ANALYZING THE EFFECTIVENESS OF CONTENT DETECTION PROGRAMS GENERATED USING AI (ARTIFICIAL INTELLIGENCE)

Marius-Cătălin MARIȘ

”Henri Coandă” Air Force Academy, Braşov, Romania

DOI: 10.19062/1842-9238.2025.23.2.6

Abstract:*Taking into account the current context in the technological field, this paper aims to emphasize the way in which, the methods of combating and detecting the use of artificial intelligence in the academic environment are or not effective and whether the old methods of approaching this environment, are still effective today.*

Keywords: *Artificial intelligence, Chatbots, AI Detection, Effectiveness, Digitalization, Technology, Educational*

1. INTRODUCTION

With the growing need for technological development and an increasing level of innovation in the field of digitization, numerous inventions have laid the foundations for a new approach on a global scale. Among these fundamental breakthroughs is the implementation of artificial intelligence in various systems and processes. Throughout this paper, we aim to obtain as objective an analysis as possible on how the presence/use of artificial intelligence in academia can be detected in a way that inspires certainty and efficiency. In more detail,the aim of this research is to analyze how renowned programs for the detection of AI-generated content manage to determine whether or not an article is generated, whether or not these methods are effective, and whether or not the results of our study can have an influence in demonstrating that current methods used in education should be rethought.

This study does not want to shift the focus away from the negative effects that the use of artificial intelligence in the realization of academic work can generate, but it wants to provide a different perspective on the methods by which to combat this topical phenomenon.

The study conducted by Vieriu and Petrea, on the impact of artificial intelligence on students in the academic environment, brings the following statistics to the fore, based on the questioning of 85 students in the fields of medical and aerospace engineering:

- 96% of surveyed students recognized the use of AI in the educational process
- 88.2% use virtual assistants (ChatGPT, Siri, etc.)
- 17.6% recognized using AI to generate content (the actual percentage is possibly much higher)
- 16.5% believe that AI has a negative effect on critical thinking (Vieriu and Petrea, 2025).

Analyzing the aforementioned statistics and quantifying the results of the study to the total number of students in the educational units, we highlight the following, according to the online statistics page, AIPRM:

- 53% of higher education students use artificial intelligence to generate content used in various materials
- 24.11% of students reported that materials made with AI have been detected and treated as fraud (6.44% in private high schools and 15.2% in public high schools)
- 51% of teachers believe that using AI in education will have a positive impact (AIPRM, 2025)

Thus,one fact can be for sure inquired: the use of AI in the accomplishment of tasks outlined in the line of education, especially in the accomplishment of work requiring content production is evident.

Although Romania's objectives in this direction are to promote the use of artificial intelligence in numerous ways that could streamline the educational process, such as: much faster browsing on the Internet, taking suggestions and implementing them in organizational and mechanization functions (making tables, graphs, statistics, etc.), the level of use of AI in content generation is becoming increasingly alarming. A consequence of this is the implementation of programs that can detect AI-generated content and their implementation in educational institutions. Our goal is to demonstrate whether or not this method is workable through the following approaches:

1. We will generate passages with different AI tools such as ChatGPT, Gemini, QuillyBot.
2. We will analyze the results of passing these passages through AI detection programs such as QuillyBot, Undetectable.AI.
3. We will apply minimal modifications to the AI generated texts to make them less detectable.
4. We will design various conclusions based on the obtained results.

2. ARTIFICIAL INTELLIGENCE AND CONTENT GENERATION TOOLS

In order to be able to analyze and discuss the results obtained through this article, it is necessary to know the conceptual elements underlying the concept of artificial intelligence, but also to be able to analyze in detail the context in which this technology has become so useful and used among young people for various purposes, with emphasis on that of generating content for school tasks.

