections, assumptions, or critique	Introduces external info not found in the text	Cites and justifies claims with traceable text excerpts
Observed in?	NotebookLM, AskYourPDF, Scholarcy	NotebookLM, AskYourPDF, ChatGPT, Scholarcy, Claude	NotebookLM, ChatGPT		NotebookLM, AskYourPDF, Scholarcy

The AI tools seem comparable in terms of their analytical performance. The procedure results indicate that all models were able to identify the central argument of the source article and remained within its content boundaries, neither hallucinations nor deviations from the source material were noted. This finding is supported by the existing research on RAG framework used by the tools, which is said to mitigate hallucination risk [14]. Despite tool-specific differences in structure or formatting, all systems were comparatively stable in identifying what they considered the essential contributions of the text. The most meaningful differences emerged not in comprehension or fidelity to the source but in the operational characteristics of each tool and the depth and transparency of their analytic processes. The generalist models, ChatGPT (GPT-5) and Claude (Sonnet 4.5) have generated adequate and consistent results, although offered no mechanisms for tracing specific claims back to the source text. This limits their usefulness in the context of academic research, which requires scrutiny and transparency. NotebookLM, Scholarcy and AskYourPDF are all document-focused tools, tailored for citation support, and allow the user to track source text passages relevant to pieces of the generated summary. This design supports more grounded summarization, as the generated summaries allow the reader to cross-check where the AI has drawn its conclusions, and seek further references in the relevant passages. NotebookLM and Scholarcy have shown the most consistent citation support, though both occasionally misaligned excerpts or overextended the highlighted text. Logical errors were observed in the Scholarcy-generated output. Although they did not interfere with summary comprehensiveness and correct identification of article insights, the emergence of such errors raises questions about the tool's reading processes.

D. Conclusions

Overall, the results point to a slight practical trade-off between transparency and readability. Tools designed for document analysis offer clearer grounding and traceability but produce rigid structure, while general-purpose LLMs generate more digestible summaries without showing their interpretive steps. All tools proved to be reliable when it comes to comprehension and summarization of the source text. Thus, tool selection should be based on whether a researcher prioritizes methodological transparency or stylistic clarity/simplicity. Document-focused tools (NotebookLM, Scholarcy, AskYourPDF) seem a more sensible choice for PhD students and researchers, as they provide the reader with more detailed and grounded information. There are many emerging tools other than the ones this study focused on, which offer various operational characteristics. This reflects both a clear market demand and the practical usefulness of automatic summarization. It seems highly beneficial for academics to explore and experiment with different AI tools to identify the solutions which fit their individual research needs.

Because this procedure functioned as a pilot case study, its results should be treated as initial comparisons of operational differences rather than a definitive evaluation of overall tool performance. Aside from shedding light on some practical aspects of the compared LLM-based tools use, the patterns observed here provide a basis for creating more targeted research questions in future, larger-scale studies.

E. Limitations

Important limitations of this study are that there was only one file for the tools to summarize and only one prompt used. This is limiting both in terms of the potential depth of analysis, as well as the potential risk of mistakes made by the AI. To speak of generalizability of these patterns, broader testing across prompt demands and text types would be needed. Moreover, there was risk of human error involved, as the summarization results were only reported and compared by the author. Nonetheless, the present analysis provides an initial, systematic comparison that can inform practical decisions about the use of AI tools for text summarizing in academic/doctoral research.

F. Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used ChatGPT (GPT-5) to reword and rephrase text. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

II. RELIABILITY OF AI DETECTORS IN PAPERS

A. Introduction

Artificial intelligence is a daily occurrence these days. From checking a Christmas cake recipe to describing complex medical research results, AI permeates every aspect of our lives and requires us to adapt to this situation. This development particularly developed in late 2022, when OpenAI released a free version of its chat service after registration on its website [15]. Currently, ChatGPT is the most popular chat service, increasing its weekly user base from 1 million to 700-800 million between November 2022 and October 2025, and plans to reach 1 trillion by the end of 2025 [16]. On the one hand, one can assume that using chat facilitates information retrieval, but what about its use by students for studying and writing term papers?

