Філософія, методологія, теорія та історія
основ національної безпеки держави

УДК 342.15

Шемаєв Володимир Миколайович
доктор військових наук, професор,
доцент кафедри інформаційної безпеки держави
Навчально-науковий інститут інформаційної безпеки
та стратегічних комунікацій НА СБ України
Shemayev Volodymyr
Doctor of Military Sciences, Professor,
Associate professor of the Department of Information Security of the State
Educational and Scientific Institute of Information Security and Strategic
Communications of the National Academy of Security Service of Ukraine
ORCID: 0009-0003-2568-5587

Єрьоміна Людмила Валеріївна
старший виклад кафедри інформаційної безпеки держави
Навчально-науковий інститут інформаційної безпеки
та стратегічних комунікацій НА СБ України
Yeromina Liudmyla
Senior Lecture of the Department of Information Security of the State,
Educational and Scientific Institute of Information Security
and Strategic Communications of the Security Service of Ukraine
ORCID: 0009-0009-1656-2546

РИЗИКИ ДЛЯ НАЦІОНАЛЬНОЇ БЕЗПЕКИ ВІД ВИКОРИСТАННЯ ТЕХНОЛОГІЇ DEEPFAKE

NATIONAL SECURITY RISKS FROM THE USE OF DEEPFAKE TECHNOLOGY

Анотація. *Вступ. Синтезовані медіа-тексти, зображення, аудіо та відео, створені або оброблені штучним інтелектом (дипфейки), наразі представляють як значні можливості, так і ризики для національної безпеки. Наміри, з якими створюються такі продукти-фальсифікати різняться за своїми етичними характеристиками: від позитивних і нейтральних (мистецтво, сфера послуг, розваги та ін.) до негативних і злочинних (дискредитація фізичних, юридичних осіб або навіть цілих держав; провокація конфліктів; правопорушення тощо). Прийняття альтернативних контрзаходів зазначеним ризикам наразі широко визнане, але на практиці впровадження цих заходів залишається на початковій стадії. Актуальності набуває проблема визначення ризиків від використання синтезованих медіа-текстів, зображень, аудіо та відео, створених або оброблених штучним інтелектом (дипфейків) для превентивного реагування на розгортання загроз національній безпеці.*

Мета. Метою статті є визначення ризиків та загроз національній безпеці від використання дипфейків та обгрунтування пріоритетів державної політики щодо превентивного захисту від зазначених загроз.

Матеріали і методи. Матеріалами дослідження є: 1) нормативно-правове забезпечення з регулювання ШІ в ЄС та Україні; 2) праці вітчизняних та зарубіжних авторів, що провадять свої науково-практичні дослідження у царині можливостей та ризиків використання ШІ, зокрема щодо його застосування іноземними країнами для здійснення протиправної діяльності (під час інформаційної війни, шахрайських дій, військового обману, провокування внутрішніх заворушень тощо).

В процесі здійснення дослідження було використано наступні наукові методи: теоретичного узагальнення та групування (для характеристики ризиків використання дипфейків та інших продуктів ШІ); формалізації, аналізу та синтезу (для характеристики поточного ландшафту синтезованих медіа-загроз для національної безпеки); логічного

узагальнення (визначення контрзаходів, які поєднують технологічну, нормативну та освітню сфери, щоб протистояти зростаючій загрозі, яку становлять синтетичні засоби масової інформації, які стають інформаційною зброєю, а також формулювання висновків).

Результати. У науковій статті визначено та проаналізовано поточні ризики для національної безпеки від застосування дипфейків: гібридна війна, шпигунство та стеження, військовий обман, розбалансування внутрішньої політики, фінансові злочини.

Встановлено зовнішні та внутрішні інструменти зниження ризиків від застосування дипфейків, зокрема, впровадження багатогранних контрзаходів, які мають поєднувати: технологічну, нормативну та освітню сфери, щоб протистояти зростаючій загрозі, що становлять синтетичні засоби масової інформації, які стають інформаційною зброєю.

