Issues in Information Systems
Volume 26, Issue 3, pp. 401-412, 2025

DOI: https://doi.org/10.48009/3_iis_2025_2025_132

Precision check: A critical look at the reliability of AI detection tools

Karen Paullet, *Robert Morris University, paullet@rmu.edu*
Jamie Pinchot, *Robert Morris University, pinchot@rmu.edu*
Evan Kinney, *Robert Morris University, kinney@rmu.edu*
Tyler Stewart, *Robert Morris University, stewartty@rmu.edu*

Abstract

Debates concerning the creativity, ethics, and the changing nature of learning have undoubtedly been triggered by the expanding application of students using generative AI in academia to complete their assignments. These tools raise questions regarding academic integrity. Differentiating between human created assignments and AI written material has become a challenge for educators and institutions. An experiment was conducted to test the efficacy and validity of AI detecting tools. Three tools were tested, GPTZero, Quillbot, and Polygraf AI, using 100 samples that were written by a human, completely written using AI or a combination of both human and AI written material to determine the tools capacity to accurately identify the content's source.

Keywords: artificial intelligence, GPTZero, ChatGPT, QuillBot, Polygraph AI, large language models

Introduction

The rapid advancements and use of artificial intelligence (AI) in academia is a critical topic that has received a great deal of attention in the field of higher education (Rafiq et al., 2025; Cotton et al., 2023). While advancements in AI have the potential to create a more engaging learning experience for students by providing personalized feedback and various types of support (Fuchs, 2023), they have also raised concerns regarding academic honesty and plagiarism (Cotton et al., 2023).

One of the most prevalent AI tools in use is ChatGPT, an AI-driven chatbot that was developed by OpenAI, an AI research company. ChatGPT was designed to generate human-like text in a conversational style, and it was introduced to the public in November of 2022. In the name ChatGPT, Chat refers to the tool's classification as a chatbot, and GPT stands for Generative Pre-trained Transformer. The name refers to the purpose of the tool, with "generative" referring to the ability to generate text and "pre-trained" referring to the content that is used to train the tool's algorithm so that it can continually improve text-generation capabilities as it absorbs additional data (Lieberman, 2024). ChatGPT is a large language model (LLM) that is trained on a vast amount of textual data and uses natural language processing (NLP) to have human-like conversations with users, answering questions or generating content based on human prompts (Fuchs, 2023).

ChatGPT can perform a wide range of writing and language tasks, including text generation, summarization, paraphrasing, grammar editing, question answering, and language translation (Uyen & An, 2025; Cotton et al., 2023). It can generate research papers, essays (Yeadon et al., 2023), stories, poems,

and other types of written content. Lieberman (2024) notes that ChatGPT has access to "an expansive knowledge base that could rival any university library in volume" (p. 70) and can produce well-written text in response to questions and prompts based on that knowledge base. It is now possible to use ChatGPT, or other tools like it, to produce quality writing that is both accurate and critical, on almost any topic (Yeadon et al., 2023).

Advancements in generative AI tools such as ChatGPT have led to a marked increase in the creation of AI-generated content (Uzun, 2023). AI-generated content is defined as any type of content that is created with the aid of machine learning algorithms or other AI techniques, and can include text, images, and video (Uzun, 2023). Students in particular are adopting AI tools at a rapid pace (Johnston et al., 2024; Nam, 2023). OpenAI reports that one-third of adults between the ages of 18 and 24 use ChatGPT (OpenAI, 2025; Leckrone, 2025) .The below research question will be explored for this experimental study.

RQ1: *How accurate are AI detection tools GPTZero, QuillBot and Polygraf.AI in determining if artificial intelligence was used to create written content?*

Student Use of AI Tools

A study by OpenAI (2025) found that more than any other kind of user, college-aged students are utilizing ChatGPT. The study notes that over one-third of young adults aged 18 to 24 use ChatGPT. Their study surveyed 1,229 college students across a mix of STEM and non-STEM disciplines and asked about the ways in which they use the tool. Students reported that the most prevalent uses for ChatGPT were: Starting papers/projects (49%), summarizing texts (47%), brainstorming creative projects (45%), exploring topics (43%), editing writing (43%), mathematical problem-solving (41%), exam preparation (38%), academic research (36%), tutoring (32%), essay drafting (31%), career-related writing (31%), exam answers (27%), science research (27%), creating lesson plans (23%), language learning (22%), career advice (18%), predicting a grade (18%), relationship advice (17%), computer programming (16%), organizing schedules (15%), grading school work (14%), facilitating student collaboration (13%), job search (12%), mental health help (12%), physical health help (12%), social media content help (10%), and visual data analysis (8%). The report notes that 23 of the 27 reported uses for ChatGPT were education and career-related, while relationship advice, mental and physical health help, and social media content help fell outside of the scope of education.

