

ЗА ЗАГАЛЬНОЮ НАУКОВОЮ РЕДАКЦІЄЮ
ВІТАЛІЯ МУДРАКОВА

ФЕНОМЕН БЕЗПЕКИ

СОЦІАЛЬНО-
ГУМАНІТАРНІ
ВИМІРИ

КОЛЕКТИВНА МОНОГРАФІЯ

ФЕНОМЕН БЕЗПЕКИ:
соціально-гуманітарні виміри

За загальною науковою редакцією
Віталія Мудракова

колективна монографія

УДК: 316.47/.48-049.5

ББК: 60.524.1/.2

Ф 42

Рекомендовано до друку Вченою радою
Хмельницького національного університету
(протокол №16 від 28.04.2022 р.)

Рецензенти:

Василь БАЛУХ

доктор історичних наук, професор, член-кореспондент
НАПН України, перший проректор Чернівецького національ-
ного університету імені Юрія Федьковича

Дмитро ШЕВЧУК

доктор філософських наук, професор, проректор з науково-
педагогічної роботи Національного університету «Острозька
академія»

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

ISBN 978-617-7600-61-8

У монографії порушуються проблеми, що у різній мірі стосуються феномену
безпеки; осмислюється своєрідність безпекових аспектів сьогодення, їх соціоку-
льтурні та освітні виміри; концептуалізується поняття «гуманітарна безпека».

Монографія буде цікава всім, хто вже працює або планує працювати у безпеко-
вій сфері різних рівнів. Зокрема представлені позиції міститимуть інтерес для
аналітиків ідеологічних царин, медіаторів, тренерів із соціальних комунікацій,
представників організацій гуманітарного та соціально-політичного спряму-
вання, політиків, академічної спільноти, а також студентів закладів вищої освіти,
спеціальності яких близькі до соціально-гуманітарної сфери.

ISBN 978-617-7600-61-8

УДК: 316.47/.48-049.5

ББК: 60.524.1/.2

Ф 42

| | |
|---|-----------|
| Передмова..... | 7 |
| Вступ..... | 9 |
| Розділ 1. ХАРАКТЕРОЛОГІЧНІ АСПЕКТИ..... | 13 |
| <i>Бородіна Наталія</i> | |
| § 1.1. Феномен безпеки у контексті толерування жорстокості.... | 15 |
| а) Безпека та права людини: інституційні та наукові інтенції..... | 16 |
| б) Безпека та права людини: український контекст усвідомлення..... | 19 |
| в) Безпека і толерування жорстокості..... | 21 |
| <i>Гончаренко Катерина</i> | |
| § 1.2. Специфіка (<i>не</i>)безпеки «суспільства ризику»..... | 29 |
| а) Вступні зауваги й теоретичні розробки..... | 30 |
| б) Постмодерністська специфіка інтерпретацій..... | 33 |
| в) Анатомія «небезпечної» безпеки..... | 37 |
| <i>Адольфі Райнер</i> | |
| § 1.3. Епоха нових ідентичностей: потреба в безпеці та боротьба за справжню історію..... | 43 |
| а) Утримання в павутинні історії (Проблемний тиск великих потрясінь)..... | 44 |
| б) Робота над історичною свідомістю..... | 46 |
| в) Конструкції ідентичності (Образ думки в розумінні ідентичності)..... | 49 |
| г) Чому безпека не є антропологічною константою (Потреби ідентичності й пропаганда ідентичності та їх приховані справжні причини)..... | 53 |
| <i>Руснак Ігор, Рупташ Ольга</i> | |
| § 1.4. Світоглядні виміри безпеки та культуротворчий потенціал технології deepfake..... | 57 |
| а) Діпфейк як феномен інформаційної доби..... | 58 |
| б) Діпфейк й епістемологічні проблеми..... | 61 |
| в) Етичні проблеми діпфейк-технології та засоби їх запобігання..... | 71 |
| г) Мистецтво та культура як контексти використання діпфейк-технології..... | 79 |
| <i>Зінченко Віктор</i> | |
| § 1.5. Тенденції та перспективи української інституалізації націст- ворення та державотворчості: особливості безпекових конс- тант..... | 85 |
| а) Історичні віхи контексту соціально-стратифікаційних процесів та феномен еліт...86 | |
| б) Національна ідея, деструктивні міфологізації та пошук форм єдності.....90 | |
| в) Небезпеки ідеологізації.....99 | |
| г) Від засновків національної культури до громадянської демократії націстворчості...102 | |

Розділ 2. СОЦІОКУЛЬТУРНІ ВИМІРИ.....109

Ганаба Світлана

- § 2.1. Історична пам'ять як чинник примирення та консолідації суспільства: безпекові конотації.....111
- а) Фактор досвіду минулого в консолідації суспільства.....112
 - б) Буттєва характерологічність історичної пам'яті.....113
 - в) Символічне кодування міського простору: випадок Хмельницького.....117
 - г) Висновкові зауваги.....124

Гапченко Олена

- § 2.2. Феномен української інтелігенції: ретроспекція модусів самоосмислення між свободою та безпекою.....127
- а) Вступні акценти.....128
 - б) Інтелігенція: культура і свобода.....129
 - в) Просвітницька тактика як стратегія безпеки.....135
 - г) Модель «стилого життя»: ціннісно-сміслові константи як модуси безпеки....140

Бродецький Олександр

- § 2.3. Ментально-ціннісне підґрунтя мовної свідомості й практики сучасних українців: безпекові аспекти.....147
- а) Суржикові колізії: виміри мови і виміри свідомості.....148
 - б) Долання наслідків мовного російщення: безпекові смисли.....152
 - в) Трансформації в мовній сфері: умови безпекової дієвості.....158

Кіріка Діана, Боднар Алла

- § 2.4. Правові важелі інституту громадського самоврядування у реалізації гуманітарної безпеки.....161
- а) Особливості проблеми та категоріального апарату.....162
 - б) Локальний контекст правових важелів.....166
 - в) Деякі аспекти функціональності.....170

Луцан Ігор

- § 2.5. Інститут військового духовенства (капеланської служби) як феномен безпеки: український контекст.....175
- а) Духовна опіка та церковно-релігійна підтримка: роль і місце в умовах війни.....177
 - б) Волонтерство/благодійність як платформа соціальних перетворень.....178
 - в) Організаційно-правові засади військового капеланства.....182
 - г) Питання професійної підготовки.....183

Семенчук Сергій

- § 2.6. Безпека об'єктів археологічної спадщини Хмельниччини на початку XXI ст.187
- а) Теоретичні особливості та регулятивні механізми.....188
 - б) Історіографічна довідка систематизацій.....191
 - в) Сутнісні ознаки та процеси.....192

Розділ 3. ОСВІТНІ КОНТЕКСТИ.....203

Парандій Валентина

- § 3.1. Освіта в системі національної безпеки: міжнародний досвід та українські перспективи.....205
- а) Постановка та характеристика проблеми.....206
 - б) Побудова системи: якісна освіта – національна безпека.....207
 - в) Безпековий вимір освітньої сфери: міжнародний досвід.....212
 - г) Алгоритм побудови майбутнього безпекового середовища.....214

Калінічева Галіна

- § 3.2. Безпековий вимір модернізації вищої освіти України.....217
- а) Актуалізаційні конотації проблеми.....218
 - б) Етапи реформацийних заходів.....219
 - в) Конкретизація інструментальності.....221
 - г) Стратегічні та загально-предметні візії.....226

Петрук Наталія

- § 3.3. Освіта як фактор формування простору гуманітарної безпеки.....231
- а) Формулювання проблеми та контекст.....232
 - б) Безпека: філософія, історія, українознавство, культурологія.....234
 - в) Безпека: комунікація, відповідальність, інтелектуальна інтеграція.....246

Островий Володимир

- § 3.4. Історична правда та історична освіта: чинники національної безпеки України.....255
- а) Контекстні засновки проблеми.....256
 - б) Функціональні характеристики шляхів і засобів.....264

Горохолінська Ірина

- § 3.5. Безпековий потенціал богословської освіти.....277
- а) Релігійний світогляд та питання безпеки: актуальні точки дотику.....278
 - б) Західноєвропейський досвід богословської освіти: безпекові можливості.....281
 - в) Безпековий потенціал богословської освіти в Україні: палітра підходів.....286
 - г) Богословська освіта та богословська просвіта на шляху гуманітарної безпеки...290

Поліщук Олександр, Мураков Віталій

- § 3.6. Самоідентифікація та освіта: до проблеми усвідомлення та впровадження гуманітарної безпеки.....291
- а) Світоглядні особливості та методологічні пріоритети.....292
 - б) Антропологічно-ціннісні аспекти осмислення.....297
 - в) Структурно-формальні шляхи реалізації.....303

Мураков Віталій

- § 3.7. Гуманітарна безпека: від проблем галузевого визначення до особливостей дидактичної визначеності.....309
- а) Самобутність кентавра або межі та своєрідність гуманітарної сфери.....310
 - б) Переоцінка самооцінки: гуманітарна безпека і метастази ідеологій.....316
 - в) Компетентності: поміж виживанням та дидактикою.....326

Ігор Руснак
i.rusnak@chnu.edu.ua
ORCID: 0000-0002-4513-9300

кандидат філософських наук,
асистент кафедри філософії і культурології
Чернівецького національного університету імені Юрія Фе-
дьковича

Ольга Рупташ
o.ruptash@chnu.edu.ua
ORCID: 0000-0001-5398-7898

докторка філософських наук, професорка,
завідувачка кафедри філософії і культурології
Чернівецького національного університету
імені Юрія Федьковича

§ 1.4.

БЕЗПЕКОВІ ВИКЛИКИ ТА КУЛЬТУРОТВОРЧИЙ ПОТЕНЦІАЛ ТЕХНОЛОГІЇ «DEEPFAKE»

а) Діпфейк як феномен інформаційної доби

«Deepfake» визначають як синтетичний медіапродукт, створений штучним інтелектом⁷⁷. «Деер» (з англ. – «глибокий») у словосполученні «deep fake» першопочатково означало заснований на «deep learning» – «глибокому навчанні». Останнє – різновид машинного навчання, яке по суті є нейронною мережею з трьома або більше рівнями. Такі нейромережі намагаються імітувати поведінку людського мозку, дозволяючи «вчитися» з великої кількості даних⁷⁸. Означення «fake» вказує на обман, шахрайство, фальшування дійсності. Цим терміном позначають увесь спектр «гіперреалістичних цифрових фальсифікацій зображень, відео чи аудіо»⁷⁹. Якщо раніше «підробки» створювались на основі синтезу *реальних* зображень, відео чи/та звукового ряду, то застосування штучного інтелекту й програм на його основі продукує їхню імплементацію для створення цифрової *псевдореальності*.

Контент, в межах якого застосовується технологія «deepfake», тісно увійшов в наше життя. Завдяки йому ми пізнаємо світ, поринаємо у вир емоцій тощо. Окрім відвертого спотворення дійсності, такі медіапродукти здатні заповнювати наявні інформаційні лакуни, доповнюючи реальність відповідно до наміру їхнього автора чи замовника. Змістове наповнення саме собою не є феноменом, а передає певну інформацію, яка обробляється реципієнтом і викликає відповідну реакцію. При цьому 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).

