Штучний інтелект для виявлення синтетичних медіа (*deepfake*) у судово-експертних дослідженнях під час військових конфліктів

Андрій Синжерян
курсант, навчально-науковий інститут № 4, Харківський національний університет внутрішніх справ МВС України, м. Кам'янець-Подільський, Україна, e-mail: razorboyone@gmail.com

Максим Сітало
курсант, навчально-науковий інститут № 4, Харківський національний університет внутрішніх справ МВС України, м. Кам'янець-Подільський, Україна, e-mail: maxim.sitalo8677@gmail.com

Денис Воропаєв
курсант, навчально-науковий інститут № 4, Харківський національний університет внутрішніх справ МВС України, м. Кам'янець-Подільський, Україна, e-mail: whytee163@gmail.com

Досліджено застосування штучного інтелекту для виявлення deepfake у судово-експертних дослідженнях під час військових конфліктів і перспективи його інтеграції для забезпечення автентичності доказів.
Ключові слова: *штучний інтелект; deepfake; автентичність доказів.*

Artificial Intelligence for Detecting Deepfakes in Forensic Research During Military Conflicts

Andrii Synzherian, Maksym Sitalo, Denys Voropaev

The paper studies the use of artificial intelligence to detect deepfake in forensic expert research amid military conflicts and prospects for its integration to ensure evidence authenticity.
Keywords: *artificial intelligence; deepfake; evidence authenticity.*

У сучасних військових конфліктах синтетичні медіа, зокрема *deepfake*, становлять серйозну загрозу інформаційній безпеці. Ці технології дають змогу створювати реалістичні відео та аудіо, які можна застосовувати для маніпулювання громадською думкою, дезінформації та шантажу. За даними досліджень, останніми роками частка *deepfake*-матеріалів зросла на 330 %, що свідчить про їх активне застосування як зброї в інформаційній війні.

Актуальність теми зумовлена необхідністю розроблення ефективних методів виявлення та ідентифікації таких матеріалів у судово-експертних дослідженнях для забезпечення правдивою інформацією під час військових конфліктів.

Метою роботи є дослідження можливостей штучного інтелекту (далі — *ШІ*) для виявлення синтетичних медіа та впровадження відповідних технологій у судово-експертну практику.

Синтетичні медіа, зокрема *deepfake*, є результатом застосування технологій ШІ — генеративно-змагальних мереж (далі — *GAN*), за допомогою яких створюють високоякісні підроблені зображення, відео й аудіо. Основою роботи *GAN* є дві нейронні мережі, що взаємодіють між собою: генератор, який створює підробки, і дискримінатор, який намагається відрізнити ці підробки від реальних даних. У міру навчання генератор стає дедалі кращим у створенні фальшивих матеріалів, що значно ускладнює їх виявлення традиційними методами [1]. Цю технологію вже застосовують як засіб дезінформації, шантажу та підриву репутації, особливо під час військових конфліктів, де об'єктивна інформація відіграє критичну роль.

Проблема *deepfake* загострюється в умовах військових конфліктів через широке поширення фальшивих відео, здатних маніпулювати громадською думкою та створювати хибні уявлення про події. Особливо небезпечно, коли такі матеріали застосовують для дискредитації військових або політичних лідерів, що може вплинути на стратегічні рішення. Ситуацію ускладнено відсутністю належної законодавчої бази, що регулює застосування та відповідальність за поширення *deepfake* [2]. Незважаючи на деякі міжнародні ініціативи, зокрема з боку Європейського Союзу та США, повноцінного юридичного механізму боротьби з цією технологією до сьогодні не розроблено. Традиційні методи виявлення фальсифікацій (аналіз цифрових артефактів або відсутність збігів у спектральних характеристиках аудіо) недостатньо ефективні через надшвидке вдосконалення алгоритмів генерації синтетичних медіа.

Окрім того, важливим аспектом є питання збереження ідентифікації на рівні біометричних ознак. Зокрема, сучасні алгоритми *deepfake* здатні з високою точністю відтворювати мімічні та фізіологічні особливості (рухи очей або міміку), що ще більше ускладнює виявлення підробок. Це потребує розроблення нових методів боротьби із цими загрозами, зокрема із застосуванням ШІ для аналізу аномалій у синтезованих матеріалах [3].

ШІ, зокрема глибокі нейронні мережі, пропонує широкий спектр можливостей для виявлення синтетичних медіа. Одним із перспективних напрямів є застосування моделей глибокого навчання для аналізу мікровиразів обличчя, руху очей та інших неявних характеристик, які складно синтезувати навіть за допомогою передових алгоритмів *deepfake*. Моделі на основі *ResNet* і *EfficientNet* можна застосовувати для детектування незначних візуальних аномалій (некоректного передавання текстури шкіри або неузгодженості рухів губ із аудіопотоком) [4]. Це дає змогу значно підвищити точність виявлення фальсифікацій.

Важливим аспектом також є створення інтегрованих систем для ав-

томатизованого виявлення *deepfake* у судово-експертних дослідженнях. Застосування таких систем під час військових конфліктів дасть змогу оперативно ідентифікувати та знешкоджувати фальсифіковані матеріали, здатні впливати на інформаційні операції супротивника. До таких систем належать проєкти на основі машинного навчання, які вже застосовують провідні *IT*-компанії для виявлення підробок на відеоплатформах [5].

Особливо важливим у судово-експертних дослідженнях є забезпечення збереження цифрових доказів. ШІ може відігравати ключову роль у верифікації автентичності відео- й аудіофайлів із застосуванням блокчейн-технологій для фіксування часу та місця створення матеріалу. Це допоможе уникнути маніпуляцій із доказами та забезпечити їх правову обґрунтованість у судових процесах. Водночас необхідно адаптувати наявні судово-експертні методи для застосування нових технологій, що передбачає впровадження спеціалізованих навчальних програм для фахівців у галузі цифрової криміналістики.

Отже, ШІ є перспективним інструментом для виявлення синтетичних медіа в умовах військових конфліктів. Розроблення та впровадження нових алгоритмів, здатних ідентифікувати *deepfake* на основі аналізу даних, є необхідною умовою для забезпечення правдивості судових доказів. Подальший розвиток законодавчої бази й адаптація технологій ШІ до судово-експертної практики сприятимуть підвищенню інформаційної безпеки під час військових дій.

Перелік джерел посилання

1. Tackling online disinformation: Commission proposes an EU-wide Code of Practice : Press release / European Commission. Apr. 26, 2018. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_18_3370 (дата звернення: 06.10.2024).
2. Understanding the source of what we see and hear online / OpenAI. Aug 4, 2024. URL: https://openai.com/index/understanding-the-source-of-what-we-see-and-hear-online/ (дата звернення: 06.10.2024).
3. Deepfake Detection Challenge Results: An open initiative to advance AI / Meta. June12,2020.URL:https://ai.meta.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/ (дата звернення: 06.10.2024).
4. Hu J., Shen L., Sun G. Squeeze-and-Excitation Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. Pp. 7132—7141 URL: https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html (дата звернення: 06.10.2024).
5. Li Y., Chang M.-Ch., Lyu S. In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS). (Hong Kong, China; 11—13 Dec, 2018). 31 Jan, 2019. URL: https://ieeexplore.ieee.org/document/8630787 (дата звернення: 06.10.2024).