https://doi.org/10.33271/nvngu/2025-4/150

S. D. Prykhodchenko*,
orcid.org/0000-0002-6562-0601,
O. Yu. Prykhodchenko,
orcid.org/0000-0001-5080-737X,
O. S. Shevtsova,
orcid.org/0000-0002-0148-5877,
I. H. Olishevskiy,
orcid.org/0000-0001-8573-3366

Dnipro University of Technology, Dnipro, Ukraine
* Corresponding author e-mail: prykhodchenko.s.d@nmu.one

SOFTWARE DETECTION OF UKRAINIAN-LANGUAGE TEXTS GENERATED BY AI: METHODS, ESTIMATIONS, CHALLENGES

Purpose. To evaluate the effectiveness of existing AI-generated text detection tools in identifying machine-generated content within students' qualifying papers. The research aims to analyze the accuracy, reliability, and limitations of these detection methods, particularly for texts written in Ukrainian, and to explore potential improvements in AI content recognition.

Methodology. The study employs a combination of statistical analysis and experimental evaluation of AI-detection tools. A dataset of AI-generated and human-written academic texts is used to assess detection accuracy, with additional experiments conducted to analyze the influence of linguistic structures, translation effects, and text formalization. The research also incorporates computational techniques to examine error rates and determine patterns in AI-generated content recognition.

Findings. The study results provide insights into the performance of current AI-detection systems, highlighting their strengths and weaknesses. The analysis reveals significant challenges in detecting AI-generated text, particularly in non-English languages such as Ukrainian. Additionally, the study identifies the impact of translation on detection accuracy and discusses the effectiveness of different linguistic and statistical approaches.

Originality. This research contributes to the ongoing discourse on academic integrity by addressing the limitations of AI-detection tools for non-English academic texts. Unlike previous studies focusing primarily on English-language AI-generated content, this study provides a unique perspective on the challenges of detecting AI-generated text in Ukrainian, offering novel insights into the adaptation of detection models for diverse linguistic contexts.

Practical value. The findings of this study are valuable for educators, researchers, and policymakers concerned with maintaining academic integrity in the era of generative AI. By identifying weaknesses in current detection systems and suggesting possible improvements, the research provides a foundation for developing more robust AI-detection methodologies that can effectively apply to academic texts in multiple languages.

Keywords: *AI-generated text, academic integrity, detection tools, higher education, natural language processing, machine learning, software modelling, algorithm, multilingual text analysis*

Introduction. In the modern era, where technology is deeply embedded in various domains, including education, we encounter new challenges and opportunities. In recent years, the global academic and educational community has faced a significant challenge posed by the rapid development of artificial intelligence technologies. Advanced algorithmic and neural-network-based solutions have now reached a level where distinguishing between a human-written essay and AI-generated text has become increasingly difficult.

This issue has become particularly evident for educators following the introduction of ChatGPT and similar AI-driven models, capable of producing high-quality content often indistinguishable from human writing. Generative algorithms now efficiently produce summaries, abstracts, reviews, and even full-length academic papers. However, this technological advancement raises concerns about fairness in education. Students who lack access to such tools or do not possess the necessary skills to use them effectively may find themselves at a disadvantage compared to peers who can leverage AI to generate academic content effortlessly.

It is essential to recognize that generative models produce text based on probabilistic predictions, where each subsequent token is determined by the highest likelihood given the preceding context. Consequently, even minor variations in input queries or prior responses can lead to significantly different outputs. This characteristic of AI-generated text means that it does not reproduce direct quotations verbatim but rather synthesizes content in a manner akin to human cognitive processing. As a result, AI-generated text is typically free from direct plagiarism as conventionally defined.

While there has been extensive research on the detection of AI-generated content, significant gaps remain, particularly concerning texts written in natural languages that are not among the world's most widely spoken. In Ukraine, for instance, academic theses – whether at the Bachelor's, Master's, or PhD level – are primarily written in Ukrainian in accordance with national regulations. Despite a notable increase in AI-generated content over the past year, the ability to reliably identify such texts remains challenging, as existing detection methods and metrics are still underdeveloped compared to established plagiarism detection techniques.

Thus, addressing the implications of generative AI for academic writing represents a crucial and timely research problem in both computer science and education. This study not only seeks to highlight the challenges posed by AI-generated texts in academic theses but also aims to explore methods for text analysis, attribution, and the key factors that influence these processes. Understanding and mitigating the risks associated with generative AI in research and education is essential for ensuring academic integrity and maintaining fairness in the evolving digital landscape.

Literature review. To begin, it is essential to examine the nature of generative text models and their capabilities. These models are primarily built on deep learning architectures and neural networks, enabling them to process vast amounts of textual data and generate new content that closely mirrors the style and structure of their training corpus.

The widespread availability of large-scale language models (LLMs) has led to a rapid surge in their adoption within academic and scientific domains. This growth has fueled not only an increase in their utilization but also a growing demand for analytical tools capable of detecting, attributing, and mitigating AI-generated content. Specifically, researchers are actively exploring methodologies to identify authorship at both the document and passage levels, addressing concerns about content authenticity and academic integrity.

The study by Shah, et al. [1] explores the detection and de-masking of AI-generated texts through artificial intelligence-based analysis of stylistic features. The research examines the inherent challenges in distinguishing AI-generated content from human-written text and seeks to establish ethical guidelines for their differentiation. To achieve this, the study analyzes a range of linguistic parameters, including syllable count, word length, sentence structure, functional word usage, and punctuation distribution. Various computational methodologies and classification approaches are evaluated to enhance the accuracy of AI-generated text detection, contributing to the broader effort to develop robust mechanisms for distinguishing between human and machine-authored content.