Основна форма діпфейк-контенту – відео, створене за допомогою набору реальних даних, яке «відтворює» події, які ніколи не відбувались. Її появи посприяли, з одного боку, технології розпізнавання патернів (образів, шаблонів), з успіхом застосовані для розпізнавання облич, а також – збільшення обчислювальних потужностей, що дозволило моделювати об'єкти за допомогою технології трасування променів (ray tracing), яка створює надзвичайно реалістичні зображення. Водночас, завдяки розвитку платформ з аудіовізуальними даними та загальному збільшенню кількості користувачів мережі Internet, збільшились можливості людей щодо їх застосування.

Інша форма діпфейк-контенту – аудіо. Завдяки технології дублювання (клонування) голосу аудіо-підробки стають більш, ніж реалістичними. Для створення звукового продукту, в якому відтворюється не лише тембр голосу, а й набір унікальних мовних конструкцій, які використовує конкретна людина в своєму мовленні, залучаються технології обробки природної

⁷⁷ Jaiman, Ashish. Debating the ethics of deepfakes. ORF. *Digital Frontiers*. 2020. URL: <https://www.orfonline.org/expert-speak/debating-the-ethics-of-deepfakes/>

⁷⁸ Deep Learning by: IBM Cloud Education. 2020. URL: <https://www.ibm.com/cloud/learn/deep-learning>

⁷⁹ Chesney, Bobby and Citron, Danielle. *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. California Law Review, 2019. Vol. 107:1753. P.1757. <https://doi.org/10.15779/Z38RV0D15J>

мови останнього покоління, що дозволяє засобами кількісного (статистичного), лінгвістичного та інших підходів визначати та відтворювати лексико-семантичний тезаурус і манеру спілкування мовця.

Будь-яка створена людиною річ – предмет або технологія, існуючи у світі культури, автоматично стає явищем цієї культури, її маніфестацією. Не випадково перед нами постають питання: «Якою є культура, в якій виникає це явище? І як це явище надалі змінює культуру?», а відтак: «Чи є якісь специфічні світоглядні характеристики культури, в якій можливе існування й використання технології “deepfake”?».

З одного боку, дідфейки не є принципово новим культурним феноменом. Аналізуючи форми його існування та загрози, які вони несуть, можна констатувати, що з подібним людство зустрічалося віддавна: лист пресвітера Іоана, фальшиві картини та купюри⁸⁰ тощо – все це можна вважати фізичними аналогами сучасного цифрового контенту. Тому варто говорити лише про новий технологічний етап існування невід’ємного культурного явища – обману. До появи цієї технології прикладами обману відеозасобами можна назвати будь-які пропагандистські фільми, які намагалися викривити чи замінити дійсність тією, яка була потрібна авторам або замовникам. До цієї ж когорти належить фальсифікація відеодоказів. Але, зважаючи на доступність інформації за допомогою сучасних технологій, небезпеки технології «deepfake», як і можливі конструктивні способи її використання, значно розширилися. Зокрема, використання дідфейк-відео ставить під загрозу окремі і без того вразливі категорії населення, зокрема дітей та підлітків. Маючи майже необмежений доступ до джерел інформації у вигляді гаджетів, підключених до мережі Internet, та не володіючи достатньо розвиненим критичним мисленням, вони можуть наражатись на небезпеку з боку цієї технології у вигляді, до прикладу, різного роду викликів (challenges). До найбільш уразливої групи населення також потрапили жінки, проти яких спрямовані майже всі дідфейк-відео порнографічного змісту, які завдають значної шкоди репутації, а також психічному стану жінки, зображення якої зімітовано в них.

З іншого боку, нові технології дозволяють досягти надзвичайної реалістичності фальсифікованих медіапродуктів, що, своєю чергою, ускладнює їхнє розпізнавання й розкриття підміни. Ми вступаємо в еру симуляції реальності, яка має ознаки гіпер-реалізму⁸¹. Зважаючи на психологічну схильність людини потрапляти у полон пізнавальних стереотипів та упереджень, а також зростання впливу соціальних мереж, які миттєво поширюють «інформацію», простежити межу між реальністю та псевдореальністю стає усе складніше. Загрози, пов’язані з дідфейк-технологією, стають усе більш обговорюваними фахівцями з різних галузей, оскільки зростає ступінь і масштаб їхньої «загрозливості». Це викликано як тим, що будь-що може бути сфальшоване

⁸⁰ Лист Пресвітера Іоана вважається підделкою, створеною, за однією із версій, щоб схилити учасників Другого хрестового походу рушити до Месопотамії.

⁸¹ Chesney, Bobby and Citron, Danielle. *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. California Law Review, 2019. Vol. 107:1753. P. 1768. <https://doi.org/10.15779/Z38RV0D15J>

завдяки удосконаленню самих технологій, так і пришвидшенню темпів і обсягів цифровізації. «Фізична реальність» поступається місцем мережі «цифрового світу», який невпинно зростає. При цьому масштаб фальшування потенційно може зростати щодо кожної з чотирьох координат Айнштайнової версії світу, враховуючи час. Охочі впливати на політику, економіку й інші соціальні процеси отримали надзвичайні можливості формувати гіпер-реалістичні образи не лише сучасності, а й минулого, тим самим задаючи напрямки й форми розвитку майбутнього.

Посадження когнітивних упереджень, притаманних усім людям, з можливостями соціальних мереж створило сприятливий ґрунт для використання дипфейків. Американські юристи, професори права Б. Чесні й К. Деніелл називають три головні цеглини в підмурку масового ураження інформаційними підробками: «Інформаційні каскади, природний потяг до негативної і нової інформації та бульбашки фільтрів створюють надто гостинне середовище, у міру дозрівання і примноження можливостей глибоких підробок»⁸². Інформаційні каскади зумовлені впливом прикладів поведінки інших людей на прийняття особистого рішення. Вони виникають у ситуації вибору, коли спостереження за вибором інших людей підштовхують індивіда до наслідування їхнього рішення, навіть, якщо він/вона має протилежну власну думку. У такому випадку поведінка більшості слугує вагомим контраргументом у внутрішній дискусії щодо раціонального вибору. «Бульбашкою фільтрів» (англ. – filter bubble) назвали своєрідне інформаційне (само)обмеження користувача інтернету, яке виникає через алгоритми персоналізації на основі її/його індивідуального досвіду користування мережею. Завдяки фільтрам, які відстежують усі запити особи в мережі, вона надалі стикається переважно лише з такою інформацією та думками, які відповідають і підкріплюють її власні переконання (цифрові стіни «печери» Платона сучасної доби).

Застосування технології «deepfake» 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 інше⁸³. Усі ці небезпеки, прямо чи опосередковано пов'язані з тенденцією до поглиблення нівелювання поняття й цінності істини.

⁸² Там само.

⁸³ Детальний розбір деяких з цих загроз висвітлений у Chesney, Bobby and Citron, Danielle. *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. California Law Review, 2019. Vol. 107:1753. P. 1777-1785. <https://doi.org/10.15779/Z38RV0D15J>.

б) Діпфейки й епістемологічні проблеми

Проблема істини не втрачала своєї значущості впродовж всієї історії людства й, починаючи з Нового часу, викликала чимало дискусій щодо можливості її досягнення засобами науки як методу пізнання світу. Дослідження пізнавальної здатності людини, від спекулятивних роздумів І. Канта й науково-теоретичних розвідок П. Дюгема до сучасних когнітивних наук, підводять до висновку, що ми сприймаємо світ крізь призму концептуальних схем, які довільно накладаються розумом на складний і багатоманітний світ природних і штучно створених речей. Такий «інструментальний» підхід до знання значною мірою зняв напруту у взаємодії суб'єкта й об'єкта у процесі пізнання (від неможливості виконати вимогу встановлення відповідності між знанням про дійсність і самою дійсністю). Своєю чергою, неокантіанець Е. Кассіер обґрунтував тезу про те, що людина – символічна істота й належить не так до світу природи, як до світу культури, утвореного із символів, значень і смислів. Сьогодні цей світ набув цифрового, мережевого виміру, що актуалізує один із аспектів названої проблеми – довіру до інформації в мережі.

Діпфейки, так само як і інші форми обману, претендують посісти якесь місце в реальності, спотворюючи її, але, будучи за формою ідентичними з фактами дійсності, вони підважують «дійсність» самих фактів: якщо немає чітких критеріїв розрізнення фактів і їхніх підробок, то будь-який факт може бути названий фейком. Учасники мережевої комунікації опиняються в середовищі з, так би мовити, *броунівським рухом*⁸⁴. Через постійні порушення правил значною кількістю учасників соціальної взаємодії добродісна дія дезорієнтована, а відтак діють ситуативно, не маючи при цьому ні типових шаблонів поведінки, ні маркерів, які б окреслювали координати ситуації, ні надійних критеріїв для оцінки своїх рішень.

Варто звернути увагу на декілька ключових моментів, пов'язаних з епістемною загрозою (epistemic threat), джерелом якої постає технологія «deepfake». Оскільки, на відміну від одиничних фальшивих відео, що існували ще в першій половині ХХ ст., вона підважує пізнавальний процес загалом, принаймні в його соціальному вимірі.

Як стверджує професор філософії та комп'ютерних наук Д. Фоліс, «діпфейки зменшують обсяг інформації, яку відео несе глядачам»⁸⁵. У такому разі проблеми та загрози, пов'язані з цією технологією, виходять далеко за межі атомарних проблем, які стосуються честі та гідності окремих людей чи провокацій, влаштованих на політичному ґрунті. Незважаючи на те, що подібні моделі обману через відеоряд вже існували на нашій планеті задовго до появи транзисторів і комп'ютерів у їх сучасному розумінні, ця технологія здатна швидко масштабувати недовіру навіть до значущих феноменів і перетворити спосіб сприйняття дійсності в тотальну недовіру до інформації, яка в

⁸⁴ Броунівський рух – невпорядкований, хаотичний рух частинок. Тут – ситуативні, невпорядковані реакції реципієнта внаслідок взаємодії з інфосферою.

⁸⁵ Fallis D. The Epistemic Threat of Deepfakes. *Philos. Technol.* 2020. URL: <https://doi.org/10.1007/s13347-020-00419-2>.

основному сприймається людиною за допомогою зору засобами сучасних технологій.

Тому потрібно розмежувати загрози технології «deepfake» на дві великі категорії залежно від намірів, яким вона слугує: 1) відео, в яких показано те, що ніколи не відбувалось, з метою переконати глядача в істинності зображуваних подій – імплантація хибних переконань, обман; а також 2) відео, в яких показано те, що ніколи не відбувалось, з метою переконати глядача в неправдивості реальних подій, що відбулись насправді і пов'язані з цим фальшивим відео – дискредитація. Сумарно – ця технологія здатна цілковито «заплутати» реципієнта, який не в змозі буде швидко встановити правдивість побаченого.

