МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МАРІУПОЛЬСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ ФІЛОЛОГІЇ ТА МАСОВИХ КОМУНІКАЦІЙ
КАФЕДРА СОЦІАЛЬНИХ КОМУНІКАЦІЙ

До захисту допустити:
Завідувач кафедри
_____ ___Іванова Т. В._____
(підпис) *(ПІБ завідувача кафедри)*
«_____»_____2025 р.

«ТЕХНОЛОГІЇ DEEPFAKE ЯК ІНСТРУМЕНТ МАНІПУЛЮВАННЯ АУДИТОРІЄЮ: РОЛЬ ЖУРНАЛІСТА У ВИКРИТТІ ТА ПРОТИДІЇ ДЕЗІНФОРМАЦІЇ»

Кваліфікаційна робота здобувача вищої освіти першого (бакалаврського) рівня вищої освіти освітньо-професійної програми «Журналістика та соціальна комунікація»_____
(назва освітньо-професійної програми)

Чулкова Артура Володимировича_____
(прізвище, ім'я, по батькові здобувача вищої освіти)

Науковий керівник:
завідувач кафедри соціальних комунікацій, доктор педагогічних наук, професор Іванова Т. В.
(прізвище, ініціали, науковий ступінь, вчене звання)

Рецензент: Юричко А. В., к. ф. н., асистент кафедри друкованих медіа та історії журналістики Навчально-наукового інституту журналістики КНУ імені Тараса Шевченка
(прізвище, ініціали, науковий ступінь, вчене звання, місце роботи)

Кваліфікаційна робота захищена
з оцінкою _____
Секретар ЕК _____
«_____»_____2025 р.

Київ – 2025

ЗМІСТ

ВСТУП

Актуальність дослідження. У сучасному цифровому світі розвиток технологій відкриває нові можливості для комунікації, але водночас створює загрози, пов'язані з маніпуляціями інформацією. Однією з таких технологій є deepfake — методика, що дозволяє створювати фальшиві відео та аудіозаписи, які важко відрізнити від справжніх. Ця технологія використовується як у розважальних цілях, так і для поширення дезінформації, що особливо небезпечно у політичному, соціальному та інформаційному контекстах.

Готовий медіапродукт представляє собою сайт «DeFake It!», який є освітнім ресурсом, спрямованим на підвищення обізнаності про deepfake, навчання аудиторії методам їхнього виявлення та протидії. Актуальність цього проєкту обумовлена тим, що суспільство все більше стикається з маніпуляціями через цифровий контент, а розпізнавання deepfake є важливою частиною сучасної медіаграмотності. Особливо це стосується журналістів, які повинні оперативно й точно виявляти фальшиві матеріали, щоб запобігти поширенню дезінформації.

Сайт має на меті не лише надати інструменти для перевірки контенту, але й навчити користувачів, як самостійно аналізувати зображення, відео та аудіо. Це сприятиме підвищенню критичного мислення серед аудиторії та зміцненню довіри до медіа.

Дослідження у сфері «глибоких фейків» задокументовують значне зростання кількості підробленого контенту в Інтернеті за останні кілька років. Відповідно до звіту «Home Security Heroes» за 2023 рік, було виявлено, що кількість підроблених відео з 2019 року зросла на 550% [11, с. 10]. У звіті також підкреслюється, що 98% цих відео були порнографічними, причому 99% були націлені на жінок. За даними DeepMedia, у 2023 році кількість підроблених відео та голосових записів, поширених у соціальних мережах у всьому світі, сягнула приблизно 500 000. Очікується, що до кінця 2025 року ця цифра зросте до 8 мільйонів, що свідчить про подвоєння кількості фейкового контенту кожні півроку. [26]

Опитування 2022 року, проведене організацією Media Literacy Now, показало, що 46% дорослих у віці від 19 до 81 року не здобували знань з медіаграмотності під час навчання в середній школі. Відсутність формального навчання може спричинити труднощі з критичною оцінкою інформації та розпізнаванням дезінформації [35]. У статті Liminal за 2024 рік висвітлюється зростаюча загроза глибоких фейків у кібербезпеці, зазначається, що ринок виявлення глибоких фейків, за прогнозами, зросте з $5,5 млрд у 2023 році до $15,7 млрд у 2026 році. Таке зростання відображає збільшення попиту на засоби для боротьби із загрозами, пов'язаними з глибокими підробками. [40]

З моменту своєї появи наприкінці 2010-х років технологія deepfake, яка передбачає використання штучного інтелекту для створення реалістичних, але сфабрикованих аудіо, відео чи зображень, стрімко розвинулася. Термін «deepfake» поєднує в собі слова «deep», що характеризує глибоке навчання — підмножину штучного інтелекту, і «fake», що вказує на оманливу природу контенту. Цю етимологію відзначає Організація з безпеки соціальних мереж, яка стверджує, що «глибокі фейки» — це «підроблені або фальшиві відео, створені за допомогою глибокого навчання, форми штучного інтелекту». Спочатку привернувши увагу завдяки несанкціонованому створенню відвертого контенту за участю публічних осіб, «дипфейки» згодом поширилися на різні сфери, що викликало значне занепокоєння щодо їхнього потенційного зловживання.

Словник Merriam-Webster визначає «глибокий фейк» як «зображення або запис, які були переконливо змінені та підтасовані, щоб представити когось як такого, що робить або говорить щось, що насправді не було зроблено або сказано». Цей термін виник у 2017 році, коли користувач Reddit під ніком «deepfakes» почав ділитися маніпулятивними відео, переважно з несанкціонованим відвертим контентом за участю публічних осіб. Це походження задокументоване Массачусетським технологічним інститутом Слоуна, який зазначає, що термін був вперше вигаданий наприкінці 2017 року користувачем Reddit з таким же ім'ям.

На основі аналізу досліджень як вітчизняних, так і зарубіжних науковців, а саме: «Діпфейк та дезінформація» (А. Вальорська) [4], «Медіаграмотність та критичне мислення в процесі викладання дисциплін комунікаційного циклу» (Т. Іванова) [5], «Deepfake: визначення, показники ефективності та стандарти, набори даних та метаогляд» (Е. Алтунку, В. Франкейра, Ш. Лі) [11], де автори розглядають природу технології deepfake та її наслідки, було зроблено висновок, що виявлення та інформування про deepfake є важливим для сучасного медіапростору.

Об'єкт дослідження — процес створення та поширення контенту за допомогою технологій deepfake.

Предмет дослідження — методи виявлення deepfake й створення освітнього сайту, який навчає розпізнавати маніпуляції.

Мета дослідження — розробка та реалізація інтерактивного сайту «DeFake It!», що надає інструменти для аналізу та перевірки deepfake, а також популяризація медіаграмотності серед цільової аудиторії.

Відповідно до поставленої мети в роботі розв'язано такі завдання:

1. дослідити технології створення deepfake та їхній вплив на суспільство;
2. проаналізувати сучасні методи виявлення deepfake, доступні журналістам та користувачам;
3. визначити концепцію, структуру та цільову аудиторію сайту;
4. розробити інтерактивний контент, що включає покрокові інструкції, тести та аналіз реальних кейсів;
5. запропонувати стратегії просування сайту «DeFake It!» для залучення аудиторії.

Наукова новизна дослідження полягає в тому, що вперше запропоновано створення інтерактивного ресурсу, орієнтованого на виявлення та аналіз deepfake у контексті сучасної журналістики.

Практична значущість дослідження:

1. Сайт «DeFake It!» сприятиме підвищенню обізнаності про небезпеку deepfake та способи їх виявлення.
2. Освітній контент ресурсу дозволить користувачам самостійно розпізнавати маніпуляції, що зміцнить медіаграмотність.
3. Використання сайту стане корисним інструментом для журналістів, студентів та викладачів, які вивчають медіакомунікації.
4. Створений сайт є прикладом інноваційного підходу до популяризації медіаграмотності, що може бути адаптовано для інших навчальних ініціатив.

Структура та обсяги кваліфікаційної роботи. Робота складається зі змісту, вступу, теоретичної частини, практичної частини, висновків до кожного розділу, загального висновку та списку використаної літератури. Зміст роботи викладено на 73 сторінках комп'ютерного тексту, включаючи 13 додатків, 3 таблиці й 3 рисунки.

РОЗДІЛ 1. ГЛИБОКІ ФЕЙКИ: ЗАГРОЗИ, МОЖЛИВОСТІ ТА РЕАЛІЇ СЬОГОДЕННЯ

1.1. Еволюція технологій створення deepfake

Поява технології «глибоких фейків» стала значним поворотним моментом в еволюції синтезу медіа, керованого штучним інтелектом. Хоча методи цифрової маніпуляції існують уже десятиліттями, глибокі фейки забезпечують безпрецедентний рівень реалістичності, що робить дедалі складнішим розрізнення автентичного та сфабрикованого контенту. Термін «глибокий фейк», що походить від слів «глибоке навчання» і «фейк», означає синтетично змінені медіа, створені за допомогою штучного інтелекту, зокрема за допомогою глибоких нейронних мереж, таких як генеративні змагальні мережі (GAN). На відміну від традиційного редагування зображень і відео, яке ґрунтується на ручних змінах, підробки використовують ШІ для автономного створення гіперреалістичного контенту, що імітує людські вирази обличчя, голоси і рухи з надзвичайною точністю.

Хоча сам термін «глибокий фейк» отримав широке визнання лише у 2017 році, його технологічні основи сягають набагато глибших часів. Від ранніх комп'ютерних зображень у голлівудських фільмах до появи машинного навчання в обробці зображень, технологія deepfake розвивалася завдяки поступовому вдосконаленню синтетичних медіа, морфінгу обличчя та аудіовізуальному синтезу, керованому штучним інтелектом. Розуміння цієї еволюції дає змогу зрозуміти як технологічні прориви, що уможливили появу «глибоких фейків», так і зростаючі етичні проблеми, пов'язані з їхнім використанням у медіа, політиці та кіберзлочинності. [34, с. 34]

До того, як технологія «глибоких фейків» набула масового поширення, різні інструменти цифрового редагування та алгоритми, керовані штучним інтелектом, заклали основу для синтетичних маніпуляцій у медіа. Епоха, що передувала дефейкам, характеризувалася ручними та напівавтоматизованими методами, які вимагали значного людського втручання, професійних навичок і спеціалізованого

програмного забезпечення. Ці методи, хоча й дозволяли створювати переконливі вигадки, не мали можливостей адаптивного навчання, якими володіє сучасний штучний інтелект, що генерує глибокі фейки.

Однією з найдавніших форм маніпуляцій у медіа була фоторетуш — техніка, яку широко використовували в журналістиці, рекламі та політичній пропаганді. Задовго до цифрової революції ретуш фотографій виконувалася вручну в фотолабораторіях, де елементи зображень фізично змінювалися за допомогою подвійної експозиції, аерографії та методів вирізання і вставки колажів. Наприкінці 20-го століття ці методи були оцифровані, проклавши шлях до сучасного програмного забезпечення для редагування фотографій і відео, такого як Adobe Photoshop, After Effects і Final Cut Pro, що дозволило робити точні зміни з більшою ефективністю.

Photoshop від Adobe зробив революцію в редагуванні зображень, забезпечивши маніпуляції на основі шарів, що дозволило користувачам легко змінювати фотографії, видаляти об'єкти і навіть реконструювати цілі фрагменти зображень. Хоча спочатку Photoshop призначався для графічного дизайну і комерційного використання, він став потужним інструментом для фотожурналістів, рекламодавців і пропагандистів. [10]

У кіноіндустрії комп’ютерні зображення стали золотим стандартом для створення реалістичних цифрових персонажів, середовищ і візуальних ефектів. Основні досягнення, такі як технологія морфінгу (використана в таких фільмах, як «Термінатор 2: Судний день» (1991)) і захоплення руху (уперше застосована в трилогії «Володар перснів» (2001-2003)), продемонстрували, як рендеринг зі штучним інтелектом може відтворювати людську анімацію і міміку з неймовірною точністю.

Індустрія візуальних ефектів відіграла вирішальну роль у формуванні технологічних попередників глибоких фейків. Такі технології, як заміна обличь, використані у таких фільмах, як «Бунтар один: Історія Зоряних воєн» (2016), для

цифрового відтворення молодих акторів або померлих фігур, дали змогу зазирнути в майбутнє синтезу глибоких фейків на основі штучного інтелекту. Однак, на відміну від сьогоднішніх фейків, ці ефекти вимагали значного людського втручання, ручного покадрового монтажу та високобюджетних виробничих команд.

Тоді як традиційні методи цифрової маніпуляції значною мірою покладалися на ручне редагування та запрограмовані ефекти, впровадження глибинного навчання в обробку зображень і відео ознаменувало собою трансформаційний момент у створенні синтетичних медіа. Прорив, який уможливив появу «глибоких фейків», стався у 2014 році, коли Ян Гудфеллоу, дослідник машинного навчання, розробив генеративні змагальні мережі (Generative Adversarial Networks, GAN) — революційну модель глибокого навчання, яка дозволяє ШІ створювати високореалістичні синтетичні зображення та відео.

До появи GAN медіа, створені штучним інтелектом, страждали від низької реалістичності, видимих спотворень і нездатності точно відтворювати риси обличчя, схожі на людські. Ранні спроби синтетичної генерації зображень часто призводили до розмитих результатів, неприродної структури обличчя або роботизованого синтезу мови. Генеративні змагальні мережі вирішили ці проблеми, дозволивши ШІ самостійно вдосконалювати, адаптувати та оптимізувати свої результати на основі безперервного зворотного зв'язку. Віднині GAN дозволяють створювати реалістичні людські обличчя, навіть генеруючи повністю вигаданих людей, які ніколи не існували (наприклад, вебсайт ThisPersonDoesNotExist.com). [31]

Генеративні змагальні мережі функціонують через подвійну мережеву систему, де дві моделі ШІ — генератор і дискримінатор — змагаються одна з одною в ітеративному процесі навчання. Генератор відповідає за створення синтетичних медіа, таких як фейкові зображення або відео. Спочатку генератор створює неякісні або нереалістичні результати, але з часом, навчаючись на помилках, він вдосконалює свою здатність створювати переконливий контент з високою

роздільною здатністю. Тоді як дискримінатор, виконуючи роль «критика», оцінює, чи є створений контент справжнім або підробленим, порівнюючи його з набором даних справжніх зображень або відео. Якщо результат роботи генератора виявляється фальшивим, він отримує зворотний зв'язок і покращує свою наступну спробу.