The study by Majhi, et al. [2] examines the application of GPT-based language models for generating academic writing samples and evaluates the challenges associated with detecting AI-generated content using traditional plagiarism detection tools. The research systematically assesses the effectiveness of both conventional plagiarism detection systems and specialized AI based detection tools in identifying machine-generated text. The findings indicate that traditional plagiarism checkers struggle to detect AI-generated content due to fundamental differences in syntactic structures and writing patterns. In contrast, AI-specific detection models demonstrate a higher success rate in distinguishing human-authored texts from those produced by generative algorithms. The study underscores the necessity of advanced AI-driven detection solutions to uphold academic integrity, emphasizing that AI-based detection tools play a critical role in preventing the infiltration of AI-generated content into scholarly communication and research publications.

The study by Kumarage, et al. [3] presents a comprehensive review of forensic text processing systems designed for analyzing AI-generated content, with a primary focus on detection, attribution, and characterization. The paper explores the challenges posed by advanced Large Language Models (LLMs), particularly their ability to generate high-quality, human-like text and the associated risks of misinformation and content authenticity. It provides an in-depth overview of current research efforts in AI-generated text forensics, introduces a detailed taxonomy of detection and attribution methodologies, and highlights key resources available for further investigation in this domain. The study discusses AI detection models, which aim to differentiate between human-authored and machine-generated texts, as well as attribution systems, which seek to trace AI-generated content back to the specific model responsible for its creation.

The study by Rashidi, et al. [4] investigates the challenges associated with misclassification of human-authored scientific articles as AI-generated content. The research reveals that up to 8 % of genuine academic abstracts were incorrectly identified as AI-generated, highlighting significant limitations in existing detection methods. The authors critically assess these shortcomings and propose new approaches to reduce classification errors and mitigate unintended consequences. The study underscores the importance of accurately distinguishing between human-written and AI-generated texts, particularly within scientific publishing and educational contexts, where misclassification could impact academic credibility and integrity.

The study by Mindner, et al. [5] investigates methods for detecting AI-generated and AI-paraphrased text, with a focus on distinguishing human-authored content from machine-generated text. The research evaluates multiple classification techniques, including XGBoost, random forests, and neural networks, demonstrating high F1 scores for identifying both simple and complex AI-generated texts. The paper emphasizes the critical importance of AI text detection, detailing key linguistic and statistical features used for classification and presenting experimental findings on the effectiveness of various detection models.

The study by Amirjalili, et al. [6] examines the extent to which artificial intelligence can meet formal academic writing standards while preserving the distinctive characteristics of scholarly texts. The research highlights key challenges in AI-generated academic content, particularly in ensuring factual accuracy, specificity, depth, and proper citation of sources. The findings suggest that while AI demonstrates potential as an assistive tool for academic writing, its current capabilities in producing high-quality, independently reliable scholarly work remain limited, emphasizing the need for human oversight and verification in academic contexts.

The study by Ma, et al. [7] analyzes the differences between AI-generated and human-written text across syntactic, semantic, and pragmatic dimensions to better understand the gap between them. The research highlights the fact that ChatGPT exhibits "reference hallucination" when generating scientific content, often producing inaccurate or non-existent citations. To evaluate these discrepancies, the study employs feature-based detection techniques alongside fine-tuned pre-trained AI detection models, providing a comparative analysis of linguistic patterns in AI-generated versus human-authored scientific texts.

The study by Chaka [8] assesses the effectiveness of AI detection tools in differentiating AI-generated and human-authored texts, based on an analysis of 17 research articles published between January and November 2023. The findings emphasize the importance of a multifaceted approach to content verification, recommending the combined use of AI-based detection systems, traditional plagiarism detection tools, and human reviewers to improve classification accuracy. The research was conducted using a modified PRISMA protocol and quality assessment criteria, addressing six key research questions to evaluate detection performance.

The study by Weber-Wulff, et al. [9] examines AI-generated text detection tools, with a particular emphasis on the risks associated with their misuse in academic settings. The research evaluates the accuracy and error patterns of these detection systems, specifically analyzing their ability to differentiate human-authored texts from those generated by ChatGPT. The findings indicate that current AI detection tools lack reliability, often demonstrating a bias toward classifying AI-generated content as human-written, which raises concerns about their effectiveness in maintaining academic integrity.

The paper [10] is a current study of the reliability of AI-generated text detection tools. The authors analyze the accuracy of the detectors, identifying their limitations, such as false positives and text recognition problems from GPT-4. The work emphasizes the need to improve detection methods and offers a theoretical framework for future developments in this area.

The study [11] is a valuable analysis of the accuracy of AI text detectors in student essays in English (L1 and L2). The paper identifies leaders (Copyleaks, Undetectable AI) and limitations (false positives), emphasizing the need to improve the reliability of such systems for academic integrity.

The article [13] investigates the accuracy of free tools for detecting AI-generated texts, using the work of Japanese students as an example. Analyses of five detectors showed their limited effectiveness, especially when partially utilizing AI. The results emphasize the need for better algorithms and cautious application of such tools in education.

An article by W. Walters [14] evaluates 16 AI text detectors, comparing their accuracy on ChatGPT-3.5, GPT-4, and student essays. Copyleaks, TurnItIn and Originality.ai showed high accuracy, especially with GPT-4. The work emphasizes progress in detection algorithms, but reveals weaknesses in most free tools, requiring further research.

The work [15] analyses the accuracy, ethical aspects and challenges of detecting such texts, highlighting the need to adapt training and assessment methods. The paper suggests ways to integrate AI to support academic environments.

The authors of the paper [16] propose FairOPT, an adaptive threshold algorithm for AI text classification that takes into account subgroups (text length, style). It improves accuracy (F1) and fairness by reducing errors. The method is promising for other modalities and integration into training.

The authors of [17] use large language models (LLMs) to analyze self-evaluation through dialogues, revealing limited accuracy when compared to humans. The work highlights the challenges of interpreting subjective data and the need to improve LLMs for such tasks.

The work [18] explores methods for detecting texts generated by large language models (LLMs) using statistical and neural network approaches. The proposed solutions show high accuracy (AUC up to 0.95) but are limited by data specificity. They are promising for the development of AI ethics and content verification.

Paper [19] investigates the application of the BERT model to classify AI-generated texts. The results show high accuracy in detecting AI-generated content through contextual analysis. As professors of computer science, we believe the approach is promising but needs to be adapted to new generative AI models.

