УДК 004.056:004.8

В. О. Яцюк

Т. Г. Кирилащук

АНАЛІЗ ВПЛИВУ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ (LLM) НА ЛАНДШАФТ КІБЕРБЕЗПЕКИ: ЗАГРОЗИ ТА МОЖЛИВОСТІ

Вінницький національний технічний університет

Анотація. *У роботі проаналізовано роль великих мовних моделей (LLM) у сучасній кібербезпеці. Метою дослідження є виявлення подвійної природи ШІ як інструменту для автоматизації захисту та засобу масштабування кібератак. Розглянуто вектори загроз (фішинг, генерація шкідливого коду) та можливості захисних систем на базі LLM. Визначено критичні вразливості самих моделей до маніпуляцій та атак на дані.*

Ключові слова: кібербезпека, великі мовні моделі, штучний інтелект, інформаційна безпека, SecureBERT, генеративний ШІ.

Abstract. *The paper analyzes the role of Large Language Models (LLMs) in modern cybersecurity. The study aims to identify the dual nature of AI as a tool for security automation and a means of scaling cyberattacks. Threat vectors (phishing, malware generation) and the capabilities of LLM-based defense systems are examined. Critical vulnerabilities of the models themselves to manipulation and data poisoning are identified.*

Keywords: cybersecurity, Large Language Models, artificial intelligence, information security, SecureBERT, generative AI.

Вступ

Стрімкий розвиток генеративного штучного інтелекту, зокрема великих мовних моделей (LLM), таких як GPT-4, LLaMa та Falcon, повністю змінили підхід до захисту даних у сфері інформаційної безпеки. За результатами систематичного аналізу літератури, проведеного у 2024 році, спостерігається вибухове зростання досліджень у цій галузі: 68% усіх релевантних наукових праць були опубліковані протягом 2023 року [1]. Це свідчить про те, що методи захисту та нападу трансформуються в реальному часі, вимагаючи від фахівців нових підходів до виявлення загроз та реагування на інциденти.

Метою даної роботи є дослідити як позитивний, так і негативний вплив використання LLM у кіберпросторі, проаналізувати ефективність використання ШІ як інструменту для кібернападів (offensive security) та як засобу захисту (defensive security), а також визначити ключові вразливості самих моделей.

Результати дослідження

Аналіз 177 наукових джерел за період 2018-2024 рр. дозволив виділити основні вектори впливу LLM на кібербезпеку.

Доведено, що LLM суттєво знижують "поріг входу" для кіберзлочинців. Основним вектором атак залишається фішинг та соціальна інженерія. Моделі здатні генерувати персоналізовані, граматично коректні повідомлення, що значно підвищує успішність атак типу Spear Phishing. Окрім цього, дослідження Gupta et al. демонструють здатність LLM генерувати функціональні фрагменти коду для вірусів-вимагачів (Ransomware), імітуючи поведінку відомих шкідливих програм, таких як WannaCry, та створювати поліморфний шкідливий код, що ускладнює його детектування традиційними сигнатурними методами [2]. Також актуальними залишаються атаки типу "Jailbreaking", які дозволяють обходити етичні обмеження моделей для отримання шкідливого контенту.

У захисному сегменті LLM демонструють високу ефективність у задачах аналізу логів, виявлення вразливостей програмного забезпечення та автоматизації процесів SOC (Security Operations Center). Спеціалізовані моделі, такі як SecureBERT та FalconLLM, показують високу точність у виявленні вразливостей коду та класифікації атак, перевершуючи традиційні методи машинного навчання [3]. Використання LLM для аналізу журналів подій дозволяє скоротити час реакції на інциденти, надаючи пояснення (Explainable AI) щодо природи аномалій, що є критично важливим для навчання нових аналітиків.

Попри ефективність, інтеграція LLM у системи безпеки несе ризики. Головною проблемою залишаються "галюцинації" (генерація неправдивої інформації), що робить повну автоматизацію небезпечною. Також виявлено вразливості моделей до атак на етапі навчання (data poisoning) та маніпуляцій через змагальні приклади (adversarial attacks), що можуть призвести до витоку конфіденційних даних або некоректної роботи систем захисту.

Висновки

Великі мовні моделі створюють як нові можливості для захисту, так і серйозні ризики. З одного боку, вони автоматизують рутинні задачі захисту та підвищують ефективність виявлення загроз. З іншого - надають зловмисникам інструменти для масштабування атак. Ключовим завданням на найближчі роки є розробка надійних методів верифікації контенту, згенерованого ШІ, та створення спеціалізованих, захищених LLM для потреб кібербезпеки, де людина залишається обов'язковою ланкою контролю (human-in-the-loop).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Hasanov I., Virtanen S., Hakkala A., Isoaho J. Application of Large Language Models in Cybersecurity: A Systematic Literature Review. *IEEE Access*. 2024. Vol. 12. P. 176751—176778. URL: https://doi.org/10.1109/ACCESS.2024.3505983 (date of access: 09.03.2026).
2. Gupta M., Akiri C., Aryal K. et al. From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE Access*. 2023. Vol. 11. P. 80218—80245. URL: https://ieeexplore.ieee.org/document/10132174 (date of access: 09.03.2026).
3. Ferrag M. A., Battah A., Tihanyi N. et al. SecureFalcon: Are we there yet in automated software vulnerability detection with LLMs? *arXiv preprint*. 2023. URL: https://arxiv.org/abs/2307.06616 (date of access: 09.03.2026).

Яцюк Владислав Олександрович − студент групи 2БС-24б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, email: vladosnova2007@gmail.com

Кирилащук Тетяна Геннадіївна − асистент кафедри захисту інформації, Вінницький національний технічний університет, Вінниця, email: tan099838@vntu.edu.ua

Yatsiuk Vladyslav O. − student of 2BS-24b group, Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: vladosnova2007@gmail.com

Kyrylaschuk Tetiana G. − Associate Professor, Department of Information Security, Vinnytsia National Technical University, Vinnytsia, email: tan099838@vntu.edu.ua