Розвиток генеративних змагальних мереж (GAN) у 2014 році заклав основу для сучасної технології «глибоких фейків». Однак лише у 2017 році технології «глибоких фейків» почали привертати увагу широкого загалу, насамперед через онлайн-спільноти, які експериментували з підміною обличь за допомогою штучного інтелекту. Те, що починалося як експериментальна новинка, швидко перетворилося на серйозну технологічну, етичну та політичну проблему, що викликало занепокоєння щодо майбутнього цифрової дезінформації, крадіжки персональних даних та маніпуляцій у ЗМІ. [20]

Дипфейки вперше з'явилися публічно наприкінці 2017 року, коли анонімний користувач Reddit під ніком «deepfakes» почав ділитися згенерованими штучним інтелектом відео, в яких обличчя знаменитостей накладалися на обличчя акторів фільмів для дорослих. З використанням програмного забезпечення з відкритим вихідним кодо, ці відео ставали дедалі переконливішими, що спонукало одночасно до інтересу й тривоги.

На той час створення «глибоких фейків» вимагало базових знань з програмування, але такі моделі штучного інтелекту, як DeepFaceLab і FakeApp, незабаром зробили цей процес широко доступним для нефахівців. За кілька тижнів з'явилися онлайн-форуми, присвячені підробленій порнографії, що призвело до етичних дебатів про цифрову згоду, порушення приватності та потенціал для експлуатації. [22]

У відповідь на суспільний резонанс такі великі платформи, як Reddit, Twitter і Pornhub, заборонили фейкову порнографію на початку 2018 року, але технологія продовжувала розвиватися. Хоча перші додатки для глибокого підроблення

використовувалися переважно для розваг, сатири та неетичного контенту для дорослих, дослідники та журналісти швидко зрозуміли, що цю ж технологію можна використовувати для політичних та соціальних маніпуляцій. [43]

У квітні 2018 року лауреат премії «Оскар» режисер Джордан Піл у співпраці з BuzzFeed створив вірусну фальшивку про колишнього президента США Барака Обаму. На відео здавалося, що Обама виголошує промову, але в дійсності аудіо та рухи губ були підтасовані так, що він говорив те, чого насправді не говорив. На відміну від зловмисних глибоких фейків, проект Піла був розроблений як кампанія з підвищення обізнаності громадськості, що попереджає глядачів про потенційну небезпеку дезінформації, створеної за допомогою штучного інтелекту. Фейк про Обаму викликав дискусії між законодавцями, фактчекерами та дослідниками ШІ, що призвело до появи перших законодавчих пропозицій, спрямованих на виявлення та врегулювання дипфейків. [45]

Демократизація технології deepfake наприкінці 2010-х років стала поворотним моментом у тому, як штучний інтелект змінив цифрові медіа. Те, що колись було експериментальною концепцією, обмеженою академічними дослідженнями та нішевими спільнотами ШІ, раптом стало широкодоступним інструментом, доступним широкому загалу. Цей зсув значною мірою був зумовлений двома паралельними процесами: зростанням кількості підробленого програмного забезпечення з відкритим вихідним кодом і появою комерційних додатків, пристосованих для розваг, маркетингу та ділового спілкування. [19]

Такі компанії, як Synthesia та Rephrase.ai, почали пропонувати згенеровані штучним інтелектом відео з речниками, що дозволило брендам створювати персоналізований маркетинговий контент без необхідності залучення професійних акторів чи студійного виробництва. Поява таких віртуальних інфлюенсерів, як Ліл Мікела, повністю створена штучним інтелектом модель в Instagram, продемонструвала потенціал брендингу на основі глибокого фейку, де синтетичні особистості можуть залучати аудиторію без непередбачуваності людської

поведінки. Більше того, інструменти на основі AI, такі як HeyGen і DeepDub, дозволили брендам безперешкодно перекладати і синхронізувати рекламу кількома мовами, що зробило глобальний маркетинг більш ефективним.

Голлівуд швидко взяв на озброєння інструменти штучного інтелекту в стилі deepfake, щоб омолоджувати акторів, воскрешати померлих виконавців і покращувати візуальні ефекти. Ця технологія спочатку вимагала коштовної та великої ручної роботи з комп'ютерною графікою й візуальними ефектами, але згодом стала значно ефективнішою завдяки синтезу обличчя на основі ШІ.

У таких фільмах, як «Ірландець» (2019), використовували реконструкцію обличчя на основі ШІ для цифрового омолодження Роберта Де Ніро, Аль Пачіно та Джо Пеші, що зменшило залежність від традиційних команд захоплення руху та VFX (візуальної графіки). У франшизі «Зоряні війни» Кері Фішер (принцеса Лея) і Пітер Кушинг (Гранд Мофф Таркін) були воскрешені за допомогою технології глибокого фейку на основі штучного інтелекту, що продемонструвало її здатність продовжувати присутність акторів у кінематографі після їхнього життя. Такі технології, як Resemble AI та ElevenLabs, дозволили кінематографістам синтезувати голоси акторів, що уможливило безперешкодне дублювання та модифікацію діалогів.

Рух за відкрите програмне забезпечення з відкритим вихідним кодом відіграв ключову роль у прискоренні інновацій у сфері дипфейків. Такі платформи, як DeepFaceLab, Faceswap і DeepFake-TensorFlow, дозволили дослідникам, розробникам і ентузіастам ШІ вдосконалювати генерацію синтетичних медіа без обмежувальних комерційних бар'єрів. DeepFaceLab, зокрема, стала найпоширенішим програмним забезпеченням з відкритим вихідним кодом, яке постійно оновлюється глобальною спільнотою, що робить його все більш витонченим і складним для виявлення. Faceswap стала альтернативною платформою для підміни облич за допомогою штучного інтелекту, спрямованою насамперед на освіту та етичні дослідження в галузі ШІ. [46, с. 2083]

Наявність відкритого вихідного коду сприяла швидкому розвитку ШІ, і комерційні програми для створення глибоких фейків скористалися зростаючим попитом на синтетичні медіа. Стартапи та технологічні компанії впроваджували рішення на основі deepfake для покращення створення контенту в різних галузях — від маркетингу до корпоративного навчання та розваг. Додаток Reface, розроблений в Україні, популяризував штучну заміну облич, дозволяючи користувачам вставляти себе у вірусне відео за лічені секунди, роблячи технологію deepfake більш цікавою та доступною. Лондонський стартап Synthesia, що базується в Лондоні, став піонером у використанні синтетичних аватарів для корпоративної комунікації, що дозволило компаніям створювати автоматизовані навчальні відео без необхідності залучення людей-ведучих. Компанія HeyGen розробила згенерованих штучним інтелектом спікерів із синхронізацією губ у режимі реального часу, змінивши підхід брендів до виробництва відео для рекламних та освітніх цілей. [29]

Зростальна складність клонування голосу за допомогою штучного інтелекту додала ще один рівень складності, оскільки такі програми, як Lyrebird і Descript's Overdub, дозволили відтворювати людські голоси з мінімальним обсягом вхідних даних. Ці інструменти спочатку були розроблені для покращення подкастингу та виробництва відео. Проте вони також породили побоювання, що підроблене аудіо може бути використане для шахрайства з видаванням себе за інших людей та сфабрикованими доказами. [41]

У міру того, як технологія deepfake ставала все більш досконалою і широко доступною, стало очевидним, що її потенціал є одночасно і революційним, і глибоко проблематичним. З одного боку, вона уможливила інновації в кіно, освіті та доступності, пропонуючи нові способи створення, комунікації та персоналізації цифрового досвіду. З іншого боку, він підняв безпрецедентні етичні дилеми, сприяючи швидкому поширенню дезінформації, породженої штучним інтелектом, шахрайству з особистими даними та порушенню цифрової приватності. Широке розповсюдження технології глибоких фейків оголило вразливі місця в цілісності

інформації, посилюючи потребу в медіаграмотності, етичній розробці ШІ та регуляторних заходах для запобігання його зловживанню.

1.2 Соціальні та етичні виклики, пов’язані з deepfake

Розвиток технології дипфейків створив безпрецедентні виклики для правових систем у всьому світі, зокрема й в Україні, де законодавство намагається встигати за швидким розвитком штучного інтелекту. На відміну від деяких західних країн, які почали впроваджувати закони, спрямовані на боротьбу з «глибокими фейками», в Україні відсутні чіткі правові норми, які безпосередньо стосуються маніпуляцій у ЗМІ, що генеруються штучним інтелектом. Ця правова прогалина створює значні труднощі при визначенні відповідальності за правопорушення, пов'язані з «глибокими фейками» — чи то в контексті дезінформації, наклепу, порушення права на приватність або загрози кібербезпеці.

Однією з головних правових проблем, пов'язаних із «глибокими фейками» в Україні, є питання відповідальності. Технологія дозволяє безперешкодно створювати та поширювати гіперреалістичні синтетичні медіа, що ускладнює визначення того, хто має нести відповідальність за потенційну шкоду. Якщо журналіст несвідомо публікує глибоко підроблене відео в рамках журналістського розслідування, що призводить до репутаційної шкоди особі, постає питання про відповідальність. Чи повинен нести відповідальність журналіст, який не перевірив контент перед публікацією, або ж вона покладається виключно на анонімного творця глибокого фейку? Ситуація стає ще складнішою, коли походження маніпульованого контенту неможливо відстежити, оскільки багато творців «дипфейків» діють в онлайн-просторі під прикриттям анонімності.

З іншого боку, журналісти самі можуть стати жертвами атак дипфейків, стаючи об'єктами сфабрикованих відеоматеріалів, покликаних зашкодити їхньому професійному авторитету або особистій репутації. З огляду на оманливу природу «глибоких фейків», виявлення джерела таких маніпуляцій залишається значним викликом, що ускладнює правозастосування. Відсутність в Україні спеціального

законодавства щодо боротьби з «глибокими фейками» загострює проблему, оскільки чинне законодавство не враховує унікальну природу обману, що генерується штучним інтелектом.

Хоча в Україні поки що немає законодавчої бази, спеціально присвяченої боротьбі з глибокими фейками, деякі чинні закони можуть бути застосовані для вирішення пов'язаних з ними питань. Наприклад, стаття 277 Цивільного кодексу України передбачає захист честі, гідності та ділової репутації, а це означає, що потерпілий від підробки, яка завдає шкоди його репутації, потенційно може звернутися до суду за захистом від наклепу. Однак доведення впливу «глибокого фейку» та визначення відповідальної сторони залишається серйозною юридичною перешкодою. Складність відстеження творця «глибокого фейку» часто робить позови про наклеп неефективними, оскільки правозастосування залежить від здатності визначити та притягнути до відповідальності особу, яка стоїть за сфабрикованим контентом. [7]

Окрім шкоди для репутації, технологія «глибоких фейків» викликає серйозні занепокоєння щодо прав на приватність і згоди. Медіа, створені за допомогою штучного інтелекту, часто використовують зовнішність людини без її дозволу, що потенційно порушує права інтелектуальної власності та захист особистого життя. Це особливо актуально для України, де стаття 32 Конституції чітко гарантує право на приватність, забороняючи збирання, зберігання та поширення конфіденційної інформації про особу без її згоди. Теоретично жертви маніпуляцій дипфейків можуть посилатися на цей конституційний захист, щоб протистояти несанкціонованому використанню їхнього зображення чи голосу. Однак відсутність механізмів правозастосування, пристосованих до синтетичних медіа, означає, що чинне законодавство про захист персональних даних не є повністю пристосованим для розгляду справ, пов'язаних із «глибокими фейками».

Питання територіальної та комунікаційної приватності також закріплено в українському законодавстві. Стаття 30 Конституції гарантує недоторканність житла,

а стаття 31 — таємницю листування і телефонних розмов. У випадках, коли фейковий контент використовується для маніпулювання особистістю, місцеперебуванням або приватним спілкуванням людини, ці конституційні гарантії теоретично можуть бути застосовані. Однак головною перешкодою залишається ідентифікація та притягнення до відповідальності порушників, оскільки «глибоко фейковий» контент може створюватися та поширюватися особами, які діють за межами української юрисдикції. [6]

Ще однією важливою проблемою у боротьбі з правопорушеннями, пов'язаними з «глибокими фейками», є роль розповсюджувачів контенту. Навіть якщо творець «глибокого фейку» залишається анонімним, платформи, які розміщують або поширюють синтетичні медіа, такі як соціальні мережі, веб-сайти для обміну відео та платформи для обміну повідомленнями, можуть значно посилити охоплення та вплив маніпулятивного контенту. Хоча в Україні наразі відсутні нормативні акти, які передбачають відповідальність платформ за розміщення фейкових матеріалів, зростаючий вплив дезінформації, створеної штучним інтелектом, може спонукати до законодавчих зусиль, спрямованих на посилення відповідальності провайдерів цифрових послуг.

Окрім творців і розповсюджувачів контенту, розробники технологій глибокого фейку також можуть зіткнутися з юридичною відповідальністю. Виникають етичні питання щодо того, чи мають дослідники ШІ та компанії-розробники програмного забезпечення моральний обов'язок запобігати неправомірному використанню їхніх технологій. Ця дискусія особливо актуальна, якщо врахувати глобальну невідповідність у регулюванні боротьби з підробками — тоді як деякі уряди рухаються в бік посилення нагляду, інструменти ШІ з відкритим вихідним кодом залишаються у вільному доступі для всіх, хто має доступ до інтернету.

Відсутність в Україні законодавчих положень, що стосуються боротьби з глибокими фейками, підкреслює ширшу проблему регулювання технологій, керованих штучним інтелектом, у цифровому ландшафті, що швидко розвивається.

Хоча чинні закони про наклеп, недоторканність приватного життя та кібербезпеку забезпечують певний рівень захисту, вони не були розроблені з урахуванням унікальних характеристик синтетичних медіа. Оскільки технологія «глибоких фейків» продовжує розвиватися, Україні, як і багатьом іншим країнам, імовірно, доведеться адаптувати свою законодавчу базу, щоб встановити чіткі правила щодо відповідальності, згоди та механізмів правозастосування для боротьби зі зловживанням контентом, створеним за допомогою штучного інтелекту.

Оскільки технологія «глибоких фейків» продовжує розвиватися, уряди країн світу також відреагували на неї низкою правових підходів — від спеціальних законів про «глибокі фейки» до ширших законів про захист приватного життя, шахрайство та наклеп. Поки деякі юрисдикції запровадили спеціальне законодавство для регулювання створення та розповсюдження «глибоких фейків», інші покладаються на існуючі правові рамки для подолання пов'язаних з ними ризиків. Така правова нерівність підкреслює складність регулювання контенту, створеного штучним інтелектом, оскільки регулятори намагаються збалансувати свободу вираження поглядів, права на недоторканність приватного життя та зростаючу загрозу синтетичного медіа-обману.