Thus, works such as [9–14] focus on analyzing the accuracy of AI-generated text detection tools.

Walters [14] focused on analyzing the accuracy of AI-generated text detection tools such as ChatGPT and GPT-4. The research covers testing between 5 and 30 detectors on student essays (L1/L2, Japanese students), identifying leaders such as Copyleaks, TurnItIn and Originality.ai, but noting problems: false positives, low reliability of free solutions and difficulties in recognizing new AI models. This points to the need for improved algorithms to ensure academic integrity. In contrast, the work of [15] emphasizes ethical aspects and the integration of AI in education, requiring the adaptation of evaluation methods. The work of [16] proposes FairOPT, an adaptive thresholding algorithm that improves F1 accuracy and fairness of AI text classification, which is promising for different domains.

An important feature of all the above works is the focus on detecting English-language texts written by humans or artificial intelligence.

Purpose. The aim of the article is to investigate the effectiveness of existing AI-generated text detection tools in identifying AI-generated content within students' qualifying papers, with a particular focus on the Ukrainian language. The study seeks to evaluate these detection systems' accuracy, reliability, and limitations, considering their dependence on linguistic and structural patterns of AI-generated text.

The main task of mathematical processing of the collected data is to find and prove hypotheses regarding the ability of current detection tools to differentiate between human-written and AI-generated content. This includes analyzing error rates, identifying linguistic and structural inconsistencies, and assessing the impact of translation on detection accuracy. By applying statistical and computational techniques, the research aims to develop insights into the strengths and weaknesses of AI detection models and propose improvements for enhancing their effectiveness in academic integrity enforcement.

Methodology. During the preliminary phase of this research, the following hypotheses were formulated:

1. Modern AI-generated text detection tools can reliably distinguish between Ukrainian-language content produced by artificial intelligence and content authored by humans.

Confirmation, occurs if the average classification error for both AI-generated and human-written Ukrainian-language texts remains consistently below 15 % across multiple detection platforms, indicating high reliability and low false positive/negative rates.

2. Modern AI-generated text detection tools can reliably differentiate between English-language AI-generated content and human-written text.

Confirmation occurs if AI-generated texts in English are correctly identified with an error rate consistently below 10–15 % and human-written texts, are misclassified as AI with error rates under the same threshold. Ideally, this should hold true across several detectors, not just isolated cases.

3. AI detection systems are trained to recognize structured patterns and linguistic constructs characteristic of AI-generated text.

Confirmation occurs if AI-generated content with rigid or template-based linguistic features is more consistently identified than less structured, non-template-based content. Human-written formal documents should yield higher false positive rates than less formal texts due to overlapping structural features.

4. The use of automatic translation tools negatively affects the ability to accurately detect AI-generated content in Ukrainian, as most of the current machine translation systems have machine learning elements that are potentially capable of generating patterned text sequences characteristic of AI-generated text.

Confirmation occurs if after translation to English, detection accuracy declines – AI-generated texts become harder to classify, or human-authored texts are more frequently misclassified.

To clarify the rationale behind these hypotheses, the initial objective of this research was to identify effective methods for recognizing AI-generated text in the Ukrainian language. The evaluation was conducted using detection tools that explicitly supported Ukrainian and did not list it among languages excluded from the analysis.

During the preliminary data analysis, key statistical measures – including mean, variance, and standard deviation – revealed that existing detection models struggled to determine authorship with high confidence. This observation led to the hypothesis that the Ukrainian language may not be well-represented in these detection models or that AI-generated content recognition's fundamental principles differ from those assumed.

Given the operational mechanisms of large language models (LLMs), we hypothesized that AI systems generate text based on the most frequent and template-based linguistic patterns found in their training data. Consequently, AI-generated text detectors may only effectively identify such patterns in language sufficiently represented in the training datasets. Therefore, we proposed that AI detection should account for the language's specific linguistic and syntactic characteristics under analysis.

To validate this hypothesis, we designed an experiment in which AI models were prompted to generate texts on both standard academic topics and unconventional, non-template topics, which were expected to contain fewer predefined structures. Additionally, highly structured human-authored texts in English – such as academic papers with rigid formatting and conventional phrasing – were selected to assess their similarity to AI-generated content.

For hypotheses 2 and 3, we compiled a test dataset consisting of excerpts from scientific papers on AI-generated text detection, which, like most scholarly writing, adhere to a highly structured and formalized style. To evaluate hypothesis 4, we performed a machine translation of the original Ukrainian test samples into English. We reanalyzed them using AI detection tools to determine the impact of translation on detection accuracy.

The first part of the study was conducted in the first half of 2024. However, the development and change of language models require constant updates and changes. Therefore, in the first months of 2025, we re-checked and updated the list of analyzers, which led to some table changes.

Findings. To evaluate the proposed hypotheses, this study aimed to assess the capability of web-based AI detection applications to differentiate between human-authored texts and AI-generated content.

As part of the analysis process, the research team developed an automated web crawler designed to systematically input each test sample into the selected detection tools, collect the corresponding classification results, and log the output for further statistical evaluation. The details of the web-based AI detection platforms used in this study are summarized in Table 1.

This system implements a pipeline architecture of data processing, in which user input passes through several successive stages of processing, interacting with external services in the mode of automated analysis.

This process represents a typical automated data processing task with a distributed architecture. It is based on a pipeline processing model, where data passes through several consecutive stages, each performing a specialized function.

In the first stage, the system receives text from a human operator through a user interface or API. This stage requires ensuring reliable data transmission, handling different text formats and possibly preprocessing the input data to remove extra characters, coding errors and other artefacts.

Next, the program automatically sends the text to several external sites for checking. For this purpose, the request routing module is used, which should support parallel processing, asynchronous calls and retry mechanisms in case of service failures. The system must be resilient to network failures and differences in the APIs of the target services. Load balancing strategies, caching of frequent requests and adaptation to different response formats can be applied here.

