

ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА МЕТОДИ СОЦІАЛЬНОЇ ІНЖЕНЕРІЇ

Вінницький національний технічний університет

Анотація. *В роботі досліджується використання штучного інтелекту для автоматизації атак соціальної інженерії.*

Ключові слова: *штучний інтелект, соціальна інженерія, фішинг.*

Abstract. *This work explores how artificial intelligence can be used to automate social engineering attacks.*

Keywords: *artificial intelligence, social engineering, phishing.*

Вступ

В умовах стрімкого розвитку інформаційних технологій соціальна інженерія залишається однією із найпоширеніших методів кібератак. З появою та впровадженням штучного інтелекту зловмисники отримали нові інструменти для автоматизації та покращення ефективності таких атак. Зокрема, штучний інтелект дозволяє генерувати персоналізовані фішингові повідомлення, імітувати голосові дзвінки, використовуючи такі технології як deepfake, та аналізувати великі обсяги даних для вибору найбільш вразливих цілей.

Ці зміни значно підвищують рівень загрози, оскільки традиційні методи протидії соціальній інженерії часто виявляються неефективними проти атак, що використовують штучний інтелект. У зв'язку з цим, виникає необхідність у глибокому аналізі впливу ШІ на методи соціальної інженерії, щоб розробити нові підходи до захисту інформації та підвищення інформаційної грамотності користувачів.

Метою дослідження є оцінка ефективності маніпуляції користувачами через фішингові повідомлення, згенеровані штучним інтелектом (ШІ), у порівнянні з реальними фішинговими повідомленнями. Шляхом опитування планується визначити, наскільки ШІ може автоматизувати методи соціальної інженерії та маніпулювати довірою користувачів, а також надати рекомендації для розробки ефективних засобів протидії.

Результати дослідження

Соціальна інженерія – це сукупність підходів прикладних соціальних наук, які орієнтовані на цілеспрямовані зміни організаційних структур, що визначають поведінку людини [1]. Також соціальну інженерію розглядають як тактику маніпулювання, впливу або обману жертви з метою отримати контроль над комп'ютерною системою або викрасти особисту та фінансову інформацію. Ця тактика використовує психологічні маніпуляції, щоб обманом змусити користувачів зробити помилки та надати конфіденційну інформацію [2].

Однією з найбільш поширених атак соціальної інженерії є фішинг. Це процес спроби отримання конфіденційної інформації, такої як імена користувачів, паролі та дані кредитних карток, шляхом видавання себе за надійну особу або організацію. Фішинг може здійснюватися через масову розсилку електронних листів, SMS-повідомлень або телефонні дзвінки. Фішингові повідомлення створюють у одержувачів відчуття терміновості, цікавості або страху, що спонукає жертв розкривати конфіденційну інформацію, переходити за посиланнями на шкідливі веб-сайти або відкривати вкладення, які містять шкідливе програмне забезпечення [2].

Фішинг є вкрай небезпечною тактикою отримання конфіденційної інформації, оскільки вона змушує робити людей необдумані дії, базуючись на емоціях та страху втратити гроші, або свою особисту інформацію.

Для дослідження, було зроблено запит до нейронної мережі Deepseek з проханням створити

прототипи трьох фішингових повідомлень, що імітують реальні зразки. Нижче наведено реальні зразки та приклади таких повідомлень, створених за допомогою ШІ:

Таблиця 1.1 – Два типи повідомлень

Реальне повідомлення	Згенероване штучним інтелектом
Пошта України: Ваша посилка надійшла на склад, але затримується через незрозумілу адресу та не може бути доставлена. Підтвердьте адресу за посиланням протягом 12 годин. [http/] (Будь ласка, дайте відповідь У. Потім зупиніть текстове повідомлення та знову відкрийте посилання для активації SMS або скопіюйте посилання в Safari та відкрийте його) Колектив Пошти України вітає Вас зі святом!	Шановний клієнте, Ваша посилка надійшла на наш склад, але ми не можемо її доставити через неправильно вказану адресу. Щоб уникнути затримки, будь ласка, підтвердьте свою адресу, перейшовши за посиланням нижче: [http] Увага! Якщо ви не підтвердите адресу протягом 12 годин, посилка буде повернута відправнику. Дякуємо за розуміння! З повагою, Команда Пошти України.
Шановний клієнте, Після перевірки нашого платіжного сервера ми виявили наступну помилку: ваш щомісячний платіжний баланс був сплачений двічі (202.42 * 2) на суму 408.84 AUD. Щоб отримати повернення коштів, вам необхідно відвідати свій обліковий запис негайно та завершити заявку. Після цього ми повернемо вам другу оплату у розмірі 202.42 AUD на ваш банківський рахунок. Натисніть кнопку нижче, щоб розпочати процес повернення коштів[3].	Шановний клієнте, Під час останньої перевірки нашого платіжного сервера ми виявили, що ваш останній платіж у розмірі 150.00 USD був здійснений двічі. Це призвело до надмірного списання коштів із вашого рахунку на загальну суму 300.00 USD. Щоб отримати повернення коштів, будь ласка, перейдіть за посиланням нижче та заповніть форму заявки на повернення: [http/] Зверніть увагу: Якщо ви не подасте заявку протягом 24 годин, ми не зможемо гарантувати повернення коштів. Дякуємо за ваше розуміння. З повагою, Відділ підтримки клієнтів.
Шановний клієнте, На вашому рахунку є поточна заборгованість, тому ваш рахунок було переведено в стан овердрафту. Щоб отримати цей платіж, вам необхідно оновити дані вашого інтернет-банкінгу. Будь ласка, натисніть на посилання нижче, щоб оновити дані: [Натисніть тут] © Westpac Banking Corporation[4]	Шановний клієнте, Ми виявили, що ваш рахунок має неоплачену заборгованість, і через це ваш рахунок було переведено в стан овердрафту. Щоб уникнути додаткових комісій та відновити доступ до вашого рахунку, будь ласка, оновіть дані вашого інтернет-банкінгу. Натисніть на посилання нижче, щоб оновити дані: [Натисніть тут] Дякуємо за ваше розуміння, З повагою, Команда підтримки Westpac.