Обґрунтовано шляхи протидії ризикам розповсюдження дезінформації від дипфейків в Україні, що передбачають необхідність: розроблення законодавства щодо регулювання та розвитку ШІ, публікації Білої книги (де має бути висвітлено підходи держави до регулювання ШІ та описано кінцеві результати, які держава планує досягнути на цьому шляху); оприлюднення рекомендацій щодо використання інструментів ШІ громадянським суспільством та бізнесом; врахування євроінтеграційних прагнень України при визначенні шляху правового регулювання штучного інтелекту з урахуванням ризиків створення дезінформації за допомогою ШІ; створення та пропагування серед населення освітніх матеріалів щодо розпізнавання та боротьби з дезінформацією, яка створюється й поширюється за допомогою ШІ.

Перспективи. Напрямом подальших досліджень у цієї сфері визначено аналіз способів поширення дипфейк-контенту з метою визначення можливостей потенційного захисту, а також специфіки його впливу у складних, мультикультурних середовищах з неоднорідними групами людей,

які по-різному (залежно від способу життя, світогляду, віро-сповідання тощо) можуть реагувати на один і той самий контент.

Ключові слова: *національна безпека, штучний інтелект, дипфейк, загрози, гібридна війна.*

Summary. *Introduction. Synthesized media texts, images, audio, and video created or processed by artificial intelligence (deepfakes) currently pose both significant opportunities and risks to national security. The intentions with which such counterfeit products are created vary in their ethical characteristics: from positive and neutral (art, services, entertainment, etc.) to negative and criminal (discrediting individuals, legal entities, or even entire states; provoking conflicts; offenses etc.). The adoption of alternative countermeasures to these risks is now widely recognized, but in practice the implementation of these measures remains at the initial stage. The problem of determining the risks from the use of synthesized media texts, images, audio and video created or processed by artificial intelligence (deepfakes) for a preventive response to the deployment of threats to national security is becoming relevant.*

Purpose. The article is aimed at identifying risks and threats to national security from the use of deepfakes and substantiating the priorities of state policy on preventive protection against these threats.

Materials and methods. The materials of the study are: 1) regulatory and legal support for the regulation of AI in the EU and Ukraine; 2) works of domestic and foreign authors conducting their scientific and practical research in the field of opportunities and risks of using AI, in particular regarding its use by foreign countries to carry out illegal activities (during information warfare, fraudulent actions, military deception, provoking internal unrest, etc.).

In the process of the study, the following scientific methods were used: theoretical generalization and grouping (to characterize the risks of using deepfakes and other AI products); formalization, analysis and synthesis (to

characterize the current landscape of synthesized media threats to national security); logical generalization (identification of countermeasures that combine the technological, regulatory and educational spheres to counter the growing threat posed by synthetic media, which are becoming information weapons, as well as the formulation of conclusions).

Results. The scientific article identifies and analyzes the current risks to national security from the use of deepfakes: hybrid warfare, espionage and surveillance, military deception, imbalance of domestic politics, financial crimes.

External and internal tools have been established to reduce the risks of deepfakes, in particular, the introduction of multifaceted countermeasures that should combine: technological, regulatory and educational spheres in order to counter the growing threat posed by synthetic media, which are becoming information weapons.

Ways to counter the risks of spreading disinformation from deepfakes in Ukraine are substantiated, which provide for the need for: development of legislation on the regulation and development of AI, publication of a White Paper (which should highlight the state's approaches to AI regulation and describe the final results that the state plans to achieve on this path); publication of recommendations on the use of AI tools by civil society and business; taking into account Ukraine's European integration aspirations when determining the path of legal regulation of artificial intelligence, taking into account the risks of creating disinformation using AI; creating and promoting educational materials among the population on recognizing and combating disinformation created and spread with the help of AI.

Prospects. The direction of further research in this area is the analysis of ways to spread deepfake content in order to determine the possibilities of potential protection, as well as the specifics of its impact in complex, multicultural environments with heterogeneous groups of people who may react differently (depending on lifestyle, worldview, religion, etc.) to the same content.

