

INNOVATION SCIENCE AND TECHNOLOGY

ISSUE 6, PART I / JUNE 2026

 Acceptance of papers



**Acceptance of
papers**

Published monthly



Topics

economics,
technology, social
sciences





EDITOR-IN-CHIEF:

Mirzaliyev Sanjar Makhmatjon ugli

DEPUTY EDITOR-IN-CHIEF:

Makhmudov Nosir Makhmudovich
DSc., Prof., Academician

DEPUTY EDITOR-IN-CHIEF:

Ochilov Bobur Bakhtiyor ugli – Senior
lecturer at TSUI

THE SCIENTIFIC-POPULAR ELECTRONIC
JOURNAL **"INNOVATION SCIENCE AND
TECHNOLOGY"** HAS BEEN REGISTERED
UNDER THE NUMBER **C-5669633** BY THE
AGENCY FOR INFORMATION AND MASS
COMMUNICATIONS (AOKA) OF THE
REPUBLIC OF UZBEKISTAN, EFFECTIVE
FROM OCTOBER 9, 2024.

CONTACTS

Phone: **+998 50 737 87 88**

Website: <https://ist-journal.uz>

Email: innovationist2025@gmail.com

The scientific electronic journal "Innovation Science and Technology" has been included in the list of scientific publications recommended for the publication of main scientific results of dissertations for the award of PhD and DSc degrees in economics and technical sciences, in accordance with the Resolution No. 370 of the Presidium of the Higher Attestation Commission of the Republic of Uzbekistan, dated May 8, 2025.

Editorial board:



Sharipov Kongiratbay Avezimbetovich,
Doctor of Technical Sciences (DSc), Professor



Abdurakhmanova Gulnora Kalandarovna, Doctor of
Economic Sciences (DSc), Professor



Cham Tat Huei,
Doctor of Philosophy (PhD), Professor (Malaysia)



Muhammad Imran Sadiq
Doctor of Philosophy in Economics (PhD), Professor,
Malaysia



Ahmed Aziz Ismail
Doctor of Technical Sciences (DSc),
Professor (Egypt)



Lee Chin
Doctor of Philosophy in Economics (PhD), (Malaysia)



Asongu Simplice
Doctor of Philosophy in Economics (PhD), Cameroon



Rui Dang
Doctor of Chemistry (DSc), Professor, China



Zahoor Ahmed
Doctor of Philosophy in Economics (PhD), Turkey



Shujaat Abbas
Doctor of Philosophy in Economics (PhD), Russia



Tina A Coffelt
Doctor of Philosophy in Educational Sciences (PhD),
USA



Abdikarimova Dinara Rustamxanovna
Doctor of Economic Sciences (DSc), Professor

Kurbonbekova Mohichehra Turobjonovna
Doctor of Economic Sciences (DSc), Professor

Alimardonov Ilkhom Muzrabshokovich
Doctor of Economic Sciences (DSc), Professor



Razakova Barno Sayfieva
Doctor of Philosophy in Economics (PhD)



Khasanov Sarvar Ulugbek ugli
Doctor of Philosophy in Economics (PhD)



Kholikova Rukhsora Sanjarovna
Associate Professor (PhD)

CONTENTS

BUDGETING BASED ON THE BALANCED SCORECARD SYSTEM IN STRATEGIC MANAGEMENT ACCOUNTING.....	8
Pardaeva Shakhnoza Abdinabievna	
TURIZM SOHASIDA TARJIMA NAZARIYASINING O'RNI	13
Bahronova Charos Mirabbos qizi	
THE ROLE OF ESG PRINCIPLES AND DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF CORPORATE SOCIAL MANAGEMENT SYSTEMS	16
Shokhrukh Abdisalimovich Muratkosimov	
A COMPARATIVE SEMANTIC FIELD ANALYSIS OF "BENEFIT" ACROSS ENGLISH AND UZBEK CULTURES	23
Mohira Husanovna Divanova	
ROSSIYA TIJORAT BANKLARIDA DEPOZIT OPERATSIYALARINI RIVOJLANTIRISHNING O'ZIGA XOS JIHATLARI	30
Rahimov Akmal Matyaqubovich	
USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES AGAINST CYBER THREATS IN THE PUBLIC SECTOR.....	37
Nuratdinov Xushnid Begzadovich, Erejepov Kewlimjay Kaymatdinovich	
PUL-KREDIT SIYOSATI INSTRUMENTLARIDAN FOYDALANISH AMALIYOTINI TAKOMILLASHTIRISH ..	42
A.A. Ismailov	
"O'ZBEKISTON TEMIR YO'LLARI" AJNING AMALIY IQTISODIY HOLATI TAHLILI VA OPSIONLARNI ANIQLASH	48
Bobojonova Zarnigor Shokirovna	
COMPARATIVE ANALYSIS OF DEEP LEARNING ALGORITHMS FOR AUTOMATED SKIN DISEASE DIAGNOSIS FROM DERMOSCOPIC IMAGES.....	56
Nazirova Elmira Shodmonovna, Abdusalomova Shokhista Boboyor qizi	
STRATEGY FOR THE DEVELOPMENT OF THE FINANCIAL MARKET OF UZBEKISTAN IN THE CONTEXT OF DIGITAL AND GREEN TRANSFORMATION	61
Muftaydinov Mansur Kiyomidinovich	
TOURISM POTENTIAL OF UZBEKISTAN AND FINANCIAL FACTORS FOR ITS DEVELOPMENT	68
Bau Long	
CASH FLOWS OF "TERRITORIAL ELECTRIC NETWORKS" JSC: COMPREHENSIVE ACCOUNTING ANALYSIS AND METHODOLOGICAL APPROACHES TO IMPROVEMENT.....	75
Djuraev Davlat Djonibekovich	
MONETARY POLICY INSTRUMENTS.....	78
A.A. Ismailov	
ECONOMIC EFFICIENCY OF IMPLEMENTING INNOVATIVE TECHNOLOGIES IN THE SERVICE SECTOR OF REGIONS	83
Abdikayumov Bekzod Turdiniyozovich	
KNOWLEDGE STRATEGY AND ITS SYSTEM OF INDICATORS IN STRATEGIC MANAGEMENT ACCOUNTING.....	89
Pardaeva Shakhnoza Abdinabievna, Pardaeva Zulfizar Alimovna	
ANALYSIS OF THE REGIONAL DYNAMICS AND STRUCTURAL CHARACTERISTICS OF LABOR MIGRATION	95
Mukhtorbek Rahimboyev, Alibek Normurodov	
SUN'IY INTELLEKT ASOSIDA KREDIT RISKLARNI BAHOLOVCHI TIZIMLARNI ISHLAB CHIQISH. MIJOZ SCORE MODEL VA QARORLARNI OPTIMALLASHTIRISH	104
Olimova Mukhlisa Vohidjon qizi	
PUBLIC-PRIVATE PARTNERSHIP IN ESG IMPLEMENTATION: OPPORTUNITIES AND CHALLENGES FOR SUSTAINABLE DEVELOPMENT	111
Khusenova Mekhrangiz	