Once the validation results have been received, the system proceeds to process them. This step requires nor-

Table 1

Web-Based Platforms for AI-Generated Text Detection

Nick	Name	Address	Declared technology
Site 1	Checkfor.ai	https://www.checkfor.ai/product/	A large neural network trained on millions of human and AI-generated texts
Site 2	AI Checker Tool	https://aicheckertool.com/	Machine learning and natural language processing
Site 3	ZeroGPT	https://www.zerogpt.com/	We use a multi-stage methodology + LLM
Site 4	GPT-2 Output Detector Demo	https://openai-openai-detector.hf.space/	Implementation of RoBERTa
Site 5	AI Content Detector	https://www.interestedinai.com/ai-content-detector-tool	Artificial intelligence-based tool
Site 6a	Dolly V2 3B	https://radar-app.vizhub.ai/	LLM
Site 6b	Camel 5B	https://radar-app.vizhub.ai/	LLM
Site 6c	Dolly V1 6B	https://radar-app.vizhub.ai/	LLM
Site 6d	Vicuna 7B	https://radar-app.vizhub.ai/	LLM
Site 7	Free AI Content Detector Tool	https://seo.ai/tools/ai-content-detector	artificial intelligence-based tool
Site 25-1	Free AI Detector | GPT-4, GPT-3, & ChatGPT AI Checker	https://www.grammarly.com/ai-detector	Machine-learning model
Site 25-2	AI Detector, AI Checker, & AI Humanizer | Undetectable AI	https://undetectable.ai/detector-humanizer	Summarizer of several AI-driven models
Site 25-3	AI Detector - Advanced AI Checker	https://quillbot.com/ai-content-detector	Machine-learning model
Site 25-4	AI Detector - Trusted AI Checker	https://decopy.ai/ai-detector/	Machine-learning models

malizing the data, bringing it into a consistent format and filtering out possible duplicate or irrelevant responses. The data may be analyzed using statistical methods, machine learning algorithms or predefined heuristics to calculate a final score. Aggregation models such as weighted averaging, consensus analysis or more sophisticated methods based on source confidence ratios may be used.

In the final step, the system generates a final report and transmits it to the human operator. The output can be presented as a textual summary, graphical analysis or structured data in a format suitable for further processing. The system must be designed so the operator can understand the decision criteria and adjust the processing parameters if necessary.

Thus, this software architecture combines elements of distributed computing, decision-making automation and human interaction, which makes it a typical representative of modern intelligent data processing systems. The architectural representation of this scripture is shown in Fig. 1

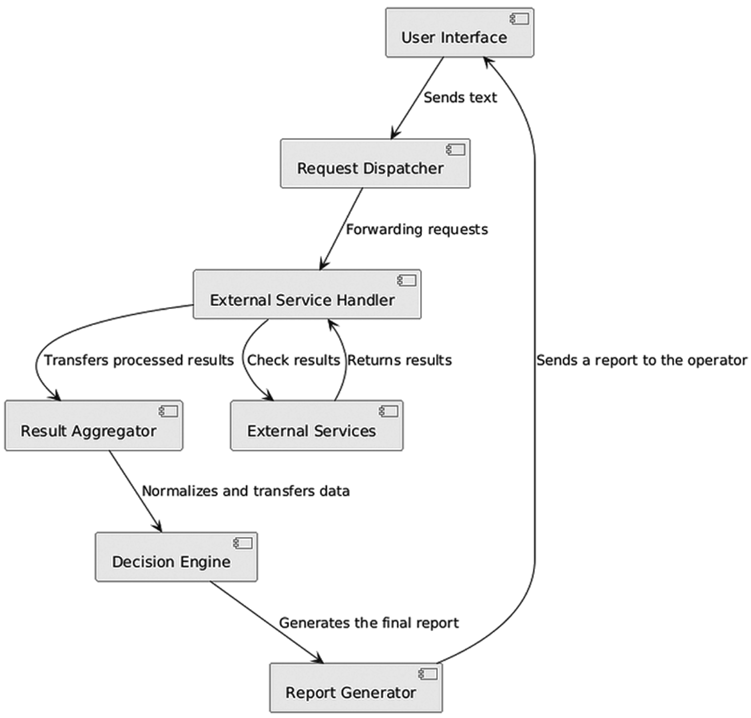

Fig. 1. The architectural model of the system is represented as a UML Component Diagram

The algorithmic sequence of the AI checking process is described below.

1. The human operator enters text into the system:

1. The text is sent for processing.
2. Sending the text for checking.

2. The program generates queries to several external sites:

1. Queries are sent asynchronously or in parallel.
2. Waiting and receiving results.

3. The system waits for respon.ses from external services:

1. As the responses are received, they are processed and stored.
2. If a service does not respond within the specified time, the system logs an error and continues processing.

4. Data aggregation and normalization:

1. Bringing responses to a consistent format.
2. Eliminating duplicates and irrelevant data.
3. Validity assessment and weight analysis of data (if applicable).

5. Decision making:

1. Analyzing the aggregated data.
2. Calculating the final result (summary score, probabilistic analysis, etc.).

6. Report generation:

1. Creating a human-readable report with results.
2. Data visualization (if required).

7. Outputting the result to the user:

1. The human operator receives the final decision.
2. The system completes processing and is ready for the next request.

Transforming this algorithmic sequence with a generalization of the parts related to the application into a single whole, called the Application, will give us the sequence diagram shown in Fig. 2.

Thus, the software architecture of the developed parsing module is a modular and multi-layered system built on the principles of encapsulation, weak coupling and clear division of responsibility. This approach allows flexibility, scalability, and ease of system maintenance and expansion. The architecture is implemented on the principle of client-server interaction between subsystems, each responsible for a strictly defined aspect of the entire software complex.

At the first level of interaction is the input subsystem, which serves as a bridge between the user and other components. It implements the graphical and console user interface and is intended for primary data input, such as keywords, filtering parameters and search settings. This subsystem implements basic validation of the entered information, after which the data is transferred to the internal subsystems in a structured form. The logical level implements business logic, where the search subsystem plays the key role. It forms search queries based on the entered parameters, accesses external information sources, including APIs of search engines or web resources, and parses the obtained results. This subsystem is designed to be scalable by connecting new data sources without changing the rest of the system.