The OpenAI (2025) study also measured student adoption of ChatGPT by state in the U.S. The states in the Top 5 for adoption by college students include: Arizona, California, New Jersey, New York, and Virginia. The states in the Top 10 for adoption additionally include: Florida, Georgia, Illinois, Utah, and Washington.

Positive Impacts

ChatGPT and other generative AI tools can be seen as a revolutionary technology that could completely reshape the ways educators teach and the ways that students learn. There are a variety of benefits to using AI tools in education ranging from research support to adaptive learning, and these tools may enable students to access learning opportunities beyond the scope of traditional instructors (Reiter et al., 2025; Uyen & An, 2025; Lukianenko et al., 2024, Basha, 2024; Xiao et al., 2023). AI tools can be used to personalize a learning experience for students (Fuchs, 2023; Lieberman, 2024; Basha, 2024). A chatbot can essentially act as a student's tutor and will be available at any time to give the student one-on-one personalized and adaptive support (Reiter et al., 2025; Lukianenko et al., 2024), which is not possible for an instructor or human tutor to offer. A chatbot can help students to develop a study plan, organize class

notes, and even quiz them based on their notes. Students can direct the chatbot to repeat content that they have not yet mastered and even point out specific concepts or materials they need to review. Students can use AI tools to help jumpstart a research project or assignment, by having the tool generate a research plan or writing outline. AI tools can also be used to aid students in brainstorming ideas for research projects or assignments. These kinds of supports can help students to systematically develop their research skills (Lieberman, 2024; Lukianenko et al., 2024; Xiao et al., 2023).

Instructors can also use AI tools to create dynamic, interactive learning experiences by tailoring their teaching approaches and interventions according to student performance. These kinds of personalized teaching approaches can also help to promote student engagement (Lukianenko et al., 2024). In addition, AI tools can be extremely beneficial to students who need additional writing supports such as grammar-checkers and spell-checkers. These tools can provide real-time feedback with recommendations to better not only grammar but also writing style and mechanics. These recommendations can be very helpful for language learning and improvement of writing skills (Uyen & An, 2025).

Negative Impacts

The ease with which text can now be generated with AI also brings with it some serious negative consequences that come with the potential for misuse. AI tools can be misused by students to avoid the academic rigor of conducting research and creating original work (Lieberman, 2024). A key concern is cheating and plagiarism when AI is used without permission or disclosure (Reiter et al., 2025; Uyen & An, 2025). Academic integrity is a fundamental principle in academia, highlighting the importance of honesty, fairness, and responsibility in scholarly work (Sullivan et al., 2023). Such misuse of AI technology by students poses a clear ethical concern and serious threat to academic integrity (Fuchs, 2023).

Another concern relates to accuracy of generated content. AI tools depend heavily on the quality, diversity, and complexity of the training data that is used to train the model. If an AI tool is not exposed to knowledge in a subject area at a sufficient level to address its complexities, it may struggle to understand the nuances of a question or prompt, potentially leading to inaccurate results (Fuchs, 2023). These inaccuracies are referred to as "hallucinations", further defined as errors with deceptive plausibility (Lee et al., 2023; Walczak & Cellary, 2023). Additionally, if the data used to train the model is biased in some way, those biases will filter into the tool's responses (Fuchs, 2023). Additionally, while AI tools like ChatGPT can provide grammatically correct answers to questions and prompts, they have been shown to lack linguistic diversity, leading to writing that is not as diverse or expressive as human writing (Chaves & Gerosa, 2022). However, this limitation may be eliminated over time as the AI tools continue to evolve.

Further, educators and scholars have begun to have serious concerns that the prevalence of academic integrity violations regarding the use of AI in schoolwork will also erode the critical thinking, writing, and comprehension skills that students are meant to build through the completion of their own academic work. Students may over time become passive learners, depending on the AI tool for responses and trusting in the veracity of its responses without questioning or fact checking. An overreliance on AI technology could allow a student to have short term success in academia but ultimately undermine their ability to progress in their analytical, problem-solving, and critical thinking skills which in turn can deteriorate their writing quality and ability to develop coherent content (Uyen & An, 2025; Lee et al., 2024; Krause et al., 2024; Lieberman, 2024; Basha, 2024; Fuchs, 2023; Bates et al., 2020). Vieriu & Petrea (2025) surveyed 85 students in Bucharest on their use of AI. Of the respondents, 48.2% showed concern in regard to AI providing incorrect or inaccurate responses to questions, 16.5% shared that they were worried about becoming too dependent on AI, and 16.5% expressed concern that their critical thinking skills would suffer due to using AI too frequently. These worries are telling and parallel the concerns that other literature has highlighted.

The prevalent use of AI in academia has introduced unprecedented challenges for educators in maintaining academic integrity (Lukianenko et al., 2024). These challenges have made it clear that there is a serious need for accurate and ethical methods for educators to distinguish AI-generated content from content generated by humans (Uzun, 2023; Cotton et al., 2023).