What is artificial intelligence? In the scientific paper done by Rituraj Mahato artificial intelligence is defined as a replica of the way human intelligence works, but does not provide biologically observable methods. (Mahato, 2022). This concept is also defined in the paper titled: „Introduction to Artificial Intelligence" as a branch of computer science that helps computers mimic human behavior in order to generate better performance in science and technology. Replicating human intelligence, solving knowledge-intensive tasks, building machines that can perform tasks that require human intelligence, and creating systems that can learn on their own are some specific goals of AI." (Ghosh et al., 2021).

What are the most widely used and available ways in which mebers of the education system can get in touch with artificial intelligence? According to the Exploding Topics page we have identified the following statistic: Over 987 million people are using an AI content generation program, of which we can identify a number of 500 million people using ChatGPT. (Cardillo, 2025). Therefore, the most widely utilized type of artificial intelligence made available to internet consumers is certainly the use of an AI product designed to

generate content, mimicking the classical communication model. This artificial intelligence product will be further referred to as ***Chatbot.***

What is a Chatbot? In his study on the functioning of artificial intelligence, Adamopoulou identifies the following representative definitions for the term Chatbot, namely:

- A program that reacts as an intelligent entity when initiating a text or voice communication and has the ability to understand one or more human languages through natural language processing.
- A program designed to simulate conversations with human users over the Internet
- A chatbot is described as the most advanced and intuitive expression of human-computer interactions (Adamopoulou *et al.* , 2020)

How is a chatbot created and how does it work? In order to build a Chatbot it is necessary to have a broad knowledge of computer science and to apply a variety of specific techniques, among which can be: representing the correct information, creating a strategy for organizing the received responses, making a set of neutral responses that are displayed at the moments when the message was not correctly received. Also, each chatbot will be configured according to its needs and purpose based on an extensive process. The same Adamopoulou identifies a logic chain through which a chatbot analyzes and generates the information requested by the user:

A first step in the development of any information system is to realize a logical schema. The whole response generation process starts with the user's request, usually represented by a question. Once this request has been processed by the artificial intelligence, the linguistic component of this chatbot analyzes it in order to be able to identify the user's intention and associate it with an information. Once the artificial intelligence has correlated the information with the user's intent it, depending on how it has been trained, chooses the most appropriate way to interpret the user's intent and chooses a course of action. It can directly generate the correlated information, it can retain the information and feed it into the training process, and it can also request more information from the user to choose the appropriate response. The information generated by this chatbot comes from a generic database that underpins all the responses generated. The response generation component is responsible for organizing and preparing the response in the form of the natural human language used by the user. Another component that is part of the structure of any chatbot is represented by the dialog management component that has the role of maintaining and updating the context of a conversation, which is the current intent, identified entities or missing entities needed to fulfill the user's requests to the users, also having the ability to request missing information by processing clarifications made by the users. (Adamopoulou *et al.* , 2020)

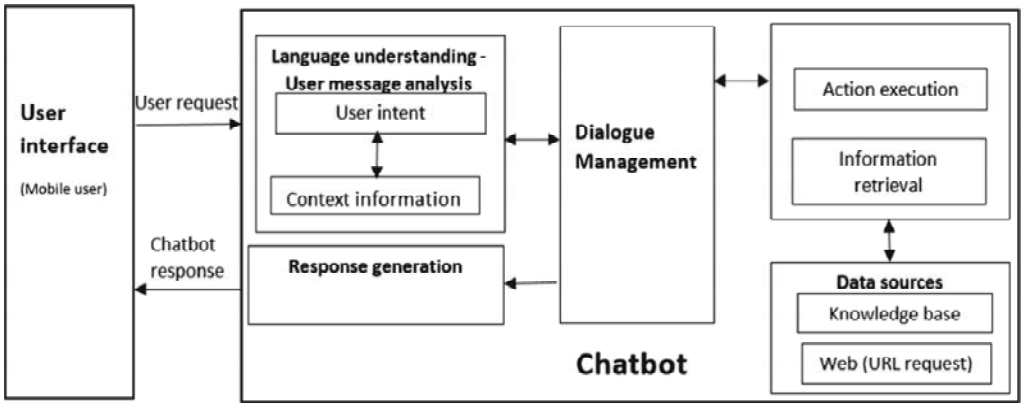


FIG.1 General architecture of a chatbot (Adamopoulou *et al.* , 2020 : 380)