Проблема технології deepfake не лише в тому, що, використовуючи відео, на якому зображені події, яких насправді ніколи не було, ми поширюємо симулякри, небезпека яких з появою цієї технології лише зростає, а ще й в тому, що якщо складно (швидко і без додаткових технічних засобів) відрізнити справжнє (відео) від несправжнього, то визначення того, що є чим, цілком лягає на плечі окремого індивіда чи соціальної групи. Тобто дідфейки не просто продукують хибні вірування в те, чого не було, а й потенційно унеможливають віру в те, що відбувалось насправді. «Однак, крім створення помилкових переконань, є й інші способи, якими дідфейки можуть перешкодити людям отримувати знання. Наприклад, навіть після перегляду справжнього відео і придбання істинних переконань людина може не отримати знання, тому що її процес формування переконань недостатньо надійний. Оскільки дідфейки стають все більш поширеними, може бути епістемно безвідповідальним просто вірити в те, що те, що зображено на відео, дійсно відбулося. Таким чином, навіть якщо хтось подивиться справжнє відео, на якому відомий політик бере хабар, і дійде до висновку, що він корумпований, можна не знати, що це так»⁸⁶.

Технологія «deepfake» базується на основі роботи специфічних нейромереж, які своєю чергою, також змінюють реальність навколо і можуть сприяти появі нових способів досягнення дійсності, які є більш ефективними, ніж класичний науковий метод, який розвивається лінійно: від проблеми – через окремі припущення щодо їх походження чи розв'язання – до гіпотези та теоретичного конструкту, який необхідно експериментально підтвердити. Паралельні обчислення, нові способи розпізнавання та створення образів, можливість планувати й передбачати в небачених раніше масштабах, завдяки квантовим технологіям – усе це виводить людство на абсолютно новий рівень обробки й використання інформації, а також породжує нові переваги та проблеми, порівняно з класичними епістемічними підходами.

Тобто, якщо ще сьогодні довіряти інформації можна завдяки самостійному її пошукові, опорі на надійні інформаційні ресурси чи зіставленню альтернативних, незалежних джерел повідомлень, то в найближчому майбутньому все може змінитись. Оскільки фальшива інформація, покликана

⁸⁶ Там само.

продувати хибні переконання, зможе існувати в одиничному, «ексклазивному» вигляді, без альтернатив і додаткових вказівок на джерела її отримання. В такому разі людина, яка отримає подібну інформацію (назвемо її подією F) буде змушена судити про неї не як про достовірну чи ні, а як таку, якій можна чи не можна довіряти. Тобто поняття істинності легко може поступитись понятию переконливості.

Фактично, класична концепція істини як відповідність наших знань дійсності може бути відкинена, повністю поступившись когерентній концепції: істинність або хибність переданої через відео інформації буде встановлюватись винятково в контексті суперечності чи несуперечності з раніше отриманою інформацією. При цьому виникає не менша загроза, оскільки немає жодних перешкод для того, щоб за допомогою технології «deepfake» був згенерований цілий масив неправдивої інформації, якій не буде суперечити окрема подія (F). І, відповідно, остання може сприйматись як те, що відбувалось насправді. Ключовим у такому когнітивному спотворенні може стати міфологічне сприйняття всієї сукупності подій-підробок. Якщо дипфейк розглядати не як атомарне явище, а як великий інформаційний кластер, який є внутрішньо несуперечливим, то цілком імовірно, що він здатен породжувати щось на кшталт міфологем, які у своїй цілісності будуть генерувати альтернативну реальність для їх реципієнтів.

Засадничий момент цього міркування – те, що замість *встановлення істинності* тієї чи тієї події, чи, радше, визнання наявності її онтологічного фундаменту, окремою людиною чи спільнотою загалом буде *встановлюватися думка про істинність* тієї чи тієї події або обчислюватиметься *ймовірність її істинності* (F). До прикладу, обчислення Т. Байєса, які ґрунтуються на появі нових «фактів», скажімо, події (F), будуть менш вірогідними, оскільки ймовірність того, що ця подія є істинною, в добу дипфейків значно знижується. Д. Фоліз порівнює дві альтернативні позиції: Ф. Дрецьке (F. Dretske)⁸⁷ та Б. Скайрмза (B. Skyrms)⁸⁸. Згідно з позицією Ф. Дрецьке, ймовірність передачі інформації строго дорівнює або 1, або 0, що, на думку Д. Фоліза висуває надто суворі вимоги до гарантій правдивості в добу постправди. Натомість автор надає перевагу позиції Б. Скайрмза, який акцентує увагу на ймовірнісній складовій істинності тієї чи тієї події на основі того чи того факту. Як зазначає Д. Фоліз, відповідно, навіть якщо відео здається справжнім, тепер існує значна ймовірність того, що подія, яка відтворюється у відео, насправді не відбулась. Тобто ймовірність помилкового судження збільшилася в результаті використання технології «deepfake». І ця ймовірність буде продовжувати зростати пропорційно з удосконаленням і збільшенням доступності цієї технології. Отже, те відео, що здається справжнім, тепер несе менше інформації, ніж колись, і, імовірно, буде нести ще менше в майбутньому⁸⁹.

⁸⁷ Dretske F. Knowledge and the Flow of Information. Cambridge: MIT Press, 1981. P. xi-273.

⁸⁸ Skyrms B. Signals. Evolution, Learning, & Information. New York: Oxford University Press, 2010. 208 p.

⁸⁹ Fallis D. The Epistemic Threat of Deepfakes. *Philos. Technol.* 2020. URL: <https://doi.org/10.1007/s13347-020-00419-2>.

Проте варто звернути увагу, що Д. Фоліз використовує припущення у прикладах, які мають слугувати аргументами на користь обґрунтування його позиції. «Якщо відео містить менше інформації, ніж раніше, це означає, що ймовірність того, що відео з цим конкретним змістом є справжнім, – нижча, ніж це було б в минулому. Зокрема, припустимо, що до дідфейків ймовірність того, що відео, на якому політик бере хабар, є справжнім, була б 0,9. Але після появи дідфейків ймовірність того, що таке відео справжнє, становить лише 0,5. Якщо ви вважаєте, що цей політик брав хабар на підставі цього відео, то переконання, яке ви встановили для себе, в п'ять разів більш імовірно буде хибним, ніж це було до дідфейків. І навіть якщо ваша віра виявиться правдою, вона, ймовірно, не буде кваліфікуватись як знання. Сьогодні ваша віра є менш безпечною»⁹⁰. В цьому контексті складно погодитись з категоричністю висновку про ймовірність справжності відео на рівні 50 відсотків. На нашу думку, це лише припущення, яке має емоційне забарвлення і покликало привернути увагу до проблеми загалом. Візьмімо до уваги недосконалість технології, її невміле використання і здатність людини спростовувати інформацію, що робить вплив дідфейків, на цей момент, менш руйнівним, ніж просте підкидання монетки під час ухвалення доленосних рішень. Водночас важко не погодитись, що з удосконаленням технології, збільшенням кількості дідфейків і зважаючи на відсутність навичок критичного мислення у значної частини суспільства щодо розпізнавання фальшивої інформації, наслідки її застосування можуть бути непередбачуваними. Саме існування цієї технології змінює «баланс сил» і призводить до збільшення недовіри до будь-якої інформації, незалежно від того – правдива вона чи ні.

Розробник і запроваджувач ефективних практичних технологічних інновацій, спрямованих на розв'язання повсякденних проблем і пошук нових можливостей А. Джайман зауважує, що дідфейки посилюють зростання недовіри до засобів інформації, що своєю чергою сприяє поширенню й закоріненню культури фактуального релятивізму⁹¹. Релятивізм протистоїть таким світоглядним позиціям, як абсолютизм, монізм і фундаменталізм, що робить його привабливим для прихильників демократичного ладу (провідна ідея якого – плюралізм і рівність політичних акторів). Окрім того, релятивізм – зручний спосіб уникнути конфлікту на ґрунті «незручних» питань, відповідь на які залежить від контексту культури, віри, національної історії тощо чи світоглядних передумов, які неможливо ні довести, ні спростувати. Наявність розбіжностей у поглядах, щодо яких неможливо виявити помилки в аргументації й водночас неможливо досягти згоди, залишаються вагомим аргументом для виправдання релятивізму. Але ця позиція протилежна й до реалізму та об'єктивізму, на засадах яких можливе універсальне ціннісне визначення свободи, правди, справедливості.

⁹⁰ Fallis D. The Epistemic Threat of Deepfakes. *Philos. Technol.* 2020. URL: <https://doi.org/10.1007/s13347-020-00419-2>.

⁹¹ Jaiman A. Debating the ethics of deepfakes. ORF. *Digital Frontiers.* 2020. URL: <https://www.orfonline.org/expert-speak/debating-the-ethics-of-deepfakes/> (дата звернення: 5.08.2021).

Прихильники діалектичного погляду на історію й, зокрема, історію філософії можуть простежити наявність певних циклів підйому й занепаду раціоналізму, реалізму й об'єктивізму як свосередних антиподів релятивізму. До прикладу, романтизм прийшов на зміну Просвітництву, започаткувавши рух Анти-просвітництва як заперечення його центральних засад: універсальності, об'єктивності, раціональності, здатності забезпечувати незмінні рішення для всіх проблем тощо⁹². Відтак ХІХ-ХХ ст. позначені радикальною критикою претензій розуму встановити незмінні універсальні засади й критерії людських суджень про світ, культуру й самих себе, починаючи від Ніцше до постмодерністів⁹³. У сучасній філософії набуває поширення рух метафізичного, чи спекулятивного, реалізму, один з напрямків якого – трансцендентальний нігілізм. Останній вже не протиставляє розум і буття як зіставлені, зв'язані поняття (як, скажімо, у А. Бадью), а загалом заперечує будь-яку кореляцію між ними, апелюючи до небуття (Р. Брасьє «Ніщо незв'язане: вимирання і Просвітництво»⁹⁴). На перший погляд, це самоочевидний розв'язок діалектичного сплетення, чи гордієвого вузла, реалізму й релятивізму: якщо ми не можемо визнати абсолютне буття, то зведемо все до ніщо, яке не містить жодних зародків суперечностей, взаємозалежностей або кореляцій. Але дідфейки мімікують під реальність, утверджують її альтернативні версії. Питання полягає в тому, чи існує кореляція між штучно створеною копією факту дійсності й самим фактом. І оскільки сучасна людина охоплена мережею копій, які формують її уявлення про реальність, здійснюють вплив на ухвалення нею рішень, здійснення вибору тощо, наповнення цієї мережі псевдокопіями призводить до дезорієнтації суб'єкта, дискредитації інформаційних мереж, а також доведення неспроможності суб'єктивно чи навіть інтерсуб'єктивно встановлювати істину.