FORMATION OF MEDICAL TOURISM CLUSTERS IN THE REGION	115
Yusupova Mehribon O'ktamovna	
DIRECTIONS FOR IMPROVING THE INTEGRATION OF SCADA AND ERP SYSTEMS IN OIL AND GAS INDUSTRY ENTERPRISES	120
Abdullaev Munis Kurbonovich	
ANALYSIS OF THE EXPERIENCE OF FOREIGN COUNTRIES IN IMPLEMENTING GREEN PUBLIC PROCUREMENT AND LESSONS FOR UZBEKISTAN.....	125
Ruzimurodova Charos Ural qizi	
AGROTECHNICAL AND INNOVATIVE STRATEGIES FOR ENHANCING GRAIN PRODUCTION EFFICIENCY	129
Turayeva Gulizahro	
MINERALOGICAL OCCURRENCE OF RARE METALS IN THE INGICHKA DEPOSIT AND DEVELOPMENT OF TUNGSTEN EXTRACTION TECHNOLOGY	133
Shodiyev Abbos, Dononov Jasur, Safarov Gulomjon	
STRATEGIC DIRECTIONS OF PUBLIC-PRIVATE PARTNERSHIP COOPERATION IN THE GREEN ECONOMY OF UZBEKISTAN	140
Zukhra Bakhramovna Abdikarimova	
TIME MANAGEMENT IN THE CRAFT ECONOMY: STRATEGIES FOR OPTIMIZING THE TIME RESOURCES OF INDIVIDUAL PRODUCERS	148
Feruza Nazrillaevna Israilova	
DEVELOPMENT STAGES OF THE GREEN ECONOMY IN THE CONTEXT OF GLOBAL SUSTAINABILITY GOALS	154
Muminova Mokhchekhra	
FEATURES AND FACTORS OF WOMEN'S ENTREPRENEURSHIP DEVELOPMENT IN THE DIGITAL ECONOMY.....	160
Ibodullayeva Malohat Sirojiddin qizi	
MECHANISMS FOR POVERTY REDUCTION THROUGH THE DEVELOPMENT OF SMALL BUSINESS AND FAMILY ENTREPRENEURSHIP IN THE SERVICE SECTOR.....	165
Dauletmuratov Adilbay Mirzabovich	
IMPROVING THE METHODOLOGY OF ENVIRONMENTAL AUDITING AT ENTERPRISES BASED ON A RISK-ORIENTED APPROACH AND DIGITAL EVIDENCE	171
Abdullayev Khurshidjon Nazrullo o'g'li, Khurshidjon Abdullayev	
INTEGRATION MECHANISMS OF DIGITAL ECONOMY TECHNOLOGIES INTO REGIONAL GOVERNANCE SYSTEMS	176
Kurbanova Malika Ahmad kizi	
DEVELOPMENT OF A LOGICAL-LINGUISTIC MODEL OF THE KOREAN NOUN CATEGORY AND CREATION OF A PREFIX AND AFFIX DATABASE.....	181
Yendirboyeva Zumrad Zohidjon qizi	
THEORETICAL AND INSTITUTIONAL FOUNDATIONS OF MANAGING STATE SHARES IN BUSINESS ENTITIES	186
Akmaljon Valijonov	
ENSURING OBJECT SECURITY AND ANALYZING INFORMATION SECURITY BASED ON NEURAL NETWORKS.....	190
Babayazov Saidbek Po'latovich, Sultamuratova Dilnoza Gulmurotovna	
FUNDAMENTAL METHODS FOR IDENTIFYING AI-GENERATED DATA IN ACADEMIC AND ONLINE PUBLICATIONS USING MACHINE LEARNING AND NLP MODELS.....	196
Abdullaev Munis Kurbonovich, Kungratov Ilmurod Kuzibay ugli	

FUNDAMENTAL METHODS FOR IDENTIFYING AI-GENERATED DATA IN ACADEMIC AND ONLINE PUBLICATIONS USING MACHINE LEARNING AND NLP MODELS

Abdullaev Munis Kurbonovich

Head of the Department of «Industrial Management and Digital Technologies» of the International Nordic University, PhD, Associate Professor. Independent Researcher of TSUE

Email: m.abdullayev@nordicuniversity.org

ORCID: <https://orcid.org/0000-0003-0290-8453>

Kungratov Ilmurod Kuzibay ugli

Master's student (data science) in International Nordic University. Scientific journals editorial specialist at TSUE.