Recently, I've been encountering frequent doubts from lecturers who receive their students' papers and wonder whether

they were written via chat or by themselves. And if so, was it entirely or only partially? And if only partially, is this acceptable or not? This raises another question: to what extent students' use of chat capabilities borders on academic support, or borders on laziness and a lack of desire to learn and remember. A meta-analysis of 51 experimental studies from 2025 demonstrated a positive impact of using ChatGPT on improving performance and a moderately positive effect on improving perception of learning. More recent studies have shown that students learning with an AI tutor do so faster and in a shorter time than during active learning in the classroom [17]. Meanwhile, an experiment was conducted at Corvinus University in Budapest to assess students' motivation and understanding of real-world material. They were divided into two groups, one of which was allowed to use AI tools during classes and exams, while the other was not. The results showed that the arbitrary use of AI tools results in a lack of engagement and poorer learning of the material by students [3]. Another study indicates that students' use of GenAI negatively impacts their academic performance and self-confidence, and further fosters an attitude of learned helplessness [18]. Yet another field experiment involving nearly a thousand students demonstrated that ChatGPT-4 did indeed significantly improve performance, but after removing access to the tool, students performed worse than those who did not use it at all. The authors emphasize that the study only illustrates short-term effects, but it raises questions about whether AI helps in knowledge retention [19].

Interesting phenomenon is the frequency of AI tool use by students depending on the time of year. Anonymous statistical data indicates that ChatGPT usage drops significantly during the summer, especially after the end of the school year. A record increase of 97.4 billion tokens was recorded at the end of May 2025, when school exams were taking place. In June, when the summer holidays begin, token production drops by almost half compared to May. Observations of data from 2023 also confirm the trend that AI tool use decreases during the holidays and increases as students return to classes [20]. Furthermore, variability is also visible during weekends, when regular drops in AI tool use are noticeable [21].

While delving into the topic of the extensive use of AI tools by students, one might consider how to address this. This article will address the issue of finding tools for checking papers in the context of human authorship versus AI tools. Currently, we know that no AI detector can 100% accurately tell us who wrote a paper [22]. However, browsing the internet, one can find many websites offering this option. The aim of this article is to provide a brief overview of how AI detectors work in scientific papers and to verify their reliability, which is defined as high sensitivity (AI detection) and high specificity (human non-detection).

B. Researching tools to verify the use of AI in papers

As mentioned above, there is no officially approved tool that can 100% confirm whether a given text was generated by AI or written by a human. Therefore, the aim of the study was to examine how well exemplary tools designed to detect AI in texts actually detect it. The process involved preparing four scientific texts – one written by the author of the article and three generated by ChatGPT: [Text 1.] a scientific text written by the author of the article; [Text 2.] a scientific text on the same topic generated by AI; [Text 3.] a scientific text generated by AI that copies the author's writing style; and [Text 4.] a text combining two AI-generated texts [Text 1 and Text 2.] modifying them so that AI cannot be recognized. The topic of the scientific text refers to the topic of an aging society and aspects related to seniority. The tools used to check the reliability of AI detectors were selected through personal research and inspired by the ranking of the best tools presented on the Writerbuddy website [https://writerbuddy.ai/blog/the-15-best-ai-plagiarism-checkers, accessed: November 26, 2025]. The free version of the study used the following tools: Text Guard [23], Grammarly [24], Justdone [25], GPTZero [26]. Two trials were conducted to deepen the results and verify the repeatability of the results. The research process was as follows:

Sample 1. Each text was entered into the tool one by one based on:

- Text 1. – Text Guard, Grammarly, Justdone, GPTZero;
- Text 2. – ChatGPT (instructions: “Write a scientific text of approximately 700 words with footnotes regarding Poland's perspective in the context of an aging society, the specificity of the generation regarding older people, and the characteristics of senior age.”), Text Guard, Grammarly, Justdone, GPTZero;
- Text 3. – ChatGPT (instructions: “Copy the author's writing style and write an article on the same topic.”*), Text Guard, Grammarly, Justdone, GPTZero;
- Text 4. – ChatGPT (instructions: “Modify these two AI-generated texts [Text 2 and Text 3] so that they cannot be recognized as AI-generated. Keep them to 700 words. Add footnotes and a bibliography.”), Text Guard, Grammarly, Justdone, GPTZero. Trial 2. Testing was performed after 15 minutes as in Trial 1, excluding ChatGPT.