Згідно із класичним зразком висловлювання сутності релятивізму, «жоден виклад правди чи реальності не може займати привілейоване становище, оскільки такі звіти – лише одна з багатьох перспектив, які панують у певний час в історії»⁹⁵. Виходячи з такої світоглядної позиції, навіть неправда постає одним із можливих різновидів опису реальності. Очевидно, що помножена на технічні можливості для створення дідфейків неправда набуває реальних просторово-часових вимірів. Підважуючи людську здатність довіряти не лише іншим, а й самим собі. На цьому ґрунті з'являється додаткова вигода для шахраїв, нечесних політиків, бізнесменів і загалом будь-кого,

⁹² Berlin I. *The Counter-Enlightenment in The Proper Study Of Mankind: An Anthology of Essays*. Random House, 2012. P. 243-268.; P. 263.

⁹³ Див.: Gordon C., Bearn F. "Nietzsche, Feyerabend, and the Voices of Relativism" in *Metaphilosophy*. 1986. Vol. 17, No. 2/3. P. 135-152.

⁹⁴ Brassier R. *Nihil Unbound: Extinction and Enlightenment*. PALGRAVE MACMILLAN. 2007. 278 p.

⁹⁵ Baghramian, Maria and J. Adam Carter, „Relativism“, *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL: <https://plato.stanford.edu/archives/spr2021/entries/relativism/> (дата звернення: 5.08.2021).

зацікавленого у спростуванні правдивої інформації. Кожного разу, коли траплятиметься «незручна» правда, її можна буде заклеїти як фейк, або дїпфейк.

Показовим прикладом збільшення подібної недовіри та її неперебачуваних наслідків стала ситуація в Габоні, початок якої припав на 2018 р. у зв'язку із хворобою чинного президента країни А. Бонго. Через стан здоров'я А. Бонго з кінця жовтня до січня 2019 р. не з'являвся на публіці⁹⁶, жодним чином не заявляючи про себе в публічному просторі. Це, звісно, породило чутки про його повну недієздатність або навіть смерть. Для того, щоб заспокоїти громадськість і запобігти політичній кризі та вакууму влади, 1 січня 2019 р. було записано відеозвернення А. Бонго до народу⁹⁷. У цьому відео президент повідомив про те, що йому вже краще, складний період позаду, і він готується до зустрічі зі своїми громадянами. Однак політичні опоненти та частина суспільства, яка не підтримувала політику чинного президента, висловили сумнів у справжності відео, зацікавившись в ньому підробку, зроблену з використанням двійника, чи продукт сучасних аудіовізуальних технологій. Це стало приводом для спроби захоплення влади військовими на початку січня 2019 р. Згодом, після проведення експертизи не було виявлено явних ознак втручання чи спроб підробки новорічного звернення очільника Габону, хоч особисто А. Бонго з'явився на публіці лише в серпні 2019 р. Критичним моментом для вірогідних сценаріїв розвитку цієї ситуації постає наявність можливості використання дїпфейк-технологій. Достатньо оголосити, що справжнє відео є дїпфейком, щоб дестабілізувати хитку політичну ситуацію в країні.

Цей випадок свідчить про одну з прихованих загроз цієї технології. Якщо зараз увага експертів прикута до ситуації, коли те, чого ніколи не було, видають за справжню подію, то згодом основну загрозу можуть становити випадки заперечення того, що було насправді, якщо з'являться можливості за допомогою дїпфейк-технологій «спростувати» те, що відбувалось насправді⁹⁸. Оголошення справжніх відео чи аудіо підробками потенційно може призвести до згорання демократичних процесів, залякування громадян, свавілля силових структур, які, підважуючи довіру до справжніх інформаційних продуктів, можуть продовжувати вести свою протиправну діяльність. До прикладу, спроби дискредитації відео, де зафільмовано події перевороту в М'янмі⁹⁹, а також суперечки користувачів мережі щодо його справжності повинні викликати велике занепокоєння. Це явище отримало назву «дивіденди

⁹⁶ President Ali Bongo of Gabon down with stroke. *Vanguard*. URL: <https://www.vanguardngr.com/2018/12/president-ali-bongo-of-gabon-down-with-stroke/> (дата звернення: 10.08.2021)

⁹⁷ I am now fine! Ali Bongo tells Gabonese in New Year message. *Africanews*. URL: <https://www.africanews.com/2019/01/01/i-am-now-fine-ali-bongo-tells-gabonese-in-new-year-message/> (дата звернення: 20.08.2021).

⁹⁸ Gregory S. Authoritarian Regimes Could Exploit Cries of 'Deepfake'. URL: <https://www.wired.com/story/opinion-authoritarian-regimes-could-exploit-cries-of-deepfake/> (дата звернення: 20.08.2021).

⁹⁹ Lytvynenko J. Yes, This PE Teacher Really Danced Through The Coup In Myanmar. *BuzzFeed News Reporter*. URL: <https://www.buzzfeednews.com/article/janeltyvynenko/myanmar-coup-dance> (дата звернення: 20.08.2021).

брехуна): «дивіденди брехуна зростатимуть разом із зернами більших тенденцій, залучаючи скептицизм щодо правди-істини “truth scepticism”»¹⁰⁰.

Якщо релятивісти, грубо кажучи, заперечують припущення про існування об’єктивної чи інтерсуб’єктивної правди-істини, то скептики першочергово ставили собі за мету поставити під сумнів «самоочевидність» поширених і «незаперечних» істин. Але останнім часом розповсюдження набуває, так би мовити, псевдоскептицизм¹⁰¹, який заперечує спроможність раціонально, науково, на підставі наявних даних довести істинність визнаних наукових теорій. Псевдоскептики з середовища вчених заперечують ті факти, які переважна більшість дослідників вважає достатніми і валідними для формулювання й визнання певної наукової концепції, теорії. Зазвичай вони при цьому спираються на альтернативний набір даних або на альтернативне теоретичне пояснення тих самих даних (напр., спростовують наукове обґрунтування кліматичних змін під впливом антропогенних чинників).

Прикладом того, наскільки наші сучасники схильні приймати на віру заяви скептиків різного роду слугує історія «спростування» приземлення на Місяць Ніла Армстронга. Як зазначає журналіст та дослідник Р. Годвін: «Щоб висадити Ніла Армстронга та Базза Олдріна на Місяць у 1969 році, знадобилося 400 000 співробітників NASA та підрядників – але лише одна людина поширила думку, що це все обман»¹⁰². Б. Кейсінг намагався довести, що США в принципі не могли здійснити висадку людини на Місяці. Його брошура 1976 р. «*Ми ніколи не їздили на Місяць: афера Америки за тридцять мільярдів доларів*», започаткувала цілий рух «розвінчання Місяце-змови». «Незважаючи на надзвичайний обсяг доказів (у тому числі 382 кг місячної породи, зібраної за п’ять місій; підтвердження з Росії, Японії та Китаю; та зображень з орбітального апарату Місяця-розвідника НАСА, на яких видно сліди, зроблені астронавтами у місячному потоці), віра в містифікацію Місяце-змови розквітла з 1969 року. Серед свідків 11 вересня, антиваксерів, хімічних трейлерів, пласкоземлів, заперечувачів Голокосту та змовників Сенді Гук, думка про те, що висадка на Місяць була підробленою, більше не є навіть джерелом гніву – це лише даний факт»¹⁰³.

Наведений псевдоаналіз відео з програми «*Apollo*» – це різновид дискредитації, створеної в доцифрову еру. Але, на відміну від очевидних причин зацікавленої сторони поставити під сумнів успіх цієї програми і можливість легкого виявлення такої зацікавленої сторони, діпфейки, з високою

¹⁰⁰ Chesney, Bobby and Citron, Danielle. *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, California Law Review, 2019. Vol. 107:1753. P. 1785. <https://doi.org/10.15779/Z38RV0D15J>

¹⁰¹ Див.: Torcello, Lawrence, The ethics of belief, cognition, and climate change pseudoskepticism: implications for public discourse, *Topics in Cognitive Science*. 2016. P. 19-48.

¹⁰² Godwin R. “One giant ... lie? Why so many people still think the moon landings were faked”. *The Guardian*. 2019. URL: <https://www.theguardian.com/science/2019/jul/10/one-giant-lie-why-so-many-people-still-think-the-moon-landings-were-faked> 17/09/2019 (дата звернення: 22.08.2021).

¹⁰³ Там само.

вірогідністю, залишатимуться анонімними, оскільки встановлення авторства чи мотивів їх створення нівелюватиме їх «цінність».

Філософська традиція віддавна намагається розмежувати дійсність (онтологічне) від випадки (психологічне). Починаючи з Платонової традиції подвосення світу, мовиться про спробу розрізнення дійсного (справжнього, істинного, сутнісного) і копії, що імітує та спотворює це дійсне. Згідно з Платоном, ми маємо справу з двома множинами: 1) світ ідей – автономний, справжній, дійсний – буття; та 2) світ речей – залежний, тимчасовий, чуттєвий, в якому живе людина – видимість.

Міф про печеру, розказаний Платоном у *«Полімеї»*, яскраво демонструє можливість припускати помилок у пізнанні дійсності через сприйняття викривленої інформації, а також складність відмовитись від створеного нею контексту – світогляду. Розглядаючи цей компонент Платонової філософії як символ оманливого «пізнання», у світлі осмислення технології *deepfake* актуалізуються такі поняття, як «симулякр» та «герменевтика підозри».

«Симулякр» для Платона – копія копії, яка нерозривно пов'язана з психологічним сприйняттям речі, події чи явища суб'єктом – його/її власна інтерпретація, яка породжує переконання. Паралелі між філософською позицією Платона й постмодерністським тлумаченням ідеї симулякрів, а відтак – і з сучасною *deepfake*-технологією більш, ніж очевидні. «Загалом, мотив теорії ідей повинен бути віднайдений у вибірковій здатності, волі до відбору чи наданні переваги. Його можна розглядати як проблему «маркування різниці» та констатації відмінностей між «річчю» як такою та її образом, оригіналом та копією, моделлю та симулякром»¹⁰⁴. Зважаючи на таку, досить однозначну позицію Ж. Дельоза, варто зауважити, що саме поняття «симулякр», а не «копія» є доречним, якщо проводити паралелі між оригінальним контентом і дідфейком, створеним на його основі.

Копія – поняття багатозначне, його зміст залежить від контексту. Якщо в юриспруденції чи діловодстві поняття копії, яка належно завірена і відповідає своєму визначенню, тобто має абсолютну схожість з оригіналом, не має негативного забарвлення, то, наприклад, у мистецтві воно має негативний підтекст, як таке, що лише повторює оригінал, але аж ніяк не здатне його замінити. Якщо розглядати поняття копії у контексті відеоматеріалів, то в сучасну цифрову епоху – це, фактично, ідентичний оригіналу контент, який відрізняється лише технічними даними і не має негативного змісту в плані підміни, додавання, вилучення або втрат контенту у фізичному плані. Симулякр, на відміну від копії, – це лише схожість у базових моментах (часто візуальна), метою якої постає не збереження автентичного образу, а, насамперед, його викривлення чи виправлення. Ж. Дельоз, міркуючи про позицію Платона щодо демаркації «справжнього» доводить, що метод, який використовує Платон, не є простим поділом на роди чи види, а є способом визначення дійсного й ілюзорного. «Він [метод – *авт.*] використовується не для того, щоб оцінити заслужених претендентів, а навпаки, для вистежування хибних претендентів

¹⁰⁴ Делез Ж. Платон и симулякр. URL: <https://kph.ffe.npu.edu.ua/le-book/clasik/data/misc/intent/07deleuze.html> (дата звернення: 24.08.2021)

як таких, з метою визначення буття (чи, радше, небуття) симулякра [...] Платон відкриває те, що симулякр – не просто хибна копія, а те, що він ставить під сумнів взагалі будь-яке зображення копії і моделі»¹⁰⁵. На нашу думку, до дівфейків, які мають стосунок до оригінального контенту подібно до того, як симулякр – до моделі, можна застосувати аналогічний підхід та виставити подібні застереження щодо небезпеки підміни реальності та її негативних соціальних наслідків. Створення дівфейк-контенту переносить нас із площини онтології – існує лише у вигляді послідовності символів у строгій прив'язці до матеріального носія інформації – до площини соціального: дівфейки створені, щоб викликати певну реакцію з боку соціуму і мають вплив винятково в ньому і через нього. А це вже цілком відповідає позиції Ж. Бодрієра і його теорії симулякрів, а не платонівській позиції, який розуміє симулякр як копію копії.

