УДК 343.98
DOI https://doi.org/10.37687/2413-6433.2021-1.7

***Найдьон Я.П.,***
аспірант кафедри інформаційного права
та права інтелектуальної власності
факультету соціології і права
*Національного технічного університету України*
*«Київський політехнічний інститут імені Ігоря Сікорського»*
*ORCID: 0000-0002-8187-6623*

# ПОРНОГРАФІЧНІ ДІПФЕЙКИ ЯК СУЧАСНИЙ ПРОЯВ КІБЕРЗЛОЧИННОСТІ

# PORNOGRAPHIC DEEPFAKES AS A MODERN MANIFESTATION OF CYBERCRIME

Сучасна людина проводить у Інтернеті все більше часу. Завдяки Інтернету і технологіям ми змогли спілкуватися з людьми з будь-якої точки Земної кулі, отримувати миттєвий доступ до інформації, працювати віддалено і робити ще багато речей, які раніше для людства були неможливі. Нас оточує безліч гаджетів і пристроїв з доступом до так званого «Інтернету речей». Але водночас Інтернет зробив нас більш вразливими. У міру комп'ютеризації суспільства і розвитку мережі Інтернет зростає кількість кіберзлочинів. По суті, Інтернет є глобальною мережею, яка використовується для зв'язку та обміну даними, а кіберзлочини, або комп'ютерні злочини, – це використання того чи іншого інструменту для досягнення злочинних цілей, таких як вчинення шахрайства, торгівля порнографією та інтелектуальною власністю, викрадення особистих даних або порушення конфіденційності інформації, фінансові злочини, торгівля контрабандним товаром, незаконне скачування файлів і т.д. На відміну від традиційних злочинів, методи здійснення кіберзлочинів легко поширюються, відтворюються і навіть продаються. Це дозволяє рекламувати злочинні техніки і пропонувати злочин як послугу. Ми не завжди можемо відокремити істину від вигадки, і право не встигає за темпом виникнення проблем. Останніми роками в темних кутках Інтернету з'явилося щось нове: оброблені фото- і відеозображення, в яких обличчя однієї людини переконливо приставлено до тіла іншої, яка бере участь у відверто сексуальних сценах. Творці, коментатори і дослідники називають такі зображення діпфейками. Хоча ці зображення фіктивні, створені за допомогою нових технологій штучного інтелекту, вони представлені як реальні зображення реальних людей, зазвичай жінок. Нині будь-хто, хто має цифрове фото, може проти своєї волі стати «порнозіркою», а в законі немає ніяких засобів перешкодити цьому. У рамках цієї статті автором розглянуто феномен діпфейку як сучасного прояву кіберзлочинності, юридичні та моральні проблеми цього явища, проблему визначення жертви діпфейку та встановлення відповідальності.

***Ключові слова:*** *діпфейк, кіберзлочинність, порнографічні відео, порнографія.*

Modern people spend more and more time on the Internet. Due to the Internet and technology, we have been able to communicate with people from anywhere in the world, get instant access to information, work remotely and do many other things that were previously impossible for humanity. We are surrounded by many gadgets and devices with access to the so-called "Internet of Things". But at the same time, the Internet has made us more vulnerable. With the computerization of society and development of the Internet, the number of cybercrimes is growing. In essence, the Internet is a global network used for communication and data exchange, and cybercrime or computer crime is the use of a tool to achieve criminal purposes, such as fraud, trafficking in pornography and intellectual property, theft personal data or breach of confidentiality of information, financial crimes, trafficking of smuggled goods, illegal downloading of files, etc. Unlike traditional crimes, cybercrime methods are easily disseminated, reproduced, and even sold. This allows you to advertise criminal techniques and offer crime as a service. We cannot always separate truth from fiction, and law does not keep up with the pace of problems. Recently, something new has appeared in the dark corners of the Internet: processed photos and videos, in which the face of one person is convincingly attached to the body of another, who participates in openly sexual scenes. Creators, commentators and researchers call such images deepfakes. Although these images are fictitious, created using new artificial intelligence technologies, they are presented as real images of real people, usually women. Today, anyone with a digital photo can become a "porn star" against their will, and there is no law in to prevent this. In this article, the author considers the phenomenon of deepfake as a modern manifestation of cybercrime, legal and moral problems of this phenomenon, the problem of identifying the victim of deepfake and establishing responsibility.