Європейський Союз ще не ухвалив спеціального закону про боротьбу з глибокими фейками, але його чинні правила щодо захисту приватності та цифрового контенту передбачають механізми для боротьби зі зловживанням синтетичними медіа. Два найважливіші правові інструменти — Загальний регламент про захист даних (GDPR) і Закон про цифрові послуги (DSA) — створюють основу для регулювання порушень, пов'язаних із «глибокими фейками», зокрема в контексті захисту даних, дезінформації та відповідальності платформ.

GDPR, який набув чинності у 2018 році, захищає персональні дані та права на недоторканність приватного життя в усіх країнах-членах ЄС. Хоча він не був спеціально розроблений для регулювання технологій глибокого фейку, він застосовується до контенту, створеного штучним інтелектом, який передбачає

несанкціоноване використання схожості, голосу або біометричних даних людини. Згідно зі статтею 6, обробка персональних даних без законних підстав є незаконною, а це означає, що використання технології deepfake для створення фальшивих ідентифікаційних даних або видавання себе за іншу особу без її згоди може вважатися порушенням [23]. Крім того, стаття 7 вимагає чіткої згоди на обробку даних, що посилює ідею про те, що люди повинні мати контроль над тим, як їхнє зображення і голос використовуються в контенті, створеному штучним інтелектом. [24]

Окрім питань конфіденційності, Закон ЄС про цифрові послуги (DSA), який був прийнятий у 2022 році, запроваджує заходи відповідальності платформ, які можуть вплинути на поширення «глибоких фейків». DSA вимагає від онлайн-платформ і компаній, що працюють у соціальних мережах, вживати проактивних заходів для видалення шкідливого або оманливого контенту, включно з маніпулятивними матеріалами, призначеними для введення в оману громадськості. Хоча закон прямо не згадує про «глибокі фейки», його положення можуть застосовуватися до платформ, які не регулюють дезінформацію, що генерується штучним інтелектом, і потенційно притягувати їх до відповідальності за нездатність запобігти розповсюдженню шкідливого синтетичного контенту.

Крім того, деякі країни-члени ЄС ухвалили національні закони, які можуть бути використані проти глибоких фейків. Наприклад, кримінальне законодавство Німеччини про наклеп, зокрема стаття 185 Кримінального кодексу Німеччини, дозволяє подавати позови проти образливого або наклепницького контенту, що може поширюватися і на шкоду репутації, заподіяну внаслідок поширення «глибоких фейків». Однак відсутність єдиної загальноєвропейської правової бази для боротьби з «глибокими фейками» означає, що правозастосування залишається фрагментарним у різних юрисдикціях. [14]

Сполучені Штати також не ухвалили федерального законодавства, яке б безпосередньо регулювало боротьбу з «глибокими фейками». Натомість правові

заходи реагування здебільшого приймаються на рівні штатів, причому деякі штати займають проактивну позицію, тоді як інші продовжують покладатися на чинні закони, пов'язані з шахрайством, наклепом і крадіжкою персональних даних.

Одна з найвідоміших законодавчих ініціатив — «Закон про відповідальність за дипфейки» (Deepfakes Accountability Act) — була представлена в Конгресі США у 2019 році. Законопроект мав на меті встановити федеральні правила, які вимагатимуть від творців «глибоких фейків» розкривати інформацію про те, що їхній контент є штучно створеним, а також розробити технологію водяних знаків, які допоможуть ідентифікувати синтетичні медіа. Однак ця пропозиція ще не стала законом, тому регулювання боротьби з глибокими фейками залишилося на розсуд окремих штатів.

Серед штатів Каліфорнія і Техас лідирують у прийнятті законів, спрямованих на боротьбу з правопорушеннями, пов'язаними з глибокими підробками, на рівні штатів:

- Каліфорнійський закон AB 602 (2019). Цей закон дозволяє людям подавати до суду на творців глибоких фейків, які виробляють і поширюють відверто сексуальний контент без згоди, надаючи правовий захист жертвам порнографії, створеної за допомогою штучного інтелекту. [12]
- Каліфорнійський закон AB 730 (2019). Цей закон забороняє поширення політично вмотивованих «глибоких фейків» протягом 60 днів після виборів, визнаючи потенціал медіа, створених штучним інтелектом, для маніпулювання громадською думкою та втручання в демократичні процеси. [13]
- Texas SB 751 (2019). Цей закон криміналізує створення та розповсюдження «глибоких фейків» з метою зашкодити виборам або вплинути на них, що робить його одним із перших законів штату, який безпосередньо стосується політичної дезінформації, спричиненої синтетичними медіа.

Попри ці зміни, відсутність всеосяжної федеральної законодавчої бази призводить до того, що регулювання боротьби з «глибокими фейками» в різних

штатах є непослідовним. Федеральні закони, пов'язані з шахрайством, крадіжкою персональних даних і кіберзлочинністю, іноді можуть застосовуватися, але їхнє правозастосування залишається дуже залежним від контексту й мети створення та поширення «глибоких фейків».

Канада ще не ухвалила спеціального закону про боротьбу з «глибокими фейками», але вже має правові норми, які можна застосувати до дезінформації, створеної штучним інтелектом, та шахрайства з використанням персональних даних. Згідно з Кримінальним кодексом Канади, «глибокі фейки» можуть переслідуватися за статтями про шахрайство, вимагання та наклеп, якщо їх використовують з оманливою або шкідливою метою.

Розділ 380(1) Кримінального кодексу визначає шахрайство як «оманливу поведінку, спрямовану на заподіяння шкоди або фінансових збитків». Якщо глибокий фейк створюється з метою ошукати людей, поширити неправдиву фінансову інформацію або вчинити цифрове шахрайство, закон може бути застосований проти зловмисників. [21]

Закони про шахрайство з особистими даними також можуть бути застосовані у випадках, коли глибокі підробки використовуються для видачі себе за іншу особу, зокрема, якщо чиєсь зображення або голос синтетично маніпулюють, щоб обдурити інших зі зловмисними цілями. Положення про наклеп теоретично можуть бути застосовані, якщо глибокий фейк завдає шкоди репутації особи, хоча правозастосування залежатиме від можливості відстежити оригінального творця синтетичного контенту.

Канада також представила законопроект С-11, який покликаний регулювати платформи цифрового контенту та дезінформацію в Інтернеті. Хоча законопроєкт стосується насамперед модерації контенту та підзвітності платформ, він може зіграти певну роль у тому, що онлайн-сервіси будуть зобов'язані виявляти та видаляти неправдиву інформацію, згенеровану штучним інтелектом, включно з глибокими фейками.

Отже, незважаючи на зростаюче занепокоєння щодо потенціалу технології deepfake для обману, шахрайства та завдання шкоди репутації, більшість країн ще не запровадили всеосяжної правової бази для регулювання синтетичних медіа, створених за допомогою штучного інтелекту. Хоча Європейський Союз, США, Канада тощо запровадили різні правові заходи, єдиного глобального стандарту для боротьби з правопорушеннями, пов'язаними з глибокими фейками, не існує.

Витонченість і доступність технології дипфейків безперервно зростає у 21 столітті, тож це створює не тільки етичні, але й суспільні ризики, що виходять за межі індивідуальної шкоди. Це може загрожувати демократії, цілісності інформації та особистої приватності. Однією з найбільш тривожних загроз, яку становлять «глибокі фейки», є їхня здатність маніпулювати політичним дискурсом і впливати на вибори. Ця технологія дозволяє створювати гіперреалістичні відео, на яких політичні діячі говорять або роблять те, чого вони насправді не говорили і не робили, що може бути стратегічно використано для впливу на громадську думку, поширення неправдивих наративів і підриву довіри до демократичних інститутів.

Політична дезінформація, створена за допомогою глибоких фейків, може набувати різних форм. У деяких випадках синтетичні відеоролики створюються з метою дискредитації політичних кандидатів, зображуючи їх у компрометуючих ситуаціях або фабрикуючи суперечливі заяви. Особливо небезпечними є фейкові атаки в останню хвилину перед виборами, оскільки неправдива інформація може швидко поширюватися до того, як фактчекери матимуть змогу її спростувати. Такі закони, як каліфорнійський AB 730, який забороняє поширення оманливих «дипфейків» про політичних кандидатів протягом 60 днів після виборів, визнають цей ризик, але правозастосування залишається складним, особливо в юрисдикціях, де немає спеціальних законів щодо дипфейків.

Окрім прямих маніпуляцій, дипфейки сприяють підриву довіри громадськості до легітимних новин і засобів масової інформації. Саме існування технології «глибоких фейків» дозволяє політичним діячам видавати автентичні кадри за

сфабриковані — явище, відоме як «дивіденди брехуна». Цей ефект дозволяє недобросовісним акторам підривати реальні докази, стверджуючи, що будь-який шкідливий відео- чи аудіозапис може бути глибоким фейком. Як наслідок, технологія «глибоких фейків» не лише створює неправду, але й забезпечує правдоподібне заперечення, що ускладнює підзвітність у політичному ландшафті. [39, с. 213]

Роль соціальних мереж у посиленні дезінформації за допомогою «глибоких фейків» також викликає серйозне занепокоєння. Такі платформи, як Facebook, Twitter (X) і TikTok, сприяють швидкому поширенню маніпулятивного контенту, часто до того, як можуть втрутитися фактчекери. У відповідь на це технологічні компанії впровадили інструменти виявлення глибоких фейків на основі штучного інтелекту. Наприклад, Deepfake Detector від McAfee використовує просунутий ШІ, щоб за лічені секунди сповіщати користувачів, якщо звук у відео згенерований штучним інтелектом, допомагаючи відрізнити справжній контент від підробленого. Однак можливості синтентичних медіа зростають швидше, аніж розвиваються технології протидії їм. [28, с. 60]

Окрім політичних і соціальних маніпуляцій, «глибокі фейки» становлять значний психологічний ризик, впливаючи на людське пізнання, пам'ять і сприйняття реальності. Здатність технології «глибоких фейків» створювати переконливий, емоційно заряджений контент робить людей більш сприйнятливими до неправдивих наративів і змінених спогадів, формуючи їхні переконання та поведінку в непередбачуваний спосіб.

Одним із найбільш тривожних аспектів психологічних маніпуляцій на основі глибоких фейків є «синдром фальшивої пам'яті» — явище, коли люди неправильно запам'ятовують або повністю вигадують спогади під впливом маніпулятивних медіа. Візуальні та слухові стимули можуть сильно впливати на спогади, а це означає, що вплив сфабрикованих історичних подій, фейкових новин або відредагованих кадрів минулого може створювати у людей абсолютно нові, але

фальшиві спогади. Це викликає серйозне занепокоєння щодо того, як технологія deepfake може бути використана для переписування історії, фабрикації подій або зміни суспільного розуміння ключових моментів. [30, с. 45]

Неправдиві наративи, створені за допомогою «глибоких фейків», можуть підживлювати колективну дезінформацію, впливаючи на цілі спільноти або демографічні показники. Теоретики змови, наприклад, уже почали використовувати технологію «глибоких фейків» для фабрикації відео, що підтверджують їхні твердження, посилюючи чинні упередження та недовіру до інституцій. З поширенням фейкового контенту здатність людського мозку розрізняти реальні та синтетичні спогади буде дедалі більше ускладнюватися, що ще більше ускладнить боротьбу з дезінформацією. Доречним прикладом є конспірологічна теорія, яку поширили російські медіа в березні 2022 року. У ній неправдиво стверджувалося, що Сполучені Штати експлуатують лабораторії з виробництва біологічної зброї в Україні, і цьому сприяють начебто «бойові гуси». Відео та текстові повідомлення поширювалися в різних медіа, зокрема Facebook, YouTube, Twitter і Telegram, щоб охопити максимальну кількість глядачів. Звісно ж, через свою експресивність, абсурдність та резонансність чудово спрацював і принцип «сарафанного радіо», що могло призводити до продукування більш безглуздих фейків через почуття переляканості в людей у розпал повномасштабної війни росії та України. [8, с. 2]

Поза тим, дипфейки використовують і для створення синтетичної порнографії без згоди («revenge porn»). Це передбачає накладання обличчя людини на відверті матеріали без її згоди, що призводить до серйозних емоційних страждань, репутаційних збитків, а в деяких випадках — до шантажу чи вимагання. Відповідно до звіту «Home Security Heroes», було виявлено, що станом на 2023 рік 98% підроблених відео були порнографічними, причому 99% були націлені на жінок [1]. Каліфорнійський закон AB 602, який дозволяє жертвам подавати до суду на творців відвертого контенту, створеного без їхньої згоди, є одним з небагатьох правових заходів, спрямованих на вирішення цієї проблеми. Однак у багатьох країнах жертви

не мають практично ніяких засобів правового захисту, оскільки підроблену порнографію важко відстежити і видалити з онлайн-платформ.

Мімікрування себе під іншу особу викликає занепокоєння у сфері кібербезпеки та фінансового шахрайства. Синтетичні голоси, згенеровані штучним інтелектом, використовуються для обходу біометричної автентифікації, а злочинці успішно обманюють банки та компанії, змушуючи їх санкціонувати шахрайські транзакції, імітуючи голоси керівників та власників рахунків. Наприклад, у березні 2019 року генеральний директор британської філії німецької енергетичної компанії RWE став жертвою шахрайства з використанням підробленого голосу, створеного за допомогою штучного інтелекту. Шахраї використали технологію синтезу голосу, щоб імітувати голос генерального директора материнської компанії, розташованої в Німеччині. Вони зателефонували британському керівнику та переконали його перевести €220 000 (приблизно $243 000) на рахунок угорського постачальника, який насправді контролювався шахраями. [38]

Хоча технологія deepfake часто асоціюється з обманом, дезінформацією та порушенням приватності, вона також має значний потенціал для етичного застосування. Синтетичні медіа, створені штучним інтелектом, за своєю суттю не є зловмисними; скоріше, їхній вплив залежить від того, як вони розробляються й використовуються. Прикладом такого застосування deepfake є галузь історії та освіти. Створені штучним інтелектом дипфейки дозволяють сучасній аудиторії бачити, чути й взаємодіяти з історичними постатями у спосіб, який раніше було неможливо собі уявити. Цю технологію можна використовувати для створення документальних фільмів, музейних експонатів, оживлення старих фотографій та архівних кадрів. Яскравим прикладом є виставка «Життя Далі» в Музеї Сальвадора Далі у Флориді, де технологія deepfake використовувалася для воскресіння художника в режимі реального часу за допомогою інтерактивного штучного інтелекту. Відвідувачі могли вести бесіди зі згенерованим штучним інтелектом Далі, вивчаючи його мистецтво, філософію та творчий процес. Такий тип оповіді за

допомогою штучного інтелекту представляє значний прогрес у тому, як можна викладати й переживати історію, дізнатися про минуле.