*interesting fact - ChatGPT's response to the request: "Below, I've prepared an article written in the style of your text – I'm maintaining the same tone: scientific, yet narratively coherent; based on statistics, smooth transitions between threads, and clear references to the relevant literature. I'm not copying the content, but rather replicating the sentence rhythm, paragraph structure, and the saturation of data and scientific references, just like in the provided excerpt."

The base text was proofread in Polish. To better illustrate its subject matter and the author's writing skills, it is included below:

Currently, in Poland, we are beginning to grapple with the challenges of aging and an aging society. Statistics presented in 2024 by the Central Statistical Office are clear. As of December 31, 2023, the number of people over 60 was 9,893,700, representing 26.3% of the Polish population, a 1% increase compared to the previous year [27, p. 14]. Demographic forecasts included in the report show an increase of almost 20% by 2060 compared to 2023, resulting in almost 12 million people over 60, who will constitute 38.3% of the total Polish population [27, p. 25]. As I mentioned at the beginning, not only is the number of older people increasing, but their life expectancy is also increasing. For men who turn 60 in 2023, the average life expectancy is almost 20 years, while for women it is almost 25 years [27, p. 21]. Furthermore, in 2060, the number of elderly people aged 60-74 will decrease, while the percentage of people aged 75 and older will increase. The largest increase is observed in the age group 85 and older, which will constitute almost 16% of the population aged 60 and older [27, p. 27]. Using statistics provides a real and crucial context for viewing this phenomenon.

This article will focus on seniors. Starting from the very beginning, it is important to understand people aged 60 and older in the context of different generations and their specific characteristics related to the period in which they were born. The most developed sociological approach to this term is attributed to Karl Mannheim, who argued that the topic of generations is "indispensable for understanding the structure of social and intellectual movements. Its practical significance becomes clear when one attempts to thoroughly understand the accelerated pace of social change characteristic of our era" [28, pp. 38-39]. Generations are also described in a relatively recent monograph by American psychologist Jean M. Twenge titled "Generations" [29]. Other terms important for exploring knowledge about seniors include their characteristics from a psychological, physiological, and social perspective [30, p. 15]. As people age, they experience significant changes related to personality, adaptation to the environment, relationships, and cognitive functions [30, pp. 48-49]. Additionally, the perception of oneself changes from the perspective of society and intergenerational relations related to the role of a grandmother/grandfather [31, p. 131], which gives him a sense of agency and being needed [32, p. 662]. The period of old age is widely described by: Joanna K. Wawrzyniak in her book, where she includes topics related to the perception of old age, discrimination, loneliness, but also activation [33]; Joanna Kliszcz, who focuses on the psychology of the needs of older people [34]; Norbert G. Pikuła, writing about the sense of senes of life of older people in the context of social change [35], Joanna Wrótniak writes about the psychosocial resources of older people [36], and interdisciplinary images of old age are included in a collective monograph [37]. Additionally, reference can be made to Erik H. Erikson's theory of identity throughout the human life cycle [38] and to his wife's later theory, which supplemented this theory with a ninth life stage following the eighth phase of crisis [39], as well as to the theory of aging [40] and issues related to the medical, psychological, and social aspects of aging [41].

To understand the functioning of older people, one must first delve into theoretical knowledge about the specifics of their age and familiarize oneself with research reports that illustrate the social situation prevailing worldwide.

