

DOI: 10.46340/eppd.2020.7.5.21

**Lina Vezhel, PhD in Philology**

ORCID ID: <https://orcid.org/0002-4025-5230>

*Taras Shevchenko National University of Kyiv, Ukraine*

## **THE STATE OF SYNTHETIC REALITIES: ASPECTS OF BRAND SAFETY**

**Ліна Вежел, к. філол. н.**

*Київський національний університет імені Тараса Шевченка, Україна*

### **СТАН СИНТЕТИЧНИХ РЕАЛЬНОСТЕЙ: БЕЗПЕКОВІ АСПЕКТИ БРЕНДУ**

Artificial intelligence algorithms can produce realistic “deepfake” videos, as well as authentic-looking fake photos and writing. Collectively called synthetic media, these tools have triggered widespread concern about their potential in spreading political disinformation. Yet the same technology can also facilitate financial and reputational harm.

The article focuses on research of the developments in AI and machine learning (ML), which are concurrently taking place with giant leaps in immersive experiences technologies, better known as augmented reality (AR) and virtual reality (VR). The marriage of these technologies with Deep Fakes might transform our understanding of reality and presence.

**Keywords:** synthetic media, synthetic reality, augmented reality (AR), virtual reality (VR), deep fake, brand safety.

**Постановка проблеми.** Сучасні технологічні рішення концентрованого опрацювання даних, розвиток та домінування систем штучного інтелекту, розшифровуючи і прогножуючи людські уподобання, здатні створювати у режимі *реального часу* захоплюючі паралельні світи, що визначають те, як ми сприймаємо, розуміємо навколишній світ, як на нього реагуємо і як взаємодіємо один з одним.

Реальності – AR, VR, MR – якісно *стирають кордони між істиною і брехнею, свободою і примусом, війною і миром*. Відбувається це не просто в проблематиках пізнання, а передусім з позиції *життєвого буття, тобто вони кидають виклик нашому загальному сприйняттю традиційних життєвих процесів*<sup>1</sup>.

**Аналіз останніх досліджень і публікацій.** Одними з перших дослідників поняття «віртуальна реальність» були : Ж. Бодріяр, А. Бюль, М. Вейнстейн, Т. Зімерманн, А. Крокер, М. Крюгер, Дж. Ланьє, М. Маклюєн, А. Сазерленд, Ш. Тьоркл, М. Фуко та інші. Дослідженню загальних питань застосування AR- технологій присвячені наукові праці Т. Коделля, П. Мілгрема і Фуміо Кісіно, А. Петланда, Л. Мановича тощо. Натомість наукові розвідки з проблем безпеки бренду в умовах синтетичних реальностей фактично відсутні. Мета статті – проаналізувати феномен Deepfakes у процесі створення синтетичних медіа і описати ініціативи з розробки засобів їх виявлення; представити авторський підхід до поняття «синтетична реальність»

**Виклад основного матеріалу.** Синтетичні всесвіти стають новим середовищем життя: вірусні відеоігри, такі як Fortnite, Animal Crossing та інші об'єднують мільйони людей і перетворюються на платформи, що копіюють правила реального світу і формують нові.

Згідно з дослідженням компанії з фінансової освіти LendEDU, більше 1000 гравців Fortnite проводять у грі від 6 до 10 годин на тиждень. Для порівняння, середній активний користувач Snapchat або Instagram витрачає приблизно 30 хвилин щодня на платформі<sup>2</sup>.

Наприкінці квітня цього року концерт Тревіса Скотта у Fortnite зібрав понад 27 мільйонів геймерів. Ця подія не могла залишитися непомітною для стратегів політичних кампаній. Ліз Сміт,

<sup>1</sup> Вежел, Л. (2020). *Стан синтетичних реальностей програмованої епохи: можливості, перспективи чи новий виклик національній безпеці*. <<https://ccs-center.com.ua/?p=82>>.

<sup>2</sup> Reisinger, D. (2018). Here's How Much Cash the Average Player Spends on 'Fortnite'. <<https://fortune.com/2018/06/27/fortnite-average-spending/>>.