After retrieving references and metadata for the retrieved scientific publications, the management moves to the download subsystem. It deals with verification of references, automatic downloading of documents, processing of formats and organization of local file storage. Downloading can take place both synchronously and in multithreaded or asynchronous mode, depending on the scale of the task and performance requirements. Once the documents are saved, the data is transferred to the analysis subsystem, which in turn is a set of algorithmically rich components.

The analysis subsystem is responsible for the intelligent text processing of scientific papers. At the first stage, text content is extracted from documents of various formats (PDF, DOCX and others), after which linguistic and statistical data processing is performed. Frequency characteristics are calculated, keywords are extracted, dictionaries and topic maps are generated. Next, clustering algorithms such as K-means, DBSCAN or hierarchical methods are used to group texts according to the degree of semantic similarity. All analysis results are stored in a database, from where they can be retrieved for visualization or exported as reports.

To ensure consistency between subsystems, a coordination layer is provided that manages data flows, task synchronization and error handling. The architecture allows each subsystem to be extended independently of the others through an interface-oriented approach and design patterns such as facade, adapter and observer. All

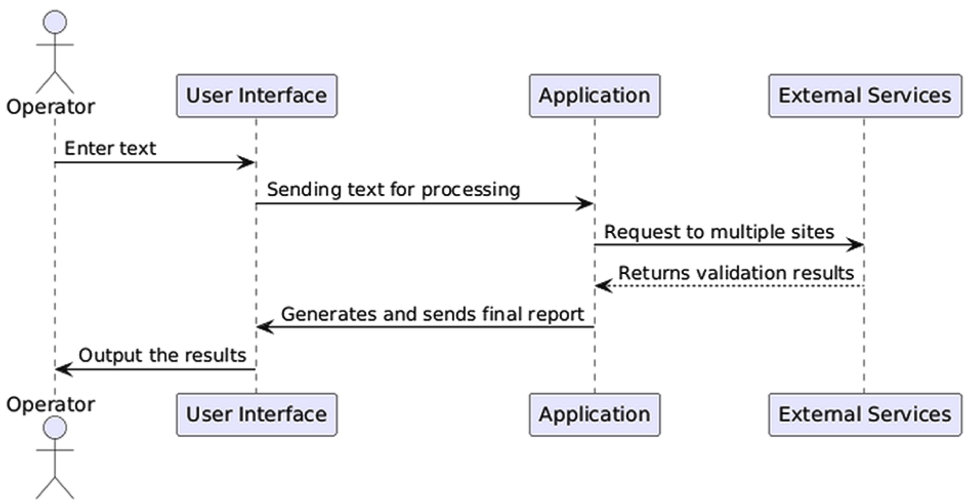

Fig. 2. Sequence of basic system operations is represented as a UML Sequence Diagram

modules are implemented so alternative implementations can be substituted without disrupting the entire system. This is particularly important for the analysis and search subsystems, where the constant evolution of technology requires regular algorithm upgrades. As a result, the proposed architecture provides a reliable and versatile platform for intelligent parsing and analysis of scientific information.

To rigorously test all hypotheses, carefully curated test datasets were required. The dataset construction followed these principles:

For Hypothesis 1, two sets of text samples were created, consisting of 34 and 32 passages, each approximately one printed page in length or shorter (a constraint imposed by some of the AI detection tools used in the study). These datasets included both AI-generated texts in Ukrainian and human-authored texts. To ensure the absence of AI-generated content in human-written samples, student diploma theses from before 2020 were selected for inclusion in the dataset.

For Hypothesis 4, the same text passages were translated into English using Google Translate's automated translation service. This step aimed to assess whether translation affects the ability of AI detection tools to differentiate between AI-generated and human-written content.

Error metrics were calculated to quantify the effectiveness of AI detection tools. The error was defined as 100 % minus the percentage of AI classification assigned by the detection tool for AI-generated texts. For human-written texts, the error was defined as the percentage of AI classification assigned by the detection tool. Thus, in both cases, an error value closer to zero indicates a more accurate classification by the detection tool.

A summary of the statistical analysis, including the expected error values and standard deviations for AI-generated text detection, is presented in Table 2.

As demonstrated in the table, translating AI-generated text into English significantly enhanced detection accuracy for platforms 1, 4, 5, 6a, 6b, 6c, and 6d. In contrast, detection performance deteriorated for platforms 2, 3, and 7. Notably, platforms 2 and 3 tended to misclassify AI-generated and human-authored texts in Ukrainian, frequently labelling human-written content as AI-generated. However, after translation into English, these platforms demonstrated improved accuracy in distinguishing human-authored texts from AI-generated ones.

A detailed statistical analysis, including the expected error values and standard deviations for human-written text detection, is presented in Table 3.

After translation into English, sites 2, 3, 6b, and 7 demonstrated improved accuracy in recognizing human-written texts, while the remaining platforms exhibited a decline in performance. Figs. 1 and 2 present histograms depicting the distribution of average classification errors for AI-generated texts in Ukrainian and their corresponding English translations. The data indicate that, translation into English reduced the average error rate across most detection sites, with the most significant improvement observed in Site 1.

For sites that appeared in 2025, the situation is that two of the analyzed sites detect the Ukrainian language and report that they are unable to process it. The remaining two show contradictory results regarding recognizing human and AI content.