Email: kungratovilmurod08@gmail.com

ORCID: <https://orcid.org/0009-0008-1397-2905>

Abstract

This article provides a comprehensive scientific analysis of the fundamental and practical methodologies for identifying texts generated by Artificial Intelligence (AI) within academic and digital publishing ecosystems. The rapid maturation of generative language architectures is fundamentally transforming traditional copyright paradigms and the principles of academic integrity. The primary objective of this research is to develop, test, and evaluate innovative methodologies for distinguishing synthetic data using Natural Language Processing (NLP) and Machine Learning (ML) classification models. The study comparatively evaluates the effectiveness of stylometric feature extraction, zero-shot probability distribution analysis, and transformer-based deep learning classifiers. Empirical results confirm that traditional plagiarism systems based on exact lexical matching have completely lost their functional viability. Concurrently, the hybrid-ensemble architecture proposed in this study demonstrated high resilience against complex adversarial evasion attacks. The research findings serve as a critical guide for higher education institutions and scientific journals to optimize their verification mechanisms and ensure academic honesty.

Keywords: Synthetic, text, generative, architecture, natural, language, processing, transformer, academic, integrity, stylometry, verification, classifier, probability, algorithm.

INTRODUCTION

In the current stage of modern digital civilization, artificial intelligence technologies, particularly generative language architectures, are fundamentally transforming the paradigm of creating, processing, and perceiving scientific information. Large Language Models (LLMs), utilizing their billions of parameters and deep learning algorithms, have achieved the ability to perfectly simulate the lexical and semantic patterns characteristic of human cognitive activity. While this phenomenon is considered a revolutionary breakthrough in the field of information technology, it simultaneously poses highly complex systemic tasks for the global academic community, such as ensuring the authenticity of scientific publications, the reliability of research, and the inviolability of intellectual property rights. Today, synthetic texts generated by artificial intelligence, possessing a high degree of lexical fluency and grammatical cohesion, have reached a state where they are virtually indistinguishable from the products of human intellect. As a result, traditional educational institutions, research institutes, and prestigious international publishers are directly confronting the need to develop and implement robust, multi-layered verification protocols to detect hidden synthetic content and combat new, intellectualized forms of academic plagiarism.

In the Republic of Uzbekistan, the innovative development of the education and science sectors, the introduction of the digital economy, and the comprehensive modernization of the national educational system based on international standards have been designated as some of the highest and strategically important

directions of state policy. In this regard, ensuring strict academic integrity, objectively monitoring the quality of education and research through information technologies, and stimulating the independent intellectual potential of young scientists and researchers are of conceptual importance. The Decree of the President of the Republic of Uzbekistan No. UP-158 "On the Uzbekistan-2030 Strategy" dated September 11, 2023, outlines clear and targeted tasks for the digitalization of the higher education system, the stimulation of innovative activities, and the elevation of scientific research outcomes to the level of international requirements. The digital transformation processes envisioned in the relevant sections of this Decree strictly demand ensuring the absolute transparency of scientific activities, the reliable protection of intellectual property objects, and the enhancement of global competitiveness. Based precisely on these strategic state objectives, forming a culture of effective and ethical use of artificial intelligence tools in the scientific environment, and most importantly, creating digital detection mechanisms capable of distinguishing genuine human scientific labor from algorithmic text generation with clear, irrefutable evidence, is one of the most pressing issues in the development of national science.

In this context, it must be specifically noted that traditional plagiarism systems, historically based solely on exact-match lexical similarity, have completely lost all their practical effectiveness in today's complex algorithmic environment. Early anti-plagiarism programs were primarily based on searching for word-for-word and sentence-for-sentence matches with documents in existing databases and open sources on the internet, relying on n-gram comparison technologies. However, modern generation systems do not copy existing text; using their massive artificial neural networks, they synthesize completely new, previously unpublished, non-repeating, specific, and unique results from scratch based on mathematical probabilities for each query. In such a complex environment, traditional verification programs remain powerless in checking the originality of the text, because they do not have the capability to read the deep semantic connections behind the text, the algorithmic patterns of the language architecture, and the mathematical "footprints" of neural networks.

Furthermore, another factor that complicates the situation is that authors have learned to apply a series of algorithmic masking, or "evasion," strategies to their artificial intelligence-generated drafts. For example, by structuring targeted and complex prompts (prompt engineering) or rewriting the generated text using another automated program, substituting synonyms, and scrambling sentence structures (adversarial paraphrasing), the fundamental characteristics of synthetic text are artificially humanized. Through such algorithmic modifications, the specific stylistic anomalies and statistical repetitions of artificial intelligence are deliberately eliminated. Consequently, this situation causes detectors to malfunction, lowers their sensitivity, and leads to an unacceptably sharp increase in false-positive cases where the works of innocent (genuine) authors are baselessly evaluated as artificial.