колишній політичний радник кандидата в президенти США Піта Буттіджіча, запропонувала спроекувати аватар Джо Байдена на Великий каньйон так, як це зробили розробники Fortnite.

Ідея реалізувалася в іншій грі – Джо Байден запустив передвиборчу кампанію в Animal Crossing: New Horizons. Користувачі отримали можливість прикрасити свою прибудинкову територію плакатами та іншою символікою кандидата від Демократичної партії.

Нещодавно Epic Games анонсував нову тритижневу серію концертів на острові Party Royale. Першим виступив репер Домінік Файк в ніч з 12 на 13 вересня. Трансляція відбувалася зі студії в Лос-Анджелесі, яка побудована спеціально для in-game концертів.

Все це частина великого плану з перетворення віртуальної сцени Fortnite у важливе місце для музикантів.

Twitch – один з найпопулярніших майданчиків в світі з аудиторією в 15 млн користувачів. На ній можна подивитися турніри з кіберспорту, а також прямі ефіри геймерів, блогерів та музикантів.

Сервісом активно користуються армія і флот США. У військових є свої команди з кіберспорту, які регулярно запускають на Twitch трансляції з іграми на кшталт Call of Duty або League of Legends. Під час цих стрімів глядачам не тільки показують ігрові навички, а й розповідають про службу в збройних силах, життя військових, а також діляться посиланнями для заповнення анкет.

Пандемія коронавірусу прискорила розвиток AR та VR-технологій. Так, вчитель із Сан-Дієго Чарльз Кумбер успішно провів урок з геометрії для 7 класу у форматі в VR- гри Half-Life: Alyx, а університет штату Арізона, який був в авангарді переходу до онлайн-навчання, активно співпрацює з VR-стартапом Dreamscape і розглядає це партнерство як наступний крок в довгостроковому прагненні використовувати технології для фундаментальної зміни навчального процесу.

AR (Augmented reality) застосовують для покращення і доповнення природного середовища за допомогою так званих «цифрових компонентів». Ці компоненти переплітаються з фізичним (реальним) світом, щоб створити імерсивний, інтерактивний і подібний до реальності досвід. У свою чергу, віртуальна реальність йде ще далі, оскільки середовище, з якою стикається користувач, є повністю штучним. Іншими словами, весь віртуальний світ – це єдине цифрове виробництво. Ключовим моментом розуміння AR / VR як технології, що за своєю суттю є імерсивною, це надання користувачам відчуття реальної «присутності» в напівприродних просторі і часі.

Як і слід було очікувати, святим Граалем у розвитку імерсивних технологій є створення повністю реалістичних цифрових компонентів, які зливаються з фізичним світом або замінюють у такий спосіб, що ми сприймаємо його як природний, кидаючи виклик нашому сенсорному сприйняттю на те, що реально, а що ні. І саме тут з'являється штучний інтелект<sup>1</sup>.

В останні роки «синтетичні медіа» стали універсальним терміном, що використовується для опису відео, зображень, тексту і голосу, повністю або частково створених системами штучного інтелекту (ШІ). Науковці прогнозують серйозні зрушення парадигми у продукуванні і споживанні контенту, яке, ймовірно, докорінно трансформує медійну галузь.

Вплив цієї нової категорії технологій на суспільство був гарячою темою в пресі протягом останніх двох років. Можливість створювати візуальні ефекти голлівудського рівня або навіть краще, не вимагаючи при цьому бюджету, часу чи навичок, які традиційно необхідні, відкриває можливості як для позитивних, так і негативних змін.

Сервіс «Смотри Mail.ru» розробив платформу з цифровими ведучими для виробництва новин і репортажів у телевізійному форматі.

Щоб зробити новинний відеоролик, потрібно завантажити текст (його прочитає ведучий) і, якщо необхідно, зображення, музику, титри і посилання. Користувач може вибрати зовнішність диктора із запропонованих варіантів і оформлення студії (наприклад, в кольорах компанії). Мова і міміка реалістичні, наприклад, читаючи новини, вони розставляють смислові акценти і відтворюють емоції. Доступ до сервісу безкоштовний та відкритий.

