УДК 343.982.9
DOI https://doi.org/10.32782/2524-0374/2022-5/142

ДІПФЕЙК В КОНТЕКСТІ ДЕКЛАРАЦІЇ ПРО МАЙБУТНЄ ІНТЕРНЕТУ

DEEPFAKE IN THE CONTEXT OF DECLARATION FOR THE FUTURE OF INTERNET

Подобний О.О., д.ю.н., професор,
завідувач кафедри кримінального права, процесу та криміналістики
Міжнародний гуманітарний університет

Слатвінська В.М., викладач кафедри кримінального права, процесу та криміналістики
Міжнародний гуманітарний університет

У статті інформаційний простір сучасного суспільства, що складається із інформаційних ресурсів, інформаційних технологій та інформаційної структури, визнано системоутворюючим фактором всього суспільного життя, фундаментальною підставою політики держави у всіх її галузях. За цих умов наголошено на шкідливості таких суспільно небезпечних явищ як дезінформація (відомості, що розраховані на введення особи в оману) та діпфейк (аудіо та відеозаписи, створені або змінені таким чином, що вони помилково сприймаються як автентичні записи реальної мови або дій індивідуума).

Зазначено, що у науковій літературі з інформаційного права, кримінології та криміналістики проблема феноменів діпфейку та дезінформації привертала увагу вчених. Водночас, якщо криміналістичне розуміння сутності і тактики використання дезінформації є сталим як у теоретичному, так і особливо в практичному сенсі, то зазначені аспекти проблеми діпфейку тільки очікують належної наукової інтерпретації.

Приділено увагу сферам протиправного застосування діпфейків у 2019–2021 роках і новим схемам 2022 року для політичної дезінформації, кібершахрайства, шантажу, отримання крипто-валют, NFT, у тому числі через фішингові сайти.

Доведено небезпеку використання технології діпфейків у формі соціальної інженерії. Досліджено норми Декларації про майбутнє Інтернету щодо протидії комп'ютерним злочинам. Поставлено питання про дозволені межі навчання щодо діпфейків в контексті цифрової криміналістики та медіа безпеки, а також розбудови етикету діпфейку.

Зауважено, що етикет не будується швидко, але технічний прогрес і Декларація про майбутнє Інтернету потребують перегляду ставлення до способів використання технології, яка базується на алгоритмах машинного навчання й штучного інтелекту. Підкреслено, якщо не дозволити цій технології існувати взагалі, то такі сфери як кіноіндустрія та медіа втратить яскраву можливість. Зроблено припущення, що для зупинення розповсюдження діпфейків може бути використана технологія чат-ботів та фільтрування інформаційних каналів.

Ключові слова: діпфейк, дезінформація, фейкова інформація, інформаційна безпека, цифрова технологія, комп'ютерний злочин.

In the article, the information space of modern society, consisting of information resources, information technology and information structure, is recognized as a system-forming factor of the entire social life, the fundamental basis of the state policy in all its areas. Under these conditions, the harmfulness of such socially dangerous phenomena as disinformation (information calculated to mislead a person) and deep fake (audio and video recordings created or modified so that they are mistakenly perceived as authentic records of real language or actions of an individual) is noted.

It is noted that in the scientific literature on information law, criminology and criminalistics the problem of the phenomena of deepfake and disinformation has attracted attention. At the same time, if the criminalistic understanding of the essence and tactics of disinformation use is stable in both theoretical and especially in the practical sense, then these aspects of the problem of deepfake just waiting for the proper scientific interpretation.

Attention is paid to the areas of illegal use of deepfakes in 2019–2021 and new schemes in 2022 for political disinformation, cyber fraud, blackmail, receiving cryptocurrencies and NFT, including phishing sites.

The danger of using deepfakes technology in the form of social engineering has been proven. The norms of the Declaration on the Future of the Internet for countering computer crimes are explored. The question of the permissible limits of deepfake training in the context of digital forensics and media security, as well as the development of deepfake etiquette, is posed.