Key words: *national security, artificial intelligence, deepfake, threats, hybrid war.*

Постановка проблеми. Термін «deepfake»/дипфек – є поєднанням двох понять: «deep learning» (глибоке навчання) та «fake» (фейк, підробка). Тобто це фото, відео або аудіо, штучно згенеровані/синтезовані з використанням знань, отриманих в процесі навчання нейронних мереж. Ці синтезовані медіа-тексти, зображення, аудіо та відео, створені або оброблені штучним інтелектом (ШІ), наразі представляють як значні можливості, так і ризики. Наміри, з якими створюються такі продукти-фальсифікати різняться за своїми етичними характеристиками: від позитивних і нейтральних (мистецтво, сфера послуг, розваги та ін.) до негативних і злочинних (дискредитація фізичних, юридичних осіб або навіть цілих держав; провокація конфліктів; правопорушення тощо). Серед ризиків, зокрема, можна відмітити:

зловмисники все частіше використовують синтезований медіа - цифровий контент, створений або оброблений штучним інтелектом (ШІ), для посилення своєї шахрайської діяльності, завдаючи шкоди окремим особам і організаціям у всьому світі;

використання синтезованого медіа-контенту почало підривати довіру людей до інформації, що створює серйозні наслідки для стабільності та стійкості інформаційного середовища;

на-сьогодні, здатність людини розпізнавати створений штучним інтелектом контент, залишається основним захисним бар'єром від того, щоб люди стали жертвою оманливо представлених синтезованих медіа. Однак нещодавнє експериментальне дослідження CSIS (Центр стратегічних та міжнародних досліджень, США) показало, що люди більше не можуть надійно розрізняти автентичні та створені ШІ зображення, аудіо та відео, отримані з загальнодоступних інструментів. Отже, те, що виявлення

людини перестало бути надійним методом ідентифікації синтезованих носіїв, лише посилює небезпеки, пов'язані з неправильним використанням технології, підкреслюючи нагальну потребу в застосуванні альтернативних контрзаходів для вирішення цієї нової загрози.

Хоча необхідність прийняття цих альтернативних контрзаходів, що охоплюють від технологічних до регулятивних, для компенсації була широко визнана критичною, на практиці впровадження цих заходів залишається в основному на початковій стадії. Таким чином, ця зростаюча вразливість означає, що усвідомлення того, коли люди більше не можуть покладатися виключно на свої очі та вуха для виявлення контенту, створеного штучним інтелектом, має вирішальне значення для того, щоб краще розпізнавати, коли виявлення людиною більше не є ефективним запобіжником проти зловживання технологією.

З безперервним удосконаленням технологій очікується, що глибина та широта випадків обману з підтримкою штучного інтелекту також розшириться та урізноманітниться. Зазначене обумовлює актуальність дослідження наукової проблеми щодо визначення ризиків та загроз національній безпеці від використання дипфейків та обгрунтування пріоритетів державної політики щодо превентивного захисту від зазначених загроз.

Аналіз останніх досліджень і публікацій. В останні роки у науковій літературі з проблем ШІ зроблено значний теоретичний доробок як вітчизняними так і зарубіжними дослідниками. З точки зору інформаційного права, проблема феномену дипфейку та дезінформації привертала увагу: М. Вальорскої [1], в контексті розуміння комп'ютерної вірусології та техніко-правової змагальності досліджувалася в роботі М. Василенка, В. Рачук [2], через призму кримінології та криміналістики розглядалася в роботах С. Денисова, Ю. Філей [3], О. Подобного, В. Слатвінської, В. Тіщенка [4-5], К. Юртаєвої [6], S. Afsana, M. Bahar, [7]

B.Chesney, D. Citron та ін. [8], в контексті розгляду технології дипфейку як феномену інформаційної доби привертають увагу роботи І. Руснака, О. Рупташ [9]. Аналіз проблем використання дипфейків для дезінформації суспільства здійснювали В. Мудраков [10], І. Березенко [11], М. Єсипович [12], О. Петрів [13]. Проблеми та шляхи державного регулювання ШІ в Україні розглядаються докладно в Білій книзі з регулювання ШІ в Україні [14] тощо.

Виділення невирішених раніше частин загальної проблеми. Водночас, недостатньо розглянутими залишаються питання визначення складових поточного ландшафту синтезованих медіа-ризиків для національної безпеки та обґрунтування способів превентивного запобігання їх перетворення у загрози.