Проте повне відмежування від філософського спадку Платона в цьому контексті було б невиправданим, оскільки залишається проблема розрізнення оригіналу, копії та симулякра. Якщо розглядати відеоконтент, то оригіналові відповідає подія, яка реально відбулась, копії – цифровий чи аналоговий контент, створений на її основі, симулякрові – дівфейк, покликаний викривити зміст цієї події для реципієнтів. Онтологія та теорія пізнання Платона потенційно вказує на напрямок розв'язання цієї складної проблеми, оскільки метою постає ідентифікація дійсного й ілюзорного на основі розрізнення внутрішніх якостей, а не порівняння зовнішнього образу. В цьому контексті можна зробити перші обережні висновки про те, що для протистояння впливу дівфейк-контенту ніколи не буде досить технічних засобів визначення подібності-відмінності, а завжди буде необхідною внутрішня пізнавальна дисципліна споживача, який повинен у будь-яку мить бути готовим вдатись до самостійного аналізу, а не лише інфантильно сприймати побачене.

У світлі цього постає цілком закономірне питання: «Чому загалом дівфейки працюють?». Відповідаючи на питання, можна виділити декілька важливих моментів, які в комплексі здатні пояснити причину наявного стану справ: (1) людина звикла довіряти побаченому, оскільки зорове сприйняття – основний спосіб отримання інформації. У світлі цього факту, дівфейки, які є не просто статичними об'єктами, а подають інформацію в динаміці, що забезпечує максимальну реалістичність спостережуваної поведінки суб'єктів, мають перевагу над іншими носіями інформації у вигляді тексту чи аудіо, чи навіть статичного фото; (2) невідне зростання кількості, швидкості й мережі охоплення інформаційного потоку, в якому ми знаходимось упродовж життя. Постійний «інформаційний шум», з одного боку, розмиває пріоритети у сприйнятті інформації в загальному плані даних, а з іншого – дозволяє дівфейкам завдяки раніше згаданим властивостям набувати такого

¹⁰⁵ Делез Ж. Платон и симулякр. URL: <https://kph.ffs.npu.edu.ua/e-book/clasik/data/misc/intent/07deleuze.html> (дата звернення: 24.08.2021)

пріоритету; (3) Д. Канеман у праці «*Thinking, Fast and Slow*»¹⁰⁶ («Мислення швидко й повільно») описує процес мислення й ухвалення рішень, який здійснюється за допомогою двох інструментів: Системи 1 та Системи 2. Технологія «deepfake» спрямована на взаємодію з Системою 1, якій притаманна така якість, як побудова причинно-наслідкових зв'язків між подіями. Дуже часто це відбувається з порушенням законів логіки: тотожності чи достатньої підстави. На думку Д. Канемана, наше сприйняття інформації побудоване таким чином, що людині набагато простіше побудувати правдоподібну історію на основі припущень, ніж перевірити її на наявність суперечностей. Оскільки контент, створений за допомогою технології «deepfake», на перший погляд, не порушує жодних логічних законів, адже все, що «відбувається» на відео – відбувається ніби насправді, то він легко маркується людиною як правдивий і, отже, може впливати на її поведінку; (4) подібний цифровий контент, як правило, поширюється за допомогою соцмереж – потужного ресурсу взаємодії між людьми. Оскільки людина у процесі соціалізації прагне до обміну інформацією з іншими, її часто не зупиняє підозра щодо якості чи істинності поширюваного ними контенту. Враховуючи також, що завдяки комбінаціям соціальних мереж, які використовуються сучасними користувачами, число Данбара¹⁰⁷ (або його аналог у цифрову еру) поступово зростає (піддається ревізії) – стрімке поширення неправдивих відео вже не ставиться під сумнів. Тож, можна констатувати, що точність і правдивість інформації не є основним критерієм, на основі якого ухвалюється рішення про поширення контенту в соціальних мережах (зокрема, щодо поширення інтернет-мемів, які часто виривають з контексту певну подію або засновані на порушенні логічних законів, чим і приваблюють своїх споживачів). Часто ці рішення ухвалюються на ґрунті емоційної реакції, або стають актом соціального обміну; (5) будь-який обман, в тому числі створений за допомогою технології «deepfake», ґрунтується на окремих чи когерентних переконаннях реципієнта – назвемо їх *базовими*. Якщо людина переконана у продажності всіх політиків, то завдяки дипфейк-відео, поширеному політичними конкурентами, вона легко повірить в те, що якийсь політик бере хабар. У такий спосіб подібні підробки здатні посилювати базові хибні переконання, а разом з використанням таргетингу¹⁰⁸ збільшувати свою ефективність.

¹⁰⁶ Kahneman D. Thinking Fast and Slow. URL: <http://dspace.vnbrims.org:13000/jspui/bitstream/123456789/2224/1/Daniel-Kahneman-Thinking-Fast-and-Slow-.pdf> (дата звернення: 25.08.2021)

¹⁰⁷ Число Данбара – приблизна оцінка кількості соціальних зв'язків, які людина може підтримувати одночасно. Найчастіше оцінюється в 150 одиниць (для стародавніх поселень). Піддається ревізії в сучасному світі у зв'язку з появою новітніх засобів комунікації.

¹⁰⁸ Таргетинг – механізм, який дозволяє виділити так звану цільову аудиторію (що відповідає заздалегідь встановленим параметрам – вік, стать, рівень освіти тощо) для демонстрації певного контенту лише для цієї аудиторії. Одним із найбільш відомих прикладів застосування таргетингу є кейс «*Facebook ma Cambridge Analytica*».

в) Етичні проблеми дідфейк-технології та засоби їх запобігання

Одне з ключових завдань теоретичного характеру, пов'язаних із технологією «deepfake», постає з'ясування, чи є вона морально прийнятною сама собою. Чи ця технологія морально нейтральна і може бути застосована як на благо, так і на шкоду, залежно від того, хто її використовує? Чи, все ж таки, її потрібно зарахувати до аморальних, як таку, в основі якої лежить обман?

А. де Руте констатує, що значна кількість робіт, присвячених проблемі наслідків використання цієї технології, майже не торкається проблеми її моральності як такої, самої по собі. «Технологія може вважатись морально неприйнятною, якщо вона порушує моральні норми (деонтологія), є причиною значних негативних наслідків (консеквенціалізм), зменшує благо і/або сприяє злу (етика чеснот) чи підриває міжособистісні стосунки і фундаментальні соціальні цінності, такі як довіра і взаємна повага (етика турботи)»¹⁰⁹. А. де Руте наполягає на тому, що «розглядаючи природу моральних норм, деонтологія безпосередньо розглядає моральну природу дій, артефактів чи технологій»¹¹⁰, а тому повинна бути пріоритетною у відповіді на питання про моральну природу дідфейків. Оскільки ця технологія може бути використана не лише з метою обману, а й має значний потенціал до конструктивного використання, дослідниця робить висновок, що «хоч технологія deepfake має великий потенціал для обману, вона не є оманливою за своєю сутністю, оскільки реалізація оманливого ефекту залежить від умов, в яких застосовується технологія»¹¹¹. Такий висновок ґрунтується на аналізі способів її використання – адже в конструктивному ключі вона може бути використана для допомоги людям з медичними проблемами, для удосконалення технології синхронного перекладу, для створення реалістичного мультимедійного контенту з навчальною чи просвітницькою метою, створення арт-об'єктів тощо. В такому контексті, конструктивний потенціал цієї технології – колосальний, однак, подібна констатація, на нашу думку, не дає остаточної відповіді на питання про її моральну природу. Якщо зважити на те, що перший контент, створений за допомогою цієї технології мав явно аморальний характер, і врахувати можливість її безконтрольного використання, то теза про нейтральність цієї технології, з морального погляду, постає, як мінімум, інтуїтивно сумнівною. Іноді можна побачити аналогії, покликані захистити моральну нейтральність технології створення дідфейк-контенту, наприклад, порівнюючи її з технологією розщеплення атомних ядер, яка може бути використана як на шкоду – як зброя масового знищення, так і на благо – для виробництва електроенергії. Однак, на нашу думку, не виправдано вбачати подібність між ними, хоча б тому, що ядерну технологію неможливо використовувати анонімно, і вибудовані реальні перешкоди для незаконного її поширення. Вона контролюється

¹⁰⁹ Ruiter de A. The Distinct Wrong of Deepfakes. *Philos. Technol.* 2021. URL: <https://doi.org/10.1007/s13347-021-00459-2>.

¹¹⁰ Там само.

¹¹¹ Там само.

та регламентується світовою спільнотою, що суттєво обмежує можливість підлітного створення чи протиправного застосування. Сутність дипфейків – діаметрально протилежна: вони мають анонімні джерела походження й, як наслідок, відповідальні – відсутні. Дії, спрямовані на підрив репутації, честі та гідності інших людей, маніпулювання громадською думкою тощо – все це від самого початку ставить під сумнів ідею моральної нейтральності цієї технології. У будь-якому випадку мета цієї технології – умисне введення в оману, що, на нашу думку, постає достатньою підставою для спростування її моральної нейтральності.

Зважаючи на потенційні ризики використання технології «deepfake», з'являються проекти на кшталт «*Prepare, Don't Panic: Synthetic Media and Deepfakes*»¹¹² (*Готуйтесь, не панікуйте: синтетичні медіа та дипфейки*). Їхні натхненники (серед яких і С. Грегорі¹¹³) прагнуть вже зараз, на зорі розвитку цієї технології, знайти ефективні методи прогнозування та запобігання негативних наслідків. Як підсумовує проєкт «*Witness*»¹¹⁴, ключовими кроками, здатними убезпечити людство від потенційної шкоди дипфейк-контенту є: 1) визнання того, що це еволюція способів обману за допомогою сучасних технологій, але зрештою не все, що ми бачимо є обманом; 2) визнання та усунення вже нанесеної шкоди; 3) формування глобального бачення прав людини на основі інклюзивного підходу; 4) визначення глобальних загроз; 5) просування міждисциплінарних підходів, які вже існують сьогодні: перевірка фактів, протистояння дезінформації, робота з відкритими джерелами тощо; 6) створення можливостей для ключових гравців – пошукових гігантів, ЗМІ та інших – для кращого розуміння загрози та більш тісної співпраці між ними задля її подолання; 7) визначення засобів координації між громадянським суспільством та ЗМІ; 8) проведення подальших досліджень розвитку технології; 9) створення ґрунту для відповідальності тих, хто створює комерційний продукт, який сприяє появі та поширенню deepfake-контенту, розробка норм і правил поведінки на ринку; 10) створення умов для рівного доступу до технологій виявлення подібного контенту, а також створення фінансового fair-play; зрештою – 11) поширення та популяризація етичних стандартів тощо¹¹⁵.