It is noted that etiquette is not built quickly, but technological advances and the Declaration on the Future of the Internet require a look at the way technology that is based on machine learning and artificial intelligence algorithms is treated. It is emphasized that if this technology is not allowed to exist at all, areas such as the film and media industries will lose a bright opportunity. It is suggested that chat-bot technology and information channel filtering can be used to stop the spread of deepfakes.

Key words: deepfake, disinformation, fake information, information security, digital technology, computer crime.

Постановка проблеми. Інформаційні ресурси, інформаційні технології та інформаційна структура в сукупності утворюють інформаційний простір сучасного суспільства, що є системоутворюючим фактором всього суспільного життя, фундаментальною підставою політики держави у всіх її галузях [1, с. 180]. Випереджальні темпи розвитку технічної складової суспільства дедалі тільки посилюються, тож підвищена увага до «комп'ютерних злочинів» не є безпідставною [2, с. 34]. Одним з таких прикладів є діпфейк (deepfake) – це аудіо та відеозаписи, створені або змінені таким чином, що вони помилково сприймаються як автентичні записи реальної мови або дій індивідуума [3, с. 96].

Стан дослідження. В останні роки у науковій літературі з інформаційного права, кримінології та криміналістики проблема феномену діпфейку та дезінформації привертала увагу М. Вальорскої, М. Д. Василенка, с. Ф. Денисова, І. І. Когутич, О. О. Подобного, В. О. Рачук, В. М. Слатвінської, В. В. Тіщенка, Ю. В. Філей, К. В. Юртаєвої, S. Afsana, M. Bahar та ін. Водночас, якщо криміналістичне розуміння сутності і тактики використання дезінформації є сталим як у теоретичному, так і особливо в практичному сенсі, то зазначені аспекти проблеми діпфейку тільки очікують належної наукової інтерпретації.

Метою статті є постановка проблеми діпфейку в контексті Декларації про майбутнє Інтернету.

Виклад основного матеріалу. З 2019 по 2021 роки діпфейки використовувалися зловмисниками для дезінформації (далі – Д.), зокрема під час створення підроблених документів тощо.

Д. – відомості, що розраховані на введення особи в оману; в оперативно-розшуковій діяльності (далі – ОРД) – різновид поведінкового акту суб'єктів, які безпосередньо здійснюють ОРД, або сторони, що їм протидіють. Д. полягає у: 1) свідомому розповсюдженні неправдивих або таких, що втратили актуальність, відомостей з метою

а) спрямування дій сторони, яка протидіє, у потрібному напрямі; б) перевірки факту та напрямів витоку інформації; в) імітування діяльності певного об'єкту відповідно до неправдивих даних; 2) здійснення спеціальних заходів, спрямованих на приховування від досліджуваної особи (сторонніх осіб) інформації, яка захищається, осіб чи об'єктів та введення їх в оману щодо справжніх рішень і завдань ОРД; Д. у кримінальному провадженні є введення в стан ілюзії (омани) конкретних осіб, запідозрених у підготовці, організації чи вчиненні злочину, і спонукання їх до певної поведінки з метою з'ясування і доказування їхньої причетності до такого злочину. Завдання з виявлення, перевірки і викриття осіб, які можуть бути причетними до задуманих, підготовлених чи вчинених злочинів, реалізують за допомогою залучення осіб, що перевіряються, у спеціальні організаційні та імітаційні ситуації, які відповідають планам слідства. У результаті створюються умови для вільного виявлення фізичного і психічного реагування, яке контролюється і фіксується правоохоронними органами. Як правило, такі завдання вирішуються за допомогою тактичних операцій із застосуванням комплексу негласних слідчих (розшукових) дій. У розслідуванні злочинів нерідко спостерігаються випадки кримінальної Д. з боку причетних до злочинної діяльності осіб у вигляді дачі неправдивих показів, спрямування підозри на невинуватих осіб, інсценування певних подій тощо з метою уникнення викриття у вчиненні злочину та ухилення від кримінальної відповідальності. У криміналістиці розробляються методи і прийоми виявлення кримінальної Д. [див.: 10].

