

Using AI to Combat AI-generated Fraud: A Semantic Detector with Blockchain Integration

Nikola Čučilović¹, Tamara Jušković²

¹ Ministry of Interior of the Republic of Serbia; nikola.cucilovic@gmail.com

² Ministry of Interior of the Republic of Serbia; tamarajuskovictamara@gmail.com

* Corresponding author: nikola.cucilovic@gmail.com

Received: April 15, 2026 • Accepted: June 4, 2026 • Published: July 10, 2026

Abstract: The rapid development of information technologies and their prevalence in the daily life of individuals, and especially the development of artificial intelligence in the last few years, have brought numerous benefits, facilitated and accelerated business activities, and enabled a wide range of modalities of use in different areas. In parallel with the development of artificial intelligence and its availability to a wide audience through increasingly numerous online tools, various modalities of its abuse have also been developed. Today, artificial intelligence has become the fuel and one of the main tools in various criminal activities on the Internet. Thus, various types of AI-assisted fraud have taken a dominant place on the ladder of high-tech crime today. One of the key questions is what measures we can take to fight against these types of abuse and whether this powerful technology can be used to fight against our own abuse. This paper presents the development and application of a semantic AI fraud detector as a conceptual model that uses NLP to recognize fraudulent messages and writes the results to the blockchain network to ensure data transparency and immutability, with suggestions for further integration and improvement.

Keywords: artificial intelligence, blockchain, semantic analysis, high-tech crime, NLP.

1. INTRODUCTION

The changes brought about in everyday life and business by the use of information technologies in general, and then the use of various tools based on artificial intelligence today, are also observed in the sphere of crime. Thus, high-tech crime today, i.e., the misuse of information technology for criminal purposes, is one of the biggest security threats, even surpassing traditional types of crime, according to the *Cognyte 2025 Law Enforcement Outlook* analysis, conducted among police organizations around the world [1]. Europol's annual analytical report, *EU SOCTA 2025 (Serious and Organized Crime Threat Assessment)*, confirms previous statements and shows that high-tech crime is one of the most profitable areas of organized crime [2]. In the aforementioned report, the abuse of



Citation: N, Čučilović¹, T, Jušković “Using AI to Combat AI-generated Fraud: A Semantic Detector with Blockchain Integration,” *Journal of Computer and Forensic Sciences*, vol. 5, no. 1, pp. 19–30, 2026, <https://doi.org/10.5937/jcfs5-66497>. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).



artificial intelligence was recognized as a catalyst for the acceleration in the evolution of high-tech crime, the application of which facilitated, accelerated, and enabled the creation of various fraudulent multimedia content, the automation of cyber attacks, personalized and targeted fraud, the quick and easy creation of fake news, etc. The report of the Special Prosecutor's Office for High-Tech Crime for the year 2025 shows that the Republic of Serbia is following the aforementioned global trends, which is expected given that it is a type of crime that is not bound by traditional borders and other geographical characteristics [3]. In addition to a constant increase in the number of reports in the last few years, the structure of the reported acts shows that various forms of fraud on the Internet account for almost a third of the total number of reports. Despite all these indicators, research conducted in the form of an online questionnaire among 829 adult citizens of the Republic of Serbia at the beginning of 2025 showed that social awareness of the possibilities of misuse of artificial intelligence is still not at an adequate level [4]. As many as 60% of respondents stated that they are not at all or only slightly familiar with examples of financial or romantic frauds that use AI, while only about 4% of citizens stated that they are very familiar with these phenomena. At the same time, the degree of self-confidence in recognizing AI fraud is low: almost a quarter of respondents (23.2%) believe that they would not be able to detect an attempt to abuse AI or that it is very unlikely, while the majority (44.4%) are not sure of their abilities.

