EVALUATING AI DETECTION TOOLS FOR ACADEMIC INTEGRITY IN HIGHER EDUCATION

Abstract. The advent of artificial intelligence (AI) in academic settings has presented educators with unprecedented challenges in maintaining academic integrity. This study examines the efficacy of three AI-powered text-checking services—Quillbot, HiveModeration, and ZeroGPT—in detecting AI-generated content and alterations in academic texts. The investigation employs five text variants: text generated by ChatGPT, original pre-AI era text, AI-rephrased text (Quillbot), Quillbot-rephrased original text, and edited (minor edits) original text by DeepL and Grammarly services. All three services exhibit 100% accuracy in detecting AI-generated text while failing to identify AI signs in original English text. Surprisingly, when examining the original Ukrainian text, ZeroGPT detects AI signs with 100% accuracy. Quillbot's paraphrasing substantially alters the text, leading to varied detection rates across the services. Notably, AI-induced alterations influence detection results, exemplified by Quillbot showing high detection rates in ChatGPT-generated and original text (79% and 60% respectively). Lastly, editing original text by DeepL and Grammarly does not trigger AI detection. In conclusion, this study offers several key findings and recommendations. Firstly, while AI detection tools excel at identifying machine-generated text, they exhibit vulnerability to bypassing through paraphrasing and specific textual styles. Secondly, false positive results underscore the unreliability of these services, highlighting the need for cautious interpretation. Thirdly, education on ethical AI use is imperative, distinguishing
between permissible and impermissible behaviors and integrating AI tools into assignments transparently. Fourthly, adapting educational systems to accommodate AI tools' use and fostering collaborative, interdisciplinary projects can enhance students' AI literacy and ethical awareness. Finally, comprehensive policies outlining ethical AI usage guidelines are essential to navigate the ethical complexities of AI integration in education effectively.

**Keywords:** AI detection, academic integrity, educational technology, ethical use of AI.

Лук’яненко Валентина Володимирівна кандидат психологічних наук, доцент кафедри англійської мови технічного спрямування №2, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Берестейський проспект (Перемоги), 37, м.Київ-56, 03056, тел.: (044) 204-83-57, https://orcid.org/0000-0003-3748-2616

Шастко Ірина Миронівна кандидат психологічних наук, доцент кафедри суспільних наук, Івано-Франківський національний технічний університет нафти і газу, вул. Карпатська, 15, м. Івано-Франківськ, 76019, тел.: (050)338-88-17, https://orcid.org/0000-0001-6963-2899

Корбут Оксана Григорівна старший викладач кафедри англійської мови технічного спрямування №2, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Берестейський проспект (Перемоги), 37, м.Київ-56, 03056, тел.: (044) 204-83-57, https://orcid.org/0000-0001-6671-3925

**ОЦІНКА ЗАСОБІВ ВИЯВЛЕННЯ ШІ ДЛЯ ЗАБЕЗПЕЧЕННЯ АКАДЕМІЧНОЇ ДОБРОЧЕСНОСТІ У ВИЩІЙ ОСВІТИ**

**Анотація.** Поява штучного інтелекту (ШІ) в академічній сфері створила виклики у вищій освіті, пов’язані з дотриманням умов академічної доброчесності. У статті досліджується ефективність трьох сервісів перевірки тексту на наявність ознак ШІ — Quillbot, HiveModeration та ZeroGPT. Дослідження охоплює п’ять текстів: текст, створений за допомогою ChatGPT; оригінальний текст; текст, написаний ШІ та перефразований за допомогою ШІ (Quillbot); перефразований оригінальний текст за допомогою Quillbot; та відредагований (незначні зміни) оригінальний текст за допомогою сервісів DeepL і Grammarly. Всі три сервіси показують 100% точність у перевірці тексту, створеного за допомогою ШІ, також вони не ідентифікують ознаки ШІ в оригінальному англійському тексті, але показують помилково позитивний результат у перевірці оригінального українського тексту. Парафраз від Quillbot значно змінює текст, що призводить до різних рівнів виявлення
Introduction. At the beginning of the AI era, educators have encountered new challenges. Everyone understands that students have a high level of technical and computer skills; they keep track of new developments and technologies and try to experiment with them. If these technologies facilitate learning, then students are likely to use them. It seems that ChatGPT has entered the academic sphere, as well as many other areas, so unexpectedly that educators have not fully understood how to adapt to these changes, how to assess student's work on one hand objectively, and on the other hand, how to prevent them from being unjustly accused of ethics violations. At the initial stage, everyone probably starts asking questions: What services can be used to check the originality of students' work? For a long time, students' work has been checked for plagiarism, and these services are reliable enough. If there are AI services that produce text, there is software that can detect signs of machine text in student work. Indeed, there are many such services. But how reliable are they? Can we trust them, can we implement them in educational activities?