AI Detection Methods and Tools

As AI-generated content becomes more common, it is important to confirm its accuracy, particularly in academic settings. The researchers tested GPTZero, Quillbot and Polygraph AI. Assessing each of the three on the same sample set may yield important information about their precision, advantages, and disadvantages.

GPTZero

GPTZero was created by Edward Tian and Alex Cui in 2023. Since its original launch, GPTZero has undoubtedly changed, expanding quickly in response to worries about academic AI-generated content. By mid-2024, its user base had grown from 30,000 early adopters to 4 million, demonstrating the growing need for AI detection tools. The tool now offers a more sophisticated perspective of writing authenticity rather than just detecting AI-generated text (Miller, 2023). It provides a more thorough understanding of the content creation process by gathering information on editing history and user interactions, which may help writers and educators strike a balance between human authorship and AI support. GPTZero is an AI detector which checks to see if a document was created using a large language model such as ChatGPT. GPTZero detects AI on sentence structure, paragraph, and document level. The accuracy of GPTZero continues to increase as more texts are submitted to the model. GPTZero analyzes patterns of writing using syntax and sentence length to identify text created by machine learning.

GPTZero uses seven machine learning components to determine the probability of the use of AI in text: Education Module, Burstiness, Perplexity, GPTZeroX, GPTZero Shield, Internet Text Search and Deep Learning (Tian, et al., 2024, Paullet, et.al. 2025). Each component provides a weighted score to the Document Classification and Document Breakdown that calculates an estimation of the amount of human-generated, AI-generated or mixed-generated writing that has been used. The Education component runs the input text against other human-written text created by students. The Burstiness component analyzes the text to see if there are patterns in the writing, whereas, Perplexity determines which words might come after one another. GPTZeroX is a component that is able to provide sentence-by-sentence classifications for human-generated, AI-generated, and mixed-generated text. GPTZero Shield defends against other tools looking to exploit the AI detector. Lastly, the Deep Learning component is used to detect the usage of AI. Human-generated text is continuously fed into GPTZero so that it is constantly learning patterns to help determine the likelihood of AI produced material. The model is trained from creative writing, scientific writing, blogs, news articles, and more. The submissions are tested against a large-scale dataset of human- and AI-generated material (Tian, et al., 2024).

GPTZero has both paid and unpaid subscription plans. The tool works the same regardless of the plan used. The difference between the two versions is the number of characters that can be scanned by GPTZero. The free version only allows for seven submissions per day, limited to 5000 characters for each submission. GPTZero tracks the IP address to determine if the limit has been reached regardless of which browser is being used. In addition to the free version there are three paid subscription plans. The Essential Plan allows users up to 150,000 words per month which is equivalent to 300 pages of scanning. The Premium Plan allows 300,000 words per month which would include 600 pages of scanning to include the AI deep scanner and is multilingual. Lastly, the Professional Plan allows for 500,000 words equivalent to 1000 pages of text,

everything listed above, and allows for 10,000,000 words overage and military grade data security (Tian, et al., 2024). All three versions use the same model to detect if text was written by a human, AI or a combination of human and AI. Each gives a breakdown for detecting the use of AI in written material with descriptions of the results.

QuillBot

QuillBot was founded in 2017 by Rohan Gupta, Anil Jason, and David Silin, computer science students at the University of Illinois Urbana-Champaign. With their first iteration of QuillBot, they created a tool to help students, particularly those learning the English language, by paraphrasing and rearranging sentences of different pieces of writing. The tool became known as QuillBot Paraphraser. This tool was such a success, that in 2019 QuillBot Paraphraser hit over one million users. In 2020, the QuillBot team began to expand its capabilities by developing QuillBot Summarizer (QuillBot, 2025). This second tool was designed to quickly condense articles, reports, and documents down to the key points with a click of a button. In July 2024, QuillBot launched a new AI detector that is able to differentiate between AI-generated text and text enhanced by AI tools, a new standard for AI usage in education. The platform has continued to grow into 2025, with now over 5.3 billion annual Paraphraser inputs and 56 million monthly active writers using the product (Amplitute, 2025).

