УДК 004.8+004.056

Л. М. Куперштейн
О. В. Залепа

ВЕЛИКІ МОВНІ МОДЕЛІ ДЛЯ СТВОРЕННЯ ВІРТУАЛЬНОГО АСИСТЕНТА З КІБЕРБЕЗПЕКИ

Вінницький національний технічний університет

Анотація

У роботі розглянуто використання великих мовних моделей (LLM) для створення віртуального асистента з кібербезпеки. Використання LLM дозволяє аналізувати великі обсяги даних, виявляти загрози та надавати рекомендації щодо захисту інформаційних систем, що значно підвищує ефективність кіберзахисту.

Ключові слова: великі мовні моделі, LLM, кібербезпека, виявлення загроз.

Abstract

The paper considers the use of large language models (LLM) to create a virtual cybersecurity assistant. The use of LLMs allows analysing large amounts of data, identifying threats and providing recommendations for the protection of information systems, which significantly increases the effectiveness of cyber defence.

Keywords: big language models, LLM, cybersecurity, threat detection.

Вступ

З розвитком цифрових технологій та збільшенням кількості кібератак попит на висококваліфікованих спеціалістів у галузі кібербезпеки стрімко зростає. Водночас штучний інтелект, зокрема великі мовні моделі (LLM), відкривають нові можливості для зміцнення кіберзахисту. Ці моделі можуть аналізувати великі обсяги даних, виявляти потенційні загрози та навіть надавати рекомендації щодо захисту інформаційних систем. Використання LLM у якості програмних консультантів з кібербезпеки може значно підсилити здатність організацій протистояти сучасним кібервикликам, забезпечуючи більшу швидкість, точність та масштабованість в обробці кібербезпекових інцидентів [1]. Це відкриває можливість не лише ефективно реагувати на вже відомі загрози, але й передбачати та запобігати майбутнім атакам, що стає критичним у контексті постійно зростаючої складності та різноманітності кіберзагроз.

Результати дослідження

Великі мовні моделі – це вид штучного інтелекту, що базується на архітектурі трансформерів, здатних генерувати та розуміти людську мову на дивовижно високому рівні. Технічна основа LLM полягає у використанні нейронних мереж глибокого навчання, які тренуються на величезних обсягах текстових даних. Це навчання дозволяє моделям виявляти закономірності, розуміти контекст і генерувати відповіді, що відповідають людським очікуванням [2].

Однією з ключових особливостей LLM є їх здатність до трансферного навчання, що дозволяє адаптувати модель до специфічних завдань без необхідності тренування з нуля. Це особливо корисно у кібербезпеці, де моделі можуть бути адаптовані для виявлення специфічних видів кіберзагроз, таких як фішингові атаки або зловмисне програмне забезпечення [3].

Також LLM можуть виконувати семантичний аналіз тексту, що дозволяє ідентифікувати потенційні зловмисні команди або небажані маніпуляції в коді, що робить їх незамінними помічниками у забезпеченні кібербезпеки. Існує багато сценаріїв використання LLM у кібербезпеці, зокрема LLM можуть аналізувати інтернет-трафік і логи системи на наявність ознак кіберзагроз, таких як віруси, трояни або спроби вторгнення. Ці моделі можуть виявляти аномалії у поведінці користувачів або системи, що можуть вказувати на зловмисні дії. Вони також можуть використовуватися для аналізу електронних листів на предмет виявлення фішингових атак, так як здатні розпізнавати характерні мовні патерни та інші ознаки фішингу, забезпечуючи додатковий рівень захисту для користувачів. Окрім того, LLM можуть використовуватись для підтримки аналітиків кібербезпеки, надаючи рекомендації щодо реагування на інциденти кібербезпеки на основі аналізу доступних даних або ж

аналізувати великі обсяги політик безпеки та рекомендувати вдосконалення на основі найновіших тенденцій у кібербезпеці та виявлених вразливостей [4].

Серед топових LLM від різних виробників, які активно використовуються в різних галузях, можна виділити наступні моделі [5]:

- GPT-4 (OpenAI): Нова версія моделі, яка продовжує бути однією з найпотужніших і найпоширеніших, використовується в численних застосунках від чат-ботів до генерації тексту та аналізу даних.
- Gemini 1.5 (Google DeepMind): Оновлена версія Gemini від Google DeepMind, яка відзначається своєю здатністю до обробки природної мови і генерації тексту.
- Claude 2 (Anthropic): Нова версія моделі Claude, розроблена компанією Anthropic, яка використовується для генерації тексту та інших завдань обробки природної мови.
- Grok (xAI): Модель Grok від xAI, компанії, заснованої Ілоном Маском, інтегрована в екосистему X (колишній Twitter) і використовується для різноманітних завдань обробки природної мови.
- LLaMA 3 (Meta): Остання версія LLaMA (Large Language Model Meta AI) від Meta, яка забезпечує покращені можливості генерації тексту та обробки природної мови.
- Mistral 7B (Mistral AI): Модель від Mistral AI з 7 мільярдами параметрів, відзначається своєю ефективністю і потужністю у вирішенні завдань обробки природної мови.
- Falcon (TII - Technology Innovation Institute): Модель Falcon від Technology Innovation Institute, яка демонструє високу продуктивність у завданнях обробки природної мови і генерації тексту.