3. STUDY ON THE EFFECTIVENESS OF AI GENERATED CONTENT DETECTION PROGRAMS

In the course of our own analysis of the concept of effectiveness, we were able to conceptually define what this term refers to:

> *"Effectiveness reflects the ability of an action, process or strategy to produce results in line with pre-set objectives. It is often associated with the successful achievement of an explicitly defined goal or mission. Furthermore, effectiveness refers to the ability to produce superior results."* (Mariș, 2024).

Therefore, the effectiveness of a computer program or a software tool (in our case: an AI content detection tool), boils down to the degree of success with which the presence of the use of artificial intelligence can be identified in the process of writing content.

With the accelerating emergence of numerous *Chatbot* tools (defined and explained in the previous chapter), which are increasingly used in the realization of specific papers in today's academic environment, the problem of developing and designing systems that can combat and detect, based on a brief analysis of the content entered, the level of use of AI technologies in content writing has arisen. These detection tools work on the basis of processing the inputted content and checking some parameters that are considered to be defining parameters in delineating the boundary between the activity of artificial intelligence and human (Odri & Yoon, 2023). Specialists have identified as useful parameters in detecting the influence of artificial intelligence, perplexity and writing rate variability. Perplexity is a measure used to evaluate the performance of language production models. Essentially, perplexity determines the ability of a system to choose the most appropriate next word in a sequence.

As mentioned by Xu and Sheng in "*Detecting AI-Generated Code Assignments Using Perplexity of Large Language Models*", perplexity is defined as a:

> *measure of how well a probability model predicts a sample. In the context of code, it's used to evaluate the predictability of the next token or line of code based on a given context.* (Xu & Sheng, 2024: 23156).

A lower perplexity score indicates that the model better predicts the next word in a sequence, while a higher score indicates that the model is less accurate. In other words, the lower the perplexity, the less likely it is that content is generated with AI.

Researches in the techonological fields consider burstiness quality as:

> *The clustering or frequent appearance of certain patterns, identifiers, or constructs in a specific section of the code. For example, in a code module dealing with database operations, there might be a burst of commands and identifiers related to database queries.* (Xu & Sheng, 2024: 23156).

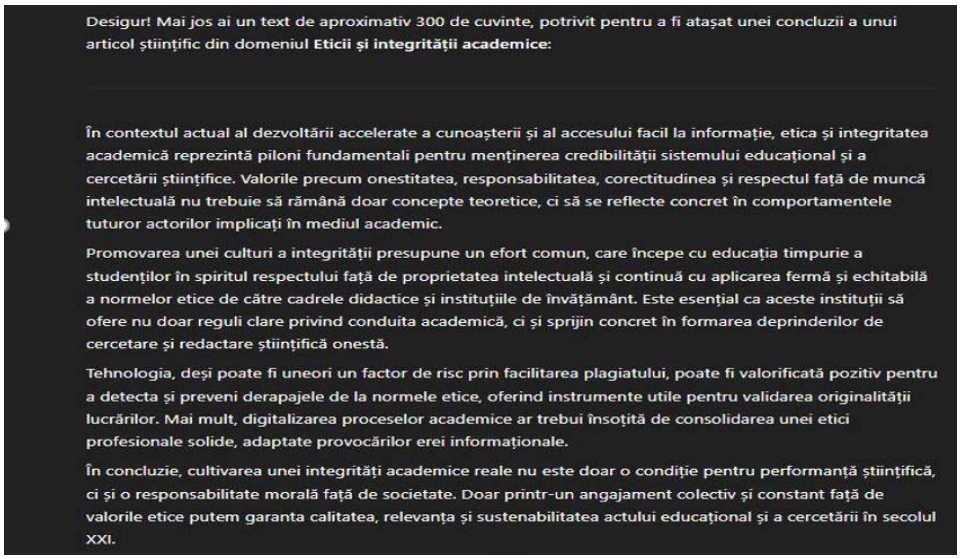
The variability in the rate of writing usually mentioned as burstiness in scintific researches, boils down to the way the structure of a piece of content is written, based on comparing the number of words and phrases used throughout the analyzed text. Therefore, a content generated with AI will be more uniform and easier to detect. (UNIC, 2025).