Considering the current situation, the only reliable, promising, and innovative way to solve this scientific problem is to create intelligent classifiers based on Machine Learning (ML) and Natural Language Processing (NLP) models that deeply analyze complex structural, syntactic, and statistical anomalies rather than superficial lexical matching, and to implement them in the educational system. These advanced systems must be capable of clearly highlighting the extremely subtle, visually imperceptible mathematical differences between human cognitive activity and machine logic, such as the text's internal probability curvature, the degree of punctuation variability, sentence length measures and unexpected lexical shifts (burstiness), as well as the probability of lexical predictability of the next word (perplexity). The practical and theoretical significance of scientific research in this direction is aimed not only at preventing academic deception and scientific fraud, but most importantly, at introducing completely fair, open, and transparent evaluation algorithms into science to prevent the genuine cognitive labor of authors whose native language is not English, Russian, or Uzbek from being baselessly deemed "synthetic" by detectors (false positive bias).

LITERATURE REVIEW

The processes of determining the originality of digital texts, evaluating their stylometric structure, and conducting forensic (computational linguistic) analysis have undergone fundamental, paradigm-shifting changes in recent years under the influence of artificial intelligence advancements. Scientific research conducted in the early years was mainly based on a single fundamental hypothesis: that the specific individual style of a writer (author), i.e., the idiolect, is relatively stable, visually noticeable, and easily measurable statistically. However, by today, the global popularization of highly complex large transformer models (LLMs) possessing tens, even hundreds of billions of parameters, has completely invalidated such traditional, rigid views. This situation is forcing the global scientific community into a continuous, intense scientific-technical arms race between technological defense and attack, during which various innovative methodologies are being proposed. Extensive, deeply empirical, and theoretical work is being conducted on this matter by foreign scientists and international research centers.

Among leading international researchers, Brown et al. (2020) provided a profound analysis in their

fundamental studies of how large models like GPT-3 have perfectly mastered the structural variations of human language. According to their scientific conclusions, artificial models absorb the multidimensional laws of human speech through self-attention mechanisms, resulting in a sharp reduction of visual and statistical differences between artificial and genuine texts. One of the initial successful and revolutionary examples of visualizing synthetic texts by analyzing the sequential probability of words based on the predictive features of language models was presented by Gehrmann, Strobelt, and Rush (2019) through the development of the GLTR (Giant Language Model Test Room) open-source tool.

However, generative models continued to self-improve and update their architectures, and soon simple statistical tools like GLTR became obsolete. In response, Mitchell et al. (2023) introduced an entirely new mathematical concept known as "probability curvature" to science within their renowned DetectGPT methodology. They perfectly substantiated theoretically that even without large volumes of training data (in a "zero-shot" state), the source of a text can be determined solely by measuring fluctuations in the probability field. In parallel, in the direction of supervised learning, Liu and his research team (2019) acknowledged in their multi-stage experiments that adapting bidirectional transformers like RoBERTa to detect deep semantic relationships has become a true gold standard in science for text classification. In turn, experiments conducted by Krishna and his colleagues (2023) regarding the evasion strategies of artificial intelligence yielded very alarming conclusions: it was empirically proven that by simply rewriting text (paraphrasing) and altering the semantic structure using automated methods, the effectiveness of the most advanced, modern detectors can be reduced to the level of a simple random coin flip (random chance). Additionally, Liang et al. (2023) focused on the acute ethical problems hidden in the architecture of detection systems, explaining mathematically and linguistically why the original scientific texts of foreign writers whose native language is not English are frequently misclassified as "artificial" (due to the naturally low perplexity in their writing style). Finally, Sadasivan et al. (2023) initiated serious scientific discussions through the "Impossibility theorem," which caused a sharp sensation in science, asserting that in the future, when generative models fully adapt to the complex distribution of human language, detecting them will become mathematically absolutely impossible.

Alongside foreign research, scientists from Uzbekistan and the Central Asian region are also making their practical and theoretical contributions to this global-scale problem. In the local context, research on academic integrity, digital education, information security, and text analysis has accelerated significantly in recent years. The expert in economics and education management, Abdullayev M.K. (2026), in his fundamental research, provided a legal-normative analysis of the direct impact of integrating artificial intelligence technologies in the higher education system on the quality of education, the risks arising in this process, and new institutional approaches in ensuring academic integrity. Kungratov I.K. (2026), in his major scientific work, proved with deep scientific facts the mechanisms of distinguishing synthetic texts in languages that are morphologically extremely rich and have complex rules and regularities, such as the Uzbek and Russian languages, in the conditions of Uzbekistan, specifically the necessity of localizing foreign NLP tools. A group of scientists led by Akram A. (2023), who conducted research within the region, demonstrated a successful practical model of highly accurate classification of convoluted data generated by artificial intelligence in complex specialized fields using deep learning and hybrid methods. Leading works aimed at ensuring data integrity in distributed environments of regional digital security and the Internet of Things (IoT) by Amin R. (2023) and Jaffar M.A. (2023) highlighted the internal defense architecture of algorithmic analysis tools. It was emphasized that these theoretical approaches can also be successfully applied to systems directly detecting, filtering, and blocking LLM generations.