Also, the widespread use of the Internet and social networks has enabled criminals to operate practically from anywhere in the world and a wide selection of potential victims, that is, the potential availability of any user of these technologies. The development of artificial intelligence and tools based on it have removed the last obstacles and indications that it is a fraud, which appeared in the case of a slightly more skilled average user, in terms of poor translation, grammatical errors, the choice of an inadequate font, etc. ChatGPT and similar tools today can very quickly prepare very sophisticated and completely grammatically, linguistically, and in every other sense accurate and precise fraudulent messages with an invitation to invest, advertisements, fake news, and other similar forms of phishing. Researchers state that natural language processing (NLP) can be used in AI-generated phishing attacks to produce highly believable and customized messages that increase the likelihood that recipients will be deceived, and also with the help of artificial intelligence, the process of sending messages can be automated and significantly accelerated [5]. The above is particularly dangerous from the aspect of inadequately trained and insufficiently educated users, i.e., older, less educated, and digitally inexperienced users, which is the profile of the most vulnerable category of the population and the category that is the least informed and able to recognize such frauds according to research. The scenario just described was confirmed by senior Europol officials in the podcast "How Crime Is Accelerated by AI" (posted June 24, 2025). "We are already seeing the use of generative AI in the creation of phishing campaigns and fake audio/video messages (deepfakes), especially in combination with vulnerable groups such as senior citizens", it was said on that occasion [6].

In addition to other mechanisms undertaken by competent authorities, especially from the aspect of preventive action, the question arises as to whether certain technical tools can help in recognizing these fraudulent patterns and indicate potential malicious activities. The European Union and the United Nations have launched initiatives for the development



of AI tools that will serve to quickly check the validity of information (fact-checking), and Google, Meta, and X are trying to implement the above on their social networks [7].

Such a tool would be particularly important for the previously described category of population, which does not have specific IT knowledge and skills, as an auxiliary tool that would indicate that a potential good opportunity is actually a fraud. All of the above points to the need for a quick reaction of the competent authorities and undertaking comprehensive measures to solve the observed trends and problems. In addition to strategic documents, legislative mechanisms, ethical guidelines, and education, it is necessary to apply technical tools and solutions based on artificial intelligence so that the fight against self-abuse is successful in this constant “game”. In this context, this paper aims to offer a concrete conceptual solution: the development of a semantic AI model for the detection of potentially fraudulent messages that uses advanced natural language processing (NLP) and transparent verification and storage of findings using blockchain technology.

2. METHODS

The proposed solution, aligned with the defined problem statements, consists of two main modules: an AI module for semantic NLP-based detection of fraudulent messages and a Blockchain module for recording detected fraud cases. By integrating the AI module with the Blockchain module, a comprehensive solution is achieved that enables early detection and secure records. It is important to emphasize that these two modules are conceptually independent: the AI detector can be used independently, while the Blockchain-based record can be integrated with other platforms. However, their combined use provides additional value and enhanced overall effectiveness. Parts of the conceptual model were originally developed within the author’s specialist thesis (2025) and have been significantly enhanced in this study through improvements in detection methodology, user interface design, feedback results, and blockchain integration.

2.1 AI module for semantic NLP-based detection of fraudulent messages

The first module is the AI module (NLP detector – Natural Language Processing), an integral component of artificial intelligence that enables machines to understand and interpret human language. This module is responsible for analyzing message content and assessing whether a message is fraudulent in nature. The use of semantic mapping of the message into the vector space of meaning (sentence embedding) enables the detection of similar phrases and avoids relying exclusively on keywords. The semantic approach to natural language processing implemented in this system ensures the identification of subtle and seemingly imperceptible differences between fraudulent and legitimate messages, allowing even linguistically similar expressions to be accurately recognized. NLP models trained on large sets of fraudulent messages can recognize linguistic features such as urgent tone, atypical communication patterns, prize phrases, and requests for confidential information, enabling early detection of fraudulent messages [8].