The steps of conducting this study, regarding the effectiveness of AI-generated text detection system sare presented below:

1. Generating a text consisting of an average of 300 words through the most popular Chatbot on the market, namely ChatGPT.
2. Using 5 AI usage detection software tools for content generation and compare the results: GPTZero, Quillbot, Undetectable.AI, Originality and Writer.

3. Manipulating and editing the text with minimal effort to provide a more natural form of AI generated content
4. Double-checking(re-checking) the results provided by the detection tools.

After using ChatGPT we generated the text sequence shown in Figure 2, in the fictitious context of producing a scientific paper in the field: Ethics and Academic Integrity. The text generated by this Chatbot tool will be analyzed on the 5 platforms mentioned above, able to detect, both the percentage of use of artificial intelligence and the Chatbot model used to generate the content.



Desigur! Mai jos ai un text de aproximativ 300 de cuvinte, potrivit pentru a fi atașat unei concluzii a unui articol științific din domeniul **Eticii și integrității academice**:

În contextul actual al dezvoltării accelerate a cunoașterii și al accesului facil la informație, etica și integritatea academică reprezintă piloni fundamentali pentru menținerea credibilității sistemului educațional și a cercetării științifice. Valorile precum onestitatea, responsabilitatea, corectitudinea și respectul față de muncă intelectuală nu trebuie să rămână doar concepte teoretice, ci să se reflecte concret în comportamentele tuturor actorilor implicați în mediul academic.

Promovarea unei culturi a integrității presupune un efort comun, care începe cu educația timpurie a studenților în spiritul respectului față de proprietatea intelectuală și continuă cu aplicarea fermă și echitabilă a normelor etice de către cadrele didactice și instituțiile de învățământ. Este esențial ca aceste instituții să ofere nu doar reguli clare privind conduita academică, ci și sprijin concret în formarea deprinderilor de cercetare și redactare științifică onestă.

Tehnologia, deși poate fi uneori un factor de risc prin facilitarea plagiatului, poate fi valorificată pozitiv pentru a detecta și preveni derapajele de la normele etice, oferind instrumente utile pentru validarea originalității lucrărilor. Mai mult, digitalizarea proceselor academice ar trebui însoțită de consolidarea unei etici profesionale solide, adaptate provocărilor erei informaționale.

În concluzie, cultivarea unei integrități academice reale nu este doar o condiție pentru performanță științifică, ci și o responsabilitate morală față de societate. Doar printr-un angajament colectiv și constant față de valorile etice putem garanta calitatea, relevanța și sustenabilitatea actului educațional și a cercetării în secolul XXI.

FIG.2 Initial text sequence generated by ChatGPT

Results:

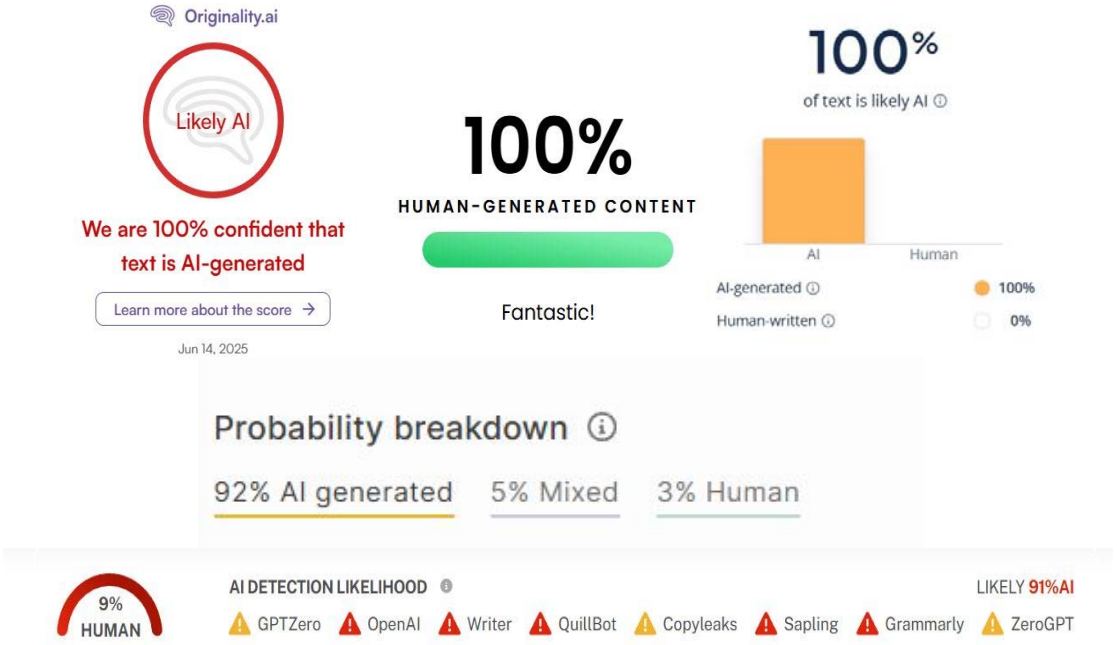


FIG.3 Results obtained on each of the 5 platforms

Figure 3 represents a centralization of the results obtained using in order the 5 AI detection platforms mentioned in the course of the paper: Originality, Writer, Quillbot, GPTzero, Undetectable.AI.

As can be seen, 4 out of 5 platforms defined the analyzed content as generated using an AI-based chatbot.

It should also be noted that the vast majority of platforms that can detect the presence of AI tools are not fully compatible with the English language, therefore the results obtained may also be influenced by this factor.

The next step is based on modifying the text initially generated by Chatgpt in order to obtain a lower degree of perplexity and a higher variability in the content structure. We will therefore train the artificial intelligence so that the language used is as natural as possible and the number of words becomes much more variable. By applying these changes we can bring to the forefront a finding related to the ease with which, also through the use of artificial intelligence, we were able to transform the initially generated text into a more natural, human and less predictable structure: in less than a minute, the ChatGpt chatbot was able to generate content with a more authentic flavor.

The new text generated by the artificial intelligence will undergo the same process of detecting the involvement of AI in its creation, and will be successively checked by all the platforms mentioned above: Originality, Writer, Quillbot, GPTzero, Undetectable.AI, in order to compare the initial and the actual results.

Trăim vremuri în care informația se mișcă... aproape cu o viteză greu de urmărit. Nu mai e clar de unde începe o idee și unde se termină alta. Sursele se suprapun. Autorii – în unele cazuri – devin greu de identificat. Și, în tot acest haos aparent, etica și integritatea academică nu mai pot fi tratate ca niște noțiuni formale sau reglementări prăfuite. Nu. Ele devin esențiale. Un punct fix. Un reper fără de care educația riscă să-și piardă direcția.

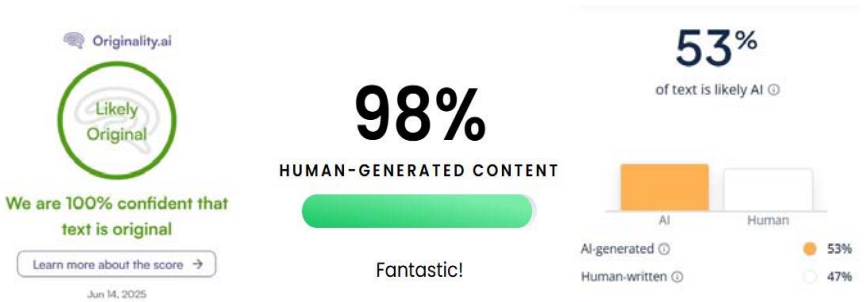
Aici nu vorbim doar de plagiat sau copiat. Nu-i așa simplu. Mai degrabă e despre relația noastră cu ideile, cu efortul intelectual – al nostru, dar și al celorlalți. Despre cum ne raportăm la cunoaștere. În mediul academic, acest respect e ceva ce se construiește greu, dar se poate pierde repede. De aceea, contează enorm.