Нейронні мережі можуть відтворювати медіаконтент з такою високою точністю, що його майже неможливо відрізнити від контенту, створеного людьми.

Мовна модель третього покоління GPT-3, розроблена компанією OpenAI, написала колонку для британської газети The Guardian про те, чому роботи не є небезпечними і не збираються захоплювати людство. Нейромережа згенерувала вісім варіантів із різними аргументами, а редактор вибрав

<sup>1</sup> Pasquarelli, W. (2019). Towards Synthetic Reality: When DeepFakes meet AR/VR.

<<https://www.oxfordinsights.com/insights/2019/8/6/towards-synthetic-reality-when-deepfakes-meet-arvr>>.

«найкращі» частини з усіх версій і зробив з них один текст. Статтю під назвою «Увесь цей текст написав робот. Тобі ще не страшно, людино?» опублікували на сайті видання<sup>1</sup>.

Компанія Microsoft вже повідомила про угоду, яка надасть їй ексклюзивну ліцензію на мовну модель OpenAI GPT-3, а експерти наголошують, що світ на порозі нової економіки штучного інтелекту.

З появою наприкінці 2017 року такого явища як deepfakes у наукових статтях все частіше використовується словосполучення «синтетична реальність» – «оточуючий світ», що формується за допомогою даних, синтезованих комбінаціями людського і штучного інтелекту (ШІ).

На нашу думку, синтетичною реальністю можна вважати реальність, що моделюється за допомогою синтетичних медіа – аудіовізуальних, текстових носіїв, згенерованих системами ШІ, для створення істотно нового інформаційного артефакту.

Агентство інтернет-досліджень (IRA) створило псевдоресурс нібито лівого спрямування – Pease Data. Для цього сайту наймалися фрілансери, які писали матеріали різної тематики: расизм, екологія, політика президента Дональда Трампа, або копіювали їх з інших ресурсів. Проаналізувавши фото профілів редакторів, аналітики з'ясували, що зображення були створені за допомогою технології deepfakes.

Існують також підозри щодо спроб шпигувати за допомогою генерованих на комп'ютері фото акаунтів, наприклад профіль «Katie Jones» на LinkedIn нібито дослідниці американського аналітичного центру.

Перш ніж експертний аналіз виявив декілька візуальних аномалій, які вказували на те, що зображення є синтетичним, профілеві вдалося з'єднатися з 52 політичними діячами у Вашингтоні, включаючи заступника помічника державного секретаря, високопоставленого радника одного із сенаторів та авторитетного економіста<sup>2</sup>.

Феномен дипфейків, що використовують різні базові технології штучного інтелекту, зокрема автокодери і генеративні змагальні мережі (GAN), є революцією в процесі створення синтетичних медіа.

GAN, що вимагають високої продуктивності обробки і графічних карт, а також інші технології deepfake на основі машинного навчання ще не досягли досить високої виробничої потужності, щоб стати суспільним надбанням. Але cheap fakes (дешеві фейки) – розважальні програми для зміни обличчя на смартфонах, елементарні методи прискорення і уповільнення з метою дезінформації та онлайн-шахрайства, використовуються в достатній кількості, щоб викликати занепокоєння. Також зростає кількість продукції на носіях deepfake voice і deepfake, які стає все важче відрізнити від оригіналів.

У розпорядженні звичайних користувачів на платформі GitHub є досить багато програмних інструментів, в тому числі FakeApp, DFaker, faceswap-GAN, faceswap і DeepFaceLab, тому неважко уявити, що цю технологію можна використовувати в політичних кампаніях та інших значущих соціальних заходах.

7 лютого 2020 року, за день до виборів в Законодавчі збори в Делі, два відео президента партії «Бхаратія Джаната» (BJP) Маноджа Тіварі стали вірусними в WhatsApp. Індійський політик використав технологію глибинних фейків для створення роликів, де він говорив на місцевому діалекті а також англійською мовою. Це був перший випадок в світі, коли політична партія використовувала дипфейк у своїй передвиборчій кампанії<sup>3</sup>.