The detection procedure consists of:



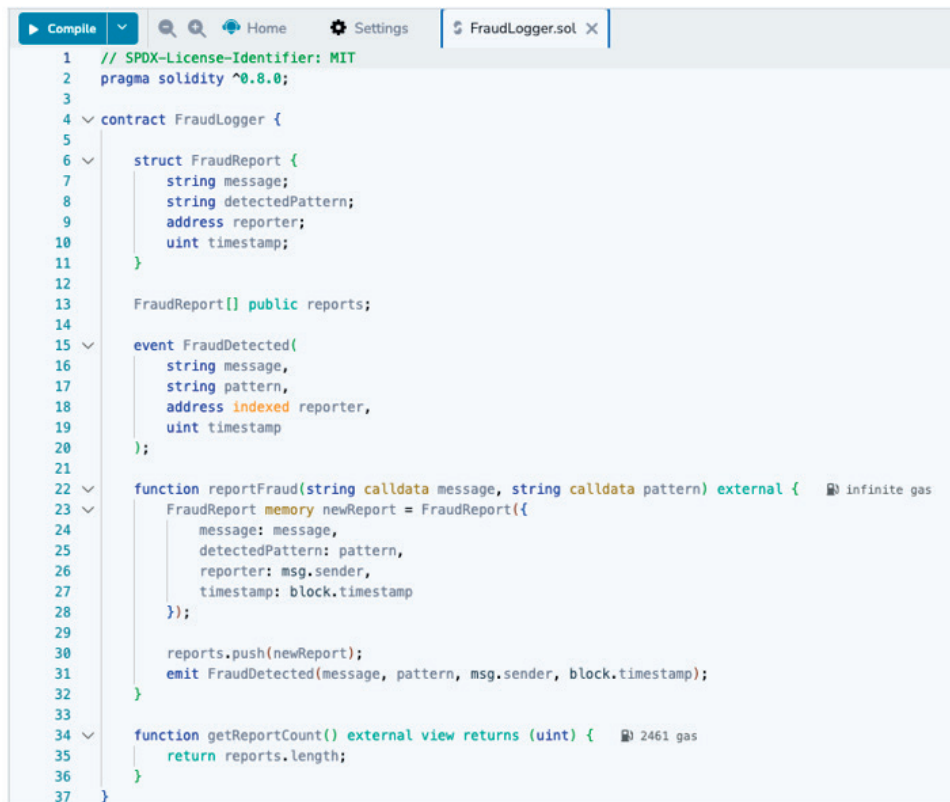
- 1) Entering a message: The user enters (copies) the text of the message he wants to check. It is assumed that the message is in an understandable language (Serbian language), although the model supports several different languages.
- 2) Using the selected model (paraphrase-MiniLM-L6-v2), the entire message is translated into a numeric vector. This vector represents the semantic meaning of the message. In this way, the model allows semantically similar texts to be placed close to each other in the vector space. For example, the sentences “You are guaranteed a quick profit without risk” and “You can easily profit, a 100% safe investment” have a high cosine degree of similarity, although they are not identical words, because they express a similar concept of a fraudulent offer.
- 3) Reference set of fraudulent phrases: The system defines a set of pre-known phrases and patterns that are typical for fraudulent messages. These phrases are collected based on the analysis of the most common cases of investment fraud and are used as an example. Our collection includes expressions in the Serbian language such as: “fast profit”, “invest now”, “easy earnings”, “click here”, “no risk”, “guaranteed profit”, “safe investment”, “overnight cashback”, and “100% safe investment”. These phrases represent the core of fraudulent content. Each of them is also encoded into vector form using the same sentence embedding model, and those vectors are precomputed.
- 4) Similarity calculation: When the system calculates a vector representation of a user’s message, it is compared to all reference vectors. The similarity metric we use is cosine similarity. Cosine similarity is a measure of similarity between two vectors in multidimensional space, defined as the cosine of the angle between them [9]. Cosine similarity gives a value of 1 for identical vector directions (maximum semantic similarity), 0 for orthogonal vectors (irrelevant contents), and -1 for completely opposite ones. The system finds the maximum similarity between the message and any reference fraudulent phrase.
- 5) Threshold decision: If the highest calculated cosine similarity is above a predefined threshold T , the message is classified as possible fraud. In the prototype, we empirically selected a threshold of 0.6 (on a scale of -1 to 1) after testing several known and fake messages. This value has been shown to be adequate to detect clear cases of semantic matching while avoiding oversensitivity. Of course, the threshold can be adjusted, lower values would increase the sensitivity with the risk of false alarms, while higher values narrow the detection to only those messages that are very similar to known scams.
- 6) Output of the AI module: The result of the analysis is a binary classification (fraudulent or legitimate message), with accompanying information whose phrase is “closest” to the reference message. For example, the system may return: “Possible fraudulent content (risk-free motif detected, similarity 0.63)”. This result is displayed to the user through the interface.