QuillBot's AI Detector allows the user to enter text to receive an in-depth report of the probability of AI tools having been used in its creation. It aims to provide a sharper and less confusing analysis compared to other detectors by distinguishing between text created with generative AI tools (like ChatGPT, Gemini, Claude) and text refined using assistive tools such as paraphrasers or grammar checkers. QuillBot claims that it is a more precise AI detector meant to avoid false positives and cultural bias. QuillBot's AI Detector works by recognizing and flagging patterns, repetitive words, and unnatural flows in writing. While QuillBot AI trains on datasets related to grammar, spelling, punctuation, tone, sentence structure, and clarity, the detector specifically analyzes text to identify different levels of potential AI involvement. It provides a percentage score for four categories: AI-generated, AI-generated and AI-refined, Human-written and AI-refined, and Human-written. The model is trained to identify text generated by tools like ChatGPT, Gemini, and Claude, as well as text refined by paraphrasers or grammar checkers, including its own. The detector requires a minimum of 80 words and analyzes the text quickly, providing a report in seconds. It supports AI detection for English, French, Spanish, German, and Dutch. QuillBot's AI Detector is available in both a Free and a Premium version. The Free version allows users to scan unlimited documents up to 1,200 words per scan and only permits uploading and reviewing one document at a time; exceeding the word limit results in an error and a blurred percentage for the detected AI text. The Premium version, included as part of the QuillBot Premium subscription, removes these limitations, offering unlimited word scanning and the ability to upload and review up to 20 documents at a time (Quillbot, 2025).

Polygraf AI

Polygraf.AI was founded in 2021. The company's founders are Yagub Rahimov and Vignesh Karumbaya. Polygraf.AI is self-described as an AI-native data integrity, monitoring and governance solution operating under a Zero Trust framework. They focus on verifying digital content and protecting sensitive data, aiming to help organizations use AI tools safely while maintaining privacy and ethics. Polygraf AI's mission is to safeguard mission-critical operations against deepfakes and synthetic content. It can be seen as a multi-layer "truth engine" that scores, explains and even quarantines suspect communications in real time. The company has received recognition, including being named 'Best in Show' at the SXSW 2025, where it also won in the Enterprise, Smart Data, FinTech & Future of Work categories (South, 2025). Polygraf AI was also recognized as a Top AI Governance Product and Top AI Content Detection Product in the 2024 Product Awards by Products That Count and is ranked as a Top Analytics Startup to Watch Globally by Dealroom.

Polygraf AI is based in Austin, Texas, United States, and is a private, pre-seed company with 11-50 employees.

To use Polygraf.AI, users must first create an account which includes the free version. The tool operates using tokens instead of words or characters. Polygraf.AI recommends submissions to be under 5,000 words in order to ensure the best processing speeds and accuracy. Both free and paid plans only support plain text input or uploading one file at a time. Polygraf.AI provides multiple tools in their AI Detector that are designed to assess different aspects of a submitted text. One of the features of Polygraf.AI's is Source Detection, which identifies the specific AI models used (such as GPT-4, Gemini, or Claude) that was used to create the writing. Another feature of Polygraf.AI's is the Writing Analysis, which evaluates text structure, vocabulary, and readability. Paid plans unlock additional tools such as the AI Highlighter, which flags any of the AI-generated content found in the text; the Content Humanizer, which rewrites text to appear more human-like; and the Copyright Detector, which identifies original source material and provides citation links.

Other capabilities of Polygraf.AI's AI Detector include a Deception Filter, which detects intentional text manipulation, and the Humanized Text Analysis feature, which analyzes how closely text mimics human writing. All features are powered by proprietary AI that is able to perform forensic linguistic analysis and real-time probability assessments (Polygraph, 2025). Once text has been submitted, Polygraf.AI's not only identifies content as human or AI-generated but also provides a linguistic forensic breakdown of the text via its Writing Analysis feature. The analyzed text is then returned with an AI Source Detection score. As of May 2025, supported models include Open AI (GPT 3.5, GPT 4), Google Gemini, Meta Llama 3, Mistral AI, Claude. Polygraph AI will state the source; for example: 74% Claude, 26% OpenAI ChatGPT, or 100% Human Created. Polygraf.AI's AI Detector also provides a Cognitive Flow Index (CFI) rating that measures how naturally flowing and readable the content is. It also breaks down Writing Analysis into different categories: Readability, Syntactic Tree Depth, Vocabulary Richness, Perplexity, Average Sentence Length, and the overall Cognitive Flow Index (Polygraph, 2025).

The tool operates on a token-based system, which varies by feature. For example, a 70-word submission may use approximately 282 tokens for AI Source Detection, 94 tokens for Writing Analysis, 235 tokens for AI Highlighting, and another 282 tokens for Copyright Detection (available only to premium and enterprise users). Polygraf.AI's Free plan provides 2,500 tokens monthly and includes AI Detector, Source Detector, Writing Analysis, and Cognitive Flow Index. The Standard plan includes 150,000 tokens and adds access to the Content Humanizer. The Premium plan offers 1 million tokens and adds two article generations per week, AI Highlighter, and Copyright Detection. The Stealth Writer plan provides 10 million tokens and adds advanced capabilities like real-time internet-connected content generation, rewriting, summarizing, referencing, and downloadable content (Polygraph, 2025).