Pe de altă parte, realitatea... ei bine, nu e tocmai ideală. Sistemul pune presiune. Multă. Rezultate. Rapoarte. Performanță. E lesne de înțeles de ce unii caută scurtături. Iar păstrarea unei conduite etice devine uneori un act de rezistență. E nevoie să începem din timp – poate chiar din școala generală, nu doar din liceu. Și nu doar cu reguli, ci cu modele. Cu profesori care arată, nu doar spun.

Tehnologia? E cu două tăișuri. Ajută, dar poate încurca. Un soft vede plagiatul, dar nu învață un tânăr ce înseamnă onestitate. Asta vine din altă parte.

Până la urmă, integritatea academică nu e despre frica de-a fi prins. Ci despre cine suntem când alegem, în liniște, să facem lucrul corect. Fără aplauze. Fără public. Doar noi și conștiința noastră.

FIG.4 The new AI generated text following the modifications according to the mentioned strategy



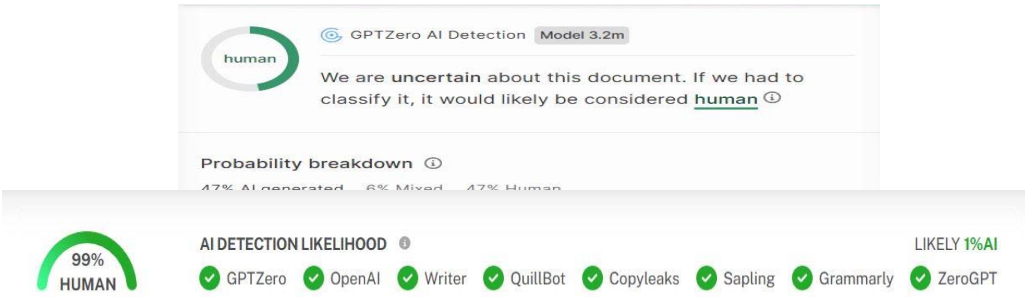


FIG.5 Results obtained after checking the new AI generated content (ChatGPT)

CONCLUSIONS

Table 1. Comparing result before and after text manipulation

AI detection software	Results *before* text manipulation	Results *after* text manipulation
Originality.AI	100% confident that text is AI generated	100% confident that text is original
Writer AI	100% Human Generated content	98% human-generated content
Quillbot	100% of text is likely AI	53% of text is likely AI
GPTzero	92% AI, 5%mixed, 3%human	47%AI - 47% human
Undetectable.AI	9% human	99% human

In the first instance, generating a text passage with the help of artificial intelligence and keeping exactly the raw form of the generated content 4 out of 5 best AI detectors returned results that, taking into account also the partial incompatibility of the software tools with the Romanian language, can be used in arguing that a text, part of a scientific article was written by means of specific Chatbots offered by companies in the field. However, by training artificial intelligence and studying the ways in which these software tools manage to detect the presence of AI, we manage, with great ease, to generate texts that can pass unnoticed by detection tools. As a proof, the example provided above (Fig. 5), where 3 out of 5 software are sure that the text generated by AI is in fact human-made, one out of 5 believes that the text is not AI-made but does not give a clear answer, and the last one believes that the text would be AI-generated but in 53% of the cases.