Розвиток технологій доповненої й віртуальної реальності (AR, VR), трасування променів, доступність і збільшення потужності домашніх комп'ютерних систем, перехід завдяки 5G-зв'язку обчислювальних потужностей до «хмар» – усе це у комплексі із застосуванням дипфейків потенційно

¹¹² Prepare, Don't Panic: Synthetic Media and Deepfakes. URL: <https://lab.witness.org/projects/synthetic-media-and-deep-fakes/> (дата звернення: 25.08.2021).

¹¹³ Gregory S. Witness. URL: https://www.witness.org/portfolio_page/sam-gregory/ (дата звернення: 25.08.2021).

¹¹⁴ Проєкт «*Witness*» створений з метою захисту прав людини за допомогою відео та сучасних технологій.

¹¹⁵ Див.: Prepare, Don't Panic: Synthetic Media and Deepfakes. URL: <https://lab.witness.org/projects/synthetic-media-and-deep-fakes/> (дата звернення: 25.08.2021).

може занурити людство не просто в *епоху постправди*, а в *постонтологічну добу*, коли відмінності реального й віртуального будуть все менш помітними.

Епістемна загроза, яку потенційно містять дідфейки, може стати загрозою і для інших фундаментальних сфер співіснування людей, наприклад, моралі чи релігії. Дискредитація понять «віра», «довіра», «переконання» та «істина-правда» матиме наслідки як для конкретної особистості, окремих соціальних груп, так і суспільства загалом. Епістемна шкода може бути відчутна, якщо «ви не зуміли набутися справжніх переконань в результаті застосування технології дідфейк. Навіть після появи дідфейків, відео з певних джерел, що заслуговують на довіру все ще можуть нести багато інформації. Наприклад, припустимо, що відео, на якому політик бере хабар, надходить з надійного джерела новин. Отже, відео з великою ймовірністю буде справжнім. Однак припустимо, що, навіть якщо ви можете сказати, що відео взято з цього джерела, ви не вирішите, чи брав політик хабар, тому що дуже занепокоєні дідфейками. Відкладаючи судження, ви уникаєте ризику помилкового переконання»¹¹⁶. У зв'язку з цим може відбуватись потенційне толерування хабарництва або інших злочинів як явищ, щодо яких неможливо щось стверджувати чи заперечувати. Оскільки на кону – ділова репутація, політична кар'єра, «доля» людини, можна припустити, що чим сильніший проступок, тим складніше буде повірити в нього (оцінити вищу ймовірність його істинності) лише на основі відео, у зв'язку з високою відповідальністю, яку потенційно братиме на себе людина, після ухвалення кінцевого рішення вірити побаченому чи ні. Отже, відеодоказ, який сьогодні продовжує відігравати значну роль під час формування переконань, скоро може втратити визначальне значення. Тобто дідфейки потенційно можуть справляти зворотній вплив, не як такі, що переконують нас в правдивості неправдивого (хибні переконання), а навпаки – змушують нас думати про неправдивість правдивого (істинні переконання, що здаються хибними). Як зазначає Б. Шварц – професор соціальної теорії коледжу Свартмор, чим складніший вибір, тим довше ми його відкладаємо, і, очевидно, моральні питання – це та сфера в якій зволікання чи ігнорування фактів цілком прийнятні»¹¹⁷.

Як стверджує докторка філософії університету Йорка Р. Ріні, «наша розумна довіра до свідчень інших людей, на диво, залежить від регулятивних наслідків постійно доступної можливості записів подій, про які вони свідчать. Оскільки дідфейки руйнують епістемну цінність записів, ми можемо зіткнутися з ще більш глобальними наслідками надійності самих наших свідчень»¹¹⁸. І якщо сьогодні тільки дактилоскопія часто схожа більше на мистецтво, ніж на науку¹¹⁹, то вже незабаром на подібне мистецтво може перетворитись встановлення фіктивності чого чи того відео для суду. Це потенційно

¹¹⁶ Fallis D. The Epistemic Threat of Deepfakes. *Philos. Technol.* 2020. URL: <https://doi.org/10.1007/s13347-020-00419-2>.

¹¹⁷ Schwartz B. The Paradox of Choice : Why More Is Less. New York : Ecco, 2004. 265 p.

¹¹⁸ Rini R. Deepfakes and the Epistemic Backstop. *Philosophers' Imprint.* 2020. 20 (24). P. 1-16.

¹¹⁹ Криміналісти в растерянности: можно ли верить отпечаткам пальцев? DW: Made for minds : веб-сайт. URL: <https://cutt.ly/7hI9LD2> (дата звернення: 10.09.2021).

ставить під загрозу сам спосіб судочинства, судову систему, систему отримання свідчень, виконання покарань тощо. Крім того, в багатьох країнах світу законодавство регулює наслідки публічного поширення реальних, наприклад, приватних відео, однак воно не поширюється на згенерований контент, зокрема за допомогою технології deepfake, що, звісно, потребує перегляду законодавчої бази з урахуванням нових реалій.

Нерозв'язаною залишається проблема анонімності творців дипфейк-контенту, що унеможливає притягнення їх до відповідальності. Якщо додати до цього складну систему юридичних зв'язків між потенційними жертвами, платформами, де розміщено контент, надавачами інтернет-послуг, розробниками софту для створення дипфейків та інших суб'єктів, то з погляду права, немає сподівань на швидке покарання винних та відповідну компенсацію збитків. Все це потребує появи більш ефективних механізмів судової практики.

Одна з проблем, яка потребуватиме розв'язання з поширенням застосування цієї технології – як забезпечити людей від негативних наслідків використання відео-контенту, згенерованого за її допомогою й водночас не зашкодити технологічному розвитку, не обмежувати розгортання його конструктивних можливостей: створення культурного продукту, удосконалення сфери послуг і розваг тощо. З погляду етики, на цьому шляху можна втрапити в пастку «подвійної моралі»: коли кожна конкретна ситуація може трактуватись за допомогою абсолютно різних світоглядних позицій чи наборів правил, що буде призводити до протилежних, несумісних висновків, які своєю чергою створюватимуть подальші непорозуміння та правові колізії як на рівні внутрішніх законів кожної окремої країни, так і на рівні міжнародних відносин між суб'єктами світової політики. Не складно передбачити, що культурні, історичні, ментальні розбіжності й конфлікти сприятимуть посиленню ефекту подвійної моралі, що поглиблюватиме напруту як між окремими країнами, так і цілими регіонами.

Використання діалектичного методу дозволяє припустити, що, поряд із подібним зволіканням (неможливістю зробити вибір тут-і-тепер) – неактивна фаза, повинна існувати й реактивна фаза – коли використання дипфейку призводить до негайних і рішучих дій у локальному (одна людина) чи глобальному (група людей) масштабах. Головна загроза в цьому випадку – те, що зараз немає жодної можливості швидко спростувати дипфейк (провести експертизу, ефективно і в повному обсязі донести її результати до громадськості тощо) і в такий спосіб зупинити реакцію спільноти на нього. А історія в Бангладеш¹²⁰, доводить, що навіть швидке спростування чи публічні заяви офіційної влади про підробку контенту не мають вирішального значення.

¹²⁰ CNN повідомляє про заворушення та смерті людей після допису у Facebook, який нібито критикував Пророка Мухамеда. Ця справа показує, що, незважаючи на зацікавленість влади у мирному розв'язанні проблеми, швидке оприлюднення інформації про злам сторінки користувача, від імені якого був поширений допис – у подібних ситуаціях, в тому числі із застосуванням дипфейків – заворушень і інших руйнівних наслідків, уникнути, найімовірніше, не вдасться. Див.: Gupta S., Regan H. Four dead in Bangladesh riot

Одним зі способів запобігання подібним ситуаціям постає маркування синтетичного медіапродукту на основі вимог законодавства, які ще потрібно ухвалити. Хоча обмеження для цифрового контенту вже існують¹²¹, 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 підрбок і попри перестороги з боку технічних служб, не забезпечить 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.

Як зазначає американська журналістка та дослідниця даних К. Хао з покликанням на Б. Періса, кейс Н. Пелосі, у відео з якою не було внесено додаткової інформації, а маніпуляції відбулись лише зі швидкістю його відтворення, показує, що методи виявлення діпфейк-контенту в цьому випадку були б неефективними. «Замість цього журналісти повинні були їх спростувати, а це означає, що люди повинні довіряти журналістам»¹²². Тому одних лише зовнішніх засобів ніколи не буде достатньо для ефективного

over offensive Facebook post. CNN.
URL: <https://edition.cnn.com/2019/10/21/asia/riot-deaths-facebook-post-intl-hnk/index.html> (дата звернення: 10.09.2021).

¹²¹ 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: 10.09.2021).

¹²² Hao K. The biggest threat of deepfakes isn't the deepfakes themselves. *MIT Technology Review*. URL: <https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/> (дата звернення: 10.09.2021).

розпізнавання подібного контенту, який з часом ставатиме все більш досконалим. Наприклад, для спростування створеного спеціалістами з МІТ діпфейк-відео зі зверненням Р. Ніксона¹²³ (зауважмо, що це та інші подібні відео використовують для демонстрації можливостей технології, а не задля введення людей в оману), достатньо звернутись до шкільного курсу історії і використати набір раніше отриманих даних і зіставити їх із побаченим у відео. З формально-логічного погляду, інформація, представлена у шкільному курсі історії звучить так: «Астронавти місії Apollo-11 повернулись на Землю» – ($S \in P$). Тоді як інформація, поширена у відео, може бути представлена у вигляді тези: «Астронавти місії Apollo-11 не повернулись на Землю» – позначимо її як ($S \notin P$). У підсумку ми отримуємо два суперечні судження, які не можуть бути одночасно істинними. Коли ($S \in P$) – істинне, то ($S \notin P$) – обов'язково хибне і навпаки – *tertium non datur*¹²⁴ (закон виключеного третього).

Будь-який діпфейк-контент можна розглядати як помилку, внаслідок порушення закону виключеного третього. У світі цього твердження, «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 вклучає 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 може стати спосіб конструювання даних, заснований на концепції фальсифікації К. Поппера. Як показав К. Поппер, на основі так званих одиничних ствердних суджень – суджень, у яких мовиться про приписування одному суб'єкту (x) певної властивості (P) неможливо побудувати загальноствердне судження $\forall x P(x)$, яке було б гарантовано істинним за умови невизначеної кількості одиничних логічних суб'єктів. Отже, наявність навіть значної кількості спостережуваних явищ q не гарантує істинного результату Q , за умови, що спостерігачу недоступний весь набір спостережуваних явищ q . Що дає змогу спершу припустити хибність будь-якого q у всій множині явищ, а тоді спростувати і загальне судження про ці явища Q .