Capabilities and Limitations of AI Detection Tools

In a study that analyzed three AI detection software packages, Quillbot, Perplexity, and ZeroGPT, Lukianenko et al. (2024) found that if the text was generated by AI, all three services indicated it with 100% accuracy. If the text was written by a human, the study found that it was identified as human writing in most cases, but there were some false positives, indicating unreliability. If the text was written by either a human or AI, but then went through editing using AI tools, the detection programs will indicate the presence of AI, in most cases, to some extent. This study utilized only English language texts in testing. Elkhatat et al. (2023) tested five AI detection tools including OpenAI, Writer, Copyleaks, GPTZero, and CrossPlag.

Findings of the study revealed that the tools were more accurate in correctly identifying AI-generated text than human-written text.

While there are a variety of techniques and tools available for detecting AI-generated content, the literature overall suggests that none are foolproof (Uzun, 2023). Lieberman (2024) notes that AI detectors never fully disclose their methods and none are 100% accurate. Some researchers argue that AI detectors are highly accurate and a reliable way to assess academic dishonesty (Karnalim, 2025; Paullet et al., 2025; Paullet et al., 2024), while others highlight the shortcomings in them, such as false positives and potential biases against non-native English writers, that can erode trust in the tools themselves (Rafiq et al., 2025; Biondi-Zoccai et al., 2025; Chaka, 2024, Liang et al., 2023). The AI-generation tools continually evolve as they are trained with more data and undergo improvements to algorithms and features. AI detection tools must keep pace with these changes and continually improve (Karnalim, 2025).

Yeadon et al. (2023) found that AI tools were capable of writing high-quality physics essays that achieved a top score, for free and within just a few seconds. However, they note that LLMs such as ChatGPT are ultimately text generation software that may struggle with answering questions that require complex numerical or graphical components. There are also ethical implications associated with detecting AI-generated content. There is a concern about false accusations if students are penalized based on probable violations rather than concrete evidence (Reinhardt & Guartuche, 2024; Deans et al., 2024). Uzun (2023) calls for further research to develop effective and ethical methods for detecting AI-generated content.

Methodology

GPTZero, QuillBot and Polygraf AI were tested to determine the capabilities and accuracy of the AI detecting tools. GPTZero claims to have a 99% accuracy rate of detecting AI usage in writing, while QuillBot has an 80-99% accuracy rate depending on the type and length of the words being analyzed and Polygraf AI has a 99% accuracy rate. In order to measure these accuracy rates the researchers created their own dataset and analyzed it using all three tools. There are free and paid versions of all three AI detection tools. To test the samples, paid versions of each of the three tools were purchased. This was done so that larger page lengths of writing could be analyzed and the number of samples could be increased. The free versions of the tools have a limit on the number and size of the writing samples that can be analyzed each day and the number of features available in each tool. A previous study by the researchers was initially carried out to evaluate the validity of GPTZero only (Paullet, et al, 2024).

A total of 100 papers were written for this study. The human samples were created from the authors prior writing. The researchers' work used for this test was written prior to ChatGPT's release date on November 30, 2022. This was done to make it impossible for any of the researchers' work to have been produced by artificial intelligence. Out of the 100 samples 99 had over 1000 words per sample. One sample was created with only 71 words on purpose to determine if the limited number of words could still be detected.

The 100 writing samples created by Researcher A were anonymized and numbered 1 through 100 and put into a Microsoft Excel spreadsheet. The samples were then given to Researcher B to run through GPTZero. GPTZero allows for a single document to be uploaded or numerous documents at one time. Researcher B uploaded the 100 documents at one time to GPTZero and ran a report showing the results of the experiment. Researcher B then uploaded the 100 documents into QuillBot. Lastly, the 100 documents were uploaded to Polygraf AI for analysis. After the three tools analyzed the writing samples the results were then loaded into Excel for comparison of the tools. The results of the study are discussed below.

Results and Discussion

To test the accuracy and validity of GPTZero, QuilBot and Polygraf AI, the researchers devised an experiment. Using 100 distinct writing samples, this study evaluated the performance and precision of the AI tools to determine text that was created by humans, artificial intelligence, and a mix of both human and AI created content in May 2025.

The study provides a distinct baseline for genuine human writing by utilizing human-generated texts from before ChatGPT and other AI tools were developed. A thorough examination of how AI detection tools comprehend various forms of writing is made possible by the inclusion of fully AI-generated samples, and human samples and a sample that had a combination of both AI and human writing. A solid basis for assessing AI detection tools is provided by this structured dataset. Researcher A created 100 writing samples to include human, AI and a combination of human and AI writings. There was a total of 36 human generated samples.

An additional 34 AI writing samples were created using ChatGPT, Gemini and Microsoft Co-Pilot and 30 writing samples were created using human samples from the authors writing and AI generated content mixed in the same document. The samples were then given to Researcher B to run through GPTZero, QuillBot and Polygraf AI respectively. Paid versions of all three tools were purchased due to the number of samples that were being tested and the limitations on the free versions of each tool. The samples were uploaded to each of the three tools being tested which are broken down below for accuracy comparison.