Клонування голосу на основі штучного інтелекту теж стало потужним інструментом для забезпечення доступності та мовної локалізації. Такі технології, як Resemble AI, ElevenLabs і DeepDub, дозволяють кінематографістам синтезувати реалістичні голоси, роблячи дубляж більш природним і захоплюючим для глобальної аудиторії. Синтез голосу зі штучним інтелектом також використовується для відновлення голосів людей, які втратили здатність говорити. Для прикладу, фізик Стівен Гокінг використовував синтезатор мовлення, розроблений Деннісом Клаттом, піонером у галузі систем перетворення тексту в мову. Клатт створив систему DECtalk, яка розробила голос «Perfect Paul». Цей голос перейняв Гокінг, що дозволило йому спілкуватися, незважаючи на його хворобу рухових нейронів. [37]

Отже, технологія deepfake створює правові, етичні та психологічні ризики, але вона також має значний потенціал за відповідального та прозорого використання. Однак етичне застосування deepfake вимагає суворого нагляду та відповідального впровадження. Прозорість має вирішальне значення — аудиторія повинна бути поінформована про використання технології глибокого підроблення. Однак потрібно розробити нормативно-правову базу, яка б заохочувала інновації, мінімізуючи при цьому ризики, пов'язані з синтетичними медіа, згенерованими штучним інтелектом.

1.3. Інформаційний фронт: як журналісти виявляють маніпуляції з дипфейками й протидіють їм

Популяризація дипфейків змусила журналістів, фактчекерів та експертів з цифрової криміналістики докладати більше зусиль до виявлення та протидії маніпуляціям у медіа, створеним за допомогою штучного інтелекту. Хоча «глибокі фейки» продовжують удосконалюватися, методи їхнього виявлення залишаються критично важливим способом захисту від поширення дезінформації.

Методи верифікації «глибоких фейків» можна розділити на дві основні категорії: ручні методи перевірки, які покладаються на людське спостереження та криміналістичний аналіз, та автоматизовані методи виявлення на основі штучного інтелекту, які використовують алгоритми машинного навчання для виявлення синтетичного контенту. Ручні методи виявлення ґрунтуються на виявленні невідповідностей, артефактів і аномалій, внесених під час процесу генерації «глибоких фейків». Ці методи вимагають уважності до деталей, розуміння криміналістики відео- та аудіоматеріалів і знання технологій виробництва фейків. [9, с. 3]

Одним із найефективніших способів виявлення підробленого контенту є візуальний аналіз, оскільки багато відео та зображень, згенерованих штучним інтелектом, містять ледь помітні артефакти та невідповідності, які відрізняють їх від справжніх медіа. Генерація глибоких фейків спирається на генеративні змагальні мережі (GAN), які часто створюють невеликі спотворення й невідповідності, особливо в низькоякісних моделях. Серед візуальних ознак маніпуляцій з глибокими підробками можна виділити такі:

- Розмиті краї та спотворені переходи. У багатьох підроблених відео намагаються бездоганно поєднати елементи замінених облич з оригінальними кадрами, що призводить до розмитих контурів, непослідовних текстур шкіри або незвичного змішування на лінії підборіддя та волосся.
- Неприродне освітлення та тіні. Обличчя, створені штучним інтелектом, часто не можуть точно відтворити природні умови освітлення, що призводить до невідповідності тіней, неприродних віддзеркалень або невідповідності яскравості між рисами обличчя та фоном.
- Аномалії рис обличчя. Асиметрія, неприродний рух очей або неправильні пропорції обличчя можуть бути характерними ознаками синтетичних матеріалів. Моделям на основі генеративних змагальних мереж іноді важко генерувати рівномірний тон шкіри та реалістичні відблиски очей, що робить

нерівності в області очей і рота важливими індикаторами маніпуляцій. Ранні моделі «глибокої підробки» мали сумнозвісну неточність у поведінці моргання очей, оскільки згенеровані штучним інтелектом обличчя часто демонстрували або надмірне моргання, або неприродну відсутність моргання. Хоча новіші алгоритми глибокого підроблення покращилися в цій сфері, нерегулярна поведінка очей залишається цінним показником для судового аналізу. [25]

Окрім візуальних ознак, невідповідності в аудіо можуть також служити «червоними прапорцями» в перевірці «глибоких фейків». Голосам, згенерованим штучним інтелектом, часто бракує природних інтонацій, ритмічної варіативності та правильного дихання, що робить їх не схожими на справжню людську мову. Розподілимо основні ознаки маніпуляцій з аудіофайлами на такі типи:

- Роботизований або неприродний тон. Синтезовані штучним інтелектом голоси можуть звучати дещо механічно, надто плавно або неприродно рівномірно за висотою та інтонацією. На відміну від природної мови, у якій є тонкі коливання, глибоко підроблені голоси часто позбавлені спонтанних варіацій.
- Характер дихання та проблеми з артикуляцією. Мова, згенерована штучним інтелектом, часто має проблеми з синхронізацією дихання, пропускаючи природні паузи або вставляючи надмірно довгі та неприродні вдихи. Крім того, деякі голоси, згенеровані штучним інтелектом, можуть неправильно вимовляти складні слова або видавати неприродні вокальні каденції, що полегшує розпізнавання для підготовлених слухачів.
- Розбіжності в синхронізації з губами. У підроблених відео, де синтетичні голоси накладаються на реальні кадри, розбіжності між рухами губ і вимовленими словами можуть свідчити про маніпуляції. Якщо мова звучить з невеликою затримкою або не ідеально відповідає рухам рота, це свідчить про те, що відео було змінено за допомогою штучного інтелекту. [18, с. 21]

Аналіз метаданих — це часто ігнорований, але дуже ефективний метод перевірки автентичності цифрових медіа. Метадані — прихована інформація,

вбудована в зображення, відео- та аудіофайли — можуть виявити ознаки маніпуляцій, невідповідності в позначках часу або незвичні методи кодування, які можуть свідчити про синтетичний вміст. Ключові елементи аналізу метаданих включають:

- EXIF-дані (Exchangeable Image File Format — обмінний формат файлів зображень). Вивчення метаданих EXIF у зображеннях і відео може виявити, коли, де і як було створено файл. Якщо відео стверджує, що це автентичні історичні кадри, але має метадані, які вказують на те, що його було створено нещодавно, це може свідчити про глибоку підробку або цифрову зміну контенту.
- Артефакти стиснення та розбіжності в кодуванні. Багато інструментів для підробки перекодовують відео з використанням незвичних параметрів стиснення, залишаючи помітні цифрові сліди, які відрізняються від стандартних форматів відеозапису. Виявлення невідповідностей у роздільній здатності, бітрейтах і параметрах кодування може допомогти викрити синтетично змінені медіа.
- Зворотний пошук зображень. При розслідуванні підозри на глибоку підробку запуск зворотного пошуку зображень за допомогою таких інструментів, як Google Images, TinEye або InVID & WeVerify, може допомогти визначити, чи було зображення або відео змінено цифровим способом з оригінального джерела.

Аналіз метаданих надає важливі криміналістичні докази, але він не є надійним: деякі творці глибоких підробок видаляють метадані або маніпулюють ними, щоб замести сліди. Проте разом з візуальним і аудіоаналізом він залишається важливою стратегією перевірки для журналістів і фактчекерів. [27, с. 35]

Ручні методи перевірки корисні для виявлення дипфейків, але вони займають багато часу й вимагають спеціальних знань. Саме тому вони не є універсальними для перевірки в режимі реального часу та в умовах високого інформаційного тиску, тож їх потрібно використовувати в поєднанні з інструментами на основі штучного інтелекту, які застосовують алгоритми машинного навчання для виявлення

неправдивого контенту в масштабах і з точністю, що виходять за межі людських можливостей.

Моделі машинного навчання, що використовуються для виявлення підробок, базуються на розпізнаванні шаблонів, виявленні аномалій і навчанні в умовах конкуренції. Серед найпоширеніших можна виділити:

- Згорткові нейронні мережі (CNN). Ці мережі спеціалізуються на аналізі зображень і відео та визнають тонкі невідповідності в піксельних шаблонах, спотворення обличчя й неприродні текстури, які можуть свідчити про глибоку підробку. Згорткові нейронні мережі особливо ефективні у виявленні підроблених візуальних медіа. [44, с. 29]
- Рекурентні нейронні мережі (RNN). В основному використовуються для аналізу аудіо та мовлення. Вони можуть виявляти часові невідповідності, неприродні мовленнєві патерни та різкі тональні зміни. Це робить їх високоефективними у виявленні маніпуляцій з аудіозаписами. [33]
- Автокодери. Ці самонавчальні алгоритми виявляють аномалії та нерівності в зображеннях або відеопослідовностях. Вони працюють, реконструюючи оригінальні зображення й порівнюючи їх з підозрюваними глибокими підробками, виявляючи спотворення, які вказують на втручання.
- Машини опорних векторів (SVM). SVM класифікують дані, вивчаючи гіперплощину, яка відокремлює реальний контент від синтетичного, пропонуючи математичний підхід до класифікації глибоких фейків.
- Генеративні змагальні мережі (GAN) для виявлення. Та сама технологія GAN, що використовується для створення «глибоких фейків», може бути використана й для їхнього виявлення. Одна нейронна мережа створює синтетичні медіа, а інша навчається виявляти та викривати шаблони маніпуляцій, що робить інструменти виявлення все більш досконалими.

Журналісти, судові аналітики та фактчекери широко використовують і автоматизовані інструменти для виявлення фейків на основі алгоритмів машинного навчання, які нині є безкоштовні та у відкритому доступі. Наголосимо на найпопулярніших сучасних системах, які покликані перевіряти та знаходити сфабрикований контент:

- **Deepware Scanner** — це інструмент для виявлення відеопідробок, призначений для аналізу кадрів і моделей руху обличчя на предмет потенційних маніпуляцій, керованих штучним інтелектом. Він використовує CNN та алгоритми виявлення аномалій для сканування відео на предмет невідповідності облич, неприродного освітлення та артефактів змішування, які є типовими ознаками підробок.
- **Resemble AI** спеціалізується на виявленні підробок голосу, ідентифікуючи синтетичну мову та аудіоімітацію, створену штучним інтелектом. Він використовує моделі на основі RNN, навчені на тисячах годин автентичних і згенерованих штучним інтелектом мовних даних, що дозволяє йому виявляти тонкі спотворення модуляції голосу, порушення вимови та неприродну тональну послідовність.
- **InVID & WeVerify** — один з найпоширеніших багатофункціональних інструментів верифікації, що містить такі функції, як зворотний пошук зображень, виявлення аномалій на основі глибокого навчання та аналіз метаданих. Цей інструмент особливо ефективний у журналістських розслідуваннях, де важливо перевіряти достовірність зображень і відео, поширених у соціальних мережах.

Пропонуємо вашій увазі таблицю порівняльного аналізу, де ми оцінили вищезазначені системи виявлення фейків Deepware, Resemble AI та InVID & WeVerify за наведеними критеріями (див. табл. 1). Для наочності даних ми за допомогою RunwayML створили авторський deepfake із зображенням Авраама Лінкольна (URL: https://youtu.be/pLn674swsyA), який і перевірили на достовірність через наведені програми.

Таблиця 1

Оцінка ефективності систем виявлення фейків за критеріями

Критерій	Deepware	Resemble AI	InVID & WeVerify
Точність виявлення дипфейків	Висока для відео; аудіо не підтримується	Висока для аудіо; відео не підтримується	Висока для відео та зображень
Підтримка форматів медіа	Відео (mp4, avi)	Аудіо (wav, mp3)	Відео, зображення (jpeg, png, mp4)
Швидкість аналізу	Швидка	Помірна	Швидка
Інтерфейс користувача	Простий, мінімалістичний	Інтуїтивний, але професійний	Інтуїтивний, орієнтований на журналістів
Тип аналізу	Відео	Аудіо	Відео, зображення, метадані
Алгоритми ШІ	Глибокі нейронні мережі	Генеративний ШІ для аудіо	Машинне навчання та аналіз контексту
Можливість інтеграції	Немає	API доступне	Немає
Мульти-платформність	Вебсайт	Вебсайт та застосунок для iOS/Android	Розширення для браузера
Захист даних користувача	Обмежена політика конфіденційно	Високий	Високий

	сті		
Додаткові функції	Немає	Генерація голосу	Аналіз метаданих
Мова інтерфейсу	Англійська	Англійська	Багатомовність (включно з англійською)
Ціна/умови ліцензії	Безкоштовно	Платна, з безкоштовним планом	Безкоштовно
Підтримка користувачів	Обмежена	Доступна	Активна підтримка
Репутація	Середня	Висока	Висока
Актуальність алгоритмів	Помірна	Висока	Висока

Важливо зазначити, що хоча інструменти виявлення на основі штучного інтелекту забезпечують критичний рівень перевірки, але вони потребують додаткового втручання. Наразі тільки людина може точно відрізнити сфабрикований контент від справжнього й остерегтися передчасних і помилкових висновків, адже автоматизовані технології ще не настільки вдосконалені.

За даними DeepMedia, у 2023 році кількість підроблених відео та голосових записів, поширених у соціальних мережах у всьому світі, сягнула приблизно 500 000. Очікується, що до кінця 2025 року ця цифра зросте до 8 мільйонів, що свідчить про подвоєння кількості фейкового контенту кожні півроку [42, с. 21]. Це свідчить про нагальну потребу в ініціативах з медіаграмотності, які допоможуть людям розпізнавати та критично оцінювати синтетичний контент. Хоча журналісти та фактчекери відіграють життєво важливу роль у розвінчуванні «глибоких фейків», величезний обсяг дезінформації, створеної штучним інтелектом, означає, що для

запобігання масштабному обману необхідна проактивна просвітницька робота з громадськістю.

Новинні організації, установи з фактчекінгу та «сторожові пси» можуть організовувати освітні семінари, які навчають людей, як виявляти аномалії, пов'язані з глибокими фейками, користуватися інструментами перевірки та здійснювати перехресну перевірку медіа-джерел [17, с. 657]. Такі ініціативи, як First Draft News, Європейська обсерваторія цифрових медіа (EDMO) та Академія української преси вже активно займаються навчанням медіаграмотності. Освітні вебсайти можуть надати покрокові інструкції, кейси та практичні вправи, як виявляти «глибокі фейки» за допомогою ручних та автоматизованих методів.