2.2 Blockchain module for recording detected fraud cases

The second important component of the system is integration with the blockchain. The idea is to create a record of detected fraudulent messages on the blockchain, thereby generating a permanent trace of the incident that is transparent and resistant to subsequent modification.



A smart contract has been implemented on the Ethereum blockchain network (Sepolia testnet), providing functionalities for storing and retrieving fraud reports. The contract is written in Solidity and includes a data structure, for example, Report, with fields such as the message hash, timestamp, and a description/category of the fraud. For the purposes of this concept, the contract is simplified and stores only minimal information: a list of textual reports (which may include the message content itself or summaries), along with the implicit timestamp of the block in which they are recorded. The available functions include ReportFraud(string report) for submitting a new report, getReportCount() for retrieving the number of records, and reports(uint index) for reading a specific report by index (Figure 1).



```

1 // SPDX-License-Identifier: MIT
2 pragma solidity ^0.8.0;
3
4 contract FraudLogger {
5
6     struct FraudReport {
7         string message;
8         string detectedPattern;
9         address reporter;
10        uint timestamp;
11    }
12
13    FraudReport[] public reports;
14
15    event FraudDetected(
16        string message,
17        string pattern,
18        address indexed reporter,
19        uint timestamp
20    );
21
22    function reportFraud(string calldata message, string calldata pattern) external {
23        FraudReport memory newReport = FraudReport({
24            message: message,
25            detectedPattern: pattern,
26            reporter: msg.sender,
27            timestamp: block.timestamp
28        });
29
30        reports.push(newReport);
31        emit FraudDetected(message, pattern, msg.sender, block.timestamp);
32    }
33
34    function getReportCount() external view returns (uint) {
35        return reports.length;
36    }
37 }

```

Figure 1. Smart contract Remix IDE.

Client-side interaction: The Gradio application interface is extended with HTML/JavaScript code integrating the Web3.js library. When the AI module detects fraud, the user is offered an option (e.g., a “Submit report to blockchain” button) to store that finding. By clicking this button, the smart contract function ReportFraud() is invoked via the Web3 interface, with the report passed as a parameter.

Authorization and transaction: When invoking the smart contract function, the MetaMask browser plugin prompts the user to confirm the transaction, i.e., to submit the report to the network. Each record receives a unique transaction identifier, which is visible on the Etherscan blockchain explorer.



Data structure and capacity: The smart contract is implemented as a simple logging mechanism. It does not contain any administrative logic and is publicly accessible for writing by anyone who calls it.

Reading and verification: All data written to the blockchain are public (since this is a public blockchain network). Any entity can call the `getReportCount` and `reports(i)` functions via a Web3 interface or view them directly on Etherscan (Figure 2).

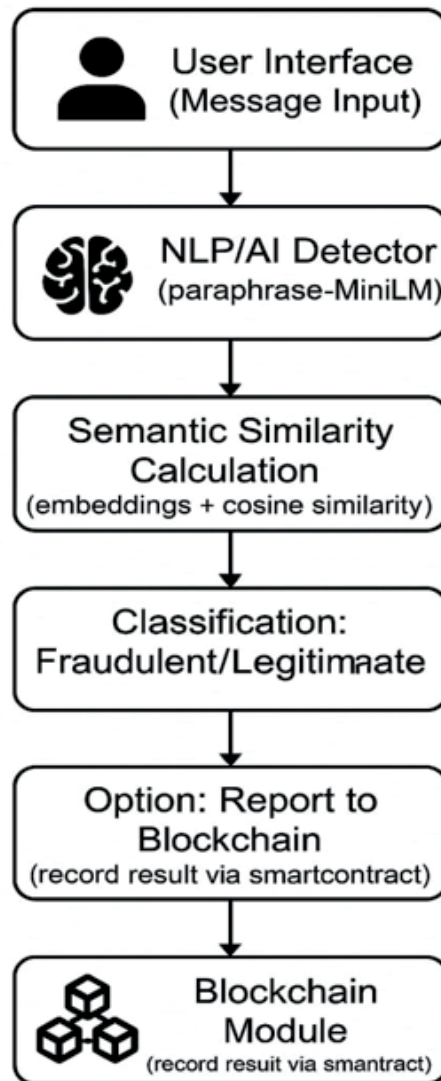


Figure 2. Workflow of the AI-based fraud detection system.