На початку ми писали про потенційну загрозу діпфейк-відео для дітей та підлітків у вигляді відео-челенджів (викликів). Спробуємо продемонструвати як може працювати механізм їхнього спростування. Припустімо, що в мережі з'явився черговий челендж, який закликає підлітків до чогось небезпечного, але з чим тотальна більшість людей ніколи не мала справи. Відсутність емпіричного досвіду тут важлива, оскільки не дозволяє одразу віднайти такі одиничні факти, які б спростовували наш челендж Q на власному досвіді. Наприклад: у мережі з'являється серія діпфейк-відео, де три молодіжні кумири

¹²³ 'Deepfake' video shows Nixon talking about moon mission disaster.
URL: <https://www.youtube.com/watch?v=EAQFI8PLvUA10>. 'Deepfake' video shows Nixon talking about moon mission disaster.
URL: <https://www.youtube.com/watch?v=EAQFI8PLvUA> (дата звернення: 10.09.2021).

¹²⁴ Третього не дано.

самі п'ють і закликають інших пити ртуть із домашніх термометрів, переконуючи глядачів в її цілющих властивостях. Для споживачів такого контенту він може видатись, на перший погляд, несуперечливим, оскільки в деяких комп'ютерних іграх, вигаданих світах ртуть може застосовуватись як корисний матеріал, який допомагає перемогти у грі. Отже, маємо просту структуру Q ($q_1 \& q_2 \& q_3$) – пити ртуть можуть всі, в тому числі й ти, оскільки твої кумири 1, 2 і 3 п'ють ртуть на відео. Для того, щоб перевірити цю конструкцію на універсальність, необхідно піддати її сумніву – пошукати одиничні приклади, які б доводили, що вона не універсальна. Наприклад, діти та підлітки могли б згадати, що на уроках з безпеки життєдіяльності чи екології вони чули про шкідливість ртутних ламп для людини і довкілля або про важкі наслідки отруєння парами ртуті золотопукачів тощо. На підставі таких прикладів формується умовивід: «Є випадки, коли ртуть шкодить людям. Оскільки я людина, вона може бути шкідлива і для мене. Отже те, в чому мене намагаються переконати, не є правдою».

Такий підхід можна вважати універсальним, оскільки для виникнення підозри в маніпуляціях з боку подібного контенту досить відшукати хоча б один приклад, який суперечить наведеному у відео «доказам». Однак, пивидше за все, простої перевірки фактів не буде достатньо, необхідно буде вибудувати причинно-наслідкові зв'язки для побудови альтернативної історії (розповіді), яка логічно й послідовно могла б спростувати дівфейк, показати його нелогічність. Це потребуватиме від кожного, хто взаємодіяв з дівфейк-контентом, значних когнітивних зусиль.

Можна припустити, що протистояння дівфейк-контенту буде ефективним на рівні колективних зусиль – коли люди, які розпізнали підробку допомагатимуть іншим, які ще цього не зробили. Зменшуючи в такий спосіб потенційну шкоду на третьому рівні загрози – суспільному (перші два – індивідуальний і груповий). З боку авторів дівфейк-контенту, відповіддю може стати створення ботоферм, які будуть допомагати просуванню підробок, створюючи інформаційний шум і не дозволяючи названим колективним зусиллям досягати успіху.

Отже, особливої уваги потребує не лише проблема продукування дівфейк-контенту, а й способи його поширення та потенційного захисту, а також специфіка 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.

Корисною, на нашу думку, у боротьбі з дівфейк-контентом може бути герменевтична методологія. Зародившись як мистецтво тлумачення текстів – екзегетика, вона поступово доповнювала свою предметну сферу ширшим колом об'єктів. Тлумачення складних місць Святого Письма згодом доповнилось потребою обстоювати власну думку у міжконфесійних богословських суперечках. Зрештою, герменевтика сформувалася як повноцінний філософський метод аналізу не лише тексту, а й інших видів повідомлень. Тлумачення поняття тексту також зазнало змін.

П. Рікер, який сконструював концепт «герменевтика підозри», вбачав її ключову роль у «запереченні можливості довіри до буквального і безпосереднього рівня тексту, який розглядається передусім як засіб маскування справжніх політичних інтересів, волі до влади, прихованих бажань і страхів. Мету інтерпретації П. Рікер вбачає у виявленні факту приховування у тексті владних інтересів та їх наступній демістифікації»¹²⁵. Для використання цієї методології потрібне розширене розуміння тексту – не як групи речень, зафіксованої на матеріальному носії, а як набір символів, які конструюють певне смислове поле. Отже, відео також може бути об'єктом інтерпретації.

К. Браян зазначає, що «словосполучення «герменевтика підозри» – це тавтологічний спосіб сказати те, що люди, які думають, завжди знали: що слова не завжди можуть означати те, що вони, як нам здається, означають»¹²⁶. З одного боку К. Браян, натякає на очевидність і принципову обмеженість цього підходу, з іншого – на можливість застосування «герменевтики підозри» за наявності в тексті специфічних форм вираження думки або коли в ньому неявно міститься прихована інформація. Тоді цей підхід постає необхідним інструментом експлікації неочевидного. «Іноді «герменевтика підозри» може бути важливою [...], коли ми підозрюємо, що тексти не кажуть нам всієї правди. Прикладом може слугувати те, що тексти Нового Завіту розповідають нам про роль жінок в ранньому християнстві (звісно, у виразі патріархальних уявлень), або те, що вони розповідають нам про “євреїв” і “фарисеїв” (явно під впливом сварок між ранньою церквою та синагогою)»¹²⁷.

Використання «герменевтики підозри» може бути корисним для подолання тотальної недовіри, яка зростає з існуванням технології «deepfake» і збільшенням скептичних настроїв пропорційно до поширення інформації про її можливості. Ефективність застосування такого методу полягає в постановці додаткових питань, пов'язаних з досліджуванним феноменом, зокрема відеоконтентом. Замість постійного сумніву – скептицизму, до якого нас схиляє припущення про те, що перед нами може бути дипфейк, «герменевтика підозри» пропонує змістити фокус з подій, які відтворюються на відео, на події, що передували його створенню. 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: 1) першоджерелом, де з'явилось відео: «Чому було обрано саме це джерело поширення інформації?»; 2) особою автора: «Яка особа його оприлюднила?», «Анонімна чи ні?»; 3) часом його появи: «Чи пов'язана поява цього відео, до прикладу, з піком діяльності його фігуранта?», «Чому саме зараз воно з'явилось?»; 4) особою, яка отримує дивіденди від його появи: «Хто виграє від дискредитації фігуранта скандалу?»; зрештою – 5) звернути увагу на себе як реципієнта – того,

¹²⁵ Соловій Р. Рецепція ідей «герменевтики підозри» Поля Рікера у філософії релігії Меролла Вестфала. *Sententiae*, 2013, № 29 (2). С. 91-100.

¹²⁶ Bryan C. The Hermeneutic of Suspicion.

URL: <https://oxford.universitypressscholarship.com/view/10.1093/acprofoso/9780199336593.001.0001/acprof-9780199336593-chapter-4> (дата звернення: 10.09.2021).

¹²⁷ Там само.

хто сприймає інформацію: «Як вона вплине на мене, на мій вибір тощо?», і в підсумку – б) поставити питання про «дивіденди брехуна»: «У чому полягає мета можливої маніпуляції?».

г) Мистецтво та культура як контексти використання дідфейк-технології

Технологія, яка так бурхливо розвивається і є настільки універсальною, не може мати лише руйнівний потенціал. Її з успіхом можна застосовувати у створенні розважального продукту, візуалізації історичних хронік, кіно з використанням образів померлих акторів, інших відомих людей. Крім того, технологія уже сьогодні дозволяє покращувати контент, який генерується у прямому ефірі, шляхом накладання реалістичної міміки під час синхронного перекладу чи використання в режимі live зображень-аватарів замість реальних тощо.

Ключовою перевагою цієї технології в контексті роботи з відео на рівні Голівуду є значно краще співвідношення ціна-якість. Голівуд і раніше «оживляв» акторів чи маніпулював їхніми тілами, однак це обходилося досить дорого, оскільки потребувало розробки нових технологічних ресурсів із залученням значної кількості вузькопрофільних спеціалістів. Проте вже сьогодні є реальні підстави сподіватись на «демократизацію» цієї технології і, відповідно, збільшення частки її застосування для створення цілком реалістичного кінопродукту. Наприклад, бельгійський художник з відеоефектів К. Уме, анонімно створивши відео з Т. Крузом, змусив мільйони людей сперечатись з приводу достовірності цих відео¹²⁸. Водночас цей контент, незважаючи на свою унікальність, не розвіяв сумнівів скептиків щодо можливості його масового застосування. Хоча К. Уме і застосовував програмне забезпечення з відкритим програмним кодом, це потребувало унікальних навиків і місяців роботи, пов'язаної з «навчанням» нейромережі, а також десятків тисяч фото реального Т. Круза, відзнятих з якомога більшої кількості ракурсів. Крім того, як зазначає сам К. Уме, «було б неможливо створити такі правдоподібні дідфейки, якщо б Фішер (актор, який грав у роликах К. Уме Т. Круза) не був би настільки схожим на голлівудського актора, а також не був би настільки хорошим актором, здатним імітувати манеру його поведінки»¹²⁹. Отже, навіть на думку К. Уме, людям не варто хвилюватись щодо масової появи подібних реалістичних відео, яке після машинного створення потребує покадрової постобробки за допомогою класичного програмного забезпечення – роботи з покращення статичного зображення. Водночас це не знімає ключової, на наш погляд, проблеми – довіри чи сумнівів суспільства з приводу готового продукту.

¹²⁸ Deepomcruise.

URL: https://www.tiktok.com/@deepomcruise?lang=en&is_copy_url=1&is_from_webapp=v1 (дата звернення: 10.09.2021).

¹²⁹ Kahn J. Deepfake master behind those viral Tom Cruise videos says the technology should be regulated. URL: <https://fortune.com/2021/03/05/tom-cruise-deepfake-creator-technology-should-be-regulated/> (дата звернення: 10.09.2021).

Оптимізму щодо масового використання дідфейк-контенту Голівудом додала промокампанія нового фільму Warner Bros. – «*Reminiscence*» («*Спогади*»)¹³⁰. «Голівуд сприймає дідфейки і ми всі можемо бути їх частиною: Warner Bros. звернулась до старту синтетичних медіа D-ID для просування свого нового фільму [...] Новий веб-сайт дозволяє кожному завантажити фотографію, яку штучний інтелект D-ID перетворює в рухоме відеопослідовність дідфейка в короткому відеокліпі, який рекламує фільм»¹³¹. Подібні промо, з одного боку, вражають доступністю (не потрібно навіть завантажувати ПЗ на свої пристрої), з іншого – принциповою можливістю створити щось подібне на основі лише однієї фотографії. Це, звісно, додає оптимізму прихильникам використання цієї технології в індустрії розваг загалом.