The study also wants to highlght that these results were obtained without subjecting the AI-generated text to a human thought filter, relying only on the content generated directly by artificial intelligence. Following the analysis obtained and the ease with which the output of artificial intelligence detection tools can be manipulated, it can be considered that by applying a series of minor modifications to a text passage generated by a *Chatbot* such as ChatGPT and adapting terms specific to human colloquial language in the content of the passage, it can be able to obtain an AI-generated text that is capable to return completely negative results in terms of the percentage of AI usage. To cancel any lack of transparency, the whole verification process of the modified AI-generated content is attached in the appendix section.

This study may represent a first step in dismantling a whole strategy to combat the use of artificial intelligence in the realization of academic work and, moreover, it could be used for a much broader research that could highlight the usefulness of using these new emerging technologies. Concerning the concept of effectiveness, the evaluated software tools were found to be vulnerable to changes in the texts generated with artificial intelligence.

In conclusion, the use of dedicated tools for the detection of the use of artificial intelligence in the content of different papers, often required in academia, is an ineffective method to combat the educational system's desire to remove this new technology.

Moreover, these techniques can not be considered effective, as the vast majority of detection tools charge money for analyzing papers with high word counts, and even in these situations, they cannot be manipulated. We propose, in this sense, a rethinking of the approach strategies used to preserve students' creativity and shifting the focus from quantity to quality. Artificial intelligence, in addition to the negative effects that are brought, more and more often, to the attention of society, can have, on the other hand, numerous ways in which it can offer improvements to educational sustems and in which it can train and keep active the attention, involvement and creativity of the educational product, the student.

Considering future research directions in the AI domain, present study will state as a base understanding on the ease of how a detection tool can be manipulated, generating new ideeas regarding the evolution of the AI usage in the academic field. In the future, the need to find new methodology and teaching practices is an imperative way to start studying how to implement Artificial Intelligence in a healthier and more ethical way.

REFERENCES

[1] E. Adamopoulou, & L. Moussiades, An Overview of Chatbot Technology. 373-383. 10.1007/978-3-030-49186-4_31, 2020;

[2] AIPRM, AI in Education Statistics. *AIPRM [online]*. URL: https://www.aiprm.com/ai-in-education-statistics/, 2025;

[3] A. Cardillo, Chatbot statistics. Exploding Topics [online]. URL: https://explodingtopics.com/blog/chatbot-statistics, 2025;

[4] M. Ghosh, & T. Arunachalam, Introduction to Artificial Intelligence. 10.1007/978-981-16-0415-7_2, 2021;

[5] G.A. Odri, D. Ji Yun Yoon, *Detecting generative artificial intelligence in scientific articles: Evasion techniques and implications for scientific integrity.* Orthopaedics & Traumatology: Surgery & Research. Vol. 109. No.8. ISSN 1877-0568, https://doi.org/10.1016/j.otsr.2023.103706. 2023;

[6] R. Mahato, *ARTIFICIAL INTELLIGENCE, WHAT IS IT?,* Outcomes of Best Practices in Classroom Research. 197-202, 2022;

