роздільна здатність і попередня обробка, які можуть змінювати статистику класу real.

Таким чином, високі валідаційні метрики всередині одного набору можуть створювати хибне уявлення про практичну придатність детектора. Об'єднання різнорідних джерел даних істотно підвищує коректність виявлення дипфейкових зображень моделлю ResNetSECBAM і може бути основою для подальших досліджень стійких систем детекції.

1. Dang H., Liu F., Stehouwer J., Liu X., Jain A. K. DFFD: Diverse Fake Face Dataset. URL: https://cvlab.cse.msu.edu/dffd-diverse-fake-face-dataset.html.
2. Kang C., Jeong S., Lee J. та ін. HiDF: A Human-Indistinguishable Deepfake Dataset // Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. – 2025. – P. 5527–5538. – DOI: 10.1145/3711896.3737399.
3. Woo S., Park J., Lee J.-Y., Kweon I. S. CBAM: Convolutional Block Attention Module // Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – P. 3–19. – DOI: 10.1007/978-3-030-01234-2_1.

Модель Claude Mythos та кібербезпека: загрози та виклики

УДК 004.8:004.056

Олег Ясній[1], Любов Цимбалюк[2], Анна Турчманович[3]

Тернопільський національний технічний університет імені Івана Пулюя ,
[1]oleh.yasniy@gmail.com, [2]lubovtsymbaliuk@gmail.com,
[3]turchmanovich101@gmail.com

Claude Mythos — одна з найпотужніших сучасних моделей ШІ від Anthropic, що має значний вплив на кібербезпеку [1, 2, 3]. Однак вона не єдина загроза: інші передові моделі, зокрема GPT-5.4 Cyber (OpenAI) та Big Sleep (Google), також володіють подібними можливостями. Ера атак із використанням ШІ настала, і організації не можуть залишатися лише реактивними [1].

Мета даної роботи – проаналізувати систему ШІ Claude Mythos та виявити основні загрози та виклики, які виникають перед бізнесом.

Багато компаній роками недофінансовували кібербезпеку, оскільки ради директорів і керівництво не надавали їй пріоритету. Це створило приховані слабкі місця, які ШІ-інструменти швидко виявляють; для частини бізнесів наслідки можуть бути критичними. Особливо вразливі галузі з розгалуженими операційними технологіями — енергетика, комунальні послуги, виробництво, водопостачання, транспорт — де багато систем працюють десятиліттями й не підлягають ефективному патчингу. Усунення інвестиційного розриву вимагатиме значного нарощення витрат, тоді як більшість організацій планує лише помірне зростання бюджетів [1].

Mythos створювався як інструмент для роботи з великими кодовими базами й автоматизації розробки, але саме ці можливості роблять його потенційно небезпечним: модель може виявляти приховані недоліки, поєднувати дрібні вразливості в складні атаки, відновлювати вихідний код із розгорнутого ПЗ і,

опинившись у мережі, швидко картографувати системи, переміщатися горизонтально та створювати інструменти для витоку даних [2].

Ключові технічні інновації Mythos: велике (практично нескінченне) контекстне вікно для одночасного аналізу великих кодових масивів; рекурсивна самокорекція, що дозволяє автоматично підлаштовувати стратегії пошуку вразливостей; інтеграція з системними інструментами (дебагери, середовища), що перетворює модель на активного агента; агентний скафолдинг, тобто автономне генерування, тестування та виконання коду.

Практичні випробування показали: Mythos виявляє тисячі вразливостей, які раніше залишалися непоміченими. ШІ радикально скорочує час між виявленням і експлуатацією помилок — те, що раніше займало тижні, тепер може бути виконано за години [1].

Відповідь організацій має базуватися не на блокуванні окремих моделей, а на посиленні захисту: створення спеціалізованих центрів протидії загрозам ШІ, які використовують ШІ для захисту; зміцнення базових механізмів кібербезпеки (контроль доступу, сегментація мережі, автоматичний патчинг, архітектура нульової довіри, виявлення аномалій); пріоритетні заходи для OT-середовищ (сувора сегментація, обмеження доступу з Інтернету, специфічні системи виявлення аномалій); підготовка до постквантових загроз і розробка дорожніх карт до 2030 року.