Однак одних лише ініціатив з медіаграмотності недостатньо. Боротьба з «глибокими фейками» вимагає підзвітності, а це означає, що соціальні мережі та цифрові платформи повинні впроваджувати суворішу політику щодо синтетичних медіа. Як приклад, соціальні мережі Twitter (X), TikTok, Instagram розробили політику маркування маніпульованих медіа й матеріалів, зроблених за допомогою ШІ. Це одна з перших спроб регулювання контенту, але для забезпечення прозорості та запобігання вірусному поширенню дезінформації, створеної штучним інтелектом, потрібні більш надійні рішення.

Висновки до Розділу 1

Поява та розвиток технології «deepfake» неминуче призводить до численних змін у сфері журналістики, що створює як переваги, так і серйозні виклики. Від своїх витоків у ранніх технологіях цифрової маніпуляції до швидкої еволюції через генеративні змагальні мережі (GAN) — синтетичні медіа змінили ландшафт створення цифрового контенту. Технологія, що започатковувалася як інновація у сфері розваг, маркетингу та історичної реконструкції на основі штучного інтелекту, стала потужним інструментом для обману, дезінформації та маніпуляції особистістю.

Етичні та правові проблеми, пов'язані з «глибокими фейками», виявилися складними та далекосяжними. Чинне законодавство таких країн, як США, Канада та Європейський Союз, пропонує часткові рішення, застосовуючи до випадків обману, спричиненого штучним інтелектом, закони про конфіденційність, наклеп та кібербезпеку. Однак відсутність у багатьох юрисдикціях спеціального законодавства про боротьбу з глибоким фейком залишає значні прогалини у сфері підзвітності. Правові рамки не встигають за розвитком штучного інтелекту, що дозволяє зловмисникам використовувати синтетичні медіа для політичної дезінформації, фінансового шахрайства та нанесення шкоди репутації.

Технологія «глибоких фейків» також постає перед нами і як глибока етична дилема. Її застосування для втручання у вибори, маніпуляцій з пам'яттю та порушення приватності викликає серйозні занепокоєння щодо цілісності медіа та цифрової довіри. Водночас відповідальне застосування дипфейків — наприклад, історичні реконструкції за допомогою штучного інтелекту та інструменти мовної локалізації — демонструє її величезний потенціал. Тож етична дилема полягає не в самій технології, а в намірах, які стоять за її використанням.

Так, журналісти, дослідники та фактчекери розробили низку стратегій виявлення та протидії. Гібридний підхід, який поєднує ручні методи перевірки з інструментами виявлення на основі машинного навчання, виявився дуже важливим для боротьби з дезінформацією, що генерується ШІ. Функціональні методи дозволяють журналістам виявляти візуальні та аудіоневідповідності й аномалії метаданих. Автоматизовані інструменти виявлення на основі ШІ, такі як Deepware Scanner, Resemble AI та InVID & WeVerify, за більш короткий час надають ефективні рішення для виявлення синтетичного контенту. Однак навіть найсучасніші методи виявлення залишаються недосконалими, оскільки творці підробок постійно вдосконалюють свої методи, щоб уникнути судової експертизи.

Окрім технологічних контрзаходів, у боротьбі з дезінформацією важливу роль відіграють ініціативи з підвищення обізнаності громадськості та медіаграмотності.

Освітні програми, кампанії з цифрової грамотності та організації з перевірки фактів відіграють життєво важливу роль у формуванні у громадськості навичок, необхідних для критичної оцінки онлайн-контенту. Підвищення обізнаності про ризики, можливості та обмеження технології «глибоких фейків» так само важливе, як і розробка інструментів для її виявлення.

Боротьба з фейковою дезінформацією — це не просто технологічний виклик. Це випробування на цифрову доброчесність, стійкість медіа та демократичну стабільність. Здатність відрізняти правду від вигадок у світі, керованому штучним інтелектом, визначатиме, як суспільства орієнтуватимуться в майбутньому цифрової комунікації. Вибір, зроблений сьогодні у сфері розвитку технологій, правової політики та етичної журналістики, визначатиме баланс між інноваціями та підзвітністю в епоху синтетичних медіа.

РОЗДІЛ 2. ПРАКТИЧНА РЕАЛІЗАЦІЯ ІНТЕРАКТИВНОГО САЙТУ «DEFAKE IT!» ЯК ІНСТРУМЕНТА ПРОТИДІЇ ДЕЗІНФОРМАЦІЇ

2.1. Аналіз цільової аудиторії та потреб у медіапродукті

Успіх DeFake It! як освітньої платформи залежить від чітко визначеної та добре сегментованої аудиторії. Оскільки технологія дипфейків впливає на різні верстви суспільства, платформа призначена для широкого кола користувачів з різними потребами та цілями. Розуміння цих різних груп користувачів має важливе значення для формування контенту, функціональності та стратегій охоплення, які гарантують, що платформа відповідатиме очікуванням її відвідувачів.

Цільова аудиторія DeFake It! сегментована на три ключові групи:

- **Первинна аудиторія.** Журналісти, фактчекери та працівники ЗМІ, яким потрібні передові інструменти перевірки та ресурси для проведення розслідувань.
- **Вторинна аудиторія**. Викладачі, студенти та політики, які шукають структуровані навчальні матеріали, академічні ресурси та політичні ідеї.
- **Третинна аудиторія.** Пересічні користувачі інтернету, які потребують базових навичок медіаграмотності та обізнаності про ризики, пов'язані з «глибокими фейками».

Кожна із цих груп відрізняється за досвідом, цифровими звичками та очікуваннями, тому потребує адаптованого контенту та функцій, щоб максимізувати освітню цінність платформи та її практичне застосування. Хоча такі фактори, як вік, професія та цифрова грамотність, впливають на те, як користувачі взаємодіють із «глибоко фейковим» контентом, гендер не є визначальним, оскільки вплив синтетичних медіа-маніпуляцій не є винятковим для будь-якої конкретної демографічної групи.

Основними користувачами DeFake It! є ті, хто працює у сфері новин, журналістських розслідувань та цифрового фактчекінгу. Ці фахівці перебувають на передовій боротьби з дезінформацією й потребують швидких, надійних і доказових методів перевірки, щоб протистояти синтетичним маніпуляціям у ЗМІ. Основна

аудиторія зазвичай перебуває у віковому діапазоні 25-50 років, включаючи журналістів-початківців, досвідчених медіа-професіоналів та ветеранів фактчекінгу. Ця група має високий рівень цифрової грамотності, звикла використовувати у своїй роботі інструменти перевірки контенту, платформи для аналізу соціальних мереж і методи розвідки з відкритих джерел (Open-source intelligence, OSINT). Їм потрібен ефективний у часі, зручний і професійний ресурс, який легко інтегрується в їхній робочий процес.

Другою за чисельністю групою користувачів є ті, хто займається освітою, науковими дослідженнями та розробкою політики. Ця аудиторія не займається верифікацією в режимі реального часу, але потребує комплексних навчальних матеріалів, структурованих навчальних програм і політичної аналітики, щоб зрозуміти ширші наслідки технології дипфейків. Ця група охоплює широкий віковий діапазон: від студентів (18-25 років), які вивчають журналістику та соціальні комунікації, до викладачів (30-60 років), які включають теми, пов'язані з дипфейками, у свої курси. Політики, часто у віці 35 років і старші, покладаються на аналіз високого рівня та регуляторні дискусії для формування законодавства та стратегій управління. Для студентів та освітян DeFake It! слугує інтерактивною освітньою платформою, а для політичних діячів — рекомендаціями в регуляторній сфері діяльності, що забезпечує прийняття обґрунтованих управлінських рішень щодо медіа, створених штучним інтелектом.

Хоча платформа призначена насамперед для професіоналів та освітян, великий сегмент інтернет-користувачів стикається з неправдивим контентом випадково через соціальні мережі, новинні платформи та онлайн-спільноти. Широка громадськість є важливою аудиторією, оскільки масове поширення медіаграмотності є найефективнішим захистом від дезінформації, чинником якої є дипфейки.

Ця аудиторія охоплює широкий демографічний діапазон — від підлітків (16+) до людей похилого віку (60+) з різним рівнем цифрової грамотності. Багато хто

споживає цифровий контент пасивно, не маючи спеціальних знань, необхідних для критичної оцінки та перевірки автентичності медіа. Вони часто покладаються на соціальні мережі та основні новинні платформи, що робить їх вразливими до обману, створеного штучним інтелектом.

Для того, щоб DeFake It! ефективно виконував свою освітню та інформаційну місію, необхідно встановити чіткі показники успіху для оцінки його впливу, залучення користувачів та загальної ефективності. Оскільки платформа обслуговує журналістів, фактчекерів, освітян, політиків і широку громадськість, потрібні різні інструменти вимірювання, щоб оцінити, наскільки добре вона обслуговує кожен сегмент аудиторії.

Успіх DeFake It! визначатиметься трьома ключовими показниками ефективності (KPI):

1. **Аналітика вебсайту, яка вимірює поведінку користувачів та їхню взаємодію з платформою**

Кількість користувачів (загальний трафік сайту), які відвідують DeFake It! протягом певного часу, відображає суспільний інтерес і популярність. Кількість часу, який користувачі проводять на платформі, свідчить про рівень залученості. Триваліші сесії показують, що користувачі активно вивчають контент, тоді як короткі відвідування можуть вказувати на потребу в більш привабливому або зручному для користувача дизайні. Високий показник відмов (коли відвідувачі залишають сайт після перегляду лише однієї сторінки) свідчить про те, що користувачі не знаходять те, що їм потрібно, або що контент недостатньо привабливий. Визначення того, які розділи вебсайту користуються найбільшим трафіком (наприклад, інструменти верифікації, навчальні модулі або глибокі тематичні дослідження), допомагає розставити пріоритети в розробці контенту, виходячи з інтересів користувачів. Відсоток користувачів, які повертаються на веб-сайт, вимірює лояльність і постійну залученість. Високі показники повернення свідчать про те, що DeFake It! стає ресурсом, якому довіряють. Ці показники можна

аналізувати, наприклад, за допомогою Google Analytics, що дозволить оптимізувати контент та покращити рівень утримання авдиторії.

2. **Рівень залученості в соціальних мережах**

Оскільки дезінформація швидко поширюється через соціальні мережі, DeFake It! використовуватиме такі платформи, як Twitter (X), Instagram, LinkedIn та YouTube, щоб ділитися інформацією про перевірку фактів, порадами щодо медіаграмотності та аналізом конкретних прикладів дезінформації. Вимірювання рівня залученості в соціальних мережах дає цінну інформацію про те, наскільки добре платформа резонує з аудиторією і чи є її повідомлення ефективними для підвищення обізнаності. Для того, щоб визначити це, необхідно враховувати такі показники, як охоплення та враження від публікації (вподобайки, коментарі, поширення), коефіцієнт кліків (CTR), темпи зростання кількості підписників. Високий рівень залученості вказує на те, що користувачі активно взаємодіють з публікаціями, що підвищує видимість алгоритмами й забезпечує більш активне поширення контенту. Відсоток користувачів, які переходять за посиланнями, що ведуть на сайт, вимірює ефективність конверсії — наскільки ефективне співвідношення публікацій у соціальних мережах та реальних відвідувань платформи. Регулярна взаємодія з авдиторією за допомогою фото, відео, інтерактивних опитувань та вправ з перевірки фактів, прямих етерів сприятиме подальшому збільшенню унікальних відвідувачів.

3. **Участь користувачів у вікторинах і тестах**

Такі платформи, як Online Test Pad, Kahoot! та Quizzes, пропонують зручні та цікаві формати для оцінювання розуміння. Високий відсоток проходження тесту свідчитиме про те, що користувачі вважають матеріал доступним, тоді як оцінки за виконання покажуть, де контент потребує подальшого роз'яснення. Крім того, такі платформи надають сертифікати з тестів на навички виявлення дипфейків, які можуть бути застосовані в професійній діяльності. Це також допоможе оцінити професійну цінність освітніх модулів.

2.2. Створення медіапродукту: вибір назви, структури контенту, візуального оформлення та інтерактивних елементів

Створення будь-якого медіапродукту починається з вибору теми, яка є актуальною та цікавою для цільової аудиторії. При розробці інтерактивного веб-сайту DeFake It! основна увага була зосереджена на висвітленні проблеми технології глибоких фейків та її наслідків для медіаграмотності. Цей вибір відповідає інтересам нашої аудиторії — журналістів, медіапрофесіоналів, студентів та споживачів цифрового контенту, які все частіше стикаються із синтетичними маніпуляціями в медіа.

Останніми роками технологія «глибоких фейків» зазнала значного зростання, що призвело до масового побоювання щодо її зловживань. У звіті Security.org за 2024 рік підкреслюється, що понад 10% компаній зіткнулися зі спробами або успішним шахрайством з використанням «глибоких фейків», а збитки в деяких випадках сягають 10% річного прибутку. Незважаючи на це, близько 25% керівників компаній практично не знайомі з технологією deepfake, а 80% компаній не мають протоколів для протидії deepfake-атакам. [3]

2023 року дослідники з Інституту Алана Тюрінга провели опитування «Behind the Deepfake: 8% Create; 90% Concerned» про вплив та сприйняття суспільством глибоких фейків у Великій Британії. 90,4% учасників висловили занепокоєння щодо поширення «глибоких фейків». Жінки, зокрема, повідомили про вищий рівень страху, пов'язаного з дипфейками, порівняно з чоловіками. Прикметно, що 91,8% репсондентів занепокоєні тим, що дипфейки можуть сприяти поширенню в Інтернеті матеріалів про сексуальне насильство над дітьми, посилювати недовіру до інформації та маніпулювати громадською думкою. [15, с. 7]

Крім того, дослідження Insikt Group, проведене у 2024 році, виявило 82 глибокі фейки, спрямовані проти публічних осіб у 38 країнах у період з липня 2023 року по липень 2024 року, що підкреслює глобальне поширення та політичні наслідки цієї технології [2]. У доповіді Сем Ріардон «The Impact of Deepfakes on

Journalism», опублікованій у 2024 році, підкреслюється, що зростання кількості контенту, створеного штучним інтелектом, у тому числі й дипфейків, розмило межу між фактами та вигадками, через що журналістам стає дедалі складніше перевіряти достовірність інформації. [36]

Основою DeFake It! є сильний, чіткий брендинг, чітко визначена місія та інтуїтивно зрозуміла назва — усе це робить наш сайт впізнаваним та ефективним інструментом для боротьби з обманом за допомогою технологій deepfake. Оскільки сайт обслуговуватиме різні групи аудиторії, зокрема журналістів, освітян і звичайних користувачів Інтернету, дизайн і стратегія наповнення мають бути структуровані таким чином, щоб сприяти ефективному навчанню, залученню та практичному застосуванню. Щоб посилити ідентичність бренду, його стратегія фокусується на цьому:

- **Ясність і прямі повідомлення.** Назва та інтерфейс сайту дають чітке, безпосереднє розуміння його призначення.
- **Інтерактивність та залучення.** Замість того, щоб бути пасивним інформаційним центром, DeFake It! побудований таким чином, щоб бути практичним. Саме тому ми пропонуємо тести, практичні вправи та інструменти перевірки в реальному часі.
- **Наукова достовірність та медіаекспертність.** Платформа містить дослідження, засновані на фактах, експертні висновки та верифікацію за допомогою штучного інтелекту. Це гарантує, що весь контент ґрунтується на доказах, а не на спекуляціях.