Щодо діпфейків, у поточному ході подій були випадки їх створення для політичної дезінформації. Наприклад, створення діпфейку впливового політика з підробленою заявою. Здавалося, подібні махінації легко припинити, якщо використовувати тільки авторитетні ЗМІ. Однак на практиці великий відсоток людей не вміють ні ранжувати джерела інформації, ні відрізняти діпфейки. Не кажучи про те, що ніхто не відкидав можливість злому навіть офіційних ЗМІ та завантаження через них штучно створених відео.

Іншою сферою використання зловмисниками технології діпфейків стали кібершахрайства. Для досягнення злочинної мети шахраї використовують незаконне використання чужих особистих даних (identity theft), зокрема голосу головних виконавчих директорів компаній. Один з найперших і водночас найбільш відомий випадок шахрайського використання технології діпфейк – фішинг відбувся у 2019 р., коли невідомі шахраї шляхом використання алгоритму глибинного навчання GAN створили високоякісну імітацію голосу директора німецької компанії і за допомогою телефонного зв'язку від його імені наказали Генеральному директору дочірньої енергетичної компанії з Великобританії відправити кошти на суму 220 тис. євро угорському постачальнику. Звісно повернути зазначені гроші не вдалося, оскільки з угорського банку гроші були миттєво переведені до Мексики, а потім розділені й переведені до інших локацій. Ще одним способом використання технології діпфейк для вчинення шахрайства в Інтернеті є створення фейкових відео з відомими особами для заманювання потерпілих на фішингові сайти. Хоча такі шахрайські схеми наразі є порівняно нечисленними, подальше вдосконалення технології Deepfake може збільшити її використання у фішингових кібератаках [4, с. 39].

Проте у 2022 році діпфейки стали використовувати для отримання криптовалют, у т.ч. жертвами стають зірки шоу бізнесу. *Нова фішингова атака в WhatsApp* – зловмисники проводять її за допомогою фейкової функції для голосових повідомлень. Починається все з того, що користувач отримує лист нібито від месенджера про отримання голосового повідомлення. У цьому листі містяться кнопка «Відтворити» і сама аудіодоріжка. Відправник використовує електронну адресу «Центру безпеки дорожнього руху області», тому його лист не блокується – адреса-то справжня. Користувач же бачить як відправника Whatsapp Notifier. Натискання кнопки «Відтворити» переспрямує сайт з трояном JS/Kryptic. Користувачеві потрібно підтвердити, що він не є роботом, і натиснути на кнопку «Дозволити». Далі в систему завантажується шкідливе програмне забезпечення. Схема фішингової атаки була виявлена фахівцями з Armoblox, компанії з інформаційної безпеки. Вони повідомляють, що шкідливе ПЗ було відправлено як мінімум на 27 655 адрес.

Шахраї отримали криптовалют на $1.6 млн завдяки фейковим стрімам знаменитостей. Фейковий стрім із співзасновником Ефіріума Віталіком Бутеріним дивилося 165 000 глядачів. Деякі з них повірили шахраям і перерахували кошти на їхню адресу. Власники активів у мережі Ефіріума перевели злочинцям монети і токени на суму $933 900.

Інша схема була орієнтована на власників NFT. В описі відео шахраї викладали посилання на фішинговий сайт і обіцяли спеціальний колекційний токен тим, хто залишить свої дані – пароль і ключ відновлення доступу до акаунту.

Водночас, варто констатувати, що ні підроблені відео, ні дезінформація як такі не є новим явищем – новим є все більша простота їх створення, підвищення їхньої якості та можливості їх розповсюдження [5, с. 33].

Початком процесу створення дієвого нормативно-правового механізму протидії досліджуваному суспільно-небезпечному явищу вбачається підписання 28.04.2022 року Україною Декларації про майбутнє Інтернету (далі – Декларація). Нинішня ситуація в державі чітко демонструє ризик серйозного порушення роботи Інтернету, зокрема у вигляді повного або часткового відключення. Існує також ризик фрагментації Інтернету, спостерігається сплеск кібератак, онлайн-цензури і дезінформації.