3. RESULTS AND DISCUSSION

As a conceptual model and an example of the practical application of an artificial intelligence-based system for detecting fraudulent messages, a software solution has been developed on the “Hugging Face” platform, available at the web address: <https://huggingface.co/spaces/NikolaCu/semanticki-ai-blockchain-KPU>.

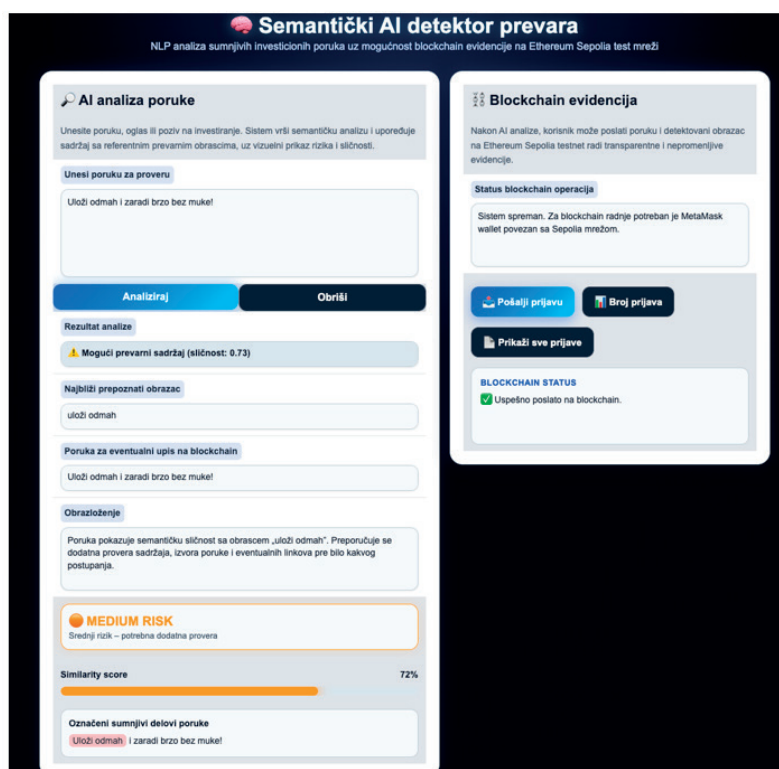


Figure 3. Platform interface with example.

By entering a phrase to be analyzed into the “Enter message” field and processing it, the system produces a result in the form of a comment indicating whether the phrase is of a fraudulent nature or not, along with a numerical value within a defined threshold. As additional information, the platform also provides a field labeled “Detected phrase,” which displays a phrase from a predefined reference set against which the new phrase has been semantically matched and identified as fraudulent content. Accordingly, for the input phrase “Earn effortlessly, quickly, and easily,” the result obtained is “Possible fraud – medium risk: additional check needed (similarity score: 72%),” while the system detected a correlation with the phrase “invest now” from the reference set (Figure 3).

