УДК 378.147

В.С.Соловей

ЦИФРОВА ГІГІЄНА В УМОВАХ МАСОВОГО ПОШИРЕННЯ AI-ГЕНЕРОВАНОГО КОНТЕНТУ

Вінницький національний технічний університет

Анотація

Робота присвячена викликам інформаційній безпеці через масове поширення AI-генерованих підробок (діпфейків). Досліджено ризики соціальної інженерії та маніпуляцій контентом. Автором запропоновано практичні алгоритми цифрової гігієни та методи верифікації інформації, що поєднують технічні інструменти та критичне мислення для захисту цифрової ідентичності користувача.

Ключові слова: цифрова гігієна, штучний інтелект, діпфейк, кібербезпека, верифікація контенту, соціальна інженерія

Abstract

This paper explores information security challenges posed by the mass spread of AI-generated content (deepfakes). It examines risks associated with social engineering and content manipulation. The author proposes practical digital hygiene algorithms and verification methods that combine technical tools and critical thinking to protect a user's digital identity.

Keywords: digital hygiene, artificial intelligence, deepfake, cybersecurity, content verification, social engineering.

Вступ

Динамічний розвиток генеративного штучного інтелекту (GenAI) радикально змінив ландшафт інформаційних загроз, зробивши створення високореалістичного маніпулятивного контенту доступним для широкого кола осіб [2]. Сьогодні технологія Deepfake вийшла за межі розважального контенту, перетворившись на інструмент цілеспрямованої дезінформації, політичної дестабілізації та складних операцій соціальної інженерії [4,8] . Дослідження підтверджують, що діпфейки здатні суттєво впливати на суспільну думку, оскільки вони експлуатують довіру до аудіовізуальних доказів, які раніше вважалися незаперечними [6].

Основним викликом для сучасної молоді є не лише володіння технологіями створення контенту, а й здатність до швидкої верифікації інформації в умовах «інформаційного банкрутства» [9]. Зростання якості AI-генерацій призводить до того, що традиційні методи перевірки стають недостатніми, що вимагає впровадження нових підходів, таких як «превентивне навчання» (inoculation) — метод, що дозволяє людям заздалегідь розпізнавати ознаки маніпуляцій шляхом ознайомлення з механізмами створення фейків, Актуальність дослідження зумовлена необхідністю формування чітких протоколів цифрової гігієни, що базуються на критичному сприйнятті медіаданих та вмінні ідентифікувати специфічні помилки нейромереж. У світі, де межа між справжнім і синтетичним стає дедалі тоншою, цифрова гігієна трансформується з набору технічних порад у фундаментальну навичку виживання в агресивному інформаційному середовищі.

Результати дослідження

У ході проведення наукового пошуку та практичного аналізу візуального контенту було отримано наступні результати:

1. **Систематизація візуальних артефактів.**

На основі аналізу Deepfake-відео виявлено низку специфічних аномалій, що свідчать про використання нейромереж. Анатомічна невідповідних рухів є одним з цих факторів. При перегляді матеріалів були виявлені різкі та механічних рух, що не відповідають природній моториці людського тіла. Неприродне кліпання та специфічні перекривлення в ділянці очей під час зміни ракурсів, що вказує на накладання ефекту на відео. А також ідеалізація персонажу - відсутність природних недоліків

(висипів, зморшок) та наявність надмірного ефекту розмиття (блюру), що створює ефект «ідеального обличчя»[7,10].

2. **Виявлення контекстуальних помилок нейромережі.**

Генеративні моделі схильні до похибок у фонових елементах та просторовій орієнтації об'єктів. Прикладом є аномальне відображення на моніторах, некоректний напрямок руху людей на задньому плані, спотворення текстур фону.

3. **Психологічний аспект дезінформації.**

Під час проведення аналізу було виявлено основу стратегію діпфейк-контенту – апелювання до сильних емоцій користувача. Мета генеративного матеріалу є дестабілізувати критичне сприйняття через апеляцію до базових афектів [8]. Висока інтенсивність спровокованих емоцій (гнів, сміх) блокує раціональний аналіз та змушує користувача сприймати синтетичний образ як істину. Отже, стан «емоційного збурення» є індикатором необхідності негайного застосування протоколів критичного мислення та крос-верифікації.