C. Results

The tables below present the results obtained from both trials, testing the reliability of AI detectors in scientific texts. The left-hand side presents the tools used in the study, with the top section showing the specific text tested by the AI detector.

TABLE IV
AI DETECTOR, SAMPLE 1

Tool/Text	Text 1.	Text 2.	Text 3.	Text 4.
Text Guard	87% AI generated text 13% Human written text	83% AI generated text 17% Human written text	85% AI generated text 15% Human written text	86% AI generated text 14% Human written text
Grammarly	13% Possible AI text detected	0% No AI-generated text detected	0% No AI-generated text detected	0% No AI-generated text detected
Justdone	88% of your text shows signs of AI generation 33% Identical 33% Minor changes 22% Paraphrased 12% Unique text	88% of your text shows signs of AI generation 33% Identical 33% Minor changes 22% Paraphrased 12% Unique text	88% of your text shows signs of AI generation 33% Identical 33% Minor changes 22% Paraphrased 12% Unique text	88% of your text shows signs of AI generation 33% Identical 33% Minor changes 22% Paraphrased 12% Unique text
GPTZero	1% AI generated 3% mixed 96% human	100% generated	100% generated	100% generated

TABLE V
AI DETECTOR, SAMPLE 2

Tool/Text	Text 1.	Text 2.	Text 3.	Text 4.
Text Guard	86% AI generated text 14% Human written text	87% AI generated text 13% Human written text	83% AI generated text 17% Human written text	84% AI generated text 16% Human written text
Grammarly	0% No AI-generated text	0% No AI-generated text	0% No AI-generated text	0% No AI-generated text
Justdone	88% of your text shows signs of AI generation 33% Identical 33% Minor changes 22% Paraphrased 12% Unique text	88% of your text shows signs of AI generation 33% Identical 33% Minor changes 22% Paraphrased 12% Unique text	98% of your text shows signs of AI generation 31% Identical 42% Minor changes 25% Paraphrased 2% Unique text	84% of your text shows signs of AI generation 29% Identical 36% Minor changes 19% Paraphrased 16% Unique text
GPTZero	1% AI generated 3% mixed 96% human	100% generated	100% generated	100% generated

The data in both tables shows the following conclusions:

- Each tool has similar results except for the last one. This suggests that the tool is indeed poor at detecting text written by humans or ChatGPT, as it always reports similar values regardless of the truth or is unable to detect it;
- Each website makes a prediction about whether the text is written by AI or not. Since the results of each tool vary, this indicates the use of different validation techniques;
- Most tools produced repeatable results for tests conducted on the input data. Lack of repeatable results would indicate a lack of reliability. GPTZero and Justdone showed the same result in both tests, Text Guard showed 1-5 percent differences, and Grammarly showed a 13% chance of using AI in the first test of the human text, while it indicated 0% for the remaining AI-generated texts, and in the second test, it indicated 0% in all texts.

However, repeatability alone is not the only factor determining the tool's reliability.

- Justdone's tool categories may indicate a) tool error; b) the assumption that the tool is actually a plagiarism detector, not an AI detector.
- the choice of tool is crucial to the success or failure of the task;
- most tools failed the task because they did not demonstrate accuracy;
- the GPTZero tool produced a result closest to the truth, but its reliability cannot be 100% determined, as in [Text 1] it did not demonstrate 100% human accuracy.

D. Discussion

Incorrect results from AI detectors are mainly due to the fact that a) each tool trains on a set of different texts, which means different results may be obtained from different detectors based on the same text; b) AI models are constantly being improved and refined, which may cause AI detectors to struggle to keep up with changes; c) AI detectors are primarily designed for use in English; the use of other languages may result in lower verifiability [22]. Generators detecting the use of AI in texts written by humans are quite common. The most popular example is ZeroGPT's verification that the United States Constitution was 94% written by AI [42]. In the study presented above, it can be concluded at first glance that GPTZero produced the result closest to the truth, demonstrating that the text was 96% written by humans. However, in one study of AI detectors, the GPTZero generator was described as performing less well in languages other than English. Therefore, to deepen the study, each text was translated into English and submitted for verification.