One approach to addressing the problem of fraudulent messages, particularly in the context of investment scams, is the application of natural language processing (NLP) for the purpose of classifying such messages as potentially fraudulent or legitimate. The use of natural language understanding techniques, combined with machine learning models, enables the system to detect even subtle linguistic patterns characteristic of sophisticated fraud [10]. Given the large number of already recorded fraudulent messages, there exists a solid foundation for leveraging previously identified and reported messages, from which



key words and patterns can be extracted and analyzed. The challenge of this problem lies in the high similarity between malicious and legitimate messages. Attackers aim to mimic the style and structure of legitimate communications in their phishing messages; therefore, the model must identify intent and perform a semantic analysis of the entire content, rather than relying solely on predefined phrases. Research indicates that traditional approaches based solely on keyword detection are outdated and insufficient for identifying sophisticated attacks, whereas semantic analysis enables the detection of such messages as well [11]. In addition to classification, it is important to consider the measurement of prediction confidence. In practice, instead of a strict binary decision (0 or 1), the model's confidence, i.e., the probability that a message is fraudulent, is often calculated. Furthermore, this is a highly dynamic environment with frequent changes, requiring the model to continuously adapt and incorporate additional techniques in order to increase the number of correctly detected messages while reducing the number of misclassifications.

Another aspect of the problem is the need for a transparent mechanism to store information about the analysis and results of detected fraud attempts in a forensically sound, immutable, and non-repudiable manner for further processing, reporting, and investigation. In the literature, one of the key ethical requirements for the application of AI-based tools is the need to record decision-making processes through so-called audit trails, ensure resistance to data manipulation, and provide decentralized control over privacy management. One of the potential solutions in this regard is the application of blockchain technology [12], whose role has been considered in the "Materials and Methods" section within the conceptual solution.

3.1 Suggestions for further improvement of the model

The presented solution represents only a conceptual model of the potential use of artificial intelligence-based tools for detecting the misuse of this technology and its possible integration with blockchain. Due to the highly sophisticated nature of such fraud, there is significant room for improvement in practice both in terms of detection accuracy and the scope of information the system can process and verify in order to achieve greater reliability and provide additional sources of verification. Below are several concrete proposals for future work and model enhancement. The real challenge lies not only in detecting such messages but also in identifying them in a timely manner while minimizing false positives [13]. Model improvements may include:

1) Introduction of an NER (Named Entity Recognition) component: The current implementation focuses on recognizing semantic context but does not extract specific entities from the text. By integrating an NER model, the system could automatically identify the names of individuals, company names, locations, dates, phone numbers, email addresses, and other entities within a message.

2) Sentiment and tone analysis of messages: Fraudulent messages often attempt to provoke a specific emotion or reaction, typically a sense of urgency, fear of missing out (FOMO), or sympathy (as in scams involving fake humanitarian appeals). By incorporating sentiment analysis, the system could assess the emotional tone of a message (positive, negative, neutral) and detect the presence of such indicators.



3) Integration with AI-driven OSINT systems for metadata verification: Perhaps the most significant improvement would involve connecting the detector with modules for collecting and verifying information from open sources (OSINT). When the system detects a suspicious message, additional confirmation or insights could be obtained by automatically analyzing other data contained within the message:

- **URL verification:** If the message contains a link, the system could use OSINT APIs (e.g., Virus Total, Phish Tank) to check whether the domain has already been reported as phishing or malicious. “NLP systems can analyze sources such as forums and social media in real time, identifying information about planned social engineering attacks or new phishing campaigns [8].”

- **Domain WHOIS analysis:** WHOIS data may reveal whether a domain was recently registered (a common trait of phishing websites) or is attempting to impersonate a legitimate service (e.g., *paypalsecurity.com* mimicking *paypal.com*).

- **IP address verification and detection of related malicious domains:** By leveraging geolocation and OSINT databases (such as Abuse IPDB), the system could flag IP addresses originating from suspicious networks or known for spam/fraudulent activity.

- **Blacklisting and sender reputation analysis:** Particularly for email, the system could integrate checks for SPF/DKIM records and overall email reputation. An AI OSINT agent could autonomously gather data on whether the sender’s domain is valid and whether the email appears in databases of compromised accounts.

- **Additional OSINT analyses:** If a fraudulent message references a specific product, investment, or institution, the system could perform a quick web search combining those terms with keywords such as “scam” or “hoax” to determine whether there are existing reports from other users on forums, social media, or in news sources. Modern AI-powered search tools could summarize such findings and provide contextual insights, which users could then evaluate alongside the semantic analysis of the message itself.

