УДК 004.8

А. В. Марущак
І. С. Зоря

ВПЛИВ AI-АВТОМАТИЗАЦІЇ НА АРХІТЕКТУРУ СИСТЕМ ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ ПРОГРАМНИХ РІШЕНЬ

Вінницький національний технічний університет

Анотація
У роботі досліджується вплив активної інтеграції систем штучного інтелекту на архітектуру забезпечення безпеки програмних рішень в умовах зменшення ролі людського фактору в IT-компаніях. Зростання використання AI-інструментів у процесах розробки, тестування та моніторингу безпеки трансформує традиційні підходи до побудови захищених систем. Особливу увагу приділено ризикам надмірної автоматизації, появі нових векторів атак (prompt injection, data poisoning), а також необхідності впровадження принципів Zero Trust щодо AI-агентів. У роботі обґрунтовано потребу перегляду архітектурних моделей безпеки з урахуванням автономності та масштабованості AI-компонентів.
Ключові слова: *AI-автоматизація; архітектура безпеки; DevSecOps; Zero Trust; програмні системи; кібербезпека.*

Abstract
The article examines the impact of the active integration of artificial intelligence systems on the architecture of security assurance for software solutions in the context of a reduced human role in IT companies. The increasing use of AI tools in development, testing, and security monitoring processes is transforming traditional approaches to building secure systems. Particular attention is paid to the risks of excessive automation, the emergence of new attack vectors (prompt injection, data poisoning), as well as the necessity of implementing Zero Trust principles for AI agents. The study substantiates the need to revise architectural security models taking into account the autonomy and scalability of AI components.
Keywords: *AI automation; security architecture; DevSecOps; Zero Trust; software systems; cybersecurity.*

Вступ

Сучасний етап розвитку IT-галузі характеризується активним впровадженням інструментів штучного інтелекту в процеси розробки, тестування та експлуатації програмних систем. AI-асистенти, автоматизовані системи аналізу коду, інтелектуальні SOC-платформи та автономні агенти дедалі частіше виконують функції, які раніше належали фахівцям з безпеки, DevOps-інженерам та аналітикам. Така трансформація спричиняє не лише зміни у кадровій структурі компаній, а й глибокий перегляд архітектури систем забезпечення безпеки програмних рішень [1].

Концепція DevSecOps передбачає інтеграцію безпеки на всіх етапах життєвого циклу ПЗ, однак із появою AI-інструментів відбувається перехід до моделі “Security as Code”, де значна частина рішень приймається автоматизовано [2]. Водночас зростає кількість специфічних загроз, пов’язаних із використанням моделей машинного навчання, зокрема атак типу prompt injection, витоку навчальних даних та маніпуляцій результатами роботи моделей [3].

Актуальність дослідження зумовлена необхідністю формування нових архітектурних підходів до побудови систем безпеки, здатних враховувати автономність AI-компонентів та мінімізувати ризики, пов’язані зі зменшенням людського контролю.

Результати дослідження

Традиційна модель забезпечення безпеки передбачала активну участь людини на етапах аналізу коду, реагування на інциденти та прийняття рішень щодо доступу до ресурсів. Проте сучасні AI-системи здатні автоматично виконувати статичний та динамічний аналіз коду, здійснювати моніторинг мережевого трафіку та виявляти аномалії в режимі реального часу [2].

З одного боку, це підвищує швидкість реагування та масштабованість системи. З іншого – створює ризик надмірної довіри до автоматизованих рішень. Помилка, допущена AI-моделлю, може бути масштабована на тисячі середовищ одночасно, що суттєво збільшує потенційні збитки.

Таким чином, архітектура безпеки повинна передбачати механізми верифікації та аудитування рішень, прийнятих AI-агентами, з використанням принципу «human-in-the-loop».

Інтеграція великих мовних моделей у корпоративні процеси породила новий клас атак. Зокрема, атаки типу prompt injection дозволяють зловмисникам маніпулювати поведінкою AI-системи через спеціально сформовані запити [3]. Крім того, актуальними залишаються загрози data poisoning – внесення шкідливих даних у навчальні набори, що призводить до некоректної роботи моделей [1].

Концепція Zero Trust Architecture (ZTA) базується на принципі «ніколи не довіряй, завжди перевіряй». Спочатку вона застосовувалася до користувачів і пристроїв, проте в умовах AI-автоматизації виникає необхідність поширення цього принципу і на автономні програмні агенти [4].

AI-агент повинен розглядатися як окремий суб’єкт доступу з мінімальним набором привілеїв. У межах архітектури Zero Trust необхідно впроваджувати: сегментацію AI-сервісів, постійну автентифікацію та авторизацію запитів, моніторинг поведінки AI-модулів, аудит журналів їхніх дій.

Застосування ZTA до AI-агентів дозволяє зменшити ризик ескалації привілеїв та неконтрольованого доступу до критичних ресурсів.

AI-інструменти дедалі частіше інтегруються в CI/CD-конвеєри, забезпечуючи автоматичне сканування коду, перевірку залежностей та виявлення конфігураційних помилок [2]. Така модель відповідає концепції «Shift Left Security», де безпека перевіряється на ранніх етапах розробки.

Водночас, як показують дослідження AI-driven DevSecOps, автоматизація не усуває повністю ризиків, а змінює їхню природу. Основною проблемою стає довіра до алгоритмів прийняття рішень. Тому архітектура повинна передбачати багаторівневу перевірку результатів AI-аналізу та можливість швидкого відкату змін у разі виявлення помилок.

На основі проведеного аналізу можна сформулювати такі архітектурні рекомендації:

1. Впровадження окремого рівня контролю для AI-агентів у загальній моделі доступу.
2. Використання принципу мінімальних привілеїв для автоматизованих сервісів.
3. Реалізація постійного моніторингу поведінки AI-модулів.
4. Забезпечення аудиту та журналювання рішень, прийнятих AI.
5. Комбінування автоматизованого аналізу з експертною перевіркою критичних операцій.

Такі підходи дозволяють мінімізувати ризики, пов’язані зі зменшенням ролі людського фактору, та забезпечити стійкість архітектури програмних систем до нових типів загроз.

Висновки

Отже, впровадження AI-автоматизації у процеси розробки та експлуатації програмних систем суттєво трансформує архітектуру забезпечення безпеки. Зменшення ролі людського фактору підвищує швидкість і масштабованість захисних механізмів, проте одночасно створює нові ризики, пов’язані з автономністю AI-агентів, специфічними векторами атак та потенційною масштабованістю помилок. У цих умовах ключового значення набуває адаптація принципів Zero Trust до AI-компонентів, впровадження багаторівневого контролю та аудитування автоматизованих рішень. Таким чином, ефективна архітектура безпеки майбутніх програмних систем повинна поєднувати інтелектуальну автоматизацію з механізмами контролю та верифікації, забезпечуючи баланс між автономністю та керованістю.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. NIST Technical Series Publications. [Електронний ресурс]. Режим доступу: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf (дата звернення: 25.02.2026).

2. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. IEEE Xplore. [Електронний ресурс]. Режим доступу: https://ieeexplore.ieee.org/document/9153877 (дата звернення: 26.02.2026).

3. OWASP Top 10 for Large Language Model Applications | OWASP Foundation. OWASP Foundation, the Open Source Foundation for Application Security | OWASP Foundation. [Електронний ресурс]. Режим доступу: https://owasp.org/www-project-top-10-for-large-language-model-applications/ (дата звернення: 26.02.2026).

4. SP 800-207, Zero Trust Architecture | CSRC. NIST Computer Security Resource Center | CSRC. [Електронний ресурс]. Режим доступу: https://csrc.nist.gov/pubs/sp/800/207/final (дата звернення: 01.03.2026).

Марущак Анастасія Віталіївна **–** студентка групи КІТС-25м, Факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: anamar349@gmail.com

Науковий керівник: ***Зоря Ірина Сергіївна*** − асистент кафедри Менеджменту та безпеки інформаційних систем, e-mail: ira.zoria@vntu.edu.ua

Marushchak Anastasiia V. – student of group CITS-25m, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: anamar349@gmail.com

Supervisor: ***Zoria Iryna S.*** − assistant of the Department of Management and Security of Information Systems, e-mail: ira.zoria@vntu.edu.ua