***Key words:*** *deepfake, cybercrime, pornographic videos, pornography.*

Помилкові новини і фейкові розсилки в месенджерах здавалися нам головною проблемою XXI століття. Але вже зараз на зміну їм приходять deepfakes (діпфейки).

Deepfake (від англ. deep – «глибокий» і fake –«фальшивий») – реалістична маніпуляція аудіо- і відеоматеріалами за допомогою штучного інтелекту. Ця технологія змушує людину говорити те, чого вона не говорила, і робити те, чого вона ніколи не робила. Технологія Deepfake йде корінням у далекі дев'яності. Тоді подібними інструментами володіли лише експерти зі спецефектів у кіноіндустрії. Згодом технологія була доопрацьована в Інтернет-співтоваристві, і програмне забезпечення для створення deepfakes з'явилося у відкритому доступі. Останнім часом технологія deepfake привертає велику увагу через її використання у фінансових махінаціях, розіграшах та фальшивих новинах (fake news). Deepfake застосовує можливості штучного інтелекту для синтезу людського зображення: об'єднує кілька знімків, на яких людина зображена з різних ракурсів і з різним виразом обличчя, і робить з них відеопотік. Аналізуючи фотографії, спеціальний алгоритм навчається того, як виглядає і може рухатися людина. При цьому працюють дві нейромережі. Перша з них генерує зображення, а друга відповідає за пошук відмінностей між ними і справжніми зразками. У разі якщо друга нейромережа виявляє підробку, зображення відправляється назад першій для удосконалення. Deepfake працює за допомогою відкритих алгоритмів машинного навчання і бібліотек, що дозволяє досягти найвищої якості контенту. Нейромережа отримує зображення з бібліотеки і навчається за допомогою роликів на відеохостингу. Штучний інтелект тим часом зіставляє фрагменти вихідних портретів з тим, що є на відео, і в підсумку виходить правдоподібний матеріал [1].

Аналізуючи сучасний стан проблематики виготовлення та поширення порнографічних діпфейків як сучасного прояву кіберзлочинності, варто зазначити, що це явище викликає занепокоєння широкого загалу та активне обговорення. Правовим проблемам цього явища приділяли увагу як вітчизняні, так і зарубіжні науковці, такі як: Аґнєшка М. Вальорска [2], О.С. Гавриш [3], О.В. Махницький [3], О.В. Одаренко [4] та інші дослідники. З огляду на те, що проблематика є досить новою, правові питання виготовлення та поширення діпфейків потребують комплексного аналізу, виявлення юридичних та моральних проблем цього явища, дослідження проблеми визначення жертви діпфейку та встановлення юридичної відповідальності за скоєне діяння.

Deepfakes вперше з'явилися в Інтернеті наприкінці 2017 року, коли анонімний користувач платформи Reddit виклав реалістичні порнографічні відео, де обличчя акторів були змінені на обличчя знаменитостей [5]. Невдовзі після цього інший анонімний користувач платформи створив і випустив безкоштовний додаток FakeApp, що дозволяє легко створювати deepfakes [6]. До створення додатку FakeApp виробництво оброблених відео було дорогим і трудомістким процесом, який потребував матеріальних витрат та людських зусиль, доступним тільки на кіностудіях Голлівуда. Творець FakeApp «зробив технологію deepfake доступною людям без технічної освіти та досвіду програмування» [7]. Виходячи з цього, натепер створення deepfakes доступно кожному, необхідно лише мати доступ до смартфона, ноутбука чи комп'ютера. Незалежно від свого віку людина за допомогою технологій спроможна за короткий проміжок часу в будь-якій точці світу зробити реалістичний порнографічний контент, який без застосування спеціальних програм та навичок досить складно відрізнити від дійсності.