A comparative analysis of average classification errors for human-authored texts and their translations re-

Table 2

Analysis of AI-generated text verification

site	Ukrainian		Ukrainian-to-English translations		Difference	
	Mean	Standard deviation	Mean	Standard deviation	delta Mean	delta Standard Deviation
site1	68.1	26.11	12.3	16.87	55.84	–9.24
site2	1.79	3.63	46.7	34.31	–44.9	30.68
Site3	2.41	3.27	46.7	33.91	–44.3	30.64
Site4	99.8	0.37	53.2	45.88	46.59	45.50
Site5	97.8	0.69	42.7	23.87	55.05	23.18
Site6a	84.9	6.88	62.4	20.37	22.50	13.49
Site6b	87.6	8.45	16.9	22.54	70.64	14.10
Site6c	85.2	15.21	31.7	19.10	53.47	3.90
Site6d	69.4	20.30	28.4	25.54	41.01	5.24
Site7	80.9	29.03	90.8	16.82	–9.92	–12.21
Site 25-1	N/A	N/A	21.25	26.23	N/A	N/A
Site 25-2	16.23	25.85	13.63	30.82	2.6	4.97
Site 25-3	N/A	N/A	6.4	21.87	N/A	N/A
Site 25-4	10.47	7.36	56.9	13.43	–46.43	6.07

Table 3

Analysis of the human-written texts verification

site	Ukrainian		Ukrainian-to-English translations		Difference	
	Mean	Standard deviation	Mean	Standard deviation	delta Mean	delta Standard Deviation
site1	0.20	0.50	0.42	2.27	–0.21	1.77
site2	65.4	29.36	13.7	19.36	51.71	–10.00
Site3	64.9	29.38	12.9	18.38	52.06	–11.00
Site4	0.14	0.25	27.2	38.68	–27.1	38.43
Site5	2.52	2.24	43.4	28.30	–40.9	26.06
Site6a	12.7	6.34	43.2	22.19	–30.5	15.85
Site6b	12.2	11.26	3.38	12.97	8.87	1.70
Site6c	12.8	15.10	43.9	30.17	–31.1	15.07
Site6d	16.9	6.37	40.6	36.60	–23.7	30.23
Site7	18.3	27.11	10.9	8.85	7.42	–18.26
Site 25-1	N/A	N/A	9.64	18.38	N/A	N/A
Site 25-2	23.6	41.28	41.35	48.23	17.7	6.95
Site 25-3	N/A	N/A	27.68	33.85	N/A	N/A
Site 25-4	88.4	22.48	67.87	13.71	–20.5	–8.77

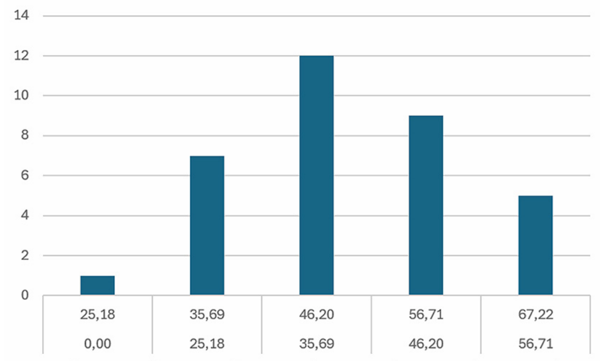

Fig. 3. Histogram of the distribution of mean scoring errors for texts generated in Ukrainian (sites 1-7, 25-4)

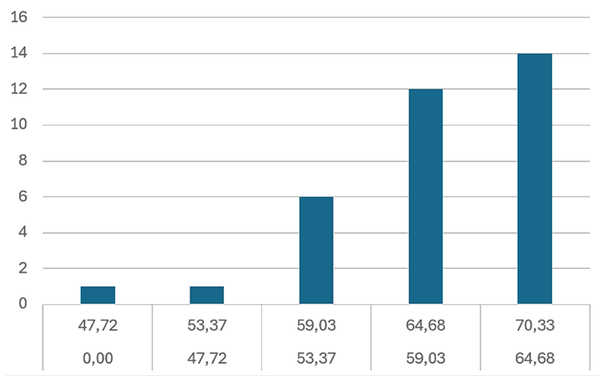

Fig. 4. Histogram of the distribution of mean scoring errors for Ukrainian-generated texts that were translated into English (sites 1-7, 25-4)

vealed no significant overall improvement. While platforms 2, 3, 6b, and 7 demonstrated increased accuracy, others showed a decline in performance.

To evaluate Hypotheses 2 and 3, three distinct English-language datasets were constructed. The first dataset consisted of passages from academic research papers cited in this study, serving as a benchmark for human-authored scholarly writing. The second dataset comprised AI-generated texts produced in response to non-template prompts similar to those outlined in Table 1. The third dataset included highly structured and formalized documents commonly used in English-language professional and legal contexts.

To build the list of standardized formal documents, queries were submitted to ChatGPT and Copilot. The resulting selection, based on linguistic rigidity and conventional structuring, is presented in Table 2. The 15 most formalized document types in English include Resume/CV, Cover Letter, Application, Contract, Presentation, Report, Email, Job Application, Business Plan, Thank You Letter, Power of Attorney, Marketing Plan, Complaint Letter, Articles of Incorporation, and Meeting Minutes [20].

Table 4 presents the mean and standard deviation of error rates for three categories of texts: AI-generated English texts, human-written texts, and highly formalized human-written documents.

The analysis of these deviations indicates that Site 1 demonstrated the most reliable performance in determining text authorship. Notably, if a single AI-generated text — incorrectly classified with 46.68 % confidence — were excluded from the dataset, the average recognition error for AI-generated texts would drop to –0.03 %, demonstrating a significant improvement in classification accuracy.

Table 4

Analysis of the error rates of AI-generated English texts, human-written texts and formal texts written by humans

	AI Generated		Human written		Human formal	
site	Mean	Standard deviation	Mean	Standard deviation	delta Mean	delta Standard Deviation
site1	2.01	10.26	20.5	37.12	–1.03	5.48
site2	30.4	37.24	17.1	26.24	–44.9	45.05
Site3	30.6	37.11	13.8	22.49	–44.6	44.63
Site4	89.	30.14	0.47	2.19	–4.16	18.84
Site5	63.1	26.10	41.9	25.09	–26.3	25.23
Site6a	76.8	21.78	53.1	21.81	–34.2	24.97
Site6b	3.95	5.98	26.3	39.50	–25.1	35.67
Site6c	69.9	21.16	29.1	23.78	–26.5	24.42
Site6d	45.8	29.53	27.1	28.19	–36.3	36.52
Site7	95.5	6.25	5.81	7.39	–14.9	25.16
Site 25-1	65.00	42.52	1.42	3.93	0.00	0.00
Site 25-2	1.00	0.00	44.12	47.57	14.92	34.32
Site 25-3	2.1	5,17	0.00	0.00	7.14	25.75
Site 25-4	31.57	11,09	34.58	13.53	52.35	22.00

Site 4 exhibited a strong bias toward misclassifying all texts as human-written, regardless of their actual origin.