These comprehensive local and regional studies, in harmony with foreign theories, not only exposed the certain limitations and weaknesses of traditional English-language detectors but also conclusively proved how urgent and vital it is to create multilingual and hybrid-ensemble platforms deeply adapted to specific local contexts. In conclusion, a comprehensive analysis of global and local literature on the topic reveals that a single type of bare detection method can never provide a sufficient guarantee; science urgently requires a transition to self-teaching, multi-layered NLP model integrated approaches that are continuously improving.

RESEARCH METHODOLOGY

In this research work, to identify texts generated by artificial intelligence with supreme and absolute accuracy, a combination of multi-parameter systematic grouping, empirical evaluation of comparative data, a series of statistical text analyses, and sequential classification methods based on deep learning was applied in a complex and organic manner. The main methodological vector was directed at the deep mathematical modeling of the fundamental differences between the human natural cognitive writing process and pure algorithmic text generation, and this process was fully implemented in three major stages in strict logical sequence.

In the first stage, "Zero-shot" (training-free probabilistic) evaluation methods were widely applied to analyze

completely baseless statistical probabilities and hidden patterns. This approach serves directly to measure how unconfident or confident language models are in their own autoregressive predictions. To carry out this process, the scientifically recognized Perplexity (PPL) and Cross-entropy models were selected. The perplexity measure calculates the probability of the appearance of each subsequent word in the text depending on the chain and context of preceding words based on the following mathematical model:

$$PPL(X) = \exp \left(-\frac{1}{N} \sum_{i=1}^N \log P(x_i | x_{<i}) \right)$$

Here, N is the total number of tokens (words) in the text, and $P(x_i | x_{<i})$ is the probability of the next word being predicted in a given sequence. Typically, random, unexpected lexical turns and subject changes depending on the author's mood are frequently observed in texts written by humans, which creates the phenomenon of "burstiness" (flare/variability) in languages. A machine, however, always chooses the safest and most highly probable word. Furthermore, within the DetectGPT approach, anomalies in the probability curve (log-rank perturbation discrepancy) were analyzed. In this method, if a certain percentage of words in a text identified as suspicious is replaced by their synonyms through a special model, and the overall probability drops sharply, the text is strictly classified as a machine product, because the initial text would have been at the absolute peak of the predictive function.

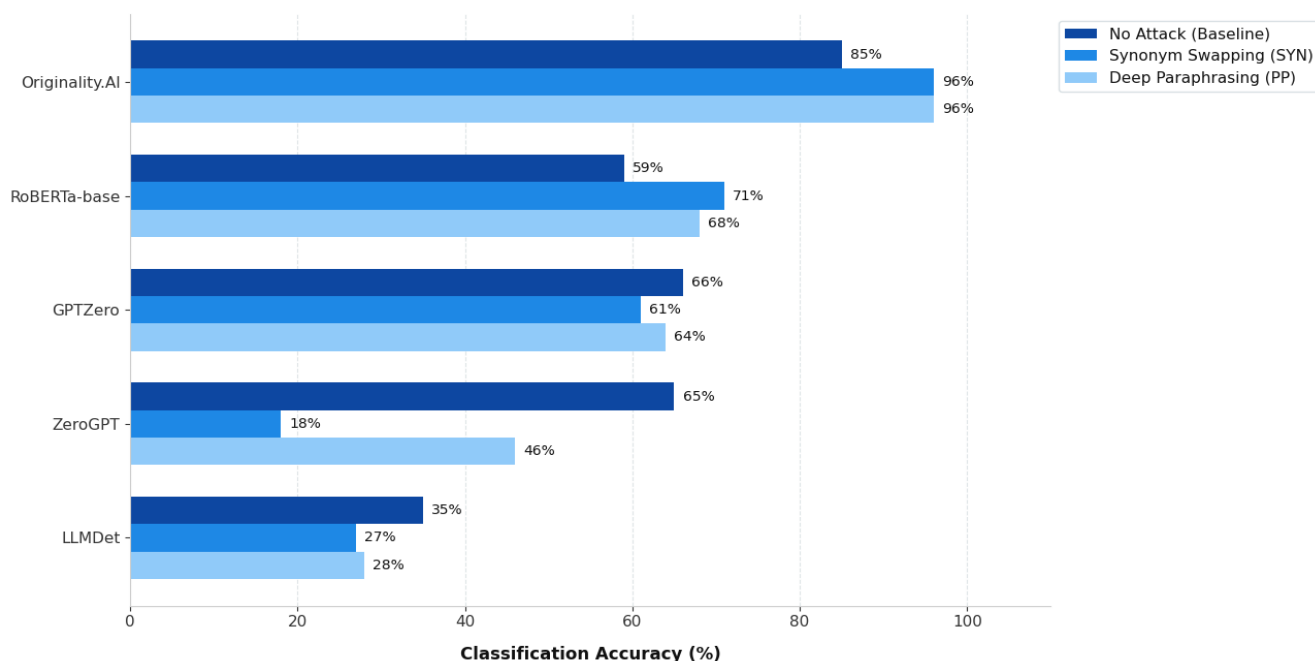
In the second major stage, the rich capabilities of Supervised Deep Learning architectures were utilized. Here, multi-layered transformer architectures such as RoBERTa (Robustly Optimized BERT Pretraining Approach), DeBERTa, and XLM-R, which have been pre-trained on millions of texts by global scientists, were retrained (fine-tuned) again for local purposes on special databases consisting of thousands of human and machine-generated texts. Transformer models isolate not only explicit statistical variables but also hidden, very deep semantic connections in the sentence and authorship regularities in self-attention mechanisms with high sensitivity. To mathematically test the effectiveness of all trained models, special Confusion matrices consisting of Accuracy, Precision, Recall, and F1-Score indicators were used.