У США компанія Market Research Media, яка проводить маркетингові дослідження для уряду, Конгресу, великих корпоративних компаній та інвесторів, опублікувала у кінці лютого звіт під назвою «Аналіз ринку технології Deepfakes», який був розміщений на платформі Marketanalysis.com. Можливості використання deepfakes у маркетингу перераховані у трьох розділах. У розділі «Розваги» зазначається, що за допомогою синтетичних людських голосів і образів, які майже не відрізняються від оригіналу, можна створювати переконливі історії, навіть якщо в фільмах відсутні творчі ефекти і реальні персонажі. У розділі "Електронна комерція" мова йде про те, що дипфейки здатні надати споживачам розширений персоналізований досвід при виборі продуктів. Під заголовком

<sup>1</sup> A robot wrote this entire article. Are you scared yet, human? (2020).

<<https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>>.

<sup>2</sup> Вальорска, М. Агнешка. (2020). Дипфейк та дезінформація

<[https://www.aup.com.ua/uploads/DEEPFAKES\\_FNF\\_AUP\\_2020.pdf?fbclid=IwAR19haHA3JOUnoXCBBRITD-61VKBkZ4AUJ5XrHUmBTVkRkQqhDJ\\_MSygdWw](https://www.aup.com.ua/uploads/DEEPFAKES_FNF_AUP_2020.pdf?fbclid=IwAR19haHA3JOUnoXCBBRITD-61VKBkZ4AUJ5XrHUmBTVkRkQqhDJ_MSygdWw)>.

<sup>3</sup> Kim, Lyons (2020). An Indian politician used AI to translate his speech into other languages to reach more voters.

<<https://www.theverge.com/2020/2/18/21142782/india-politician-deepfakes-ai-elections>>.

«Комунікація» увага звертається на можливість отримувати вигоду з deepfake обличчя бренду, які можуть говорити будь-якою мовою за допомогою синтезу голосу і маніпуляцій з обличчям<sup>1</sup>.

Як відомо, виробництво відео – дуже витратний процес з орендою камер, студій та оплатою роботи акторів. Після того, як рекламний ролик про товар або продукт знято, його дуже складно швидко і за доступною ціною відредагувати або перекласти на інші мови. Але замість того, щоб знімати нове відео, можна змонтувати одне і за допомогою deepfake-технології отримати різні версії.

Завдяки співпраці компаній Ridley Scott Associates і Synthesia, у соціальному роликів Девід Бекхем заговорив на 9 мовах.

Клієнтами Synthesia, стартапу, який є одним з лідерів Catalogue Deepfake Sales, стали такі світові бренди, як Accenture, McCann Worldgroup, Dallas Mavericks і Axiata Group.

Alethea.ai запустила перший торговий майданчик для дїпфейків на основі технології блокчейну для фіксації записів на право власності та дозволів на використання, а також з гарантією того, що творці володіють правами на програмне забезпечення штучного інтелекту, яке вони застосовують. Платформа працює на власному токени Alethea, який покликаний стимулювати різні взаємодії між гравцями екосистеми.

Рекламні агентства, продюсерські компанії, або будь-яка особа чи установа, які хочуть створити дїпфейк-відео, можуть вибрати профілі, що підходять для їхнього сценарію, з цифрового каталогу на ресурсі Alethea.ai.

Баскетболіст Дем'єн Ліллард через коронавірус не зміг знятися в рекламі Hulu, тоді сервіс навчив нейромережу на його відеозаписах з Zoom і наклав обличчя спортсмена на дублера<sup>2</sup>.

Hulu – не перша компанія, яка під час пандемії використовує маніпуляції зі ШІ у своїх рекламних роликах.

State Farm, одна з найбільших страхових груп в США, запустила рекламу для промоції документального фільму «Останній танець», в якому були представлені професійно оброблені кадри за участю ведучого ESPN Sports Center Кенні Мейна<sup>3</sup>.