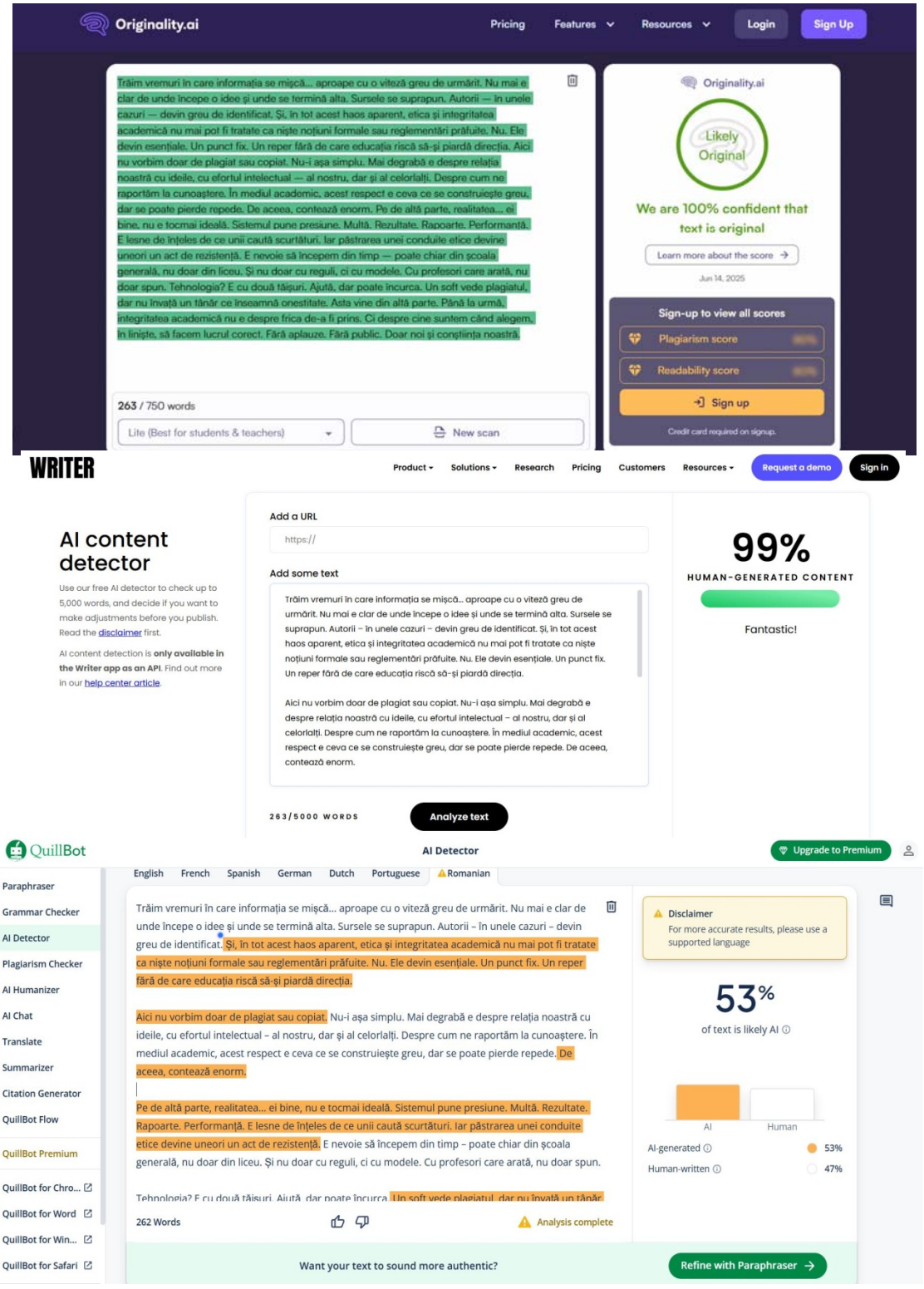
[7] C. Mariș, *The effectiveness of aviation employment in military operations*, License Thesis. Air Force Academy "Henri Coandă",Brașov, 2024;

[8] UNIC, *Perplexity and Burstiness in AI and Human Writing: Two Important Concepts.UNIC[online]*. URL: https://www.unic.ac.cy/telblog/2023/04/11/perplexity-and-burstiness-in-ai-and-human-writing-two-important-concepts/, 2025;

[9] A. M.Vieriu & G. Petrea, The Impact of Artificial Intelligence (AI) on Students' Academic Development. *Education Sciences*, *15*(3), 343. https://doi.org/10.3390/educsci15030343 https://doi.org/10.3390/educsci15030343, 2025;

[10] Zhenyu Xu, V. S. Sheng, *Detecting AI-Generated Code Assignments Using Perplexity of Large Language Models.* The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24). 23155-23161, 2024.

Appendix.1. Verification of AI-generated text in the OriginalityAI, Writer, QuillBot detection tool OriginalityAI.



Checking AI-generated text in the GPTzero and Undetectable.AI detection tool