Хоча спочатку deepfakes були спрямовані проти знаменитостей, то нині звичайна людина також може побачити своє зображення у фейковому відео. Тепер, коли соціальні медіа широко поширені і багато людей викладають свої фотографії в Інтернеті, користувачі FakeApp можуть використовувати зображення будь-якої людини з Instagram або Facebook. Так, користувачі Reddit відкрито обговорювали створення deepfakes, що зображують друзів, викладачів або колишніх партнерів. Оскільки технологія deepfake дозволяє створювати реалістичні порнографічні відео без згоди зображених людей і оскільки ці відео можуть широко поширюватися в мережі Інтернет, вони належать до галузі кіберзлочинів. Американська дослідниця Ребекка Делфіно до категорії кіберзлочинів відносить такі злочини, як порно з помсти і порнографія, створена без згоди учасників. Порно з помсти – це злочин, що складається в «наданні приватного, сексуально відвертого зображення особі, для якої воно не призначалося» без згоди учасників. Тут ідеться про справжні, а не оброблені фото- і відеозображення. Зазвичай партнери діляться подібними зображеннями за взаємною згодою, а пізніше один з них поширює їх, бажаючи помститися іншому. Однак матеріали порно з помсти можуть також бути краденими або отриманими сторонніми людьми шляхом злому чи крадіжки мобільного телефону або комп'ютера. Хакери можуть розміщувати ці матеріали в Інтернет разом з інформацією, що ідентифікує жертву або без такої. Більш широке поняття – «порнографія, створена без згоди учасників» – включає сексуально відверті фото- і відеозображення, записані без згоди учасника, наприклад приховано. Порнографічні deepfakes і порно з помсти мають низку тривожних спільних рис. По-перше, обидва прояви переважно стосуються жінок. Один журналіст описував deepfakes як «спосіб для чоловіків мати необмежену уявну владу над жіночими

тілами». По-друге, обидва явища припускають поширення сексуально відвертих матеріалів без згоди. Жертви порно з помсти не погоджувалися на публічне поширення зображень, так само як жертви deepfakes не погоджувалися на маніпулювання з їхніми зображеннями. Таким чином, і те, і інше порушує очікування людини, що її сексуальна активність буде заснована виключно на добровільній згоді [8, с. 110–111].

Обидва ці явища можуть мати довгострокові ефекти на жертв і викликати схожі емоційні та психологічні негативні наслідки. Примітно, що порно з помсти не тільки завдає шкоди репутації та емоційному благополуччю людини, але і несе в собі ризики значного законодавчого, професійного і навіть фізичного збитку. Проблеми з фізичною безпекою включають насильство, переслідування та інші злочинні діяння.

Швидке поширення фейкових порнографічних відео і стрімкий розвиток технологій, що дозволили з'явитися феномену deepfake, порушують складні юридичні та моральні проблеми. Ці проблеми ведуть до необхідності прийняття адекватного кримінального законодавства.

Розвиток нових технологій порушує проблему визначення жертв deepfakes, позаяк вони зображують двох людей: того, хто брав участь у сексуальному акті, і того, чиє обличчя було використано. Незважаючи на це, в більшості статей і коментарів жертвою називають того, чиє обличчя було використано. Однак якщо відео було використано без згоди учасника акту, то чи не є ця людина також жертвою злочину? Крім того, як у законі може бути визначена заподіяна шкода, якщо deepfake не зображує справжніх епізодів з життя жертви? За що повинна настати відповідальність, якщо не була показана жодна реальна людина? Людина, чиє обличчя було використано, ймовірно, не давала згоди на участь у порнографії, але це не так однозначно щодо того, чиє тіло було використано. Актор, чиє тіло з'являється на відео, міг погодитися на оригінальне порнографічне відео, але, швидше за все, не на те, щоб чиєсь обличчя було приставлене до його тіла. Це також жертва злочину. Таким чином, обидві людини, зображені за допомогою deepfake, повинні вважатися жертвами.

Deepfakes порушують також складні проблеми визначення правопорушника і видалення відео після публікації в Інтернеті. Після того як відео опубліковане в Інтернеті, його складно видалити, і, як правило, репутаційний і емоційний збиток уже нанесено. В ідеалі винним за законом повинен вважатися той, хто створив і опублікував фейкове відео. Однак визначити цю особу буває складно з двох причин. По-перше, багато майданчиків, наприклад Reddit, Twitter і Pornhub, передбачають анонімне використання; не дивно, що deepfakes процвітають саме на цих площадках. По-друге, творці deepfakes можуть використовувати такі програми, як Tor, що приховують IP-адресу, пов'язану з deepfake.