На основі цих та інших LLM у сфері кібербезпеки було реалізовано багато ефективних рішень, серед яких [6]:

- Darktrace: Darktrace використовує генеративний AI та LLM для аналізу інцидентів у реальному часі, захисту даних і управління ризиками. Вони впровадили нові моделі ризику та відповідності, щоб допомогти своїм клієнтам залишатися захищеними під час використання генеративного AI та LLM інструментів.
- Cylance (BlackBerry): Cylance використовує штучний інтелект, включаючи великі мовні моделі, для запобігання кібератакам на основі поведінкового аналізу. Їхні рішення можуть передбачати та запобігати загрозам у реальному часі.
- Vectra AI: Vectra AI використовує машинне навчання та LLM для виявлення та реагування на загрози в мережевому трафіку, забезпечуючи захист у режимі реального часу та покращений аналіз інцидентів.
- SentinelOne: SentinelOne інтегрує AI та LLM для автономного виявлення та реагування на загрози. Їхні рішення здатні автоматично аналізувати дані та реагувати на загрози без втручання людини.
- Cybereason: Cybereason застосовує AI та LLM для виявлення складних загроз та управління інцидентами. Їхні рішення забезпечують глибокий аналіз загроз і автоматичне реагування на них.
- McAfee MVISION: McAfee MVISION використовує AI та LLM для розширення можливостей захисту та управління кібербезпекою. Вони застосовують ці технології для прогнозування загроз і вдосконалення заходів захисту.
- FortiAI (Fortinet): FortiAI використовує AI та LLM для автоматичного виявлення загроз і реагування на них. Їхні рішення забезпечують миттєву ідентифікацію та усунення загроз у мережах.

Ці розробки демонструють різноманітні підходи до використання великих мовних моделей у сфері кібербезпеки, надаючи організаціям ефективні інструменти для виявлення, аналізу та реагування на кібератаки.

Водночас існують певні виклики та обмеження використання LLM у кібербезпеці, адже такі мовні моделі вимагають великої кількості даних для ефективного навчання. Якість та релевантність цих даних безпосередньо впливають на їх точність. Збір та обробка відповідних даних може бути складним та витратним завданням. Окрім того, як і всі машинні алгоритми, LLM можуть містити вбудовані упередження, які впливають на їх рішення та висновки [7]. Невірно навчена модель може ігнорувати або неправильно ідентифікувати реальні загрози. Ще одним викликом є підтримка і оновлення таких моделей, так як вони вимагають значних ресурсів, включаючи обладнання для обробки великих обсягів даних та кваліфікований персонал для моніторингу та налаштування систем.

Висновки

Великі мовні моделі відкривають значні можливості для зміцнення кібербезпеки через їх здатність аналізувати великі обсяги даних, розпізнавати складні мовні патерни та надавати швидкі реакції на потенційні загрози. Застосування LLM як віртуальних асистентів з кібербезпеки може радикально змінити підходи до захисту цифрового простору, підвищуючи ефективність та швидкість реагування на кібератаки. Проте, разом з можливостями, LLM вносять і виклики, включаючи потребу в великих обсягах якісних даних, ризик упередженості, високі витрати на підтримку, проблеми з конфіденційністю, обмежену контекстуальну сприйнятливість та питання відповідальності. Розгортання LLM вимагає балансу між інноваційними можливостями та обмеженнями, а також ретельного регулювання і стратегічного підходу до управління ризиками.

Враховуючи ці аспекти, можна стверджувати, що великі мовні моделі є перспективним інструментом у галузі кібербезпеки, який може радикально змінити підходи до захисту цифрового простору, але водночас вимагає обачливого та виваженого застосування.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Куперштейн Л.М., Примаков Б.С. Про використання ChatGPT в кібербезпеці // Матеріали LII Науково-технічної конференції підрозділів ВНТУ, Вінниця, 21-23 червня 2023 р. – Електрон. текст. дані. – 2023. URL: https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2023/paper/view/18653 (дата звернення 09.05.24).
2. Large Language Model for Chatbot / Prof. Trupti Farande et al. International Journal of Advanced Research in Science, Communication and Technology. 2024. P. 291–293.
3. Large Language Models (LLMs) with Google AI. Google Cloud. URL: https://cloud.google.com/ai/llms (дата звернення: 10.05.2024).
4. What are Large Language Models? | NVIDIA Glossary. NVIDIA. URL: https://nvidia.com/en-us/glossary/large-language-models/ (дата звернення: 14.05.2024).
5. Guinness H. The best large language models (LLMs) in 2024. Automate your work today | Zapier. URL: https://zapier.com/blog/best-llm/ (дата звернення: 17.05.2024).
6. Top 7 AI tools for cybersecurity in 2024. GeeksforGeeks. URL: https://www.geeksforgeeks.org/ai-tools-for-cybersecurity/ (дата звернення: 17.05.2024).
7. Large Language Models for Cybersecurity | AI Research Journal. 2024. URL: https://airesearchjournal.com/llm-cybersecurity (дата звернення: 18.05.2024).

Куперштейн Леонід Михайлович — к. т. н., доцент кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця, email: kupershtein@vntu.edu.ua

Залепа Олександр Вячеславович — студент групи 1БС-20б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, email: sashasun2002@gmail.com

Kupershtein Leonid — PhD (eng), associated professor of information protection department, Vinnytsia National Technical University, Vinnytsia, email: kupershtein@vntu.edu.ua.

Oleksandr Zalepa — student of group 1БС-20б, Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: sashasun2002@gmail.com