Замість очікування спеціалізованих інструментів найефективнішим є системне зміцнення фундаменту кібербезпеки. Кібербезпека має бути на порядку денному ради директорів: потрібна персональна відповідальність керівництва, постійні інвестиції та оперативні рішення [1].

Тактичні пріоритети для практичної оборони

- Автоматичний патчинг — прискорити усунення відомих вразливостей;
- Архітектура нульової довіри — постійна перевірка користувачів і пристроїв;
- Виявлення аномалій — орієнтація на нетипову поведінку замість сигнатур;
- Модернізація контролю ідентифікації — багатофакторна автентифікація, стійка до фішингу;
- Скорочення технічного боргу — планове оновлення застарілих систем;
- Управління ризиками ланцюга постачань — оцінка та моніторинг безпеки постачальників;
- Посилення середовища — обмеження шкоди у разі компрометації (сегментація, найменші привілеї).

Ера ШІ-атак настала — сучасні моделі (Claude Mythos, GPT-5.4 Cyber, Big Sleep) значно прискорюють виявлення й експлуатацію вразливостей; реактивна позиція неприйнятна.

Головна вразливість — технічний борг і недофінансування — застарілі системи та недостатні бюджети створюють критичні «вхідні точки», які ШІ швидко виявляє.

Пріоритет — зміцнення базових засад захисту — контроль доступу, сегментація, автоматичний патчинг, нульова довіра та виявлення аномалій мають бути негайно посилені.

Проактивна оборона з використанням ШІ — створення центрів протидії загрозам ШІ, що застосовують ті самі інструменти для захисту, ефективніше за спроби блокувати окремі моделі.

Управління ризиками та стратегічні інвестиції — кібербезпека має стати пріоритетом ради директорів; потрібні чіткі дорожні карти, збільшення фінансування і підготовка до постквантових загроз.

1. URL :https://www.bain.com/insights/claude-mythos-and-ai-cybersecurity-wake-up-call/ (дата звернення: 07.05.2026)
2. URL https://www.mindstudio.ai/blog/claude-mythos-vs-opus-4-6-cybersecurity-gap (дата звернення: 07.05.2026)
3. Bell, David. (2026). Mythos and the Question Nobody Wants to Answer.

Алгоритми забезпечення безпеки інформаційних ресурсів в системах електронних платежів

УДК 004.056:336.74

Людмила Бабала[1], Андрій Сенюк[2]

Західноукраїнський національний університет[1,2]
l.duma@wumu.edu.ua[1], *seniukandrii2003@gmail.com*[2]

Сучасний розвиток цифрових технологій сприяв активному впровадженню систем електронних платежів у фінансовій сфері. Використання інтернет-банкінгу, мобільних застосунків, цифрових гаманців та електронних платіжних платформ значно підвищує швидкість проведення фінансових операцій, однак водночас створює нові виклики у сфері забезпечення безпеки інформаційних ресурсів [1]. Особливої актуальності набуває проблема захисту конфіденційних даних користувачів, банківської інформації та фінансових транзакцій від несанкціонованого доступу, кібератак та шахрайських дій.

Системи електронних платежів є складними інформаційно-комунікаційними структурами, які об’єднують банки, платіжні сервіси, торговельні платформи та кінцевих користувачів. Через значну кількість учасників процесу та постійний обмін інформацією виникає необхідність застосування ефективних алгоритмів захисту інформаційних ресурсів [2].

Одним із базових інструментів забезпечення інформаційної безпеки електронних платежів є криптографічні алгоритми. Їх застосування спрямоване на захист конфіденційності, цілісності та автентичності інформації. Найбільш поширеним є використання симетричних і асиметричних методів шифрування. Симетричні алгоритми, зокрема AES (Advanced Encryption Standard), забезпечують швидке шифрування великих обсягів даних, що є важливим під час обробки платіжних транзакцій [3]. Асиметричні алгоритми, такі як RSA, використовуються для безпечного обміну ключами та електронного цифрового підпису [4].