While existing literature acknowledges the emergence of AI detection tools in academia, few studies offer comprehensive evaluations of their reliability and effectiveness, particularly in diverse linguistic contexts and across various forms of text manipulation. This study aims to fill this gap by conducting a rigorous analysis of AI detection tools' performance, considering factors such as language diversity, text manipulation techniques, and implications for academic assessment.

Literature review. As education continues to evolve, new technologies are emerging to revolutionize the learning process. Tan [1] states that ChatGPT as a transformative technology could completely reshape the way universities educate
and learn. With its advanced natural language processing capabilities, ChatGPT holds immense potential for teaching and learning at universities. It can effectively interact with students, respond to their queries, and provide personalized support [2]. Moreover, ChatGPT can foster collaborative learning, creativity and the development of research skills [3]. By using ChatGPT, teachers can create dynamic, interactive learning environments that promote student engagement [4; 5]. This technology can significantly enhance students' learning processes, enabling educators to tailor their teaching approaches and interventions accordingly.

Incorporating AI tools into university teaching and learning can also help bridge the gap between technology and education, providing a more efficient learning experience for students. They can help educators develop course plans and generate tests, lists of questions at different levels, and evaluation rubrics in a short period of time. The integration of ChatGPT and other services into university teaching and learning has the potential to provide personalized support, encourage collaboration and promote a more engaging and effective learning environment. However, AI services should not be seen as a replacement for human interaction in learning and achieving their academic goals. In addition, educators must be cautious when using ChatGPT to ensure the credibility of the information sources provided by this language model. Educators should also be aware of potential limitations and challenges that may arise when using ChatGPT in education, such as concerns that students may become over-reliant on AI services, leading to diminished critical thinking and problem-solving skills. These limitations and challenges highlight the need for a balanced approach, where ChatGPT is used as a tool to support and enhance student learning while encouraging creative and critical thinking and active engagement.

The ethical use of ChatGPT, particularly in the context of students cheating by having the AI tool write essays for them, is a significant concern. It is crucial for educators and institutions to implement measures that detect and prevent such unethical use of AI tools [6; 7; 8]. This can be done through a combination of measures, such as AI detection software, and educational strategies that promote academic integrity. Educators should emphasize the importance of original thinking, research, and ethical writing practices to students, discouraging them from relying on AI tools for the sole purpose of cheating. By providing assignments that require personal analysis, creativity, and application of knowledge, it is possible to create assessments that are difficult for AI tools to replicate. This approach encourages students to engage in their own learning actively, promotes academic integrity, and reduces overreliance on AI services like ChatGPT for unethical purposes [7; 9]. The responsible integration of AI services into university education requires ongoing reflection, dialogue, and collaboration among policymakers, researchers, educators, and technology experts.

It is an arguable question whether universities should or shouldn’t use AI detection tools in the evaluation of student work. Weber-Wulff et al. [10] checked
12 publicly available tools and two commercial systems (Turnitin and PlagiarismCheck) and came to the conclusion that they are neither accurate nor reliable and have a main bias towards classifying the output as human-written rather than detecting AI-generated text. On one side, these tools can efficiently identify potential instances of AI-generated content, helping educators to maintain academic integrity. On the other side, it's important to recognize that these tools are not always reliable. Therefore, educators should approach the use of AI detection tools with a critical mindset and supplement their findings with human judgement and contextual understanding.

This work aims to analyze the results of three text-checking services for AI detection (Quillbot, HiveModeration, ZeroGPT). Five text variants are proposed to be checked: text generated by ChatGPT; original text written before the AI era; rephrased text by AI service (Quillbot); Quillbot-rephrased original text; and edited (minor edits) original text by DeepL service.

Results and Discussion
Our results are presented in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>HiveModeration</th>
<th>QuillBot</th>
<th>ZeroGPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChatGPT generated text</td>
<td>99.9%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Original text</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>AI-generated paraphrased text (Quillbot)</td>
<td>1.9%</td>
<td>79%</td>
<td>39.93%</td>
</tr>
<tr>
<td>Original paraphrased text (Quillbot)</td>
<td>0%</td>
<td>60%</td>
<td>18.11%</td>
</tr>
<tr>
<td>Original proofread text (Deeple and Grammarly)</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