The tables below list the results of the output of each tool provided by Researcher B compared to Researcher A's sample. Table 1 shows that GPTZero was able to identify the 35 of the 36 human created samples as being "human". Of the 35 samples, 35 showed a 99% chance of being written by a human. As mentioned previously, Researcher A created one human sample under 100 words to see if the AI tools could analyze the document. The sample that was under 100 words GPTZero labeled as "not determined". This was due to the sample size being too small. GPTZero next appropriately identified all 34 AI generated writing samples. Of the 34 samples, 28 resulted in a 99% chance of being generated by AI and 6 samples showed 95-98% probability of being generated using AI.

The final 30 samples were a combination of human and AI writing mixed in each sample. Of the 30 samples 8 had a 99% accuracy of detecting which parts of the sample were written by a human and which were created using AI. For instance, one sample was created by using 62% human written content and 38% AI generated content in the sample. By dividing the total number of words in the document by the total number of words used in the human written content and then by the total amount of AI generated content, the percentages were determined. Six of the samples that Researcher B tested matched the results with a 99% accuracy rate, sixteen samples showed a 95-98% accuracy rate and 6 samples showed an 80-94% accuracy rate of matching the original samples from Researcher A.

When creating the mixed documents, Researcher A either included all of the human content in the sample first, then the AI content, or the other way around. The accuracy rate of those samples that followed the same pattern was 95% or higher. The samples with AI content interspersed between paragraphs of human-generated content were those with an 80- 95% success rate. Even at 80%, the result was sufficiently high enough to demonstrate the difference between AI-generated and human-created content.

Table 1. Validated Dataset for GPTZero

Samples – Researcher A	GPTZero	1% Margin	2-5% Margin	5-20% Margin	Not able to Analyze
Human	36	35	0	0	1
AI	34	28	6	0	0
Mix of Human & AI	30	8	16	6	0

Table 2 demonstrates that QuillBot correctly classified 35 of the 36 samples that were created by humans as "human." Out of 35 samples, 29 had a 99–100% chance of being written by a human, 3 had an accuracy of 95–98%, and 3 had an accuracy of 80–95%. QuillBot marked the sample that was under 100 words as "not enough text". The sample was too small for QuillBot to analyze. All 34 AI-generated writing samples were then correctly identified by QuillBot. Of the 34 samples, 25 had a 99% chance of being AI-generated, 8 had a 95–98% chance, and 3 had an 80–95% accuracy rate. Each of the final 30 samples included a mix of AI and human writing. Six of the thirty samples matched the numbers produced by Researcher A with 99% accuracy. Five samples matched the original samples from Researcher A with an accuracy rate of 80-94%, while 19 samples had an accuracy rate of 95-98%.

Table 2. Validated Dataset for Quillbot

Samples – Researcher A	QuillBot	1% Margin	2-5% Margin	5-20% Margin	Not able to Analyze
Human	36	29	3	0	1
AI	34	25	8	3	0
Mix of Human & AI	30	6	19	5	0

Table 3 shows that 36 of the 36 human-created samples were correctly identified as "human" by Polygraf AI. Twelve samples had an accuracy of 96–98%, thirteen samples had an accuracy of 80–95%, and eleven samples had a 99–100% chance of being written by a human. Despite the short text, Polygraf AI was the only tool test that was able to accurately identify the sample which was less than 100 words, as being written by a human. All 34 AI-generated writing samples were then correctly identified by Polygraf. Six of the 34 samples had an accuracy rate of 80–95%, 15 had a 95–98% chance, and 6 had a 99% chance of being AI-generated. A combination of AI and human writing was present in each of the final 30 samples. Of the thirty samples, two had a 99% accuracy rate in matching the numbers generated by Researcher A. With an accuracy rate of 80–94%, three samples matched the original samples from Researcher A, while 13 samples had an accuracy rate of 95–98%. Polygraf AI matched the greatest number of samples with accuracy to the original samples developed by Researched A.

Table 3. Validated Dataset for Polygraf AI

Samples – Researcher A	Polygraf AI	1% Margin	2-5% Margin	5-20% Margin	Not able to Analyze
Human	36	11	12	13	0
AI	34	13	15	6	0
Mix of Human & AI	30	2	13	3	12

This experimental study answered the research question by demonstrating the similar outcomes of all three tools when tested against the original 100 samples created by Researcher A. GPTZero and QuillBot have correctly identified 35 out of the 36 human created samples. The sample that was less than 100 words was not recognized by either tool, which marked it as inconclusive. However, Polygraf AI did analyze the sample and correctly identified it as being created by a human. Despite the fact that all three tools could

identify which samples were written by humans, AI, or a combination of both, Polygraf AI was the most accurate in identifying the writing origin of all samples.