1) Multimodal detection: Future development may include multimodal analysis capabilities, enabling the detection of fraudulent content across text, images, audio, and video, particularly in the context of AI-generated scams.

3.2 The use of AI detectors to combat high-tech crime

Development of information technologies and the Internet on a global scale has led to the emergence of increasingly sophisticated fraud schemes, which are becoming more advanced on a daily basis and pose a significant challenge for investigative authorities in cybercrime investigations. AI detectors can serve as an important tool and a preventive mechanism in combating high-tech crime, thereby contributing to a safer digital environment. Artificial intelligence, within the context of high-tech crime that includes fraudulent messages, phishing attacks, identity theft, and similar activities, has substantial practical and preventive value, as it enables the recognition of communication patterns of suspects and potential fraud.

The primary and most significant role of AI detectors in high-tech crime relates directly to the core subject of this research paper, namely, the detection of fraudulent messages.



Based on previously identified characteristic fraud patterns, it is possible to warn potential victims before financial or other types of damage occur, which provides considerable preventive value. Additionally, by analyzing message content, AI can identify suspicious phrasing, such as urgency in tone, for example, “act immediately”, as well as other manipulative elements typical of phishing attacks. However, AI detectors, in combination with unsupervised learning models, are also capable of detecting previously unknown forms of phishing attacks, the so-called “zero-day” attacks. By utilizing NLP models to understand the semantic context of messages, AI detectors can recognize seemingly well-crafted phishing messages that involve impersonation or social engineering techniques.

Given the fact that AI detectors analyze the content and intent of a message, they enable a deeper understanding of communication and, consequently, the recognition of social engineering. Social engineering represents one of the most significant applications of semantic AI detectors, as such attacks are based on human psychology rather than on the technical parameters of computer systems. What makes the detection of social engineering using traditional methods particularly challenging is that the messages received by victims are usually well-written and tailored to the target. A semantic AI detector, utilizing natural language processing (NLP), can identify manipulative intent and thereby detect elements such as urgency (e.g., “act quickly”), authority (e.g., “the service requires a password reset”), or emotional manipulation. In practice, such detectors can be integrated into email services and employee communication platforms in order to enable real-time response and prevent potential damage.

AI detectors can also play a significant role in the daily work of investigative authorities, given that state institutions are often faced with processing large volumes of data. By rapidly identifying suspicious communications or networks of organized cybercrime, AI enables faster data collection, detection, and processing, thereby significantly improving the efficiency of investigations in terms of time. AI systems can analyze vast amounts of data in a short period, unlike human analysts who require more time for such tasks. Furthermore, AI systems have high practical value in linking and analyzing digital evidence. Over the years, the patterns of criminal groups evolve, and to investigative authorities, their elements may appear unconnected. However, artificial intelligence can quickly detect behavioral patterns of criminal groups across different situations. With such capabilities, investigative authorities can act preventively and protect citizens from potential harm.

From a psychological perspective, AI detectors also contribute to increased public trust in state institutions. The implementation of systems that promptly recognize fraudulent messages and other forms of high-tech crime enhances the sense of security within the community when using electronic government services, online communication, internet payments, and e-banking.

Although there is no official global estimate of the share of cybercrime that utilizes artificial intelligence, research by the platform *Total Assure* [14], based on data from *IBM (International Business Machines Corporation)*, indicates that AI-assisted attacks have increased by 72% compared to 2024. Furthermore, according to *Tech Advisors*, AI-generated phishing attacks have surged by as much as 1,265%. It is also important to note that the same platform reports that, with the assistance of artificial intelligence, approximately 81% of passwords can be cracked within one month, which significantly complicates the fight against high-tech crime.



AI detectors could significantly influence the success rate of detecting high-tech criminal offenses and prosecuting their perpetrators. Their application within state institutions would enable more efficient direction of investigations, allowing authorities to respond in a timely manner. The primary goal of implementing AI detectors in everyday operations is the prevention of new damage and the reduction of already incurred harm. Despite all the advantages, it is important to remain aware that the sophistication of fraud is increasing daily, along with the emergence of new forms of high-tech crime. Therefore, faster responses by competent authorities to new forms of crime represent a key solution, which is why the active use of artificial intelligence constitutes a fundamental basis for successful investigations in the digital era.