Метою статті є визначення ризиків та загроз національній безпеці від використання дипфейків та обгрунтування пріоритетів державної політики щодо превентивного захисту від зазначених загроз.

Виклад основного матеріалу дослідження. Реальною загрозою для світу, а особливо для України в нинішніх умовах, стали новітні засоби пропаганди, залякування та спотворення інформації. Найбільш серйозною зброєю в руках любителів спотворити інформацію, видати брехню за правду, на разі є технологія дипфейку, яка дозволяє, за допомогою нейромереж (фактично нинішніх потужностей навчання штучного інтелекту) реалістично накласти обличчя однієї людини, на обличчя іншої і зробити від її особи будь-яку заяву.

Застосування технології дипфейку містить потенційні загрози не лише окремим індивідам, групам чи установам, а й цілим державам і міждержавним об'єднанням. До них належать: порушення алгоритмів і технік демократичних процедур: передвиборчих перегонів, політичної агітації, дебатів і дискусій; розмивання довіри до демократичних інститутів: незалежних ЗМІ, партійних осередків і лідерів, виборчої системи в цілому;

поглиблення суспільних розколів на ґрунті політики, релігії, раси й національності, гендеру тощо; загроза національній безпеці та міжнародній співпраці шляхом спотворення цілей, ідеалів і цінностей державної політики, дискредитації політичних еліт, лідерів партій, впливових громадських діячів або «лідерів думок» та інше [8].

Поточний ландшафт синтезованих медіа-загроз для національної безпеки стосується: гібридної війни, шпигунства та стеження, військового обману, розбалансування внутрішньої політики, фінансові злочини, зокрема:

- *гібридна війна* (інформаційна війна, кібератаки, політичний і економічний примус). Приклади: інформаційні кампанії навколо основних політичних центрів або важливих подій, включаючи вибори в Європі, Тайвані, США, російсько-українська війна та конфлікт між Ізраїлем та ХАМАС тощо; програмне забезпечення для зміни обличчя в режимі реального часу було використано, щоб успішно видати себе за мера Києва Віталія Кличка в серії відеодзвінків з кількома мерами великих європейських міст; ймовірно згенероване штучним інтелектом відео секс-запису кандидата в президенти Туреччини на виборах 2023 року, було широко поширене, що призвело до зняття кандидата з перегонів;

- *шпигунство та стеження* з боку держав та приватної індустрії кіберспостереження для отримання конфіденційної інформації від цілей. Приклади: приватні компанії кіберрозвідки використовували сотні підроблених облікових записів у соціальних мережах, видаючи себе за активістів, журналістів і молодих жінок, щоб таємно збирати інформацію від цілей, зокрема IP-адреси та особисту контактну інформацію; пов'язані з державою суб'єкти використовували соціальну інженерію за допомогою великих мовних моделей (LLM), щоб маніпулювати цілями та сприяти збору й аналізу інформації з відкритих джерел;

- *військовий обман* для цілеспрямованих військових операцій - для отримання переваги на полі бою (створення повністю фіктивних подій для зміни або викривлення розвідки противника, видавання себе за військового персоналу для фальсифікації або плутання наказів, а також створення інформаційного шуму для маскування своїх дій від супротивника. Приклади: згенерований ШІ контент із зображенням президента України Володимира Зеленського був опублікований і широко поширений у соціальних мережах, щоб посіяти плутанину та розбрат, включаючи синтетичне відео , на якому він закликає війська негайно скласти зброю та здатися російським силам; транслювання вигаданої екстреної трансляції президента росії путіна про оголошення воєнного стану через вторгнення українських військ на територію росії тощо;

- *створення дисбалансів у внутрішній політиці країни* - для створення оманливого політичного контенту, переважно щодо політичних виборів. Приклади: уряд Венесуели публікував фейкові новини, в яких фігурували диктори новин, створені штучним інтелектом, у рамках широкомасштабної внутрішньої пропагандистської кампанії впливу на своїх громадян; проізраїльські і пропалестинські облікові записи в соціальних мережах поширювали синтетичні зображення триваючого конфлікту в Газі; синтетичні засоби масової інформації про політиків були фальшиво представлені як автентичні, включно з відео прем'єр-міністра Великобританії Кіра Стармера, президента США Джо Байдена, кандидата в президенти Словаччини; Британські ультраправі актори та політики широко поширювали антимігрантський та ісламофобний синтетичний контент у соціальних мережах напередодні виборів 2024 року тощо;