Технології штучного інтелекту вже успішно застосовують у *документалістиці*. У фільмі «Ласкаво просимо в Чечню» замість темних силуетів і зацензурованих обличчя з неприродно низькими голосами творці використовують технологію deepfake, щоб не наражати на небезпеку чеченців-геїв і лесбіянок, які зазнали жорстоких гонінь та переслідувань.

Минулого року фільм «In Event Of Moon Disaster» виграв спеціальний приз журі Creative Technology на конкурсі цифрового сторітейлінгу від IDFA DocLab. Ще одна нагорода була отримана від Mozilla Creative Media Awards.

Перший відзначений нагородами дїпфейк-фільм, у якому український стартап Respeecher «змусив» президента Ніксона зачитати промову про провалення місії «Аполлон-11», з'явиться ще на трьох кінофестивалях у цьому році.

Існує тонка межа між дїпфейками для розваг, реклами, кіно і продукування мемів та їх використанням для створення компрометуючих фото і відео, подробиці голосу або поширення дезінформації.

На сьогодні ініціативи у сфері кібербезпеки з розробки заходів проти дїпфейків і рішень для них реалізуються за двома напрямками: під керівництвом і за підтримки світових онлайн-гігантів (зокрема, Facebook і Google, Microsoft) науково-дослідні інститути розробляють моделі виявлення синтетичних зображення і звуків. З іншого боку, установи, що займаються питаннями безпеки, намагаються створити персоналізовану систему кіберзахисту, орієнтовану на конкретні цілі, за допомогою біометричних відео і аудіо, щоб усунути корпоративні ризики від дїпфейків.

Microsoft представила дві технології, аутентифікація відео Video Authenticator, який здатний аналізувати фото чи відео для визначення ймовірності та оцінки довіри до того, чи медіафайли не були штучно змінені. У випадку відео дана технологія може показувати відсоток ймовірності справжності файлу у реальному часі для кожного кадру під час відтворення відео<sup>4</sup>.

<sup>1</sup> Deepfake Technology, Market Analysis (2020). <<https://marketanalysis.com/deepfake-technology-market-analysis/#:~:text=Emergence%20of%20deepfake%20technologies%20has,government%20and%20media%20market%20segments.&text=Deepfakes%20rely%20on%20artificial%20neural,that%20recognize%20patterns%20in%20data>>.

<sup>2</sup> Sutton, K. (2020). Hulu Deepfakes Football Stars' Faces Onto Body Doubles in New Sellouts Ad. <<https://www.adweek.com/tv-video/hulu-deepfakes-football-stars-faces-onto-body-doubles-in-new-sellouts-ad/>>.

<sup>3</sup> Hsu, T. (2020). An ESPN Commercial Hints at Advertising's Deepfake Future. <<https://www.nytimes.com/2020/04/22/business/media/espn-kenny-mayne-state-farm-commercial.html>>.

<sup>4</sup> Burt, T. (2020). New Steps to Combat Disinformation. <<https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/>>.

Інша технологія допоможе, як визначити контент, який використовується в маніпулятивних цілях, так і переконати людей у справжності переглянутих медіафайлів. Вона складається з 2 компонентів: перший – це інструмент, вбудований у Microsoft Azure, а інший – зчитувач.

Агентство перспективних досліджень в галузі оборони США (DARPA) створило програму криміналістичної експертизи. Та ж технологія GAN, що генерує синтетичні зображення, тепер використовується і для їх виявлення.

А компанія Adobe у співпраці з ученими Каліфорнійського університету в Берклі розробила інструмент на основі машинного навчання, який автоматично виявляє маніпуляції, що були зроблені на фотографії за допомогою Liquify (функції Photoshop, яка використовується для коригування форми і зміни виразу обличчя).

З розвитком і вдосконаленням технологій дипфейків, які стає все важче виявляти, необхідно впроваджувати протоколи захисту, що містять кілька елементів управління і аутентифікації. Це означає «сканування» багаторівневої безпеки на декількох пристроях для перевірки того, чи є аудіо і зображення, отримані від цифрових інструментів і онлайн-каналів, синтетичними або підробленими.