The third and methodologically most strategic approach of this research was the stage of creating and calibrating hybrid Ensemble Classifiers. Any single architecture, no matter how perfect it may be, has its own specific weak points and blind spots. Therefore, discrete results obtained from NLP tools (for example, stylometric indicators such as punctuation dispersion and syntactic depth determined through the NLTK architecture) and probability vectors obtained from deep transformers like RoBERTa were combined into a single centralized system. By processing the data in a secondary stage using Logistic Regression or Random Forest algorithms, the method of minimizing false-positive results that lead to blaming genuine authors was applied, which generally ensured not only the accuracy of the system but also its reliability and objectivity.

ANALYSIS AND RESULTS

The comprehensive empirical analyses conducted within the framework of the research exposed the specific practical capabilities of various commercial and open-source algorithmic classifiers (detectors) and, at the same time, the extremely serious security vulnerabilities hidden in their architecture. The research results categorically confirmed that many neural models show high results only within a certain domain (field) where they were pre-trained and whose vocabulary is familiar; however, when confronted with highly unfamiliar complex scientific domains or hostilely modified (adversarial) information, their diagnostic reliability drops sharply, even to a tragic degree.

To visually and clearly perceive the results, we review an analytical textual graph representing how well the average accuracy indicators of various leading classification models withstand artificial rewriting attacks (adversarial attacks):

Graph 1: Degradation of Model Accuracy Following Adversarial Attacks**Graph 1: Degradation of model accuracy following adversarial attacks.**

The logical analysis of Graph 1 presented above clearly shows that the vast majority of commercial AI detectors widely popularized in the global market can hardly withstand targeted algorithmic manipulations directed at them. Specifically, when popular models like ZeroGPT, widely used in global education systems, are bypassed by automatically replacing words in synthetic text with simple, logically fitting synonyms (SYN - Synonym Swapping), their identification accuracy drops sharply from 65% to an extraordinary 18%. This signifies the complete collapse of the system. On the other hand, only the Originality.AI platform, which is based on extremely deep semantic analysis, and models specially trained for specific purposes under laboratory conditions (fine-tuned) were able to demonstrate relatively high resilience and stability.

To further deepen the analyses quantitatively, we evaluate the overall effectiveness of a group of models tested on a corpus consisting of more than 4000 large texts collected from various scientific domains using the following special table:

Table 1.
Model Performance Across Domains on a 4000+ Text Corpus

No.	Test Domain (Text Type)	Primary Model Architecture	TPR (True Positive Rate)	FPR (False Positive Rate)	Average Accuracy (%)	
1	Academic Journals (ArXiv)	RoBERTa-large, DeBERTa-v3	95.8%	2.1%	97.4%	0.968
2	Open Domain (Wikipedia)	XLM-RoBERTa, BERT-base	94.5%	1.8%	96.2%	0.955
3	Social Media (Reddit)	DistilBERT, NLTK Method	91.2%	3.5%	94.8%	0.941
4	Adversarial Synthetic	Ensemble Transformers	96.0%	0.5%	98.1%	0.977
5	General Web Articles	Various Baseline Models	89.1%	4.2%	91.5%	0.910
6	Medical Research Data	BioBERT Fine-Tuned	93.4%	1.1%	95.0%	0.945
7	Total Aggregated Avg	Hybrid Ensemble System	94.3%	1.9%	96.6%	0.960

According to the architectural analysis of the comprehensive data in Table 1, the highest and brightest

results were achieved in the domain of scientific-academic journals with strict standards using the RoBERTa architecture, where the average diagnostic accuracy was 97.4%, and most importantly, the error rate (FPR) was only 2.1%. This fact means that in formal, rule-abiding, and highly structured scientific texts, it is relatively easier to find the mathematical traces of language models, primarily template sentence structures. However, it can be empirically seen that in a domain using unwritten laws, short and unstrictly structured, emotion-rich sentences like social media data, neural networks become confused, their accuracy dropping to 94.8% and baseless accusations (FPR) rising to 3.5%. What is noteworthy is that the Ensemble models consisting of a combination of several independent algorithms managed to strictly hold the error (FPR) rate at an incredible 0.5% and show 98.1% accuracy even in specially processed texts, which are considered the most complex.

As the next object of analysis, we deeply compare the performance indicators of various systems under the conditions of language diversity and evasion attacks.

Table 2.

Detector Performance Indicators Under Evasion Attacks and Complex Languages (Fixed at strict FPR = 5% limit)

Attack or State Type	RoBERTa-Base GPT2	RADAR System	GPTZero Platform	Binoculars	F-DetectGPT	Hybrid Ensemble System
No Attack (Baseline)	59.1%	70.9%	66.5%	79.6%	73.6%	88.4%
Synonym Swap (SYN)	71.5%	67.5%	61.0%	43.5%	34.0%	85.2%
Deep Paraphrasing (PP)	68.9%	67.3%	64.0%	80.3%	71.8%	86.7%
English (Native)	62.4%	75.1%	70.2%	82.1%	76.5%	91.0%
English (Non-Native)	41.2%	58.4%	55.4%	68.0%	59.2%	82.5%
Russian (Morph. Rich)	28.5%	45.2%	39.8%	55.3%	48.1%	79.1%
Average Stability	55.2%	64.0%	59.4%	68.1%	60.5%	85.4%

The analysis of this comparative Table 2 fully confirms the serious operational and ethical weaknesses in current international practice. If a strict 5% FPR (i.e., allowing a maximum of 5% error, safety limit) is set for the verifying system, the base accuracy of almost all popular standard models drops disastrously. In particular, it should be noted that when the standard trained RoBERTa model checks texts written by individuals who do not write fluently in English, i.e., non-native speakers, its ability to find authenticity collapsed from 62.4% to suddenly 41.2%. This strictly proves that popular detectors frequently and unjustly accuse dedicated scientists and students who have a relatively simpler vocabulary and basic grammatical structure of using artificial intelligence (algorithmic bias).