Site 6b showed a high capability for correctly identifying AI-generated texts. However, it also tended to misclassify human-authored texts as AI-generated, leading to concerns about false positives.

Of the sites that appeared in 2025, only site 25-3 shows good results in recognizing English-language content. However, the downside of this system is the inability to recognize Ukrainian-language texts.

Based on the findings of the experimental studies, the following key conclusions can be drawn:

1. No web application demonstrated reliable authorship recognition for Ukrainian-language texts. All evaluated detection tools exhibited an average error rate exceeding 15 %, indicating that these systems struggle with accurate classification in Ukrainian.

2. Translation into English significantly improved detection accuracy for Site 1, the only platform to achieve an error rate below 15 %. Specifically, after translation, Site 1 produced an average error of 12.77 %, with a standard deviation of 17.07 %, suggesting that these detection models perform better in languages with stronger representation in training data.

3. Hypotheses 2 and 4 were confirmed exclusively for Site 1. When analyzing English-language texts, this tool yielded an average error rate of 23.54 %, with a standard deviation of 52.86 %. However, after excluding extreme outliers, the adjusted average error decreased to

6.19 %, with a standard deviation of 19.31 %, confirming that translation enhances AI detection effectiveness.

4. Hypothesis 3 was not confirmed within acceptable statistical thresholds. The findings indicate that current AI detection tools do not consistently rely on identifying structured linguistic patterns, contradicting the assumption that such systems detect AI-generated text primarily based on syntactic or stylistic regularities.

Therefore, it can be concluded that Site 1, while restricted to English-language content, demonstrates a reasonable capability for detecting AI-generated text. However, its performance on Ukrainian-language texts remains inconsistent, producing ambiguous and contradictory results that prevent its use for reliable authorship attribution.

Additionally, the other AI detection tools evaluated in this study exhibited high variability and lack of reliability, making them unsuitable for practical application in cases where individuals need to verify the authenticity of their authorship claims.

Conclusions. This study has demonstrated that translating AI-generated text from Ukrainian to English significantly improves detection accuracy. This finding underscores the fact that large language models (LLMs) are predominantly trained in English, while their multilingual capabilities often rely on lower-ranked translation-based models. This raises an important concern regarding linguistic bias in AI development, as non-English languages risk being reduced to simplified or anglicized versions of their original structures. Over time, this could contribute to a homogenization of linguistic diversity, limiting the richness and complexity of non-English academic and scientific communication.

Another critical issue is the emerging "AI vs. AI" paradigm, wherein artificial intelligence systems are being deployed to detect content generated by other AI models. As AI-generated text becomes increasingly sophisticated, distinguishing between human-written and machine-generated content has become an inherently complex task – one that is often delegated to automated detection tools, many of which are based on machine-learning techniques. This cycle of AI-driven detection highlights the evolving challenges in ensuring academic integrity, particularly as legacy detection models struggle to keep pace with advancements in large language models.

This research examined whether AI-generated scientific texts – produced using models such as GPT-3 (e.g., ChatGPT) and GPT-4 (e.g., Copilot) – can be effectively distinguished from human-authored works. The primary focus was on textual analysis of Ukrainian-language content, but experimental results indicate that existing AI detection tools are largely ineffective in reliably identifying machine-generated text.

Given the rapid advancements in LLMs, software utilizing these models can potentially revolutionize academic writing. Current AI tools like ChatGPT can produce highly sophisticated scientific articles, conference papers, and other academic documents. This trend is becoming increasingly apparent in higher education, where AI-generated text can seamlessly integrate into the academic workflow, raising ethical and methodological concerns.

Findings from this study further reveal that traditional plagiarism detection tools fail to identify AI-generated content as plagiarism in the conventional sense. Moreover, the effectiveness of detection systems is highly language-dependent, with many existing tools being optimized primarily for English text.

Therefore, it is crucial to systematically assess and refine AI detection methodologies, considering language-specific characteristics and computational constraints. The results of this research highlight an urgent need to develop advanced AI detection tools capable of recognizing machine-generated content across multiple languages, ensuring fairness and academic integrity in an increasingly AI-driven scholarly landscape.

References.

1. Shah, A., Ranka, P., Dedhia, U., Prasad, S., Muni, S., & Bhowmick, K. (2023). Detecting and Unmasking AI-Generated Texts through Explainable Artificial Intelligence using Stylistic Features. *International Journal of Advanced Computer Science and Applications, 14*, 110. https://doi.org/10.14569/IJACSA.2023.01410110

2. Majhi, D., & Santra, P. (2023). Scholarly Communication and Machine-Generated Text: Is it Finally AI vs AI in Plagiarism Detection? *SRELS Journal of Information Management, 60*, 173-181. https://doi.org/10.17821/srels/2023/v60i3/171028

3. Kumarage, T., Agrawal, G., Sheth, P., Moraffah, R., Chadha, A., Garland, J., & Liu, H. (2024). *A Survey of AI-generated Text Forensic Systems: Detection, Attribution, and Characterization*. https://doi.org/10.13140/RG.2.2.23893.60645

4. Rashidi, H., Fennell, B., Albahra, S., Hu, B., & Gorbett, T. (2023). The ChatGPT conundrum: Human-generated scientific manuscripts misidentified as AI creations by AI text detection tool. *Journal of Pathology Informatics, 14*, 100342. https://doi.org/10.1016/j.jpi.2023.100342

5. Mindner, L., Schlippe, T., & Schaaff, K. (2023). *Classification of Human- and AI-Generated Texts: Investigating Features for ChatGPT*. https://doi.org/10.1007/978-981-99-7947-9_12

6. Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: a comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education*. https://doi.org/10.3389/feduc.2024.1347421

7. Ma, Y., Jiawei, L., & Yi, F. (2023). *Is This Abstract Generated by AI? A Research for the Gap between AI-generated Scientific Text and Human-written Scientific Text*. https://doi.org/10.48550/arXiv.2301.10416

8. 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*, 1-12. https://doi.org/10.37074/jalt.2024.7.1.14

9. Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*. https://doi.org/10.1007/s40979-023-00146-z

10. Gotoman, J., Luna, H., Sangria, J., Cereneo, C., & Barbuco, D. (2025). Accuracy and reliability of AI-generated text detection tools: A literature review. *American Journal of IR 4.0 and Beyond, 4*(1), 1-9. https://doi.org/10.54536/ajirb.v4i1.3795

11. Chaka, C. (2024). Accuracy pecking order – How 30 AI detectors stack up in detecting generative artificial intelligence content in university English L1 and English L2 student essays. *Journal of Applied Learning & Teaching, 7*(1), 1-13. https://doi.org/10.37074/jalt.2024.7.1.33

12. Ladha, N., Yadav, K., & Rathore, P. (2023). AI-generated content detectors: Boon or bane for scientific writing. *Indian Journal of Science and Technology, 16*(39), 3435-3439. https://doi.org/10.17485/IJST/v16i39.1632

13. Price, G., & Sakellarios, M. (2023). The effectiveness of free software for detecting AI-generated writing. *International Journal of Teaching, Learning and Education, 2*(6), 31-38. https://doi.org/10.22161/ijtle.2.6.4

14. Walters, W. (2023). The effectiveness of software designed to detect AI-generated writing: A comparison of 16 AI text detectors. *Open Information Science, 7*. https://doi.org/10.1515/opis-2022-0158

15. László, B. (2024). Artificial Intelligence-Generated Text in Higher Education – Usage and Detection in the Literature. *Interdisciplinary Description of Complex Systems, 22*, 238-245. https://doi.org/10.7906/indecs.22.3.1

16. Jung, M., Panizo, C., Dugan, L., Fung, M., Chen, P.-Y., & Liang, P. (2025). *Group-adaptive threshold optimization for robust AI-generated text detection*. arXiv Preprint. https://doi.org/10.48550/arXiv.2502.04528
17. Ji, J., Guo, J., Qiu, W., Huang, Z., Xu, Y., Lu, X., Jiang, X., Li, R., & Li, S. (2025). *I know myself better, but not really greatly: Using LLMs to detect and explain LLM-generated texts*. arXiv Preprint. https://doi.org/10.48550/arXiv.2502.12743
18. Shivarkar, P. (2025). LLM AI-generated text detection. *International Journal of Scientific Research in Engineering and Management, 9*(1), 1-9. https://doi.org/10.55041/IJSREM41593
19. Walker, E., Evans, L., Mitchell, A., Zhang, Z., Patel, R., & Chen, I. (2024). *Text classification in detection of AI-generated content using BERT*. ResearchGate Preprint. https://doi.org/10.13140/RG.2.2.30193.90722
20. Krois-Lindner, A. (2011). *International legal English: a course for classroom or self-study use*. Cambridge University Press. Cambridge professional English. ISBN 9783125351905.

Програмне виявлення україномовних текстів, згенерованих ШІ: методи, оцінки, виклики

С. Д. Приходченко[*], *О. Ю. Приходченко*, *О. С. Шевцова*, *І. Г. Олішевський*

Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна

[*] Автор-кореспондент e-mail: prykhodchenko.s.d@nmu.one

Мета. Оцінка ефективності існуючих інструментів виявлення тексту, згенерованого штучним інтелектом, для виявлення машинно-генерованого контенту у кваліфікаційних роботах студентів. Дослідження має на меті проаналізувати точність, надійність і обмеження цих методів виявлення, зокрема для текстів, написаних українською мовою, а також вивчити потенційні покращення розпізнавання вмісту ШІ.

Методика. У дослідженні використовується поєднання статистичного аналізу й експериментальної оцінки інструментів виявлення ШІ. Набір даних академічних текстів, створених ШІ і написаних людиною, використовується для оцінки точності виявлення з додатковими експериментами, проведеними для аналізу впливу мовних структур, ефектів перекладу та формалізації тексту. Дослідження також включає обчислювальні методи для вивчення рівня помилок і визначення закономірностей у розпізнаванні контенту, створеного ШІ.

Результати. Результати дослідження дають зрозуміти продуктивність поточних систем виявлення ШІ, підкреслюючи їхні сильні й слабкі сторони. Аналіз виявив значні проблеми із виявленням тексту, створеного штучним інтелектом, особливо не англійською мовою, таких як українська. Крім того, дослідження визначає вплив перекладу на точність виявлення та обговорює ефективність різних лінгвістичних і статистичних підходів.

Наукова новизна. Це дослідження робить внесок у поточний дискурс про академічну доброчесність, розглядаючи обмеження інструментів виявлення ШІ для неангломовних академічних текстів. На відміну від попередніх досліджень, що зосереджені в основному на англомовному контенті, створеному штучним інтелектом, це дослідження пропонує унікальний погляд на проблеми виявлення тексту, створеного штучним інтелектом, українською мовою, пропонуючи нові погляди на адаптацію моделей виявлення для різноманітних мовних контекстів.

Практична значимість. Результати цього дослідження є цінними для викладачів, дослідників і політиків, які займаються підтриманням академічної доброчесності в епоху генеративного ШІ. Виявляючи слабкі місця в поточних системах виявлення та пропонуючи можливі вдосконалення, дослідження створює основу для розробки більш надійних методологій виявлення ШІ, які можна ефективно застосовувати до академічних текстів кількома мовами.

Ключові слова: *згенерований ШІ текст, академічна доброчесність, інструменти виявлення, вища освіта, обробка природної мови, машинне навчання, багатомовний аналіз тексту*

The manuscript was submitted 20.02.25.