TABLE VI
GPTZERO, AI DETECTOR PAPER IN ENGLISH

Text 1.	Text 2.	Text 3.	Text 4.
100% human	56% AI generated 0% mixed 44% human	100% AI generated	18% AI generated 1% mixed 81% human

The results show that GPTZero performs differently in English than in Polish. Comparing the results, they are completely different. Accuracy is higher for Text 1 in English, while for the remaining texts, the generator performed better in Polish. In two English texts, the AI detector indicated correct verification (Text 1 and Text 3). In Text 2, which was generated on a specific topic, it was rated 56% as AI-generated. Surprisingly, the text that combined two AI-generated texts (Text 2 and Text 3) but modified them to prevent AI from being recognized was rated 81% as human-written by GPTZero. This begs the question: is it that GPTZero is unable to detect 100% of AI use in works, or is ChatGPT, thanks to its experience and history, becoming increasingly human-like when instructed to write a text as if it were a human? Using AI-based tools should be approached with common sense. It's clear that they're becoming increasingly common and are undoubtedly very helpful. However, a distinction must be made between using them and actually using them.

It's also fair to say that AI detectors need to be more refined. I wonder if they'll even be able to keep up with different chats. I think it would be best if a single chat provider also had the ability to check texts, but we don't know if the text we received from someone was written by the same party, and the cycle continues. The question is whether tools like detectors actually work or are simply a marketing ploy, as there are many such tools on the market, but none are 100% reliable.

CONCLUSION

This analysis makes one thing clear: while AI has firmly embedded itself in the modern research workflow, it remains an epistemically fragile partner. Tools designed for summarization or authorship detection can certainly speed up the early, often tedious stages of a literature review, but our findings warn against treating them as objective authorities. They are prone to bias and a kind of 'confident' opacity that can easily mislead a researcher. In practice, calls for transparency and responsible use remain largely aspirational, as most contemporary AI research tools operate as black boxes that systematically resist meaningful methodological scrutiny.

We found that AI's value in academia is strictly tied to human oversight. When we offload the heavy lifting of interpretation or evaluation to an algorithm, we aren't just saving time—we are risking the very rigor that defines scholarly work. Our proposed framework, therefore, isn't about rejecting AI, but about disciplined integration where the machine assists rather than replaces. However, it is important to be realistic about the scope of this study. We have focused on a specific subset of AI tools, and the fast-moving nature of this technology means our findings might not cover every emerging platform. We also have to acknowledge the lack of broad empirical validation across different academic fields, which remains a clear limitation. Looking ahead, the next step for researchers should be to investigate how these risks vary from one discipline to another and to what extent institutional policies actually change the way scholars interact with these tools. Ultimately, keeping the human researcher at the center of the process is the only way to ensure that the integrity of our knowledge remains intact.

REFERENCES

[1] M. W. Romaniuk, A. Szarfenberg, I. Pawłowska and K. Choszczyk, "Doctoral Theses in the Digital Age – ICT use by Social Sciences PhD Students of The Maria Grzegorzewska University," International Journal of Electronics and Telecommunications, vol. 70, no. 1, pp. 199-204, 2024.

[2] M. W. Romaniuk, J. Gierzyński, M. M. Pietrzak and J. Zbróg, "Integrating Technology in Social Science Research: Emerging Trends and Ethical Considerations," International Journal of Electronics and Telecommunications, vol. 71, no. 1, pp. 171-179, 2025.

[3] M. W. Romaniuk, P. Mika, J. Apanasewicz and E. Duda-Maciejewska, "Enhancing Research Practices: Digital Technologies in the Social Sciences and Practical Tools for Doctoral Students," International Journal of Electronics and Telecommunications, vol. 71, no. 1, pp. 181-188, 2025.

[4] A.M. Sami, Z. Rasheed, K.-K. Kemell, M. Waseem, T. Kilamo, M. Saari, A. N. Duc, K. Systä and P. Abrahamsson, "System for systematic literature review using multiple AI agents: Concept and an empirical evaluation," arXiv, 2025.