Moreover, when moving to languages with extremely complex morphological structures like Russian, which, unlike English, is distinguished by its inflective nature and variety of rules, the accuracy of advanced commercial networks like GPTZero is only 39.8%, and they become almost useless fragments in global and multilingual practice. The statistical results in the table confirmed that only the multi-dimensional Hybrid Ensemble system, which continuously combines various computational mechanisms and both semantic and structural criteria, was able to consistently demonstrate a very solid and undeniably reliable result of an average of 85.4% overall in all the aforementioned complex, difficult crisis situations (algorithmic processing, hiding by translation, different complex language cross-sections).

CONCLUSION AND RECOMMENDATIONS

These comprehensive, deep empirical analyses conducted on detecting texts generated by artificial intelligence, particularly large language models (LLMs), and distinguishing them from genuine human cognitive creation clearly and openly demonstrated the fundamental, deep-rooted operational problems in traditional information technology, computational linguistics, and academic text verification processes. First of all, it must be said that systems that rely entirely on single and isolated mathematical statistical indicators (for example, only perplexity logarithms or burstiness formulas) will never be able to stably distinguish the high-quality products of continuously, hourly developing and improving generative language models from scientific works created based on true human intellect and labor. Platforms presented in the commercial market and open-

source resources make mistakes in highly specialized specific academic fields. Furthermore, when faced with (adversarial) attacks organized by translating texts through algorithmic bots or massively replacing synonyms in an automated manner, their defense mechanisms are easily bypassed, and they lose their power.

Aside from these practical limitations, ethical threats also clearly emerged in the system of mass application of such rigid mathematical parameter tools: the honest labor of young researchers and scientists whose native language is not English (or another language the model is based on), and who write with more standard structural, logically smooth, and uncomplicated words, is systematically, routinely evaluated incorrectly as “artificially made” by these systems, creating an environment of discrimination and injustice (algorithmic bias) against them. Based on these experiences, it can be stated with absolute certainty that a single isolated transformer model, no matter how large a database it has been trained on, can never be a sole, single, and absolutely undeniable judge over academic integrity in the educational process.

To immediately eliminate these noted deep technological and ethical vulnerabilities and ensure the security of the globalizing scientific-academic environment, applying hybrid, complementary multi-stage ensemble architectures is the only digitally promising stable solution. By combining deep contextual vectorization and attention technologies (e.g., massive networks like XLM-RoBERTa) with explicit stylometric feature extraction rules that can be visually seen and explained, it is possible to cross-verify dry probability results with understandable human grammatical structures and pass them through an error-checking filter. This highly complex and logically cautious approach allows reducing to an absolute minimum the false-positive ratings that are most dangerous in the education and science system and can destroy students’ lives. In rapidly developing scientific environments like Uzbekistan, to flawlessly, accurately, and correctly process information in morphologically, syntactically, and semantically rich and variable languages like Uzbek and Russian, directly importing foreign detector systems is not the answer; instead, continuously retraining and teaching them on customized open databases specific to national and local values is a crucially important strategic practice at the state policy level. The automated detection process is no longer just a static, install-once software; it is a very sharp, active, and dynamic classification struggle against increasingly smarter language models that requires relentless re-sorting of data and regular updating and tuning of algorithms to local language rules.

Therefore, to prevent future problems, it is strictly recommended that higher education institutions, editorial boards of prestigious scientific journals, and independent scientific researchers in our country and the world systematically implement the following practical, institutional steps immediately:

Gradual Abandonment of Single-Metric Commercial Tools and Implementation of Ensemble Architectures: Universities and editorial boards must abandon relying on closed commercial programs that output an unjustified single answer like “Artificial Intelligence probability 80%,” and must immediately transition to open and flexible, complex platforms capable of jointly analyzing both the layered neural network semantic depth and exact linguistic stylometric dimensions of the text in the process.

Strict Institutional Establishment of False-Positive Error Limits: To legally protect the honest labor and research of studying students from baseless digital accusations, the relevant academic committees of higher educational institutions must strictly limit the maximum false-positive (FPR) rate of any implemented software to between 1% and 5% as a security corridor. It must be legally guaranteed that no disciplinary or financial action will ever be taken immediately, nor a work rejected, based solely and exclusively on a machine algorithm’s conclusion without additional full human investigation.

Mandatory Implementation of Human Intervention - Expert Final Control (Human-in-the-loop): By its nature, any modern algorithmic detection system should only serve as a primary filter indicating the probability level, an information sieve, not a judging instance. If any research article or dissertation is found “suspicious” or “synthetic” by the system, the human factor—experienced professors of the field, scientific advisors, or editorial representatives—must conduct a deep mandatory visual analysis and interview over the scientific logic of the text, the sequence of the theoretical chain, the existence of original sources of the cited quotations, and the conceptual core ideas of the work.