- *фінансові злочини* з підтримкою штучного інтелекту швидко стали одним із найпоширеніших зловживань синтетичними носіями. Злочинці використовували генеративні інструменти штучного інтелекту для видавання себе за іншу особу, вимагання та зламу для здійснення

шахрайських дій, персоналізовані фішингові електронні листи штучного інтелекту та шахрайство з голосовими телефонами. Приклади: Голову британської енергетичної компанії особисто обманом змусили перевести майже 250 000 доларів шахраї, які використовували клонування голосу, щоб видати себе за генерального директора материнської компанії; згенеровані штучним інтелектом тисячі синтетичних відео знаменитостей, таких як Ілон Маск і MrBeast, які просувають фальшиві фінансові схеми, широко поширювалися в соціальних мережах.

Одним зі способів запобігання подібним ситуаціям постає маркування синтетичного медіапродукту на основі вимог законодавства, які ще потрібно ухвалити. Хоча обмеження для цифрового контенту вже існують [9], не зовсім зрозуміло, як їх застосовувати у випадку з діпфейк-контентом.

Оскільки вимога маркувати «покращений» цифровий контент не повинна бути вибірковою, вона має стосуватись усіх фото- та відеоматеріалів, завантажених до мережі, враховуючи приватні сторінки: портфоліо акторів, моделей та ін.; фото, відео в соцмережах; анкети на сайтах знайомств тощо. Зважаючи на те, що переважна більшість цифрового матеріалу проходить пост-обробку заради його покращення, подібне маркування не матиме особливого значення (майже все в мережі виявиться маркованим). Окрім цього, порівняно зі «звичним» маркованим цифровим матеріалом, немаркований діпфейк-контент матиме певну перевагу, і викликатиме більше довіри у споживачів, що створить зворотний ефект від вимоги маркування.

У зв'язку з цим актуалізується питання методів швидкої та якісної перевірки інформації. Одним із відносно ефективних варіантів постає автоматизація процесу, наприклад, завдяки алгоритмам на основі нейромереж (GAN-системи), які можуть стати системами автоматичного розпізнавання діпфейк-контенту. Такий спосіб можна назвати *зовнішнім* щодо споживача інформації.

Проте, зважаючи на раніше окреслені пункти, зовнішній інструмент маркування діпфейків, навіть за умови своєї високої ефективності у розпізнаванні підробок і попри перестороги з боку технічних служб, не убезпечить людину від реактивної емоційної поведінки. Іншими словами, певна частина споживачів, перебуваючи під впливом, наприклад, теорії змови щодо контролю інтернет-простору якимись таємними групами, не звертатиме уваги на «сумнівні», з цього погляду, маркування. З іншого боку, подібне маркування дозволило б розв'язати певні колізії, пов'язані зі свободою слова, свободою поширення інформації тощо.

Внутрішніми засобами боротьби з подібним контентом є вивчення законів логіки, а також оволодіння якомога ширшим і глибшим набором знань про дійсність. Обидва засоби можуть бути досягнуті завдяки ґрунтовній освіті, яка включає не лише статичні знання про світ, які необхідно засвоїти, а й вміння знаходити, зіставляти, інтерпретувати, спростовувати інформацію тощо.

Ще одним внутрішнім інструментом у боротьбі з впливом діпфейків на особистість є розвиток критичного мислення. Наприклад, якщо ми бачимо на відео асоціальну поведінку відомого політика, замість питань про ймовірність такої події можна сконцентрувати свою увагу на питаннях, пов'язаних з:

- першоджерелом, де з'явилось відео: «Чому було обрано саме це джерело поширення інформації?»;

- особою автора: «Яка особа його оприлюднила?», «Анонімна чи ні?»;

- з часом його появи: «Чи пов'язана поява цього відео, до прикладу, з піком діяльності його фігуранта?», «Чому саме зараз воно з'явилось?»;

- з особою, яка отримає дивіденди від його появи: «Хто виграє від дискредитації фігуранта скандалу?»;

- звернути увагу на себе як реципієнта – того, хто сприймає інформацію: «Як вона вплине на мене, на мій вибір тощо?»;

- поставити питання про «дивіденди брехуна»: «У чому полягає мета можливої маніпуляції?» [10].