Візуальний дизайн та макет відображатимуть цю філософію брендингу — сучасний, інтуїтивно зрозумілий та професійний, при цьому уникаючи надмірної складності в сприйнятті. Місія DeFake It! — надати людям знання, інструменти та навички, необхідні для виявлення та протидії синтетичним маніпуляціям у медіаіндустрії. Платформа прагне заповнити прогалину в обізнаності громадськості

та професійних ресурсах з верифікації фейків, зокрема «глибоких підробок». Ця місія має п'ять основних завдань:

- навчити користувачів технологіям боротьби з «глибокими фейками» — надання доступних, але науково обґрунтованих пояснень, як створюються та перевіряються «глибокі фейки», як вони функціонують і які їхні етичні наслідки.
- запропонувати практичні інструменти перевірки — інтеграція ручних методів виявлення та ресурсів зі штучним інтелектом для практичного аналізу зображень, відео- та аудіоматеріалів.
- сприяти медіаграмотності — заохочити користувачів критично ставитися до онлайн-контенту, розпізнавати тактики маніпуляції та приймати поінформовані й відповідальні рішення.
- підтримати журналістів та освітян — наша платформа слугує ресурсом для професійного розвитку, пропонуючи готові до використання матеріали й тести для підвищення рівня знань у сфері виявлення фейків, зокрема «глибоких підробок».
- підвищити обізнаність авдиторії (журналісти, фактчекери, студенти, викладачі та широка громадськість) про етичні проблеми — висвітлення ключових питань конфіденційності, згоди та дезінформації, пов'язаних із технологією «глибоких фейків», допомагаючи користувачам зрозуміти ризики та етичну відповідальність.

Вибір правильної назви для нашого сайту — це фундаментальний крок у визначенні його ідентичності, мети та впливу. Сильна назва має не лише відображати місію проекту, але й запам'ятовуватися, залучати та бути інтуїтивно зрозумілою для аудиторії. Назва DeFake It! була ретельно підібрана, оскільки вона точно відповідає цілям платформи, ефективно передаючи її основну функцію — виявляти та протидіяти глибоким фейковим маніпуляціям. Прикметно, що

особливості назви закладені не лише на семантичному, а й на мовному рівні, що посилює її ефективність та впізнаваність.

Назва побудована у вигляді дієслівної фрази наказового способу, що безпосередньо спонукає користувачів до дії. Дієслово керує займенником «it» — так назва віддзеркалює мотиваційні вирази на кшталт «Зроби це!». Фраза за своєю суттю передбачає негайність і глобальну актуальність, надихаючи користувачів долучатися до виявлення дипфейків так, ніби від них залежить цілісність цифрової правди.

Вибір DeFake It! гарантує, що назва виконує функцію слогана, передаючи чіткий, точний і зрозумілий меседж, який не потребує пояснень. Він миттєво передає тему розвінчання фейків і сприяє миттєвому впізнаванню. На відміну від розпливчастих або занадто технічних назв, цей слоган є простим і орієнтованим на дію, що робить його доступним навіть для тих, хто не знайомий з методами виявлення глибоких фейків.

Слова «deepfake» і «defake» мають схожу фонетичну структуру, що робить їх майже взаємозамінними за звучанням. Схожість звучання створює інтуїтивний асоціативний зв'язок, що дозволяє користувачам миттєво зрозуміти, на що спрямована платформа. Транскрипція «deepfake» /diːp feɪk/ та «defake» /diː feɪk/ підкреслює м'який перехід між двома словами. Ледь чутний звук «р» у слові «deepfake» плавно зливається з «defake», посилюючи семантичний та звуковий зв'язок між термінами.

Префікс «de-» має вирішальне значення для значення назви, оскільки він універсально означає скасування, видалення або заперечення. Так само, як «декодувати» означає розшифровувати закодовану інформацію, а «деактивувати» — видаляти шкідливі речовини, «defake» означає процес викриття та протидії синтетичним маніпуляціям у медіа. Хоча «defake» технічно є оказіоналізмом, значення цього слова одразу зрозуміле, оскільки воно створене префіксальним способом (de- + fake — неправдива інформація).

Розробка DeFake It! проходитиме у шість ключових етапів, що забезпечить її функціонування як цілісного та доступного ресурсу з медіаграмотності:

1. **Створення інфраструктури вебсайту**

- Вибір платформи. DeFake It! буде розроблено на Flamer — гнучкій, масштабованій та широко підтримуваній системі керування контентом.
- Домен і хостинг. Буде придбано власний домен, щоб надати платформі авторитетність і довгострокову стабільність. Домен має бути простим для написання й запам'ятовування, відображати назву платформи та її місію й характер (у нашому випадку — некомерційний та освітній). За цими вимогами ми обрали домен defake-it.framer.website.

2. **Визначення структури вебсайту та інтерфейсу користувача (UI).** Структурно сайт складатиметься з п'яти основних розділів: «Про нас», «Статті», «Інструменти», «Тести», «Контакти». Кожен розділ спроектований для максимальної зручності та залучення користувачів, що дозволяє їм легко орієнтуватися та знаходити необхідну інформацію:

- Головна сторінка (розділ «Про нас»). Знайомить із місією та цілями платформи, висвітлює роль штучного інтелекту та виявлення дипфейків у сучасній журналістиці та медіаграмотності. Містить FAQ (часті питання) щодо deepfake та верифікації контенту.
- Розділ «Матеріали». Містить навчальні матеріали про методи створення й виявлення дипфейків, практичні приклади та лайфхаки тощо. Матеріали фільтруються за тегами, датою або темою для швидкого пошуку.
- Розділ «Ресурси». Надає доступ до ресурсів штучного інтелекту та ручних методів перевірки deepfake. Покрокові інструкції пояснюють принципи роботи кожного інструменту.
- Розділ «Тести». Наразі включає два інтерактивні тести: теоретичний тест, що дозволяє оцінити свої знання технології deepfake, та практичний тест на

розпізнавання реальних та ШІ-генерованих зображень. Підходить як для початківців, так і для професіоналів, є можливість отримати сертифікат після успішного проходження тестів. Реалізовано на зручних платформах (Kahoot! та Online Test Pad).

- Розділ «Контакти». Містить форми для зворотного зв’язку, можливості для співпраці та партнерства. Також передбачено посилання на соціальні мережі для залучення спільноти.

3. Розробка контенту та структуризація інформації

Контент-план DeFake It! має структуру, яка гарантує, що користувачі поступово формують своє розуміння технології дипфейків, методів їх виявлення та етичних міркувань. Пости ретельно відібрані, щоб охопити як теоретичні, так і практичні аспекти, і завершуються інтерактивними тестами для оцінки та застосування здобутих знань. На момент розвитку проєкту буде опубліковано дев’ять матеріалів, які демонструють потенціал DeFake It! як інтерактивної платформи для верифікації глибоких фейків та підвищення медіаграмотності. Ці пости будуть виходити приблизно кожні 1-2 тижні, що забезпечить постійний потік освітнього контенту й дасть змогу проводити глибокі дослідження та аналіз. Крім того, цей контент-план слугує прототипом моделі для майбутніх матеріалів, які будуть опубліковані на сайті. З розвитком технологій і засобів протидії «глибоким фейкам» будуть додаватися нові статті та кейси з верифікації, що дозволить DeFake It! залишатися динамічним і актуальним освітнім ресурсом.

Пости розташовані в логічній послідовності, де кожна стаття ґрунтується на раніше представлених концепціях. Дві останні публікації — це інтерактивні тести, що дозволяють користувачам застосувати те, що вони вивчили за допомогою теоретичних оцінок і практичних вправ на перевірку.

Теоретичний тест буде містити 20 запитань, а практичний — 15. Завдання охоплюють основи виявлення підробок, методи виявлення ШІ, тематичні дослідження та стратегії ручної перевірки. Ці інтерактивні тести дозволять

користувачам визначити сильні сторони та сфери, які потребують подальшого навчання. Наповнення сайту складатиметься з таких публікацій:

Публікація 1. Створення дипфейку в освітніх цілях за допомогою інструментів штучного інтелекту на прикладі Авраама Лінкольна (частина 1). У цій статті розповідається про розробку сценарію та створення дипфейку за підготовленим сюжетом в освітніх цілях за допомогою ChatGPT 4.0, DeepL і RunwayML. На авторському прикладі демонструється, як ШІ може відтворювати суспільні події та історичних постатей. ChatGPT і DeepL використовуються для розробки послідовного сценарію дипфейку, адаптуючи згенеровану мову до історичного контексту й керованої ШІ механіки RunwayML. Кінцевий продукт — підроблене відео промови Авраама Лінкольна, який на власному прикладі демонструє небезпеку фальсифікацій у ЗМІ та важливість цифрової безпеки у медіапросторі. Публікація показує, як ШІ створює глибоко підроблений контент, інформуючи користувачів про процес, потенціал та ризики, що стоять за цифровим обманом.

Публікація 2. Перевірка дипфейку за допомогою інструментів штучного інтелекту на прикладі Авраама Лінкольна (частина 2). У цій статті досліджується, як інструменти виявлення підробок на основі штучного інтелекту, зокрема Deepware, Resemble AI та InVID & WeVerify, можуть перевірити автентичність синтетичного контенту. Підробка Авраама Лінкольна з публікації 1 аналізується за допомогою поетапних методів перевірки, демонструючи, як різні системи виявлення виявляють маніпуляції у відео, створених штучним інтелектом. Матеріал забезпечує користувачів практичними інструментами та методологіями для верифікації фейків, підвищуючи їхню здатність виявляти маніпуляції в медіа в реальних умовах.

Публікація 3. Створюємо дипфейк за текстом і фото: інструменти AI для кожного та кожної. У цій статті розглядаються передові технології створення фальшивих відео із синхронізацією губ і голосу за допомогою ElevenLabs і HeyGen.

У кейсі наводиться відеоанонс заходу Академії української преси «Медіамайстерність 4.0: викладання крізь призму штучного інтелекту», створений за допомогою штучного інтелекту, в якому Тетяна Іванова та Любов Цукор з'являються в цифровому відео, промовляючи текст, який вони насправді не записували. У публікації пояснюється, як працює технологія синхронізації губ на основі штучного інтелекту та чому стає дедалі складніше виявити такі маніпуляції.

Публикація 4. Етичні інструменти на основі штучного інтелекту для медіапрофесіоналів на всі випадки життя. Основна увага приділяється таким застосункам, як: інструменти для перетворення тексту у відео для створення візуального ряду новин; персоналізація контенту на основі штучного інтелекту для покращення цифрового сторітелінгу; інструменти штучного інтелекту для обробки великих даних для аналізу та перевірки медіа в масштабах; програмне забезпечення для автоматизованої візуалізації для створення діаграм, графіків і GIF-файлів на основі текстових даних.

Публікація 5. Туторіал зі створення освітнього дипфейку: від сценарію до візуалізації за допомогою інструментів штучного інтелекту. Цей покроковий туторіал є практичним посібником для всіх, хто бажає створити **освітнє дипфейк-відео** на основі історичних персонажів. У ньому докладно описано, як на кожному етапі застосовуються інструменти ШІ: від **генерації сценарію в ChatGPT** з урахуванням стилістики історичних постатей — до **створення голосів в ElevenLabs**, **анімації в HeyGen**, **перекладу через DeepL** та **монтажу фінального відео в Canva**. Туторіал охоплює підбір релевантних персонажів та теми; формулювання ефективних промптів; синтез голосу з урахуванням тону, акценту та емоційності; синхронізацію обличчя з аудіо; адаптацію для україномовної аудиторії; етичний супровід: маркування фейку, пояснення мети. Матеріал орієнтований на **медіапрофесіоналів, викладачів, тренерів з медіаграмотності, журналістів та креаторів**, які хочуть **використовувати deepfake-технології у легальний та просвітницький спосіб**. Він демонструє, що **глибокі фейки можуть бути не лише**

загрозою, а й потужним інструментом для навчання, якщо супроводжуються прозорістю та критичним аналізом.

Публікація 6. Перевірка дипфейків вручну та за допомогою штучного інтелекту: Зеленський, Залужний, Міддлтон. Ця публікація зосереджена на **техніках верифікації дипфейків** через аналіз **трьох реальних кейсів**, що стали гучними у світі цифрової безпеки та медіаграмотності. У кожному з них було використано штучний інтелект для створення **візуально та аудіально переконливого фейку**, який мав на меті маніпулювати громадською думкою, дестабілізувати ситуацію або посіяти паніку.

Кейс Зеленського. У березні 2022 року, на ранніх стадіях російського вторгнення в Україну, з'явився дипфейк, на якому президент Володимир Зеленський нібито наказує українським солдатам здаватися в полон. Сфабриковане відео було поширене через різні онлайн-платформи і навіть з'явилося на зламаному сайті українського телеканалу «Україна 24». У відео за допомогою штучного інтелекту маніпулювали обличчям і голосом Зеленського, щоб створити оманливе повідомлення, спрямоване на підрив морального духу українців.

Кейс Залужного. У листопаді 2023 року з'явилося фейкове відео за участю генерала Валерія Залужного, головнокомандувача Збройних сил України. На змонтованому відео Залужний закликає до військового перевороту проти президента Зеленського, що свідчить про внутрішні розбіжності в українському керівництві. Цей високоякісний глибокий фейк був поширений через соціальні мережі з метою посіяти розгубленість і недовіру серед українського населення та міжнародних союзників. [16]

Кейс Міддлтон. У березні 2024 року Кетрін, принцеса Уельська, публічно повідомила про свій діагноз раку й про те, що проходить курс хіміотерапії. Після цієї заяви з'явилися конспірологічні теорії, які стверджували, що подальші публічні виступи та відеозвернення принцеси були дипфейками, створеними для приховування її справжнього стану. Ці твердження набували популярності на

платформах соціальних мереж, де користувачі ретельно перевіряли відео на наявність ознак цифрової маніпуляції. Незважаючи на офіційні заяви, що підтверджують автентичність зовнішності принцеси, спекуляції не припинилися. [32]

Публікація 7. Інтерактивний тест на виявлення дипфейків — оцінка теоретичних знань. Тест з декількома варіантами відповідей, що оцінює розуміння користувачами технології виявлення підробок, етичних проблем, інструментів виявлення штучного інтелекту та методів верифікації. Тест складатиметься з 16 запитань, що охоплюють історію та еволюцію технології deepfake, поширені методи маніпуляцій, що використовуються в медіа, створених за допомогою штучного інтелекту, різноманітні практики фактчекінгу та верифікації. Дозволяє користувачам оцінити свої теоретичні знання, визначаючи сфери, у яких вони потребують подальшого навчання.