4. **Метод «психологічного щеплення»**

Користувачі, які були ознайомлені або використовували алгоритми створення фейків мають більшу здатність до самостійної ідентифікації артефактів штучного інтелекту та підвищену когнітивну стійкість до маніпуляцій.

5. **Протоколи верифікації.**

5.1. Першочерговим етапом є детальний аналіз фізичних ознак об'єкта, що передбачає виявлення технічних артефактів нейромереж, таких як анатомічна невідповідність рухів, специфічні перекривлення в ділянці очей (окулографічні дефекти) та неприродно ідеальна текстура шкіри з надмірним ефектом розмиття.

5.2. Другий напрямок охоплює ретельний моніторинг аномалій оточуючого середовища, де фокус уваги зміщується на пошук просторових викривлень фону, некоректного відображення світла й тіней на обличчі порівняно з оточенням, а також логічних помилок у русі об'єктів другого плану.

5.3. Завершальним етапом є критична перевірка джерел поширення, яка поєднує в собі репутаційний аналіз платформи публікації, оцінку емоційного впливу контенту як інструменту соціальної інженерії та використання методів зворотного пошуку зображень для встановлення первинного контексту зйомки. Такий комплексний підхід дозволяє нівелювати вплив цілеспрямованої дезінформації шляхом перетворення пасивного споживання інформації на активний процес технічної та логічної верифікації.

Висновки

Дослідження підтверджує, що в умовах стрімкого розвитку GenAI традиційна цифрова гігієна має трансформуватися у систему активної верифікації контенту. Систематизація візуальних артефактів (анатомічних аномалій, дефектів шкіри та фону) у поєднанні з методом «психологічного щеплення» дозволяє значно підвищити стійкість користувачів до маніпуляцій. Запропонований протокол комплексної перевірки фізичних ознак та джерел поширення є ефективним інструментом нейтралізації дезінформації та збереження інформаційної безпеки особистості.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Про основні засади забезпечення кібербезпеки України : Закон України від 05.10.2017р.№2163-VIII. URL: https://zakon.rada.gov.ua/laws/show/2163-19.
2. Динамічний розвиток генеративного штучного інтелекту (GenAI) та ландшафт інформаційних загроз : аналіт. звіт. 2026.
3. Adopting a Risk-Based Approach to Generative AI. *NIST Special Publication*. 2025. URL: https://csrc.nist.gov/.
4. Deepfakes and National Security : Research Note / PERSEREC. 2024. URL: https://www.cdse.edu/Portals/124/Documents/perserec/perserec-deepfakes-research-note.pdf.
5. ENISA Threat Landscape 2025: Evolution of AI-generated misinformation. European Union Agency for Cybersecurity, 2025.

6. Inoculation helps people spot political deepfakes, study finds / *The Conversation*. 2024. URL: https://theconversation.com/inoculation-helps-people-spot-political-deepfakes-study-finds-273739.
7. IEEE Xplore Digital Library. Advanced methods for deepfake detection using neural network artifacts. 2025.
8. MITRE ATT&CK® Knowledge Base. Techniques for social engineering and impersonation. URL: https://attack.mitre.org.
9. Google Scholar : академічна пошукова система. Цифрова гігієна та медіаграмотність молоді. URL: https://scholar.google.com.ua.
10. SANS Institute. Endpoint Protection and AI Content Verification Resources. 2025. URL: https://www.sans.org/white-papers.

Соловей Вероніка Сергіївна — студентка групи 1БС-24б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, email:nikasolovey1 @gmail.com

Науковий Керівник: ***Радченко Євгеній Валентинович*** – асистент кафедри захисту інформації, Вінницький національний технічнийуніверситет, м. Вінниця, email: eradchenko@vntu.edu.ua

Solovei Veronika Sergiivna — student in Group 1BS-24b, Faculty of InformationTechnologyand Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: nikasolovey1@gmail.com

Supervisor: ***Yevhenii Radchenko V.*** – Assistant professor,Department of Information Security, Vinnytsia National Technical University,Vinnytsia, email: eradchenko@vntu.edu.ua