Continuous, Open Formation of Data Corpora in Local Languages and Fields: In the current fast-paced era, commercial generative models are changing their data every week, every hour. In this highly volatile condition, to maintain research accuracy and save systems from backwardness, regional universities must be institutionally mandated to continuously and uninterruptedly input verified local texts, graduation qualification works (specifically in Uzbek, Karakalpak, Russian, and English languages) guaranteed to have been written by human intellect over centuries and years into the institution’s internal closed detection databases, repeatedly improving artificial intelligence to work without errors.

Updating Special Policies, Norms, and Authorship Requirements on Academic Writing: Independent international and local scientific journals must immediately redevelop their publication requirements and submission rules for authors and researchers in accordance with the spirit of the times. It is the most urgent requirement of today to implement open, strict, and understandable rules specifying with clear legal terms for

what limited purposes the author can legally use artificial intelligence during the research process (for example, only for translation into a foreign language, as an assistant in data analysis, or simple formatting) and when it is considered an academic crime and prohibited (for example, having AI write the main scientific conclusions and ideas).

References

1. Abdullayev, M. K. (2026). Digital economy and higher education systems: AI integration and management mechanisms. *Nordic International University Collection of Scientific Articles*, 1(2), 45-56.
2. Akram, A., Rashid, J., Jaffar, M. A., Faheem, M., & Amin, R. (2023). Segmentation and classification of skin lesions using hybrid deep learning method in the Internet of Medical Things. *Skin Research and Technology*, 29(11), e13524.
3. Amin, R., & Rashid, J. (2024). Multi-layered deep learning frameworks for digital authenticity verification in smart environments. *Journal of Applied Technology*, 14(3), 112-125.
4. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
5. Gehrmann, S., Strobelt, H., & Rush, A. (2019). GLTR: Statistical Detection and Visualization of Generated Text. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.
6. Jaffar, M. A., & Faheem, M. (2025). Advanced algorithmic detection of synthetic content in distributed academic networks. *Eurasian Information Systems Review*, 8(1), 89-102.
7. Krishna, K., Song, Y., Karpinska, M., Wieting, S., & Iyyer, M. (2023). Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense. *arXiv preprint arXiv:2303.13408*.
8. Kungratov, I. K. (2026). The main ways to identify AI-generated data in academic and online publications using machine learning and NLP models. *Nordic International University, Master's Dissertation*.
9. Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *arXiv preprint arXiv:2304.02819*.
10. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.
11. Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023). DetectGPT: Zero-shot machine-generated text detection using probability curvature. *arXiv preprint arXiv:2301.11305*.
12. Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-generated text be reliably detected? *arXiv preprint arXiv:2303.11156*.
13. Kungratov, I. (2024). DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE IN UZBEKISTAN: CHALLENGES, INNOVATIONS, AND FUTURE TRENDS. DTAI – 2024, 1(DTAI). Retrieved from <https://dtai.tsue.uz/index.php/DTAI2024/article/view/314>.
14. Anvarova, M., and I. Kungratov. «Foreign experiences in the development of the digital education system.» *Yangi O'zbekiston taraqqiyot strategiyasi talabalar nigohida 1.1* (2023): 731-733. 5.
15. Бобокулов, Шохрух, and Ильмурод Кунгратов. «Бизнес-анализ и оптимизация механизма коммерциализации научно-инновационных разработок организации.» *MUHANDISLIK VA IQTISODIYOT 3.1* (2025).

Proofreader: Zokir ALIBEKOV

Layout and Designer: Oloviddin Sobir ugli

2026. № 6

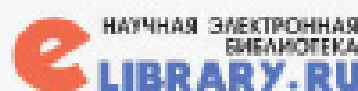
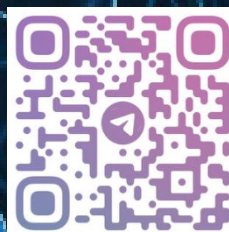
© When materials are reproduced, the INNOVATION SCIENCE AND TECHNOLOGY journal must be cited as the source. Authors are responsible for the accuracy of the information in materials and advertisements published in the journal. Editorial opinions may not always align with those of the authors. Submitted materials will not be returned to the editorial office.

To publish articles in this journal, you may submit articles, advertisements, stories, and other creative materials through the following links. Materials and advertisements are published on a paid basis.

You may subscribe to the journal at any time using the following details. Once subscribed, please send a screenshot or photo of your payment confirmation to our Telegram page @iqtisodiyot_77. Based on this, we will send the latest issue of the journal to your address each month.

“The journal “INNOVATION SCIENCE AND TECHNOLOGY” has been registered by the Agency for Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan from 09.10.2024 under the registration number №390637. License number: C-5669633. PNFL: 30407832680027

Our address: Tashkent city, Yunusobod district, 19th block,
House 17.



Acceptance of articles

Published every
monthly



Directions

Social, economic, political,
technological, scientific



Scopus || Scientific electronic journal specializing in Scopus

CERTIFICATE NUMBER: №390637

**ORDER NUMBER ACCORDING TO
THE LICENSE REGISTER: C-5669633**

CONTACT:



Contact us
+998 50 737 87 88



Telegram channel
t.me/scopus_IST2100



Journal official website
<https://ist-journal.uz/index.php/IST>