4. CONCLUSION

From all of the above, it is clear that a comprehensive fight that includes legislative, ethical, educational, and technological measures is necessary for a successful fight against the abuse of artificial intelligence and high-tech crime supported by it. The development of various software solutions that combine the power provided by artificial intelligence, this time for the purposes of detecting and preventing such frauds in combination with blockchain, which enables transparency, immutability, and preservation of analysis results, represents a powerful weapon for further struggle. Of course, as technology is constantly advancing, and as fraud techniques change and adapt, it is necessary to continuously improve and adapt technological solutions intended for defense. Further improvements of the model through API integration with AI OSINT modules, integration of NER modules, or semantic analysis of attached messages are good directions on this path.

The aforementioned conceptual model, additionally improved, could be an integral part of an already available online platform for reporting incidents (e.g., National CERT), which would provide citizens with the opportunity to quickly and easily check certain invitations to invest and other messages they receive before taking concrete actions, all with the aim of identifying potential fraudulent activities.

FUNDING:

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

CONFLICTS OF INTEREST:

The authors declare no conflict of interest.



REFERENCES

- [1] Cognyte, 2025 Law Enforcement Outlook: Intelligence Challenges and Cyber Threat Landscape, Cognyte Research Division, Israel, 2025. [Online]. <https://www.cognyte.com/resources>
- [2] Europol, EU Serious and Organised Crime Threat Assessment (SOCTA 2025). Den Haag: Europol, 2025.
- [3] Draft Program for Combating High-Tech Crime – Report of the Special Department for Combating High-Tech Crime of the Higher Public Prosecutor's Office in the Republic of Serbia, Belgrade, 2025.
- [4] N. Čučilović, Research on User Perception of Artificial Intelligence Misuse, Specialist Thesis, Faculty of Organizational Sciences, Belgrade, 2025.
- [5] A. Arif, M. I. Khan, and A. R. A. Khan, "An overview of cyber threats generated by AI," *International Journal of Multidisciplinary Sciences and Arts*, vol. 3, no. 4, pp. 67–76, 2024.
- [6] Europol, "How Crime Is Accelerated by AI," Europol Podcast (Season 3, Episode 4), Andrei Linta, Edvardas Šileris, Tamara Šote, Marijn Schuurbiers, and Danny Van Althuis, 24 June 2025. [Online]. <https://www.europol.europa.eu/media-press/europol-podcast/episode-21-how-crime-is-accelerated-by-ai>
- [7] F. Pilati and T. Venturini, "The use of artificial intelligence in counter-disinformation: a world wide (web) mapping," *Frontiers in Political Science*, vol. 7, art. 1517726, 2025.
- [8] N. Mohamed, "Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms," *Knowl. Inf. Syst.*, vol. 66, pp. 101–158, 2024. doi: 10.1007/s10115-025-02429-y.
- [9] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008, p. 122.
- [10] X. Yang, C. Zhang, Y. Sun, K. Pang, L. Jing, S. Wa, et al., "FinChain-BERT: A High-Accuracy Automatic Fraud Detection Model Based on NLP Methods for Financial Scenarios," *Information*, vol. 14, no. 9, art. 499, 2023.
- [11] A. Mukherjee, N. Agarwal, and S. Gupta, "A survey on automatic phishing email detection using natural language processing techniques," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 11, pp. 1881–1885, 2019.
- [12] M. de la Roche and E. Voloder, *Report on Artificial Intelligence and Blockchain Convergences*. Brussels: INATBA, July 2024.
- [13] Z. Benetti, "Fraud Detection in Ethereum Using Web-scraping and Natural Language Processing Techniques," M.S. thesis, ETH Zürich – University of Zürich, 2021.
- [14] Total Assure, "AI cybersecurity statistics in 2025: Comprehensive data on threats, detection, and defense," Jan. 13, 2026. [Online]. Available: https://www.totalassure.com/blog/ai-cybersecurity-stats-2025?utm_source=chatgpt.com.