Наведемо цитату з Декларації: «Підтвердити наше зобов'язання, щоб дії, що вживаються урядами, органами влади та цифровими сервісами, включаючи онлайн-платформи для скорочення незаконного і шкідливого контенту і діяльності в мережі, повинні відповідати міжнародному законодавству про права людини, включаючи право на свободу вираження поглядів, заохочуючи різноманітність думок і плюралізм без страху цензури, переслідування або залякування» [6]. Як фільтрувати незаконний та шкідливий контент, на кого цей обов'язок буде покладено – не відомо. Яким чином відбуватиметься процес заохочення теж не зрозуміло, оскільки грань між різноманіттям думок і цензурою, переслідуванням або залякуванням дуже тонка.

Декларація містить норму з протидії комп'ютерним злочинам: «Утримуватися від використання Інтернету для підриву виборчої інфраструктури, виборів і політичних процесів, у тому числі за допомогою таємних кампаній з маніпулювання інформацією». Цілком ймовірно, що це забезпечить заборону політичних діпфейків.

Хоча технологія діпфейк здається цікавою для створення підроблених відео або зображень чогось або окремих людей, вона, як правило, поширюється як дезінформація через Інтернет. Зміст діпфейків може бути небезпечним як для окремих людей, так і для спільнот, організацій, релігій, країн тощо. Оскільки для створення фальшивого контенту потрібна висока кваліфікація і поєднання декількох алгоритмів глибокого навчання, він здається майже справжнім і правдоподібним, його важко відрізнити [7, с. 13]. Виявлення глибоких підробок – одне з важливих завдань цифрової криміналістики та медіабезпеки. Глибокі підробки являють собою значний ризик для автентичності та безпеки сучасних інформаційних

засобів. Вони можуть використовуватися як інструменти політичної пропаганди, поширення дезінформації, шахрайства з ідентифікацією особистості та шантажу. Глибокі підробки піддають область машинного інтелекту етичним ризикам і є яскравим прикладом згубного впливу сучасних систем штучного інтелекту [8, с. 376]. Небезпека діпфейків полягає в тому, що технологія може бути використана для того, щоб змусити людей повірити в те, що щось реально, коли це не так. Настільні програми для смартфонів, такі як FaceApp і Fake App, побудовані на цьому процесі. Ці відео можуть вплинути на сприйняття чесності людини. Тому ідентифікація та класифікація цих відео стала необхідністю [9, с. 813].

Натомість Декларація проголошує наступне: «Сприяти захисту споживачів, зокрема вразливих споживачів, від шахрайства в Інтернеті та іншої недобросовісної практики в Інтернеті, а також від небезпечних товарів, що продаються в Інтернеті» [6]. Постають питання: який механізм захисту, де той перелік прикладів шахрайства в Інтернеті? З однієї сторони, Декларація обіцяє інтернет-мета всесвіт, що прискорить впровадження технологій nft, web3, а з іншої цікаво як це буде відбуватись на практиці, бо ці та інші нюанси потребують додаткового тлумачення.

Крім того, постає питання про дозволені межі навчання щодо діпфейків в контексті цифрової криміналістики та медіабезпеки. Наведемо наступні приклади: якщо під час навчання розповідати про законодавчі аспекти діпфейків і створювати їх в методичних цілях, щоб на практиці закріпити навички розпізнавання реального відео від підробного. Чи коректно ділитися діпфейк-відео з попереджувальним написом «діпфейк»? Чи сприяє методика занурення у практику навичкам розрізняти «чорне» від «білого»? Чи дозволено комусь ще, окрім криміналістів, проводити навчання з цієї теми? (можливо, це має бути сертифікований спеціаліст). Чи дозволено розміщувати такі ролики, наприклад на Ютубі? Якщо ні, то чому вони там є?