Conclusion and Future Research

Access to accurate AI detection tools has become critical as there is a clear need to distinguish between AI-generated and human created content. This is particularly important in the field of education, where academic integrity is at stake (Cotton et al., 2023). The purpose of this study was to ascertain whether GPTZero, QuillBot, and Polygraf AI could reliably identify whether writing samples were produced by AI, humans, or a combination of the two. Based on the experiment's results, the researchers concluded that GPTZero, Quillbot and Polygraf AI could reliably make this determination as of May 2025. Of the three tools that were examined, Polygraf AI is the most recent, having been developed in 2021 yet was the most accurate. It's crucial to remember that this could alter at any time based on how quickly developers can update the tools to reflect the most recent developments in AI and how quickly AI advances.

Using only 100 samples could be seen as a limitation to the study. Future research is needed to determine the accuracy of content created with a combination of human and AI generated content. In the future the researchers could test a larger number of AI detecting tools. Artificial intelligence is here to stay. It is crucial for educators to adopt the technology and come up with ways to stop the improper use of AI in the classroom. In addition to learning how to express themselves in writing, students need to be taught about the dangers of using AI. As AI capabilities advance, these tools will require constant updates, particularly as generative models get better at simulating human writing. Students' interactions with technology may significantly change if AI detection tools are framed as instructional tools rather than punitive measures. Teachers can assist students in understanding when and how AI tools can be used responsibly while preserving originality and critical thinking by incorporating discussions about AI-assisted writing into their curricula. Institutions could think about policies that encourage the ethical application of AI rather than completely opposing it. A balanced approach might result, for instance, from openness in AI-assisted work, organized talks about its function in writing and research, and explicit guidelines for student contributions.

References

Amplitute. (2025). Using data to stay close to the customer. https://amplitude.com/case-studies/quillbot

Basha, J. Y. (2024). The negative impacts of AI tools on students in academic and real-life performance. *International Journal of Social Sciences and Commerce, 1*(3), 1-16. https://doi.org/10.51470/IJSSC.2024.01.03.01

Bates, T., Cobo, C., Marino, O., & Wheeler, S. (2020). Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education, 17*(42). https://doi.org/10.1186/s41239-020-00218-x

Biondi-Zoccai, G., Cazzaro, A., Cobalchin, E., D'Auria, D., Ardizzone, G., Giordano, S., Mirzoyev, U., Seferovic, P.M., Bajraktari, G., & Muraru, D. (2025). Artificial intelligence tools for scientific writing: The good, the bad and the ugly. *Top Italian Scientists Journal.* https://journal.topitalianscientists.org/Artificial_Intelligence_Tools_for_Scientific_Writing_The_Good_The_Bad_and_The_Ugly

Chaka, C. (2024). Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. *Journal of Applied Learning & Teaching, 7*(1), 3-12. http://journals.sfu.ca/jalt/index.php/jalt/index

Chaves, A.P., & Gerosa, M.A. (2022). *Lecture notes in computer science: Vol. 13171. The impact of chatbot linguistic register on user perceptions: a replication study. Chatbot Research and Design.* Springer. https://doi.org/10.1007/978-3-030-94890-0_9

Cotton, D.R., Cotton, P.A., & Shipway, J.R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Deans, K.R., Jones, J., & Harvey, J.B. (2024). Artificial intelligence in higher education: A comparative study on the performance and detectability of AI-generated graduate-level coursework utilizing ChatGPT. *IngentaConnect.* https://doi.org/10.1016/j.iheduc.2024.100986

Elkhatat, A.M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. International Journal for Educational Integrity, 19(17). https://doi.org/10.1007/s40979-023-00140-5

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse? *Frontiers in Education*, 1-4. https://doi.org/10.3389/feduc.2023.1166682

Johnston, H., Wells, R., Shanks, E., Boey, T., & Parsons, B. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity, 20*(2), 1-21. https://doi.org/10.1007/s40979-024-00149-4

Karnalim, O. (2025). Identifying AI-generated code with parallel KNN weight outlier detection. In Vendral, V.E., Cukierman, U.R., & Auer, M.E. (Eds.), *Advanced Technologies and the University of the Future. Lecture Notes in Networks and Systems.* Springer. http://doi.org/10.1007/978-3-031-71530-3_29

Krause, S., Panchal, B.H., & Ubhe, N. (2024). The evolution of learning: Assessing the transformative impact of generative AI on higher education. https://doi.org/10.48550/arXiv.2404.10551

Leckrone, B. (2025, February 27). *One-third of college-age young adults use ChatGTP: Report.* Best Colleges. https://www.bestcolleges.com/news/students-embrace-chatgpt/

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. Computers and Education: Artificial Intelligence, 6(100221). https://doi.org/10.1016/j.caeai.2024.100221

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7). https://doi.org/10.48550/arXiv.2304.02819