Проте є ще одне застереження, пов'язане з використанням цифрових алгоритмів та їхньою здатністю змінити сприйняття реальності реципієнтами. Не секрет, що актори та Голівуд завжди впливали на моду та зміну культурних патернів загалом¹³². Окремі кінострічки та образи акторів впливали на покоління людей, які прагнули бути подібними до них. Зв'язок мистецтва, моди, індустрії розваг можна проілюструвати на прикладі творчості І. Сен-Лорана¹³³. І яким би незвичним чи запуганим не був цей зв'язок – джерелом творчості та натхнення завжди була людина. З появою новітніх технологій все може змінитись – штучний інтелект нині має змогу сортувати або «підказувати» людині «правильний» напрям руху в мистецтві. На початку 2020 р. Warner Bros. підписала угоду з Cinelytic щодо використання алгоритмів штучного інтелекту під час ухвалення рішень про запуск того чи того продукту на основі великої кількості змінних, адаптованих до окремих регіонів світу. Представники компанії запевняють, що кінцеве рішення буде залишатись за людиною, проте складно повірити, що подібні алгоритми не будуть серйозно впливати на передумови цих рішень, якщо надана ними інформація буде суперечити фінансовим очікуванням босів компанії. Отже, замість того, щоб змінювати дійсність, впливаючи на смаки людей, сучасна кіноіндустрія може під них підлаштовуватись.

Інший вектор розвитку технології «deepfake» скерований на інтерактивну взаємодію. Наприклад, проєкт «*Dali Lives*»¹³⁴ пропонує взаємодіяти з дідфейком С. Далі, створеним на основі відзнятого під час життя контенту з митцем. Сьогодні класичне питання: «З якою історичною постаттю ви хотіли

¹³⁰ The AI Face Platform. URL: <https://www.d-id.com/> (дата звернення: 10.09.2021).

¹³¹ Roetgers J. Warner Bros. Is Using Personalized Deepfakes For Its Latest Movie Promo. URL: <https://www.protocol.com/newsletters/next-up/warner-bros-deepfake-movie-promo?rebellitem=1#rebellitem1> (дата звернення: 10.09.2021).

¹³² Sardone A. Hollywood's influence on fashion. URL: <https://www.universityoffashion.com/blog/hollywoods-influence-on-fashion/> (дата звернення: 10.09.2021).

¹³³ The Mondrian collection of Yves Saint Laurent. URL: https://en.wikipedia.org/wiki/The_Mondrian_collection_of_Yves_Saint_Laurent (дата звернення: 10.09.2021).

¹³⁴ Behind the Scenes : Dali Lives. URL: <https://www.youtube.com/watch?v=BiDaxl4xqJ4> (дата звернення: 10.09.2021).

би зустрілись?» – може втілитися в життя саме завдяки цій технології. Підсилює ефект інтерактивної взаємодії «пропозиція митця» зробити спільне селфі. Очевидно, подібний вектор здатен зацікавити людей та зробити відвідування музеїв, інших закладів культури більш привабливим, особливо для молоді. Актуальними подібні аспекти застосування технології постають і для України, де на 30-у річницю Незалежності держави «заспівав» Кузьма Скрыбін¹³⁵.

Відкритим залишається питання щодо моральної сторони подібних заходів. Якщо образ С. Далі та інших відомих осіб і після смерті є ввічаними, їх можна назвати публічним надбанням, то деталі «воскресіння» звичайних людей – досить запутані. Сфера, де переплітаються релігійні, юридичні, звичаєві, культурні відносини – надзвичайно складна, і передбачити наслідки впровадження цієї технології в ній майже неможливо. Можна лише припустити, що реакція суспільства повинна значно корелювати з його світоглядом і поглядами на смерть. Наприклад, в Південній Кореї, де приблизно половина населення не вважають себе релігійними, і близько 15% – буддисти, документальний фільм під назвою «*Зустріч з тобою*»¹³⁶ отримав схвальні відгуки, однак, ризикнемо припустити, що в більш консервативному культурному середовищі сприйняття подібного контенту було б не таким однозначним.

Ще одним вектором застосування технології постає створення мистецьких об'єктів. Цікавою є робота мисткині Д. Верінг, яка осмислює «ідентичність, саморозкриття та сучасну медіа культуру», ставить за мету «дослідити суперечності між суспільним та приватним життям, прагнення розкрити власні секрети та знати секрети інших, а також розмиті межі між задокументованою та сконструйованою позицією»¹³⁷. У такий парадоксальний спосіб технологія дозволила під новим кутом зору подивитись на проблеми, пов'язані з довірою до інформації, за допомогою технології, яка цю довіру здатна остаточно зруйнувати.

Л. Флоріді у праці «*Штучний інтелект, дінфейки та майбутнє екстени*»¹³⁸ розкриває присутність подібних технологій в мистецтві з іншого боку – як таких, що здатні змінити напі уявлення про «справжність» та «оригінальність». Проєкт «*Next Rembrandt*»¹³⁹ покликаний переглянути наше розуміння мистецтва та дозволяє буквально переосмислити аспекти смерті автора та цінності творів мистецтва загалом. Професор філософії та етики інформації

¹³⁵ На концерті до Дня Незалежності “заспівав” віртуальний Кузьма Скрыбін. URL: <https://kyiv.media/news/na-konzerti-do-dnya-nezalezhnosti-zaspivav-virtualnyj-kuzma-skryabin> (дата звернення: 10.09.2021).

¹³⁶ Park M. South Korean mother given tearful VR reunion with deceased daughter. URL: <https://www.reuters.com/article/us-southkorea-virtualreality-reunion/south-korean-mother-given-tearful-vr-reunion-with-deceased-daughter-idUSKBN2081D6> (дата звернення: 10.09.2021).

¹³⁷ Beer J. This new deep fake video is both advertising and a piece of art. URL: <https://www.fastcompany.com/90279597/this-new-deep-fake-video-is-both-advertising-and-a-piece-of-art> (дата звернення: 10.09.2021).

¹³⁸ Floridi L. Artificial Intelligence, Deepfakes and a Future of Ecstypes. *Philos. Technol.* 2018. 31. P. 317-321. <https://doi.org/10.1007/s13347-018-0325-3>.

¹³⁹ The Next Rembrandt. URL: <https://www.nextrembrandt.com/> (дата звернення: 10.09.2021).

Л. Флоріді пропонує називати подібні роботи «ектипами». Ектип – «це копія, але не будь-яка, а, радше копія, яка має особливий стосунок до свого джерела, архетипу [...] Це не оригінальна річ, але вона значним і достовірним чином пов'язана зі справжньою річчю»¹⁴⁰. Подібні проєкти потенційно здатні не просто змінити баланс сил на мистецькому ринку, а й змінити чинники, за якими відбувається оцінка творів мистецтва. «Чи може, до прикладу, визначальну роль грати смерть «автора картини» або витрати в тому числі й часові, пов'язані з виробництвом ектипу?» – сьогодні на такі питання немає однозначної відповіді.

Л. Флоріді пропонує розділяти автентичні неоригінальні та неавтентичні оригінальні ектипи – приклад з Рембрандтом належить до першого виду ектипів. Другим Л. Флоріді вважає останню промову Д. Кеннеді, що так і не була ним проголошена¹⁴¹. «Незважаючи на те, що це була звичайна промова [...], вона несподівано потрапила в заголовки новин. Це була промова для Dallas Trade Mart від 22 листопада 1963 р., яку Кеннеді повинен був прочитати, якби не був вбитий за кілька хвилин до того. Текст оригінальний – взятий з першоджерела. Але голос, який декламує, не є автентичним, оскільки був синтезований програмним забезпеченням, яке проаналізувало 831 запис промов Кеннеді, щоб «навчитись» говорити як він»¹⁴². З погляду комунікативного потенціалу, ця промова не містить жодного повідомлення, оскільки не має свого реципієнта, не здатна здійснювати вплив тощо. З погляду історичного дослідження, також немає потреби в її озвучуванні. Чим же постає цей продукт, окрім своєрідної «розваги»? Чи є моральним таке розважання? З одного боку, воно цілком нейтральне, оскільки не містить добрих чи злих намірів, не приносить ні користі, ні шкоди. А з іншого – це обман, спотворення дійсності, яке на відміну від мистецтва, не володіє самоцінністю як витвір людського духу, реалізації творчих інтенцій. Діпфейк – продукт складного, але лише технологічного процесу, який, у цьому випадку, не характеризується ні оригінальністю, ні новизною.

Отже, розвиток і поширення сфер використання технології «deepfake» викликає чимало дискусій з приводу різного роду можливостей і загроз, пов'язаних з виробництвом нових синтетичних медіапродуктів, способом їх розповсюдження, впливом на окремих індивідів, груп населення і суспільство загалом. У фокусі уваги філософів – передусім епістемні й етичні проблеми цього нового типу інформації (чи дезінформації). До епістемних викликів належить проблема істини, яка набуває нових значень, поширення релятивізму та скептицизму, до етичних – проблема відповідальності та морального і правового регулювання, захисту прав найбільш вразливих категорій

¹⁴⁰ Floridi L. Artificial Intelligence, Deepfakes and a Future of Ectypes. *Philos. Technol.* 2018. 31. P. 317-321. <https://doi.org/10.1007/s13347-018-0325-3>

¹⁴¹ Artificial intelligence used to recreate JFK's Dallas speech that he never gave. URL: https://www.washingtonpost.com/video/business/technology/artificial-intelligence-used-to-recreate-jfks-dallas-speech-that-he-never-gave/2018/05/10/caad5f8a-3f33-11e8-955b-7d2e19b79966_video.html (дата звернення: 10.09.2021).

¹⁴² Floridi, L. Artificial Intelligence, Deepfakes and a Future of Ectypes. *Philos. Technol.* 2018. 31. 317–321. <https://doi.org/10.1007/s13347-018-0325-3>

населення та суспільства загалом. Виявлені головні запобіжники негативного впливу діпфейків на особистість і суспільство: зовнішні – у вигляді засобів розпізнавання, які будуть з'являтися з розвитком самої технології, яка як і будь-який інший феномен науково-технічного прогресу має два боки: потенціал для породження нових проблем і засоби для їхнього розв'язання; та внутрішні – у вигляді освіти й розвитку критичного мислення як її неодмінного елементу. Водночас наявні значні сподівання на креативний потенціал застосування названої технології, втілення її культуротворчих можливостей.

НАУКОВЕ ВИДАННЯ

ФЕНОМЕН БЕЗПЕКИ:
соціально-гуманітарні виміри

За загальною науковою редакцією
Віталія Мудракова

колективна монографія

Підписано до друку 17.05.2022 Формат 60*84/16
Наклад 200 прим. Зам. № 6/22 Обл.-вид. арк. 19,3
Надруковано: ФОП Мельник А.А.
м. Хмельницький, вул. Чорновола, 37