Висновки і перспективи подальших досліджень. Таким чином, поряд із розвитком технології генеративного штучного інтелекту зростає й потенціал її негативного використання: лише за кілька останніх років ландшафт синтетичних медіазагроз кардинально змінився, починаючи від гібридних воєн й закінчуючи фінансовими шахрайствами тощо. Введення синтетичних медіа у зброю не тільки почало завдавати реальної та суттєвої шкоди людям і організаціям у всьому світі, але також загрожує підірвати довіру суспільства до всієї інформації в Інтернеті, що має тривожні наслідки для національної безпеки.

Ці ризики загострюються через те, що широко доступна генеративна технологія штучного інтелекту просунулася настільки, що люди більше не можуть покладатися на свої очі та вуха для надійного виявлення синтетичного вмісту, з яким вони можуть зіткнутися у своєму повсякденному житті.

Тому зацікавлені сторони в приватному та державному секторах повинні працювати разом, щоб впроваджувати багатогранні контрзаходи, які мають поєднувати: технологічну, нормативну та освітню сфери, щоб протистояти зростаючій загрозі, яку становлять синтетичні засоби масової інформації, які стають інформаційною зброєю.

Для протидії ризикам розповсюдження дезінформації від дипфейків в Україні, державі запропоновано:

- розробити нормативно-правову базу з регулювання та розвитку ШІ в Україні, опублікувати узгоджений з суспільством варіант Білої книги, де має бути висвітлено дорожню карту держави до регулювання ШІ, кінцеві результати, які держава планує досягнути на цьому шляху. Враховувати євроінтеграційні прагнення України при визначенні шляху правового

регулювання штучного інтелекту з урахуванням ризиків створення дезінформації за допомогою ШІ;

- опублікувати рекомендації щодо використання інструментів ШІ громадянським суспільством та бізнесом;

- створювати та пропагувати серед населення інклюзивні освітні матеріали щодо розпізнавання та боротьби з дезінформацією, яка створюється й поширюється за допомогою ШІ.

Отже, боротьба проти дезінформації та інших ризиків від використання дипфейків для національної безпеки, а також врегулювання створення та розвитку ШІ є одними з головних стратегічних завдань держави. Розробка та прийняття відповідного законодавства, яке регулюватиме створення та використання ШІ в Україні, а також поширення дезінформації за його допомогою разом із просвітою населення у сферах дезінформації та штучного інтелекту будуть дієвим комплексним підходом, що гарантуватиме гармонійний розвиток суспільства в епоху ШІ.

Напрямом подальших досліджень у цієї сфері є аналіз способів поширення діпфейк-контенту з метою визначення можливостей потенційного захисту, а також специфіки його впливу у складних, мультикультурних середовищах з неоднорідними групами людей, які по-різному (залежно від способу життя, світогляду, віро-сповідання тощо) можуть реагувати на один і той самий контент.

Література

1. Вальорска М. А. Діпфейк та дезінформація : практ. посіб. / А.М. Вальорска; пер. з нім. В. Олійника. Київ: Академія української преси; Центр Вільної Преси, 2020. 36 с.

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. URL: http://npnuola.onua.edu.ua/index.php/1234/article/view/693/757 (дата звернення 20.03.2025).

3. Денисов С. Ф., Філей Ю. В. Порнографічні фейки: проблеми протидії. *Вісник пенетенціарної асоціації України.* 2020. № 2(12). С. 94–102.

4. Подобний О. О., Слатвінська В.М. Основні завдання інформатизації правоохоронної діяльності. *Юридичний науковий електронний журнал.* 2021. № 9. С. 180–182.

5. 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.

6. Юртаєва К.В. Кримінологічний аналіз використання технології Deepfake: коли фейк стає злочином. *Вісник кримінологічної асоціації України.* 2021. № 1(24). С. 31–42.

7. Bahar M., Afsana Sh. Deep Insights of Deepfake Technology: A Review. 2021. URL: https://arxiv.org/abs/2105.00192 (дата звернення 20.03.2025).