Lieberman, G. (2024). The use and detection of AI-based tools in higher education. *Journal of Instructional Research, 13*(4), 70-78. https://doi.org/10.9743/JIR.2024.13.4

Lukianenko, V.V., Shastko, I.M., & Korbut, O.H. (2024). Evaluating AI detection tools for academic integrity in higher education. *Scientific Innovations and Advanced Technologies, 5*(33), 970-978. https://doi.org/10.52058/2786-5274-2024-5(33)-970-978

Miller, W. (2023, February 16). These technologies are brilliant, but we need safeguards against them: Meet the Etobicoke born inventor of the ChatGPT detector. *Toronto Life.* https://torontolife.com/city/these-technologies-are-brilliant-but-we-need-safeguards-against-them-meet-the-etobicoke-born-inventor-of-the-chatgpt-detector/

Nam, J. (2023, November 22). *56% of college students have used AI on assignments or exams.* Best Colleges. https://www.bestcolleges.com/research/most-college-students-have-used-ai-survey/

OpenAI. (2025). *Building an AI-ready workforce: A look at college student ChatGPT adoption in the US.* https://cdn.openai.com/global-affairs/openai-edu-ai-ready-workforce.pdf

Paullet, K., Pinchot, J., Kinney, E., & Stewart, T. (2025). Utilizing GPTZero to detect AI-generated writing. *Information Systems Education Journal, 23*(6), 44-52. https://isedj.org/2025-23/n6/ISEDJv23n6p44.pdf

Paullet, K., Pinchot, J., Kinney, E., & Stewart, T. (2024). Can GPTZero detect if students are using artificial intelligence to create assignments? *Issues in Information Systems, 25*(3), 165-174. https://doi.org/10.48009/3_iis_2024_113

Polygraph (2025). AI content detector. https://polygraf.ai/ai-content-detector

Rafiq, S., Ain, Q., & Afzal, A. (2025). The role of AI detection tools in upholding academic integrity: An evaluation of their effectiveness. *Contemporary Journal of Social Science Review,* 3(1), 901-915. https://contemporaryjournal.com/index.php/14/article/view/379

Reinhardt, K.S., & Guartuche, O. (2024). Curriculum integration of AI technology for student autonomy and self-efficacy. *The Texas Forum of Teacher Education, 16*(2166-0190), 41-46. https://txate.org/resources/Documents/The%20Forum%20Volume%2016%20Winter%2024.pdf#page=40

Reiter, L., Jorling, M., Fuchs, C., & Bohm, R. (2025). Student (mis)use of generative AI tools for university-related tasks. *International Journal of Human-Computer Interaction*, 1-14. https://doi.org/10.1080/10447318.2025.2462083

South by Southwest (2025). SXSW announce winners of 2025 pitch competition. https://fox59.com/business/press-releases/cision/20250309NY36989/sxsw-announces-winners-of-2025-pitch-competition/#:~:text=SXSW%20Pitch%202025%20also%20honored,for%20%22Best%20In%20Inclusivity.%22

Stanford, J. (2025). How disruptive is AI in students' academic writing in higher education? *Studies in Technology Enhanced Learning, Special issue: Generative AI and education, 4*(1). https://doi.org/10.21428/8c225f6e.0390b853

Sullivan, M., Kelly, A., McLaughlin, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. https://doi.org/10.37074/jalt.2023.6.1.17

Tian, E. (2023). Ways to by-pass AI detection? https://gptzero.me/news/gptzero-by-passers

QuillBot (2025). About QuillBot. https://quillbot.com/about

Uyen, B.T.T., & An, T.T.V. (2025). The impact of AI writing tools on academic integrity: Unveiling English-majored students' perceptions and practical solutions. *AsiaCALL Online Journal, 16*(1), 83-110. https://doi.org/10.54855/acoj.251615

Uzun, L. (2023). ChatGPT and academic integrity concerns: Detecting artificial intelligence generated content. *Language Education & Technology, 3*(1), 45-54. https://www.langedutech.com/letjournal/index.php/let/article/view/49

Vieriu, A.M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences, 15*(3), 343. https://doi.org/10.3390/educsci15030343

Walczak, K., & Cellary, W. (2023). Challenges for higher education in the era of widespread access to generative AI. *Economic and Business Review, 9*(2), 71-100. https://doi.org/10.18559/ebr.2023.2.743

Xiao, P., Chen, Y., & Bao, W. (2023). Waiting, banning, and embracing: An empirical analysis of adapting policies for generative AI in higher education. https://doi.org/10.2139/ssrn.4458269

Yeadon, W., Inyang, O., Mizouri, A., Peach, A., & Testrow, C. (2022). The death of the short-form physics essay in the coming AI revolution. *Physics Education, 58*(3), 1-13. https://doi.org/10.1088/1361-6552/acc5cf