[5] Z. Dar, M. Raheel, U. Bokhari, A. Jamil, E. M. Alazzawi and A. A. Hameed, "Advanced Generative AI Methods for Academic Text Summarization," in 2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI), Mt Pleasant, 2024.

[6] M. Perkins and J. Roe, "The use of Generative AI in qualitative analysis: Inductive thematic analysis with ChatGPT," Journal of Applied Learning and Teaching, vol. 1, pp. 390-395, 2024.

[7] P.A. Christou, "How to use artificial intelligence (AI) as a resource, methodological and analysis tool in qualitative research?," Qualitative Report, vol. 7, 2023.

[8] A. Kumar Yadav, Ranvijay, R. S. Yadav and A. K. Maurya, "State-of-the-art approach to extractive text summarization: a comprehensive review," Multimedia Tools and Applications, p. 29135–29197, 2023.

[9] H. Shakil, A. Farooq and J. Kalita, "Abstractive text summarization: State of the art, challenges, and improvements," Neurocomputing, 2024.

[10] M. Abazari Kia, A. Garifullina, M. Kern, J. Chamberlain and S. Jameel, "Question-driven text summarization using an extractive-abstractive framework.," Computational Intelligence, vol. 3, 2024.

[11] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel and D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems, Vancouver, 2020.

[12] T. Zhang, F. Ladhak, E. Durmus, P. Liang, K. McKeown and T. Hashimoto B., "Benchmarking Large Language Models for News Summarization," Transactions of the Association for Computational Linguistics, p. 39–57, 2024.

[13] Y. Dai, Y. M. Tak, W. Chen and J. Hou, "How organizational trust impacts organizational citizenship behavior: Organizational identification and employee loyalty as mediators," Frontiers in Psychology, vol. 13, 15 Listopad 2022.

[14] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, Q. Bing and T. Liu, "Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions.," ACM Transactions on Information Systems, no. 1, 2024.

[15] J. Cholewa and Z. Rak, "ChatGPT jako przyjaciel pokolenia Z [ChatGPT as a friend of Generation Z]," 2023. [Online]. Available: https://jows.pl/brepo/panel_repo_files/2024/01/02/xnwdvd/jows-04-2023-rak-i-cholewa.pdf [Accessed 6 12 2025].

[16] [Online]. Available: https://explodingtopics.com/blog/chatgpt-users [Accessed 06 12 2025].

[17] M. Benedeka and B. R. Sziklaib, "Impact of AI Tools on Learning Outcomes: Decreasing Knowledge andOver-Reliance," [Online]. Available: https://www.researchgate.net/publication/396714748_Impact_of_AI_Tools_on_Learning_Outcomes_Decreasing_Knowledge_and_Over-Reliance [Accessed 06 12 2025].

[18] S. Azeem and M. Abbas, "Personality correlates of academic use of generative artificial intelligence and its outcomes: does fairness matter?," [Online]. Available: https://link.springer.com/article/10.1007/s10639-025-13489-6 [Accessed 06 12 2025].

[19] H. Bastani, O. Bastani, A. Sungu, H. Ge, O. Kabakcı and R. Mariman, "Generative AI Can Harm Learning," [Online]. Available: https://download.ssrn.com/2024/7/15/4895486.pdf [Accessed 06 12 2025].

[20] J. Wilkins, "OpenAI Usage Plummets in the Summer, When Students Aren't Cheating on Homework," [Online]. Available: https://futurism.com/openai-use-cheating-homework [Accessed 06 12 2025].

[21] T. Burman, "ChatGPT Usage Skyrockets as Kids Return to School," [Online]. Available: https://www.newsweek.com/chatgpt-use-skyrockets-school-kids-homework-2120753 [Accessed 06 12 2025].