8. Chesney B., Citron D. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*. 2019. Vol. 107. P. 1753. P. 1777-1785. https://doi.org/10.15779/Z38RV0D15J.

9. Gray J. Norway passes law requiring influencers to label retouched photos on social media. URL: https://www.dpreview.com/news/1157704583/norway-passes-law-requiring influencers-to-label-retouched-photos-on-social-media?utm_source=ixbtcom (дата звернення: 10.03.2025).

10. Феномен безпеки: соціально-гуманітарні виміри / за заг. наук. редакцією В. Мудракова. Хмельницький: ФОП Мельник А.А., 2022. 332 с.

11. Ілля Березенко. Нова зброя дезінформації під назвою DEEPFAKE (Дипфейк). URL: https://mil.in.ua/uk/blogs/nova-zbroya-dezinformatsiyi-pid-nazvoyu-deepfake-dipfejk/(дата звернення 20.03.2025).

12. М. Єсипович. Дипфейк — веселий тренд чи правопорушення? URL: https://ain.ua/2023/10/06/deepfake-veselyj-trend-chy-pravoporushennya/ (дата звернення 20.03.2025).

13. О. Петрів. Дезінформація та штучний інтелект: (не)видима загроза сучасності. URL: https://cedem.org.ua/analytics/dezinformatsiya-shtuchnyi-intelekt/ (дата звернення 20.03.2025).

14. Біла книга з регулювання ШІ в Україні: бачення Мінцифри. URL: https://thedigital.gov.ua/storage/uploads/files/page/community/docs/%D0%A0%D0%B5%D0%B3%D1%83%D0%BB%D1%8E%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F%20%D0%A8%D0%86.pdf

15. Declaration for the Future of Internet 2022. URL: https://digital-strategy.ec.europa.eu/en/library/declaration-future-internet (дата звернення 20.03.2025).

16. Seferbekov D. S., Datta A., Islam M., Amin M. Towards Solving the DeepFake Problem : An Analysis on Improving DeepFake Detection using Dynamic Face Augmentation. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW).* Montreal, BC, Canada, 2021. P. 3769-3778.

17. Karthik P. C., Sanjana S., M. P. Adithya Vijayan, Thushara P., Wilson A. Review of Deepfake Detection Techniques. *International Journal of Engineering Research &Technology.* 2021. Vol. 10, Issue 05. P. 813-816. URL: https://www.ijert.org/research/review-of-deepfake-detection-techniques-IJERTV10IS050425.pdf (дата звернення 20.03.2025).

References

1. Val′orska M. A. (2020). Dipfeyk ta dezinformatsiya [Deepfake and disinformation]: prakt. posib. / A. M. Val′orska; per. z nim. V. Oliynyka. Kyiv: Akademiya ukrayins′koyi presy; Tsentr Vil′noyi Presy, 36 p. [in Ukrainian].

2. Vasylenko M.D., Rachuk V.O., Slatvins′ka V.M. (2021). Shkidlyvi prohramy v konteksti rozuminnya komp″yuternoyi virusolohiyi ta tekhniko-pravovoyi zmahal′nosti: mizhdystsyplinarne doslidzhennya [Malicious programs in the context of understanding computer virology and technical and legal adversarialism: an interdisciplinary study]. *Naukovi pratsi Natsional'noho universytetu «Odes'ka yurydychna akademiya».* T. 28. P. 28–36. URL: http://npnuola.onua.edu.ua/index.php/1234/article/view/693/757 [in Ukrainian].

3. Denysov S. F., Filey YU. V. (2020). Pornohrafichni feyky: problemy protydiyi [Pornographic fakes: problems of counteraction]. *Visnyk penetentsiarnoyi asotsiatsiyi Ukrayiny.* № 2(12). P. 94–102 [in Ukrainian].

4. Podobnyy O. O., Slatvins′ka V.M. (2021). Osnovni zavdannya informatyzatsiyi pravookhoronnoyi diyal′nosti [The main tasks of informatization of law enforcement activities]. *Yurydychnyy naukovyy elektronnyy zhurnal.* № 9. P. 180–182 [in Ukrainian].