У якості можливого *висновку* слід зазначити, що вище наведено лише незначний айсберг питань, які ніяк законодавчо не врегульовано. Цим порушуємо важливе питання розбудови етикету діпфейк. Сама по собі технологія гарна, але якщо її використовувати під час соціальної інженерії, то це одне, а якщо як пам'ятку про навчання – це інше. Етикет не будується швидко, але технічний прогрес, який спостерігаємо, і Декларація потребують перегляду ставлення до способів використання технології, яка базується на алгоритмах машинного навчання й штучного інтелекту.

Якщо не дозволити цій технології існувати взагалі, то такі сфери як кіноіндустрія та медіа втратить яскраву можливість. Загальновідомо, що для зупинення розповсюдження фейків створено чат-боти, фільтруються інформаційні канали. Те саме чекає і на діпфейки? Де та грань між множинністю думок та тоталітарним контролем? Заборона щось викладати в мережу має зазвичай зворотну сторону медалі, як і будь-яка заборона. Це лише посилює бажання, та набуває стрімкого зросту інша вразливіша технологія.

Вважаємо, що проблема не у проблемі, а у ставленні до неї. Тож вбачається доречним осмислено користуватися технологіями задля благополуччя Інтернету майбутнього.

ЛІТЕРАТУРА

1. Подобний О. О., Слатвінська В.М. Основні завдання інформатизації правоохоронної діяльності. *Юридичний науковий електронний журнал*. № 9. 2021. С. 180–182. DOI: https://doi.org/10.32782/2524-0374/2021-9/43. URL: http://www.lsej.org.ua/9_2021/45.pdf
2. Василенко М.Д., Рачук В.О., Слатвінська В.М. Шкідливі програми в контексті розуміння комп'ютерної вірусології та техніко-правової змагальності: міждисциплінарне дослідження. *Наукові праці Національного університету «Одеська юридична академія»*. 2021. Т. 28. С. 28–36. DOI: https://doi.org/10.32837/npnuola.v28i0.693 URL: http://npnuola.onua.edu.ua/index.php/1234/article/view/693/757
3. Денисов С. Ф., Філей Ю. В. Порнографічні фейки: проблеми протидії. *Вісник пенетенціарної асоціації України*. № 2(12). 2020. С. 94–102.
4. Юртаєва К.В. Кримінологічний аналіз використання технології Deepfake: коли фейк стає злочином. *Вісник кримінологічної асоціації України*. 2021. № 1(24). С. 31–42.
5. Вальорска М. А. Діпфейк та дезінформація : практ. посіб. / Аґнєшка М. Вальорска ; пер. з нім. В. Олійника Київ : Академія української преси ; Центр Вільної Преси, 2020. 36 с.
6. Declaration for the Future of Internet 2022. URL: https://digital-strategy.ec.europa.eu/en/library/declaration-future-internet
7. Mahmud, Bahar & Sharmin, Afsana. Deep Insights of Deepfake Technology: A Review. 2021. URL: https://www.academia.edu/76656464/Deep_Insights_of_Deepfake_Technology_A_Review
8. Das, S. Seferbekov, A. Datta, M. Islam and M. Amin, "Towards Solving the DeepFake Problem : An Analysis on Improving DeepFake Detection using Dynamic Face Augmentation", in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, 2021 pp. 3769-3778. doi: 10.1109/ICCVW54120.2021.00421
9. Karthik P. C., Sanjana S., M. P. Adithya Vijayan, Thushara P., Wilson A. International Journal of Engineering Research & Technology. Vol. 10 Issue 05. 2021. P. 813-816. URL: https://www.ijert.org/research/review-of-deepfake-detection-techniques-IJERTV10IS050425.pdf
10. Подобний О.О., Тіщенко В.В. Дезінформація. *Велика українська юридична енциклопедія* : у 20 т. Харків : Право, 2016. Т. 20. Криміналістика, судова експертиза, юридична психологія / редкол. В. Ю. Шепітько (голова) та ін. ; Нац. акад. прав. Наук України ; Ін-т держави і права ім. В. М. Корецького НАН України ; Нац. юрид. цн.-т ім. Ярослава Мудрого. 2018. С. 158.