[22] "Czy wykrywacze AI mówią prawdę? Test najpopularniejszych detektorów [Do AI Detectors Tell the Truth? A Test of the Most Popular Detectors]," [Online]. Available: https://non.agency/blog/czy-wykrywacze-ai-mowia-prawde-test-najpopularniejszych-detektorow/ [Accessed 07 12 2025].

[23] "TextGuard AI," [Online]. Available: https://textguard.ai/ [Accessed 1 12 2025].

[24] "Grammarly," [Online]. Available: https://app.grammarly.com/ [Accessed 01 12 2025].

[25] "JustDone," [Online]. Available: https://justdone.com/ai-detector [Accessed 01 12 2025].

[26] "GPTZero," [Online]. Available: https://gptzero.me/ [Accessed 01 12 2025].

[27] GUS, "Sytuacja osób starszych w Polsce w 2023 r. [The situation of older people in Poland in 2023]," Głowny Urząd Statystyczny, Warszawa, Białystok, 2024.

[28] K. Szafraniec, Pokolenia i polskie zmiany. 45 lat badań wzdłuż czasu [Generations and Polish Changes: 45 years of research across time], Warszawa: PWN SA, 2022.

[29] J. Twenge, Pokolenia [Generations], Sopot: Wydawnictwo Smak Słowa, 2024.

[30] S. Steuden, Psychologia starzenia się i starości [Psychology of aging and old age], Warszawa: Wydawnictwo Naukowe PWN SA, 2014.

[31] J.C. Cavanaugh, "Starzenie się," in Psychologia rozwojowa [Developmental psychology], Poznań, Zysk i S-ka Wydawnictwo s.c., 1997.

[32] A.I. Brzezińska and S. Hejmanowski, "Okres późnej dorosłości. Jak rozpoznać ryzyko i jak pomagać?," in Psychologiczne portrety człowieka. Praktyczna psychologia rozwojowa [Psychological portraits of humans. Practical developmental psychology], Gdańsk, Gdańskie Wydawnictwo Psychologiczne sp. Z o.o., 2021.

[33] J.K. Wawrzyniak, Starość człowieka - szanse i zagrożenia. Implikacje pedagogiczne [Human old age - opportunities and threats. Pedagogical implications], Warszawa: CEDEWU, 2017.

[34] J. Kliszcz, Psychologia potrzeb osób starszych. Potrzeby psychospołeczne po 65. roku życia [Psychology of the needs of older people. Psychosocial needs after age 65], Warszawa: Difin SA, 2019.

[35] N. G. Pikuła, Poczucie sensu życia osób starszych. Inspiracje do edukacji w starości [A sense of purpose in life for older people. Inspirations for edukaction in old age], Kraków: Wydawnictwo Impuls, 2019.

[36] J. Wrótniak, Zasoby psychospołeczne osób w podeszłym wieku z poczuciem samotności [Psychosocial resources of elderly people with a sense of loneliness], Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 2015.

[37] M. S. P. Bełtowska, Obrazy starości. Analiza interdyscyplinarna [Images of old age: an interdisciplinary analysis], Toruń: Wydawnictwo Adam Marszałek, 2024.

[38] E. H. Erikson, Tożsamość a cykl życia [Identity and the Life Cycle", Poznań: Wydawnictwo Zysk i S-ka, 2004.

[39] B. Bugajska, Tożsamość człowieka w starości. Studium socjopedagogiczne [Human identity in old age. A sociopedagogical study], Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, 2005.

[40] N. A. Pachana, Starzenie się [Aging}, Łódź: Wydawnictwo Uniwersytetu Łódźkiego, 2021.

[41] K. Szostakowska, Rodzinna opieka nad seniorem - trudna lekcja życia [Family care for seniors - a difficult life lesson], Warszawa: Wydawnictwo Akademii Pedagogiki Specjalnej im. Marii Grzegorzewskiej, 2021.

[42] "AI wrote the US Constitution, says AI content detector," [Online]. Available:https://medium.com/@michellehwd/ai-wrote-the-us-onstitution-says-ai-content-detector-f24681fdc75f [Accessed 08 12 2025].