УДК 316.6:[004.8:004.056]

ВПЛИВ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ НА ЕВОЛЮЦІЮ СОЦІАЛЬНОЇ ІНЖЕНЕРІЇ: ПСИХОЛОГІЯ ДОВІРИ В ЕПОХУ DEEPFAKE

Новік Т.О.
e-mail: taisiia.novik@nure.ua
Харківський національний університет радіоелектроніки, каф. СГН
м. Харків, Україна

This study examines the profound impact of generative AI, specifically deepfakes and vishing, on the evolution of social engineering by exploiting human cognitive vulnerabilities. By analyzing how AI manipulates biological trust mechanisms and authority bias, the research highlights why traditional security architectures fail against psychological manipulation. To counter these advanced threats, the paper advocates for the implementation of a "Human Zero Trust" framework, requiring strict multi-factor verification for all critical communications regardless of their audiovisual authenticity.

Генерований штучний інтелект (ШІ) трансформував соціальну інженерію, замінивши текстовий фішинг високотехнологічним клонуванням голосу та відео-діпфейками в реальному часі. Синтетичний контент став настільки реалістичним, що традиційні методи розпізнавання загроз втратили ефективність. Це ставить перед кібербезпекою фундаментальне питання: чому людський мозок залишається беззахисним перед машинною імітацією знайомих біометричних ознак?

Відповідь лежить на перетині нейробіології та інформаційних технологій [1]. Історично людська психіка еволюціонувала так, щоб миттєво довіряти візуальним та аудіальним сигналам від «своїх». Ідентифікація за тембром голосу чи мімікою відбувається на підсвідомому рівні за мілісекунди. Коли працівник отримує голосовий виклик нібито від керівника з вимогою надати доступ до системи, його мозок фіксує знайомий тембр і запускає стресову реакцію підпорядкування авторитету (Authority Bias). У цей момент аналітичне мислення блокується.

Яскравим прикладом експлуатації цієї вразливості є безпрецедентний інцидент на початку 2024 року в Гонконзі, коли фінансовий працівник транснаціональної корпорації переказав зловмисникам 25 мільйонів доларів США. Під час відеоконференції він бачив і чув свого фінансового директора (CFO) та інших колег, які давали вказівки щодо транзакцій. Усі присутні на дзвінку, окрім самої жертви, виявилися діпфейками, згенерованими ШІ на основі публічних відео та аудіозаписів [2]. Цей кейс доводить: коли людина опиняється у звичному корпоративному контексті, «евристика доступності» змушує мозок ігнорувати будь-які сумніви щодо реальності співрозмовників.

Масштаби проблеми зростають надзвичайно швидко. Згідно з

аналітичними даними, протягом 2023–2025 років кількість успішних атак із використанням технологій deepfake зросла в десятки разів, що робить цей вектор одним із найефективніших для прориву корпоративного периметра. А ось рівень виявлення людиною цих атак становить лише 24,5% [3]. Зловмисники усвідомили, що найслабшою ланкою сьогодні є не програмний код, а людський фактор. Постійне інформаційне перевантаження свідомості знижує ефективність прийняття рішень, роблячи кібербезпеку найважливішим елементом цифрової взаємодії [4]. Саме на цьому тлі втоми та зниження пильності використання генеративних нейромереж дозволяє кіберзлочинцям легко пробивати психологічний захист користувачів. ШІ створює ідеальну «ілюзію присутності», яка змушує жертву переносити свій минулий досвід довіри на згенеровану алгоритмом ситуацію. Використання генеративних нейромереж дозволяє кіберзлочинцям максимально автоматизувати та масштабувати психологічний тиск.

Отже, генеративний ШІ успішно «зламує» не програмне забезпечення, а фундаментальні біологічні механізми довіри. Оскільки машинна імітація стала невідрізною від реальності, архітектура кібербезпеки потребує перегляду: захист інформації більше не може покладатися на людські органи чуття як на надійний інструмент верифікації.

Ефективним рішенням є інтеграція концепції Human Zero Trust у корпоративну культуру. Це передбачає обов'язкову багатофакторну верифікацію критичних подій через незалежні канали зв’язку (Out-of-Band Authentication). Лише симбіоз жорстких технічних протоколів і психологічної культури перевірки дозволить нівелювати загрози в епоху Deepfake.

Список використаних джерел:

1. Personality Development in the Paradigm of Current Neuropedagogy / Y. Ribtsun et al. BRAIN. Broad Research in Artificial Intelligence and Neuroscience. 2023. Vol. 14, Iss. 4. P. 388–403. DOI: https://doi.org/10.18662/brain/14.4/512.

2. Finance worker pays out $25 million after video call with deepfake 'chief financial officer'. CNN Business. 2024. URL: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk (дата звернення: 12.03.2026).

3. Deepfake Trends and Cybersecurity Threat Report 2024-2025. Identity Security Insights. URL: https://deepstrike.io/blog/deepfake-statistics-2025 (дата звернення: 12.03.2026).

4. Коробкіна Т. В., Дашенкова Н. М. Освіта довіри: нейропедагогіка безпеки в епоху невизначеності. *Проблеми сучасних трансформацій. Серія: педагогіка та психологія. 2025. № 8*. DOI: https://doi.org/10.54929/2786-9199-2025-8-01-02.