Публікація 8. Інтерактивний тест на виявлення підробок — практична перевірка зображень та відео. У цій практичній вправі користувачам буде представлено 15 зображень і відео, які потрібно буде проаналізувати та визначити, чи є вони справжніми, чи згенерованими штучним інтелектом. Тест буде зосереджений на візуальних невідповідностях та індикаторах виявлення ШІ. Тест імітує реальні сценарії виявлення дипфейків, дозволяючи користувачам застосувати отримані знання на практиці.

4. Створення візуальної айдентики DeFake It!

Візуальний стиль DeFake It! відіграє ключову роль у впізнаваності платформи, залученні користувачів і відповідності її місії. Центральним елементом бренду є робот-талісман, який візуально втілює ідею перевірки медіа за допомогою штучного інтелекту. Робот — це не просто уособлення штучного інтелекту, а відображення взаємодії технологій із людськими процесами. У сучасному світі ШІ дедалі більше автоматизує завдання, які традиційно виконували люди — від створення контенту

до фактчекінгу. Проте, хоча ШІ може загрожувати журналістиці, він також може стати потужним союзником за умови відповідального використання.

Щоб відповідати освітній направленості платформи, надамо роботу певну візуальну характеристику, яка допоможе нам згодом згенерувати правильний логотип, що відповідає іміджу, стилю і потребам нашої платформи:

- Футуристичний, але дружній дизайн. Уникає занадто механістичного або загрозливого вигляду, що робить його доступним для всіх груп користувачів.
- Мінімалістичний і професійний стиль. Чіткі лінії та простота форми забезпечують його придатність для різних цілей — від брендингу сайту до рекламних матеріалів.
- Нейтральний, але інтелектуальний вираз обличчя. Робот передає аналітичну точність, яка асоціюється із фактчекінгом.
- Тонкі цифрові елементи. Наприклад, скануюче око, м'яке підсвічування чи цифрові лінії. Це підкреслює його зв’язок зі штучно-інтелектуальними технологіями.
- Кольорова схема. Синій, білий і сріблястий — у поєднанні вони символізують інтелект, довіру та ясність.

За допомогою Lexica (платформа генерації зображень ШІ) було створено візуальний рендер робота за заздалегідь визначеними характеристиками. Використовуваний запит містив:

«Сучасний, професійний робот-асистент для виявлення deepfake. Футуристичний дизайн зі скануючим оком, цифровим інтерфейсом і гладким металевим корпусом. Кольорова палітра: синій, сріблястий і білий. Фон має бути простим і однотонним для зручності використання в брендингу».

Після генерації кількох варіантів було обрано найкращий прототип для подальшого доопрацювання. На фінальному етапі ми видалили фон для зручного

використання логотипу за допомогою RemoveBG (ШІ-інструмент для видалення фону), що дозволяє легко адаптувати зображення для різних потреб.

Тепер оберемо цілісну колірну палітру, яка буде зосереджена на нашій платформі. Це також невіддільна частина створення сайту, оскільки має ледь не основний вплив на функціональність та користувацький досвід. Ключовими кольорами платформи є:

1. Основний колір: синій (#007BFF). Використовується для інтерактивних елементів (кнопок, посилань, підсвічувань), щоб створити візуальний зв'язок з надійними інструментами на основі штучного інтелекту. Асоціюється з технологіями, інтелектом і цифровою безпекою. Підсилює відповідність платформи принципам доброчесності медіа та аналітичним інструментам, оскільки повторює колірні схеми, що використовуються LinkedIn, IBM та Microsoft.

2. Другий основний колір: білий (#FFFFFF). Створює нейтральне, чисте тло, гарантуючи, що контент залишається в центрі уваги. Підвищує контрастність і читабельність, запобігаючи зоровій втомі та покращуючи користувацький досвід.

3. Акцентний колір: сріблястий/сірий (#A8A8A8). Використовується для обвідки, ефектів наведення та другорядних елементів.

4. Контрастний і акцентний колір: темно-синій/чорний (#212529). Покращує розбірливість і фокусується на ключових розділах, таких як заголовки, банери та важливі сповіщення. Використовується економно для створення ієрархії та зручної навігації користувача, виділяє лише найбільш важливий контент.

5. Оптимізація користувацького досвіду (UX) та доступності. Гарантує повну адаптивність та зручність на різних пристроях. Реалізація логічної та інтуїтивної структури, щоб мінімізувати можливі труднощі у використанні. Спочатку платформа буде доступна українською мовою з можливістю розширення на інші мови в майбутньому.

6. Просування, залучення аудиторії та постійне вдосконалення. Це означає інтеграцію освітньої платформи DeFake It! у соціальні мережі, заохочення

медіапрофесіоналів і науковців до внеску в платформу через аналітику, кейси та дослідження, регулярні оновлення публікацій, інструментів перевірки та інтерактивних функцій відповідно до відгуків користувачів і нових тенденцій у технології deepfake.

2.3. Методи просування сайту «DeFake It!»

Ефективність «DeFake It!» залежить не лише від якості освітнього контенту, а й від його здатності охопити та залучити цільову аудиторію. Для того, щоб платформа стала широкодоступним і авторитетним ресурсом для виявлення фейків, необхідна комплексна стратегія просування.

Охоплення журналістів, фактчекерів, освітян, студентів і широкої громадськості вимагає багатогранного підходу, який містить залучення соціальних мереж, партнерство з освітніми установами, співпрацю з фактчекінговими організаціями та стратегічний цифровий маркетинг. Основна мета — перетворити DeFake It! на платформу для глибокої перевірки фейків та медіаграмотності, зробити її інструменти та контент широко впізнаваними та активно використовуваними.

Щоб ефективно позиціонувати DeFake It! у наявному ландшафті платформ з медіаграмотності та фактчекінгу, украй важливо проаналізувати та порівняти подібні ініціативи. Розуміння того, як працюють інші платформи, їхні сильні та слабкі сторони, а також їхній вплив на аудиторію, допоможе вдосконалити унікальну ціннісну пропозицію DeFake It! Чотири відомі українські платформи з медіаграмотності — «Детектор медіа», StopFake, «Без брехні» та «По той бік новин» — слугують орієнтирами. Ці платформи надають ресурси для перевірки фактів, викривають дезінформацію та пропонують освітні матеріали з медіаграмотності. Хоча вони відіграють важливу роль у боротьбі з дезінформацією, їхні підходи відрізняються від інтерактивної методології DeFake It!, що ґрунтується на штучному інтелекті.

Представляємо порівняльну таблицю аналізу, яка оцінює різні освітні платформи про дипфейки за ключовими критеріями: інтерактивність, доступність, популярність та освітній вплив (див. табл. 2). Це порівняння підкреслює сильні та слабкі сторони кожної платформи, показуючи, як DeFake It! вирізняється завдяки інтерактивним тестам, інструментам верифікації на основі ШІ та навчальним матеріалам.

Таблиця 2

Оцінка українських освітніх платформ з фактчекінгу за критеріями

Платформа	**Основний фокус**	**Рівень інтерактивності**	**Доступність та зручність**	**Популярність та охоплення**	**Освітній вплив та ефективність**
Detector Media	Медіааналіз, журналістські розслідування, фактчекінгові звіти.	**Низький**. Переважно статті, звіти та експертний аналіз.	Безкоштовна, доступна українською, переважно текстовий формат.	Добре відома в Україні, активна медіаспівпраця.	**Високий.** Сильна репутація у сфері медіааналізу та фактчекінгу.
StopFake	Викриття кремлівської пропаганди, моніторинг дезінформа	**Середній**. Відеозвіти та навчальні сесії.	Безкоштовна, сайт і соціальні мережі, фокус на відеоматеріа	Міжнародне визнання, широко використовується для фактчекінгу.	**Високий.** Значний вплив на викриття пропаганди.

	ції, навчання медіаграмотності.		лах.		
Без Брехні	Політичний фактчекінг, аналіз фейкових новин у медіа та соцмережах	**Середній.** Пояснення фейків покроково.	Безкоштовна, переважно україномовний контент.	Висока популярність в Україні, довіра у сфері політичного фактчекінгу.	**Середній.** Фокусується на політичній дезінформації, але не включає дипфейки.
По Той Бік Новин	Швидке спростування міфів, аналіз дезінформації та інфографіки	**Середній**. Фактчекінг з використанням інфографіки та коротких постів.	Безкоштовна, орієнтація на соцмережі, активне залучення аудиторії.	Популярна в соцмережах, висока взаємодія з користувачами.	**Середній.** Корисна для загальної медіаосвіти, але без структури навчання про дипфейки.
DeFake It!	Верифікація дипфейків за допомогою AI,	**Високий.** Інтерактивні тести, AI-інструмент	Безкоштовна, орієнтація на різні засоби мультимедіа:	Новий проєкт, аудиторія зростає, унікальна	**Високий.** Поєднує AI-верифікацію з

	інтерактивне навчання, практичний аналіз	и перевірки, навчальні туторіали.	фото, відео; імплементація ШІ-технологій та інтерактивних тестувань	спеціалізація на дипфейках.	повноцінною навчальною програмою про дипфейки.

Отже, на відміну від інших платформ з медіаграмотності, DeFake It! використовує інтегрований зі штучним інтелектом практичний підхід до виявлення дипфейків. Тоді як конкурентоспроможні українські освітні ресурси зосереджені переважно на аналізі фейкових новин, моніторингу дезінформації та ручній перевірці фактів, ця платформа пропонує інструменти для перевірки фейків на основі штучного інтелекту в режимі реального часу, які дозволяють користувачам активно аналізувати маніпульований контент. Платформа містить інтерактивні тести та вправи з інструкціями, які дають змогу користувачам практикуватись у виявленні «глибоких фейків» у різних медіаформатах — зображеннях, відео, авдіо. Крім того, платформа пропонує покрокові навчальні посібники зі створення «глибоких фейків», які допомагають користувачам глибше зрозуміти методи маніпуляцій та їхню механіку.

Для ефективного просування DeFake It! також буде застосовано багатоканальний підхід, що поєднує цифрові та традиційні медіа. Соціальні медіа-платформи, такі як Instagram, X (колишній Twitter), TikTok та YouTube, послужать основними рушіями залучення, забезпечуючи охоплення широкої та різноманітної аудиторії за допомогою інтерактивного контенту, створеного на основі штучного інтелекту. Традиційні методи PR — радіо, телебачення, газети та блоги — допоможуть підвищити довіру до платформи та розширити охоплення професійних кіл, освітян і політиків.

Кожна платформа вимагає індивідуальної стратегії просування для досягнення максимальної ефективності. Намалюємо таблицю, яка представляє аналіз різних медіаканалів, оцінюючи рівень взаємодії, охоплення аудиторії, гнучкість контенту, рентабельність та ефективність для DeFake It! (див. табл. 3). Це допомагає визначити найкращі платформи для охоплення цільової аудиторії та ефективної комунікації:

Таблиця 3

Оцінка каналів просування освітньої інтерактивної платформи DeFake It!

Медіаканал	**Рівень взаємодії**	**Охоплення аудиторії**	**Гнучкість контенту**	**Рентабельність**	**Ефективність для DeFake It!**
Instagram	**Високий**. Інтерактивний контент (сторіз, рілз, пости).	**Широке**. Молодь (18-34), медіапрофесіонали, загальна аудиторія.	**Дуже висока**. Підтримує візуальний, відео- та текстовий контент.	**Висока**. Безкоштовне використання, можливість платної реклами.	**Дуже висока**. Візуальний формат ідеальний для аналізу дипфейків.
X (Twitter)	**Середній**. Оперативні дискусії, менша взаємодія на пост.	**Широке**. Журналісти, політики, медіааналітики.	**Висока**. Текстові оновлення, можливість вкладення зображень.	**Висока**. Безкоштовне використання з можливістю платної	**Висока**. Корисно для оновлень у режимі реального часу.

				реклами.	
TikTok	**Дуже високий**. Короткі відео, висока ймовірність вірусного поширення.	**Масове**. Молодь та покоління Z.	**Дуже висока**. Креативні відеоролики, сторітелінг.	**Висока**. Органічне зростання, можливість платного просування.	**Дуже висока**. Ідеально для коротких освітніх відео.
YouTube	**Високий**. Довгі формати для поглибленого аналізу.	**Широке**. Аудиторія різних поколінь.	**Висока**. Підтримка відеоуроків і документальних форматів.	**Середня**. Потрібні інвестиції у створення контенту, але сильне органічне охоплення.	**Висока**. Відмінно підходить для освітніх матеріалів.
Радіо	**Низький**. Пасивне слухання, мінімальна взаємодія.	**Нішеве**. Аудиторія старшого віку, новинні слухачі.	**Низька**. Тільки аудіоформат.	**Середня**. Потрібні витрати на рекламу, але є лояльна аудиторія.	**Середня**. Корисно для експертних інтерв’ю.
Телебачення	**Низький**. Традиційн	**Широке**. Переважно	**Низька**. Фіксований	**Дуже висока**.	**Низька**. Корисно

	ий формат, мінімальна взаємодія.	середній та старший вік.	формат, обмежена гнучкість.	Високі витрати на виробництво і рекламу.	для репутації, але не для взаємодії з аудиторією.
Газети	**Низький**. Статичний формат, обмежена взаємодія.	**Зменшується**. Орієнтація на старшу аудиторію.	**Низька**. Текстовий формат.	**Дуже висока**. Витрати на друк і розповсюдження.	**Низька**. Корисно для авторитетності, але без інтерактивності.
Блоги й форуми	**Середній**. Можливість поглибленого аналізу.	**Нішеве**. Орієнтація на медіаграмотність та технології.	**Середня**. Аналітичні статті, але без гнучкості в реальному часі.	**Середня**. Потребує витрат на SEO та маркетинг.	**Середня**. Корисно для детальних пояснень, але потребує активного пошуку аудиторією.