5. Podobnyy O.O., Tishchenko V.V. (2018). Dezinformatsiya. Velyka ukrayins′ka yurydychna entsyklopediya: u 20 t. [Disinformation. Great Ukrainian Legal Encyclopedia: in 20 volumes.]. Kharkiv: Pravo, 2016. T. 20. Kryminalistyka, sudova ekspertyza, yurydychna psykholohiya / redkol. V. YU. Shepit′ko (holova) ta in.; Nats. akad. prav. Nauk Ukrayiny; In-t derzhavy i prava im. V. M. Korets′koho NAN Ukrayiny; Nats. yuryd. tsn.-t im. Yaroslava Mudroho. P. 158 [in Ukrainian].

6. Yurtayeva K.V. (2021). Kryminolohichnyy analiz vykorystannya tekhnolohiyi Deepfake: koly feyk staye zlochynom [Criminological analysis of the use of Deepfake technology: when a fake becomes a crime]. *Visnyk kryminolohichnoyi asotsiatsiyi Ukrayiny.* № 1(24). P. 31–42 [in Ukrainian].

7. Bahar M., Afsana Sh. (2021). Deep Insights of Deepfake Technology: A Review. URL: https://arxiv.org/abs/2105.00192.

8. Chesney B., Citron D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*. Vol. 107. P. 1753. P. 1777-1785. https://doi.org/10.15779/Z38RV0D15J.

9. Gray J. Norway passes law requiring influencers to label retouched photos on social media. URL: https://www.dpreview.com/news/1157704583/norway-passes-law-requiring influencers-to-label-retouched-photos-on-social-media?utm_source=ixbtcom.

10. Fenomen bezpeky: sotsial′no-humanitarni vymiry [The phenomenon of security: social and humanitarian dimensions] / za zah. nauk. redaktsiyeyu V. Mudrakova. Khmel′nyts′kyy: FOP Mel′nyk A.A., 2022. 332 p. [in Ukrainian].

11. I. Berezenko. (2022). Nova zbroya dezinformatsiyi pid nazvoyu DEEPFAKE (Dypfeyk) [A new weapon of disinformation called DEEPFAKE]. URL: https://mil.in.ua/uk/blogs/nova-zbroya-dezinformatsiyi-pid-nazvoyu-deepfake-dipfejk/ [in Ukrainian].

12. M. Yesypovych. (2023). Dypfeyk – veselyy trend chy pravoporushennya? [Deepfake – a fun trend or a crime?]. URL: https://ain.ua/2023/10/06/deepfake-veselyj-trend-chy-pravoporushennya/ [in Ukrainian].

13. Petriv O. Dezinformatsiya ta shtuchnyy intelekt: (ne)vydyma zahroza suchasnosti [Disinformation and artificial intelligence: the (in)visible threat of modernity]. URL: https://cedem.org.ua/analytics/dezinformatsiya-shtuchnyi-intelekt/ [in Ukrainian].

14. Bila knyha z rehulyuvannya SHI v Ukrayini: bachennya Mintsyfry. [White Paper on AI Regulation in Ukraine: Vision of the Ministry of Digital Economy].URL: https://thedigital.gov.ua/storage/uploads/files/page/community/docs/%D0%A0%D0%B5%D0%B3%D1%83%D0%BB%D1%8E%D0%B2%D

0%B0%D0%BD%D0%BD%D1%8F%20%D0%A8%D0%86.pdf [in Ukrainian].

15. Declaration for the Future of Internet 2022. URL: https://digital-strategy.ec.europa.eu/en/library/declaration-future-internet.

16. Seferbekov D. S., Datta A., Islam M., Amin M., (2021). Towards Solving the DeepFake Problem: An Analysis on Improving DeepFake Detection using Dynamic Face Augmentation//IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada. P. 3769-3778.

17. Karthik P. C., Sanjana S., M. P. Adithya Vijayan, Thushara P., Wilson A. (2021). Review of Deepfake Detection Techniques. *International Journal of Engineering Research &Technology*. Vol. 10, Issue 05. P. 813-816. URL: https://www.ijert.org/research/review-of-deepfake-detection-techniques-IJERTV10IS050425.pdf.