Далі було розроблено та проведено опитування серед студентів, щоб дослідити, яке фішингове повідомлення є більш переконливим. Участь в опитуванні взяла 101 особа.

Учасникам було запропоновано шість фішингових повідомлень, три з яких згенеровані штучним інтелектом, а інші три є реальними повідомленнями. На рисунку 1 наведені результати опитування.

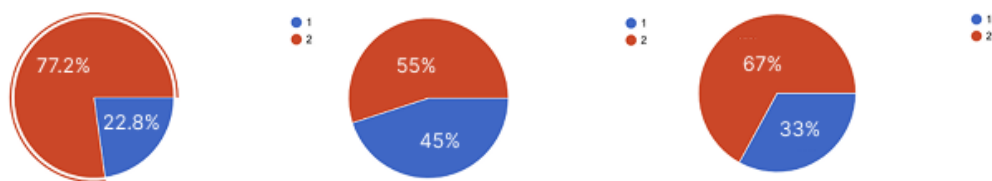


Рисунок 1 - Результати опитування

Результати дослідження вказують на те, що в першому та третьому завданнях переважна більшість учасників висловила упередження на користь повідомлень, створених штучним інтелектом, визнаючи їх більш переконливими.

На даний момент ШІ є потужним інструментом для створення фішингових повідомлень, здатних обманювати навіть досвідчених користувачів, що зумовлює необхідність розробки контрзаходів на основі аналогічних технологій. Розробникам рекомендується створення систем, які аналізують тексти, виявляють підозрілі патерни (наприклад, аномальний тон, підозрілі гіперпосилання або надмірну терміновість) та блокують фішингові повідомлення до їх отримання користувачами. Для цього доцільно використовувати алгоритми, що навчені ідентифікувати ключові ознаки фішингу. Також рекомендується впровадження суворих регуляційних норм, включаючи обмеження доступу до потужних моделей ШІ для неавторизованих користувачів та впровадження механізмів моніторингу для виявлення зловживань. Важливим є розвиток етичних стандартів використання ШІ для запобігання його застосуванню в шкідливих цілях.

Висновки

Отримані результати дадуть змогу оцінити рівень загрози, яку становлять фішингові повідомлення, згенеровані штучним інтелектом, та розробити практичні рекомендації для підвищення кібербезпеки. Це включає вдосконалення систем захисту, розробку ефективних алгоритмів для виявлення фішингу, а також підвищення обізнаності користувачів щодо нових методів соціальної інженерії, що використовують ШІ.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. СИСТЕМНИЙ АНАЛІЗ ТА СОЦІАЛЬНА ІНЖЕНЕРІЯ ЯК МЕТОДИ ПРОЕКТУВАННЯ ОСВІТНЬОГО ПРОСТОРУ [Електронний ресурс]. Режим доступу до ресурсу: <https://crust.ust.edu.ua/server/api/core/bitstreams/37f9bc81-0869-4af2-a0ab-fe955fea76b5/content> (дата звернення 01.12.2024)
2. Social Engineering. What is social Engineering? [Електронний ресурс]. Режим доступу до ресурсу: <https://www.cmu.edu/iso/aware/dont-take-the-bait/social-engineering.html> (дата звернення 01.12.2024)
3. Cyber Criminal poses as Telstra In New Phishing Scam [Електронний ресурс]. Режим доступу до ресурсу: <https://www.mailguard.com.au/blog/breaking-telstras-domain-hacked-in-new-phishing-scam> (дата звернення 20.12.2024).
4. Westpac Scam Targets Phone banking credentials [Електронний ресурс]. Режим доступу до ресурсу: <https://www.mailguard.com.au/blog/breaking-westpac-scam-targets-phone-banking-credentials> (дата звернення 20.12.2024)

Кубіря Варвара Іванівна – студентка групи 1KITC-23б, факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, Вінниця, e-mail: vkubiria@gmail.com

Kubiria Varvara – student of group 1KITS-23b, Faculty of Management and information security, Vinnytsia National Technical University, Vinnytsia, email: vkubiria@gmail.com

Науковий керівник: **Шелепало Галина Василівна** – к. фіз.-мат. н., доцент кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця.

Supervisor: **Shelepalo Halyna V.** – PhD, Candidate of Physical and Mathematical Sciences, Docent of the Department of Information Protection, Vinnytsia National Technical University, Vinnytsia, Ukraine.