Проблеми з визначенням жертви і правопорушника, з відповідальністю платформи ще більше заплутуються через неясності у застосуванні чинного законодавства до deepfakes. Цьому заважають усе ті ж проблеми зі зміною зображення і анонімністю, позаяк залишається неясним, хто саме повинен нести відповідальність за deepfakes. Деякі заходи спрямовані на творців контенту, інші – на соціальні мережі, що нездатні відстежити поширюваний на них контент, треті – і на те і на інше.

Підбиваючи підсумок, варто зазначити, що такі платформи, як Pornhub, загалом не змогли видалити наявний контент такого роду і перешкоджати розміщенню нового; вже одне це демонструє гостру необхідність прийняття законодавчих рішень, які зачіпають як творців, так і провайдерів такого контенту [9]. Однак нещодавно інструмент для визначення зображень людей із заміненими обличчями, які згенерував комп'ютер, розробила корпорація Microsoft. Програмне забезпечення аналізує фото і відео та дає оцінку ймовірності, що матеріал створений штучно. Таким чином, компанія сподівається, що технологія допоможе «боротися з дезінформацією». Microsoft Video Authenticator виявляє фейки за знаками, які не помітні людині, наприклад тонкі пікселі у відтінках сірого на тому місці, де створена комп'ютером версія особи була об'єднана з оригіналом. Варто сподіватись, що цей інструмент буде мати практичне значення та позитивний результат у боротьбі з явищем deepfakes.

**Список використаних джерел:**

1. Панасенко А. Технологии Deepfake как угроза информационной безопасности. URL: https://www.antimalware.ru/analytics/Threats_Analysis/Deepfakes-as-a-information-security-threat (дата звернення: 8.08.2020).

2. Вальорска М. Аґнєшка. Діпфейк та дезінформація : практичний посібник / пер. з нім. В. Олійника. Київ : Академія української преси. Центр Вільної Преси, 2020. 36 с.

3. Махницький О.В., Гавриш О.С. Аналіз кіберзагроз: найближчі перспективи. *Міжнародна та національна безпека: теоретичні і прикладні аспекти :* матер. III Міжнар. наук.-практ. конф. (м. Дніпро, 15 березня 2019 р.). Дніпро : ДДУВС, 2019. С. 277–279.

4. Медіазнавчі студії в європейському діалозі: освітній та науковий дискурс: міжнародна науково-практична онлайн-конференція. URL: https://ij.kubg.edu.ua/images/phocagallery/Podii2020/verstka_program.pdf (дата звернення: 04.02.2021).

5. Технологии 96% дипфейков – не с политиками. Их используют в порно. URL: https://vctr.media/deepfake-eto-bolshe-pro-porno-29521/ (дата звернення: 28.08.2020).

6. Hawkins D. Reddit Bans 'Deepfakes', Pornography Using the Faces of Celebrities Such as Taylor Swift and Gal Gadot. *Wash. Post.,* Feb. 8, 2018. URL: https://www.washingtonpost.com/news/morning-mix/wp/2018/02/08/reddit-bans-deepfakespornography-using-the-faces-of-celebrities-like-taylor-swift-and-gal-gadot[https://perma.cc/2TPE-AYBG].

7. Cole S. We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now. *VICE: Motherboard,* Jan. 24, 2018/ URL: https://motherboard.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley [https://perma.cc/EN85-JSM7].

8. Делфино Р.А. Порнографические дипфейки: следующий трагический акт феномена «порно из мести» и необходимость принятия уголовного закона на федеральном уровне. *Актуальные проблемы экономики и права*. 2020. Т. 14. № 1. С. 105–141.

9. Clark B. Pornhub Promised to Ban "Deepfakes" Videos. And It Failed Miserably. *TNW*, Apr. 18, 2018/ URL: https://thenextweb.com/insider/2018/04/19/pornhub-promised-to-ban-deepfakes-videos-and-it-failed-miserably/ [https://perma.cc/GYC8-48AF].