# INTEGRATION OF AI ALGORITHMS IN THE VULNERABILITY TESTING PROCESS

**CZU: 004.056:[004.83+004.89]**
**DOI: https://doi.org/10.53486/csc2025.19**

**HÎNCU VERONICA**
Academy of Economic Studies of Moldova
soltanici.veronica@ase.md
**ORCID ID:** 0009-0002-9291-088X

**Abstract.** Artificial intelligence's rapid development is changing cybersecurity, particularly in the areas of vulnerability assessment and detection. Scalability and accuracy issues plague traditional vulnerability testing techniques, which is why AI-powered solutions are starting to look like a desirable substitute. This paper examines the integration of AI algorithms into vulnerability assessment, focusing how they can improve mitigation techniques, highlight risks, and improve threat detection. We investigate how AI-driven methods can improve system vulnerability detection, lower false positives, and expedite reaction times using a real-world case study. The findings demonstrate how AI could transform risk assessment and improve the intelligence, speed, and adaptability of security solutions.

**Keywords:** Artificial intelligence, Vulnerability scan, Risk analysis, Threat detection, Cybersecurity.

**JEL Classification: C63, D81, L86, O33.**

## INTRODUCTION

The accelerated evolution of artificial intelligence (AI) is transforming cybersecurity, especially in identifying and assessing vulnerabilities. With the surge in cyber threats, organizations demand more robust and scalable security measures. Conventional vulnerability scanning techniques struggle with scalability and precision, driving interest in AI-powered solutions. Machine learning (ML) and deep learning models can refine risk analysis, streamline remediation efforts, and enhance the precision of threat identification. This study investigates AI's role in vulnerability discovery through a real-world example, demonstrating its effectiveness in minimizing false alarms and speeding up incident resolution.

## MAIN RESULT

This research aims to explore the integration of Artificial Intelligence (AI) into vulnerability detection and assessment processes within cybersecurity. The rapid advancement of AI technologies has introduced new potential solutions to address long-standing challenges in the field of cybersecurity, such as scalability, accuracy, and efficiency of traditional vulnerability testing methods.

The primary objective is to analyze the most prevalent vulnerabilities, particularly those highlighted in the OWASP Top 10, to identify patterns that AI models can detect more accurately and efficiently than traditional methods. The research also seeks to evaluate the performance of various AI algorithms, specifically those based on Large Language Models (LLMs), such as GPT and Claude, in real-world environments. These models will be tested for their ability to identify and assess vulnerabilities in software systems, providing insights into their capacity to reduce false positives and expedite the vulnerability detection process.

Furthermore, the study aims to propose an optimized approach for implementing AI-driven models into cybersecurity workflows. This includes a comprehensive evaluation of the different AI models tested, considering factors like execution time, detection accuracy, and contextual analysis. By doing so, the research will demonstrate how AI can enhance the efficiency and reliability of vulnerability assessment processes, ultimately improving the overall cybersecurity resilience of organizations.

**1. Case Study: AI-Powered Vulnerability Assessment**

**1.1. Analysis of Cybersecurity Vulnerabilities.** Cybersecurity threats are continuously evolving, with attackers exploiting weaknesses in software systems. According to OWASP (Open Web Application Security Project), the most prevalent vulnerabilities include:

- **Injection Attacks (SQLi, Command Injection, XXE)**: Exploiting unvalidated input to manipulate application logic.
- **Cross-Site Scripting (XSS)**: Injecting malicious scripts into web applications.
- **Insecure Direct Object Reference (IDOR)**: Unauthorized access to protected resources.
- **Server-Side Request Forgery (SSRF)**: Exploiting server-side misconfigurations to access internal resources.
- **Broken Authentication & Security Misconfiguration**: Weak access controls leading to unauthorized access.

Understanding these vulnerabilities is essential for designing AI-driven security solutions.

**1.2. AI Model Implementation: Vulnhuntr** Vulnhuntr, an advanced static analysis tool, was utilized to test AI algorithms in vulnerability detection. This application integrates Large Language Models (LLMs) such as GPT, Claude, and Ollama to analyze source code, identify security weaknesses, and generate Proof-of-Concept (PoC) exploits.

The AI-driven testing process involved:

1. **Project Initialization:** Vulnhuntr analyzed repositories to identify critical files containing potential vulnerabilities.
2. **Primary Analysis:** AI models examined entry points for user input, detecting insecure patterns.
3. **Iterative Evaluation:** Context-aware processing refined initial findings, ensuring accurate identification of security threats.
4. **Report Generation:** A structured JSON report outlined detected vulnerabilities, PoCs, and recommended mitigations.

**2. Evaluating Performance in Real Conditions**

**2.1. AI Model Comparison.** To evaluate AI effectiveness, multiple models were tested under real-world conditions. The study examined their accuracy, false-positive rate, and execution speed.

**Table 1. Comparative Analysis of AI Models for Vulnerability Detection**

| Model | Execution Time | Confidence Score | Key Observations |
|---|---|---|---|
| **GPT-3.5** | **9-11s** | **8-9** | **Fast, but limited contextual analysis** |
| **GPT-4o** | **20-40s** | **9-10** | **High precision, detailed assessments** |
| **GPT-4o-mini** | **15-25s** | **8-9** | **Balanced speed and accuracy** |
| **Claude Sonnet** | **25-50s** | **9-10** | **Best in-depth analysis, slowest execution** |

**Source:** *Made by the author based on results after testing AI models*.

### 2.2. AI Performance in Vulnerability Detection

The comparative testing of AI models revealed nuanced differences in their ability to detect specific types of vulnerabilities. For example, in the case of Cross-Site Scripting (XSS), all models demonstrated strong detection capabilities. However, GPT-4o stood out by offering a more comprehensive assessment of the associated risks and mitigation measures. Its detailed contextual analysis enabled a better understanding of how such vulnerabilities could be exploited in different environments.

When examining SQL Injection vulnerabilities, again all models successfully identified the issue, but Claude Sonnet offered the most detailed and technically robust remediation strategies. It provided structured suggestions for sanitizing inputs, enhancing query logic, and introducing layered protection mechanisms—an indication of its deep contextual capabilities.

As the analysis extended to more complex vulnerabilities like Server-Side Request Forgery (SSRF) and Insecure Direct Object Reference (IDOR), the strengths of advanced models became even more apparent. While GPT-3.5 and GPT-4o-mini flagged potential issues, they lacked the depth of explanation required to fully understand the exploitation path. In contrast, GPT-4o and Claude Sonnet not only pinpointed the vulnerabilities but also contextualized their potential impact, highlighting relevant attack scenarios and proposing practical mitigation actions.

This multi-model testing approach underlined a key finding: while lighter models can effectively identify basic security flaws, high-stakes vulnerabilities benefit significantly from the analytical depth of more advanced AI. This insight supports a tiered deployment of AI models depending on the complexity and severity of the target vulnerabilities.

## 3. Recommendations

Based on the research and findings, integrating AI into the vulnerability detection process can significantly enhance the speed and accuracy of identifying security threats. However, to ensure the effectiveness of AI-driven solutions, certain strategic approaches should be adopted by organizations.

First, it is recommended to **adopt a tiered approach** to vulnerability detection. Early scans can be quickly performed using AI models like GPT-3.5. This model excels at providing rapid analysis, making it ideal for identifying high-priority issues early in the process. However, for more complex vulnerabilities or cases requiring deeper analysis, advanced models like GPT-4o or Claude Sonnet should be used. These models are better equipped to handle intricate attack vectors, providing more detailed insights and mitigation strategies.

Next, organizations should consider **parallel processing** to optimize the detection process. Running multiple AI models simultaneously, especially using faster models for preliminary scans alongside more advanced models for detailed analysis, can help reduce total processing time. This method allows security teams to identify vulnerabilities rapidly while still benefiting from in-depth analysis where necessary. This parallel approach balances both speed and precision, ensuring that critical vulnerabilities are addressed quickly while maintaining high-quality results.

Another important recommendation is to implement **conditional escalation** of AI models based on the complexity and severity of the vulnerability. For routine, straightforward threats like XSS or SQL injections, simpler models like GPT-3.5 are sufficient. However, for more sophisticated vulnerabilities, such as SSRF or IDOR, advanced models like GPT-4o or Claude Sonnet should be triggered. This approach allows organizations to economize resources while ensuring that higher-risk vulnerabilities receive the attention they deserve.

Furthermore, the implementation of AI in cybersecurity should always be **complemented by expert human oversight**. While AI can identify vulnerabilities and suggest potential solutions, the interpretation of these findings and the decision-making process must remain in the hands of experienced security professionals. AI-generated insights are powerful tools, but human expertise is essential in understanding the broader context and in applying the most effective remediation strategies.

Finally, to continually improve the effectiveness of AI in cybersecurity, it is crucial to **invest in model updates and data quality**. The AI models used in vulnerability detection must be trained on high-quality, up-to-date datasets to ensure that they remain effective against emerging threats. Regular updates, combined with the use of real-world test cases, will help ensure that AI systems continue to evolve and improve, staying ahead of evolving cyber threats.

By following these recommendations, organizations can maximize the potential of AI in vulnerability detection, improving their overall cybersecurity posture while ensuring that the technology is used effectively, responsibly, and efficiently.

## CONCLUSIONS

The integration of Artificial Intelligence (AI) into cybersecurity practices is revolutionizing the field of vulnerability detection and assessment. The AI algorithms tested in this study demonstrate significant improvements in detecting complex vulnerabilities faster and with greater accuracy compared to traditional methods. AI's ability to process large volumes of data, recognize complex patterns, and predict potential risks has proven crucial in enhancing the overall security of information systems.

The research confirms that AI-driven solutions, like the models tested in this study, are invaluable for detecting vulnerabilities such as Cross-Site Scripting (XSS), SQL Injection (SQLi), Server-Side Request Forgery (SSRF), and Insecure Direct Object Reference (IDOR). These models not only improve the accuracy of detection but also reduce the time required for threat identification, allowing organizations to respond more swiftly to potential cyberattacks.

In conclusion, AI's potential to transform cybersecurity is undeniable. However, its effective integration into existing systems requires continuous refinement of AI algorithms, the development of high-quality datasets for training, and regular validation by human experts. As AI evolves, its role in vulnerability detection will become even more essential in fortifying digital infrastructure against increasingly sophisticated cyber threats.

## REFERENCES

1. Almashaqbeh, G., & Mohapatra, P. (2022). Artificial Intelligence for Vulnerability Assessment in Software Systems. *ACM Computing Surveys*, 55(4).
2. OWASP Foundation. (2024). OWASP Top Ten Web Application Security Risks. https://owasp.org/www-project-top-ten/
3. Rajabi, E., & Hashemi, S. (2021). A Comparative Study of Machine Learning Techniques for Vulnerability Detection. *Journal of Cybersecurity and Privacy*, 1(3).
4. Hussain, F., & Watters, P. (2023). Automating Security Testing Using AI: Case Studies and Tools. *Cybersecurity Research Review*, 12(2).
5. OpenAI. (2024). GPT-4 Technical Report. https://openai.com/research/gpt-4
6. Anthropic. (2024). Claude Model Overview. https://www.anthropic.com/index/claude
7. ISO/IEC 27001. (2022). Information Security Management Systems. https://www.iso.org/isoiec-27001-information-security.html