**Висновки.** Термін «синтетичні медіа» може використовуватися для опису широкого діапазону контенту, створеного, повністю або частково, з використанням машинного навчання чи інших автоматизованих засобів. Такий контент охоплює в першу чергу відео, аудіо, зображення і текст, але він також поширюється і на цифрові об'єкти (наприклад, віртуальні інфлуенсери), доповнену реальність і повністю імерсивні віртуальні середовища. Це, безумовно, призведе до стимулювання творчості та демократизації редагування і створення різноформатного контенту за допомогою систем штучного інтелекту. Однак атрибути цього ж процесу перетворюють синтетичні медіа на потенційну загрозу безпеці, дозволяючи маніпулювати і вводити в оману людей, що викликає серйозну стурбованість у світі, сповненому фейковими новинами та дезінформацією. Не менш важливо, що поява штучних засобів інформації змушує переосмислити права інтелектуальної власності, право на образ та голос людини, а також інші способи монетизації такого контенту і запобігання неправомірного його використання.

#### References:

1. Vezhel, L. (2020). Stan syntetychnykh realnostei proqramovanoi epokhy:mozhlyvosti, perspektyvy chy novyi vyklyk natsionalnoi bezpetsi [The state of synthetic realities of the programmable age: opportunities, prospects and a new challenge to national security]. *CCS-center*. <<https://ccs-center.com.ua/?p=82>>. [in Ukrainian].
2. A robot wrote this entire article. Are you scared yet, human?. *The Guardian*. <<https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>>. [in English].
3. Burt, T. (2020). New Steps to Combat Disinformation. <<https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/>>. [in English].
4. Deepfake Technology, Market Analysis (2020). <<https://marketanalysis.com/deepfake-technology-market-analysis/#:~:text=Emergence%20of%20deepfake%20technologies%20has,government%20and%20media%20market%20segments.&text=Deepfakes%20rely%20on%20artificial%20neural,that%20recognize%20patterns%20in%20data>>. [in English].
5. Hsu, T. (2020). An ESPN Commercial Hints at Advertising's Deepfake Future. <<https://www.nytimes.com/2020/04/22/business/media/espn-kenny-mayne-state-farm-commercial.html>>. [in English].
6. Kim, Lyons. (2020). An Indian politician used AI to translate his speech into other languages to reach more voters <<https://www.theverge.com/2020/2/18/21142782/india-politician-deepfakes-ai-elections>>. [in English].
7. Pasquarelli, W. (2019). Towards Synthetic Reality: When DeepFakes meet AR/VR. <<https://www.oxfordinsights.com/insights/2019/8/6/towards-synthetic-reality-when-deepfakes-meet-arvr>>. [in English].
8. Reisinger, D. (2018). Here's How Much Cash the Average Player Spends on 'Fortnite'. <<https://fortune.com/2018/06/27/fortnite-average-spending/>>. [in English].
9. Sutton, K. Hulu Deepfakes Football Stars' Faces Onto Body Doubles in New Sellouts Ad. <<https://www.adweek.com/tv-video/hulu-deepfakes-football-stars-faces-onto-body-doubles-in-new-sellouts-ad/>>. [in English].
10. Valorska, M. Agnieszka (2020). Dipfeik ta dezinformatsiia [Deepfake and misinformation]. <[https://www.aup.com.ua/uploads/DEEPFAKES\\_FNF\\_AUP\\_2020.pdf?fbclid=IwAR19haHA3JOUoXCBBRITD-61VKBkZ4AUJ5XrHUmBTvkRkQqhDJ\\_MSygdWw](https://www.aup.com.ua/uploads/DEEPFAKES_FNF_AUP_2020.pdf?fbclid=IwAR19haHA3JOUoXCBBRITD-61VKBkZ4AUJ5XrHUmBTvkRkQqhDJ_MSygdWw)>. [in Ukrainian].