It is important to mention, that only English texts were checked. Our results show that AI detection tools detect machine-generated text with 100% accuracy. Checking the original text showed that the three services did not detect AI signs in the English text. Surprisingly, after checking the original Ukrainian text, one of the three services (ZeroGPT) detected AI signs in the text (100%). The Quillbot paraphrase significantly alters the text. We observe very different results in checking AI-generated text by three services (79% Quillbot, 1.9% HiveModeration, 39.93% ZeroGPT). Analysis of the paraphrase of the original text also showed different results in checking by different services. However, we observe that changing the text by the AI service affects the checking result, for example, the Quillbot service showed a high percentage (79% and 60%) in two checks. The last check of the text edited by the DeepL service (minor edits) did not detect AI signs in the original text, which was confirmed by using another editing service - Grammarly.
On the one hand, AI detection programs are vulnerable. They can be bypassed by paraphrasing, using specific styles, rhetorical questions, etc. On the other hand, positive results in checking original texts also occur in cases of queries not in English, checking legal documents, plans, programs, or in the case of checking essays written by non-native English speakers, who use standardized phrases and language clichés in speech. Our research and literary sources [7; 10; 11; 12] indicate that AI plagiarism-checking programs are unreliable. These programs use a different approach, which differs from plagiarism checking (finding matches between the text and sources published online). They assess the degree of text conformity to the most common results of large language models and provide no means for independent verification of the originality of student work.

According to Harwell's [13] calculations, if the error rate of such services is at least 1% (as indicated by Turnitin developers), considering that, on average, a student writes 70 written works during their studies, there is a 50.5% chance that they will be unjustly accused of AI usage at least once during their entire education.

The study [12] revealed significant discrepancies in AI detectors' ability to differentiate between human and AI-generated essays. Over half of the essays written by non-native speakers for the TOEFL were flagged as AI-generated, while over 90% of essays written by native English-speaking eighth graders were classified as human-generated. This discrimination arises from the detectors' reliance on text perplexity, a measure of how predictable the next word in a sentence is for a generative language model. Large language models like ChatGPT, trained to produce low perplexity text, risk being mistaken for AI-generated when humans use common words in familiar patterns. Non-native English speakers are particularly susceptible due to their tendency to employ simpler word choices. Interestingly, the researchers found that by rewriting TOEFL essays using more sophisticated language, they could evade detection by AI detectors, highlighting the potential for unintended consequences in using such technology.

Conclusions and recommendations:

Based on the analysis of three AI plagiarism-checking services (Quillbot, Perplexity, ZeroGPT), the following conclusions can be drawn:

1. If the text is generated by AI, the services indicate it with 100% accuracy.
2. If the text is written by a human, in most cases, the programs will show that the text does not contain AI signs. However, there are cases of false positive results, indicating the unreliability of these services.
3. If the text is written by either a human or AI but undergoes editing or paraphrasing by AI services, then AI detection programs will, in most cases, indicate the presence of AI in the text to some extent.

Based on the literature review and our research results, we have outlined the following recommendations for higher educational institutions:

1. Educate students on the ethical use of AI services, distinguishing between ethical and unethical behaviors such as writing essays and generating code...
versus data fabrication. Emphasize the ethicality of activities like planning, generating ideas, and editing when utilizing AI tools.

2. Provide assignments where students will use AI tools at one of the stages, taking into account ethical norms and conditions of service usage. Students should indicate which part of the work they performed using AI.

3. Adapt the education and assessment system to new realities. Written works can be done in class, or if learning occurs online, defenses of such works can be conducted; students should use Google Docs and show the editing history of the work.

4. Encourage collaborative projects that involve the interdisciplinary use of AI tools. Facilitate partnerships between students from different disciplines to work on projects that leverage AI for problem-solving, research, or creative endeavors. Emphasize the importance of diverse perspectives and skill sets in leveraging AI technology to address complex challenges and foster innovation.

5. Provide faculty members with training and resources on integrating AI tools into their teaching practices. Offer workshops, seminars, or online courses to enhance faculty proficiency in utilizing AI technologies effectively and ethically.

6. Incorporate a comprehensive policy that outlines guidelines for the ethical and responsible use of AI in task execution and assessment within educational programs. This policy should address issues such as proper citation and acknowledgment of AI-generated content, guidelines for integrating AI tools in assignments and examinations, and procedures for handling instances of suspected AI usage.

**Limitations:**

One limitation of this study is the focus on English text, which may not fully capture the complexities of AI detection in other languages. Additionally, the evaluation of AI detection tools is limited to a select number of services and may not represent the entire landscape of available tools. Furthermore, the study does not explore the potential biases embedded within AI detection algorithms, which may impact their performance and reliability.

**References:**


Література:


