Кавецький М. С., Руженцев В. І.

ЕВОЛЮЦІЯ ВЕБ-АТАК В ЕПОХУ ШТУЧНОГО ІНТЕЛЕКТУ: ВІД ТРАДИЦІЙНИХ ЗАГРОЗ ДО АВТОНОМНИХ LLM-АГЕНТІВ

Актуальність теми зумовлена безпрецедентним зростанням кількості, швидкості та складності кібератак на веб-інфраструктуру, що зазнала радикальної трансформації під впливом розвитку штучного інтелекту (ШІ) у 2022–2024 роках. Інтеграція великих мовних моделей (LLM) у хакерський інструментарій призвела до феномену «інфляції кіберзагроз» – ситуації, коли вартість та технічний бар'єр для проведення складних атак радикально знижуються, а їхні масштаби експоненційно зростають. Традиційні статичні засоби захисту (наприклад, сигнатурні WAF) стають неефективними проти автономних ШІ-агентів, здатних генерувати поліморфні експлойти та адаптуватися до захисних бар'єрів у реальному часі [1]. Це вимагає негайного перегляду парадигми кібербезпеки та переходу до проактивних методів оборони.

Об’єктом дослідження є сучасна екосистема кібербезпеки веб-додатків та мереж у контексті безперервної еволюції ландшафту кіберзагроз. Предметом дослідження виступають методи, алгоритми та інструментарій використання систем штучного інтелекту (зокрема, LLM та агентних архітектур) для автоматичної генерації, масштабування та маскування веб-атак, а також новітні підходи до протидії цим загрозам.

Попри вдосконалення процесів розробки, найнебезпечнішими загрозами залишаються класичні вразливості, такі як міжсайтовий скриптинг (XSS), SQL-ін'єкції (SQLi) та обхід каталогів (Path Traversal). Водночас спостерігається стрімка зміна головних векторів: API-інтерфейси стали первинними цілями хакерів (зафіксовано 150 мільярдів API-атак з початку 2023 року) через масове впровадження мікросервісів та підключення ШІ-додатків до зовнішніх ресурсів. Також змінилася природа DDoS-атак: сучасні ботнети (наприклад, Aisuru) здатні генерувати гіпер-об'ємні атаки прикладного рівня (L7) потужністю понад 31 Тбіт/с, використовуючи ШІ для імітації поведінки реальних користувачів та обходу CAPTCHA [2].

Кіберзлочинці перейшли від експериментів з ШІ до його повної операціоналізації. Великі мовні моделі (LLM) активно використовуються для написання скриптів експлуатації (PoC/PoV), пошуку вразливостей у коді (fuzzing) та «just-in-time» генерації шкідливого програмного забезпечення. ШІ стирає межі між стадіями атаки: у 2025 році зафіксовано першу масштабну кібершпигунську кампанію, яку на 80-90% було реалізовано повністю автономними ШІ-агентами (на базі модифікованих інструментів типу Claude Code) з мінімальним втручанням людини [3]. Крім того, використання генеративного ШІ спричинило сплеск гіперперсоналізованого фішингу та соціальної інженерії на 1265% [4].

Зі впровадженням інтелектуальних Web Application Firewalls (WAF), зловмисники почали використовувати методи AML (Adversarial Machine Learning) для "обману" алгоритмів захисту. Атаки ухилення (evasion attacks) дозволяють змінювати синтаксис шкідливого payload-запиту (за рахунок кодування, варіацій пробілів чи коментарів) так, щоб він зберігав свою руйнівну семантику (наприклад, для виконання SQLi), але класифікувався нейромережею WAF як легітимний трафік.

Оскільки швидкість кібератак (час від проникнення до розгортання в мережі) скоротилася з тижнів до хвилин, традиційне ручне реагування втратило свою дієвість.

Сучасні стратегії захисту (2026–2030) базуватимуться на концепції «агентних центрів безпеки» (Agentic SOC), де автономні ШІ-системи захисників здійснюють розвідку, виявляють аномалії та ізолюють загрози у реальному часі без участі людини [5]. Перемога в сучасному кіберпросторі залежатиме від того, чий штучний інтелект швидше адаптується до поведінки супротивника.

В якості висновку можна сказати, що у роботі розглянуто еволюцію веб-атак у контексті стрімкого розвитку технологій штучного інтелекту. Показано, що сучасний ландшафт кіберзагроз характеризується зростанням автоматизації атак, використанням великих мовних моделей та появою автономних агентних систем, здатних адаптуватися до механізмів захисту. Встановлено, що традиційні статичні підходи до кіберзахисту поступово втрачають ефективність у протидії новим типам загроз. У зв'язку з цим перспективними напрямами розвитку систем кібербезпеки є впровадження інтелектуальних механізмів виявлення аномалій, автоматизація процесів реагування на інциденти та використання методів машинного навчання для аналізу поведінки мережевого трафіку. Подальші дослідження доцільно спрямувати на вдосконалення адаптивних підходів до виявлення та нейтралізації кіберзагроз у середовищах сучасних веб-систем.

Список використаних джерел

1. Sultana, Sharmin & Rahman, Akib. (2025). A Multi-Layered Defense Framework Against Adversarial Attacks on ML-based Web Application Firewalls. Open Access Research Journal of Science and Technology. 15, P. 2. 10.53022/oarjst.2025.15.2.0137

2. Yoachimik O., Pacheco J. *Cloudflare's 2025 Q3 DDoS threat report – including Aisuru, the apex of botnets* [Електронний ресурс]. – Cloudflare Blog, 03 Dec 2025. – Режим доступу: https://blog.cloudflare.com/ddos-threat-report-2025-q3/ (дата звернення: 08.03.2026).

3. Disrupting the first reported AI-orchestrated cyber espionage campaign [Електронний ресурс]. – Anthropic. – Режим доступу: https://www.anthropic.com/news/disrupting-AI-espionage (дата звернення: 08.03.2026).

4. Khalil M. AI Cyber Attack Statistics 2025, Trends, Costs, Defense [Електронний ресурс]. – DeepStrike Blog, 10 Oct 2025. – Режим доступу: https://deepstrike.io/blog/ai-cyber-attack-statistics-2025 (дата звернення: 08.03.2026).

5. Israel A. *2026 Cybersecurity Predictions: The Cyber Horizon… From Cloud to AI* [Електронний ресурс]. – CyberProof Blog, 28 Jan 2026. – Режим доступу: https://www.cyberproof.com/blog/2026-cybersecurity-predictions-cyber-horizon-from-cloud-to-ai/ (дата звернення: 08.03.2026).