Зважаючи на те, що DeFake It! зосереджується на виявленні глибоких фейків за допомогою штучного інтелекту та інтерактивному навчанні, соціальні мережі стануть основним рушієм просування завдяки високому рівню залучення, адаптивності та охопленню аудиторії. Instagram, TikTok, X та YouTube

слугуватимуть основними платформами для просування, дозволяючи проводити наочні демонстрації, інтерактивне навчання та швидке залучення аудиторії.

Водночас традиційні PR-методи (радіо, телебачення, газети та блоги) сприятимуть підвищенню довіри до кампанії, що забезпечить визнання DeFake It! серед політиків, дослідників та медіа-професіоналів. Хоча традиційним медіа бракує інтерактивності, вона залишається важливою для встановлення інституційної довіри та охоплення аудиторії, яка не може взаємодіяти із цифровими платформами.

Розширення впливу DeFake It! вимагає й стратегічної співпраці з провідними освітніми організаціями, які просувають медіаграмотність та журналістську доброчесність в Україні. Кілька інституцій, серед яких Академія української преси (АУП), Український інститут медіа та комунікації, Фонд розвитку ЗМІ та Тренінговий центр «Детектор медіа», активно проводять тренінги, воркшопи та інформаційні кампанії, спрямовані на підвищення медіаграмотності та виявлення дезінформації. З-поміж них Академія української преси (далі — АУП) має найбільші можливості для співпраці, зважаючи на її визнану роль у навчанні журналістів, освітян та студентів медіаграмотності та технік фактчекінгу. Інтегрувавши DeFake It! у воркшопи, тренінги та навчальні посібники АУП, платформа стане важливим ресурсом для виявлення глибоких фейків в українському журналістському та академічному середовищі.

Щоб забезпечити практичне застосування, DeFake It! буде включено до семінарів з медіаграмотності та просвітницьких програм для журналістів, які надаватимуть практичний досвід у виявленні та протидії синтетичним медіа. Під час тренінгів АУП з медіаграмотності учасники аналізуватимуть реальні випадки глибоких фейків, застосовуватимуть методи верифікації, керовані штучним інтелектом, та вивчатимуть ручні методи перевірки фактів.

Окрім тренінгів, DeFake It! буде включена в навчальні матеріали та посібники, що розповсюджуватимуться в рамках програм з медіаграмотності АУП.

Інтерактивні тести та інструменти перевірки платформи слугуватимуть практичними вправами в рамках навчальних програм з журналістики, допомагаючи студентам та професіоналам вдосконалювати свої навички виявлення глибоких фейків у реальному контексті.

Розгалужена мережа освітян та медіатренерів АУП сприятиме просуванню DeFake It! у школах, університетах та медіаорганізаціях, гарантуючи, що широка аудиторія отримає доступ до верифікаційних інструментів та освітніх ресурсів на основі штучного інтелекту.

Завдяки репутації та інституційному охопленню АУП платформа DeFake It! набуде популярності серед журналістів, політиків та освітян, зміцнивши свою роль як провідної платформи для підвищення обізнаності про глибокі фейки та навчання медіаграмотності.

Висновки до Розділу 2

Розвиток DeFake It! як інтерактивної освітньої платформи для боротьби з глибокими фейками вимагає стратегічного й добре структурованого підходу, що забезпечує як технічну досконалість, так і ефективне охоплення аудиторії. Поєднуючи інструменти верифікації на основі штучного інтелекту, практичні навчальні матеріали та аналіз фейків у режимі реального часу, платформа покликана заповнити критичну прогалину в медіаграмотності, надаючи користувачам практичні навички виявлення та протидії синтетичним медіаманіпуляціям.

Цілеспрямована стратегія контенту гарантує, що користувачі поступово накопичуватимуть знання, просуваючись від базових знань про створення «глибоких фейків» до передових методів верифікації та реальних кейсів. Включення інтерактивних вправ, структурованих навчальних посібників та інструментів на основі штучного інтелекту підвищує залученість, роблячи виявлення «глибоких фейків» доступним і застосовним для різних груп користувачів — від журналістів і викладачів до студентів і широких верств населення, що споживають медіа.

Ефективний візуальний і структурний дизайн посилює вплив DeFake It!, надаючи пріоритет ясності, зручності та інтерактивній взаємодії. Завдяки вибору чітко визначеного доменного імені, структуруванню контенту на окремі розділи та впровадженню цілісної кольорової палітри, що асоціюється з технологіями та цифровою довірою, платформа підтримує довіру до себе та доступність.

Багатоканальна стратегія просування гарантує, що DeFake It! досягне своєї цільової аудиторії завдяки поєднанню маркетингу в соціальних мережах, інституційної співпраці та традиційних засобів масової інформації. Платформи соціальних мереж, такі як Instagram, TikTok, X та YouTube, надають широкі можливості для залучення завдяки короткометражним відео, інтерактивним завданням та навчальним посібникам з виявлення глибоких фейків, тоді як партнерство з радіо, телебаченням та академічними установами зміцнює довіру до кампанії та розширює охоплення професіоналів та політиків. Співпраця з Академією української преси (АУП) ще більше посилює освітню цінність платформи, дозволяючи інтегрувати її інструменти та ресурси в навчальні семінари, курси журналістики та програми з медіаграмотності.

Завдяки комплексному та структурованому впровадженню DeFake It! позиціонується як провідна платформа для підвищення обізнаності про фейки та медіаграмотності. Поєднання інтерактивного навчання, верифікації за допомогою штучного інтелекту та стратегічного просування гарантує, що платформа залишатиметься актуальною, доступною та ефективною, надаючи користувачам можливість впевнено та професійно орієнтуватися у зростаючих викликах синтетичних маніпуляцій у медіа. в Інтернеті на спеціальних платформах. Він є модернізованим аналогом радіомовлення.

ВИСНОВКИ

Стрімкий розвиток штучного інтелекту створив як можливості, так і значні виклики в сучасному інформаційному просторі. Технологія «глибоких фейків», яка колись була нішевою дослідницькою концепцією, перетворилася на потужний інструмент як для творчих інновацій, так і для широкомасштабної дезінформації. Хоча контент, створений штучним інтелектом, пропонує переваги в таких сферах, як освіта, розваги та доступність, його зловживання створює серйозні етичні та соціальні ризики, особливо в журналістиці, політиці та кібербезпеці. Можливість маніпулювати реальністю за допомогою медіа, керованих штучним інтелектом, загрожує суспільній довірі, демократичним процесам і журналістській доброчесності, що робить розробку ефективних методів верифікації та освітніх ресурсів нагальною необхідністю.

Розуміння технології «глибоких фейків» вимагає глибокого вивчення її еволюції, методів створення та застосування в реальному світі. Поєднання візуального та аудіосинтезу на основі штучного інтелекту, керованого генеративними змагальними мережами (GAN) та моделями глибокого навчання, значно підвищив реалістичність синтетичних медіа. Демократизація інструментів створення глибоких фейків розширила доступ до цих технологій, збільшивши їх використання в політичній пропаганді, дезінформаційних кампаніях і кібершахрайстві.

Методи верифікації також розвивалися у відповідь на ці досягнення, включаючи як ручні підходи, так і підходи, керовані штучним інтелектом. Традиційні методи, такі як аналіз метаданих, судова експертиза зображень і поведінкові невідповідності у відео- та аудіоконтенті, залишаються актуальними, але все частіше доповнюються інструментами глибокого виявлення підробок на основі штучного інтелекту. Такі платформи, як Deepware, Resemble AI та InVID & WeVerify, використовують машинне навчання для виявлення невидимих для людського ока слідів маніпуляцій, що робить їх важливим ресурсом у боротьбі з

синтетичними медіа-обманами. Поділяючи підходи до верифікації на ручні та автоматизовані, стає можливим побудувати комплексну стратегію протидії глибоко фейковій дезінформації, гарантуючи, що як окремі особи, так і працівники в медіаіндустрії зможуть ефективно аналізувати підозрілий контент.

Хоча теоретичні дослідження «глибоких фейків» дають цінну інформацію про їхні ризики та заходи протидії, ефективна боротьба з ними потребує доступних і практичних рішень для працівників медіаіндустрії та широкої громадськості. Розробка інтерактивної освітньої платформи DeFake It! задовольняє цю потребу, бо пропонує структурований, практичний підхід до верифікації «глибоких фейків».

Поєднуючи покрокові навчальні інструкції, реальні кейси та вправи з верифікації за допомогою штучного інтелекту, DeFake It! надає користувачам як теоретичні знання, так і практичні навички. Поступовий процес навчання гарантує, що користувачі перейдуть від базової обізнаності про технології глибоких підробок до просунутих можливостей перевірки, застосовуючи свої знання в інтерактивних тестах і практичних вправах з виявлення фейків. Включення гучних кейсів, таких як дипфейки з Володимиром Зеленським, Валерієм Залужним та Кейт Міддлтон, посилює реальне значення перевірки глибоких фейків, роблячи навчання цікавим і безпосередньо застосовним до поточних викликів у медіасфері.

Чітко визначений брендинг та візуальна ідентичність посилюють доступність та залученість платформи. Вибір робота як маскота символізує поєлнання технологій штучного інтелекту та журналістської доброчесності й відповідальності. Це наголошує на важливій місії етичного використання AI для перевірки інформації. Обрана кольорова гама та зручна структура сайту ще більше підвищують зручність використання, гарантуючи, що DeFake It! слугує практичним та цікавим ресурсом як для професіоналів, так і для споживачів медіа.

Зі зростанням витонченості дезінформації, генерованої штучним інтелектом, медіаграмотність повинна розвиватися, щоб включити розпізнавання фейків, зокрема «глибоких підробок», як основну компетенцію. Традиційні методи

перевірки фактів, хоча і є важливими, вже не є достатніми в епоху, коли ШІ може створювати гіперреалістичний синтетичний контент. Такі платформи, як DeFake It!, представляють нове покоління інструментів медіаосвіти, що поєднують виявлення фейків у режимі реального часу зі штучним інтелектом та інтерактивним навчанням для підвищення цифрової стійкості до маніпуляцій.

Завдяки співпраці з установами, що займаються медіаграмотністю, такими як Академія української преси (АУП), DeFake It! стає більше, ніж просто вебсайтом. Він слугує інтегрованим освітнім інструментом у програмах підготовки журналістів, семінарах з медіаграмотності та академічних навчальних планах. Надаючи журналістам, студентам та викладачам знання та навички верифікації, необхідні для навігації в інформаційному ландшафті, керованому штучним інтелектом, DeFake It! зміцнює журналістську доброчесність та довіру громадськості до фактологічних репортажів.

У майбутньому штучний інтелект продовжуватиме розвиватися, з'являтимуться нові розробки у створенні синтетичних медіа, алгоритмів виявлення та інструментів верифікації в режимі реального часу. Проблема полягає в тому, щоб технології верифікації не відставали від обману, породжуваного штучним інтелектом. Подальші досягнення в галузі машинного навчання, автентифікації контенту на основі блокчейну та виявлення дезінформації на різних платформах можуть забезпечити додаткові рівні безпеки в боротьбі з глибокими маніпуляціями з фейками. Однак постійна просвітницька робота з громадськістю залишається критично важливою, оскільки технологічні рішення самі по собі не можуть повністю протидіяти суспільному впливу синтетичної дезінформації.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. 2023 State of Deepfakes: Realities, Threats, and Impact. *Security Hero*. URL: https://www.securityhero.io/state-of-deepfakes/ (date of access: 10.04.2025).
2. 2024 Deepfakes and Election Disinformation Report: Key Findings & Mitigation Strategies. *Advanced Cyber Threat Intelligence | Recorded Future*. URL: https://www.recordedfuture.com/research/targets-objectives-emerging-tactics-political-deepfakes (date of access: 24.04.2025).
3. 2024 Deepfakes Guide and Statistics | Security.org. *Security.org*. URL: https://www.security.org/resources/deepfake-statistics/ (date of access: 24.04.2025).
4. Вальорська, А. Діпфейк та дезінформація / А. Вальорська ; переклад з німецької мови В. Олійника. – Київ : Академія української преси ; Центр вільної преси, 2020. – 36 с.
5. Іванова Т. Медіаграмотність та критичне мислення в процесі викладання дисциплін комунікаційного циклу : навч. посіб. Київ : Акад. укр. преси. Центр вільн. преси, 2024. 178 с. URL: https://www.aup.com.ua/uploads/Mediagramotnist_ta_kritichne_AUP24.pdf (дата звернення: 08.05.2025).
6. Конституція України : від 28.06.1996 р. № 254к/96-ВР : станом на 1 січ. 2020 р. URL: https://zakon.rada.gov.ua/laws/show/254к/96-вр#Text (дата звернення: 15.04.2025).
7. Кримінальний кодекс України. *Офіційний вебпортал парламенту України*. URL: https://zakon.rada.gov.ua/laws/show/2341-14#Text (дата звернення: 15.04.2025).
8. Чулков А. Deepfake технології та вплив на журналістику: розвиток технологій глибокого фейку та їхній потенційний вплив на відеомонтаж у журналістиці. м. Київ, 5 берез. 2024 р. Київ, 2024. С. 1–3.

9. Чулков А. Технологія deepfake та маніпулювання аудиторією: роль журналіста у викритті та протидії дезінформації. м. Київ, 24 жовт. 2024 р. Київ, 2024. С. 1–4.
10. AI Verification Tools: The Authenticity of AI-Generated Content. *Bora*. URL: https://welcometobora.com/blog/ai-verification-tools-authenticity-of-ai-generated-content/ (date of access: 08.05.2025).
11. Altuncu E., Franqueira V. N. L., Li S. Deepfake: definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*. 2024. Vol. 7. URL: https://doi.org/10.3389/fdata.2024.1400024 (date of access: 24.04.2025).
12. Assembly Bill No. 602 : of 11.02.2021 no. 602 : as of 28 September 2021. URL: https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202120220AB602 (date of access: 14.04.2025).
13. Assembly Bill No. 730 : of 19.02.2019 no. 730 : as of 3 October 2019. URL: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730 (date of access: 14.04.2025).
14. Beleidigung : Strafgesetzbuch, StGB vom 03.04.2021 Nr. 185. URL: https://www.buzer.de/s1.htm?a=185-187&ag=6165#:~:text=§%20185%20Beleidigung,-§%20185%20hat&text=Die%20Beleidigung%20wird%20mit%20Freiheitsstrafe,Jahren%20oder%20mit%20Geldstrafe%20bestraft (Zugriff am: 14.04.2025).
15. Deepfake Detection using Biological Features: A Survey / K. Patil et al. *ArXiv*. 2023. P. 2–8. URL: https://doi.org/10.48550/arXiv.2301.05819 (date of access: 15.04.2025).
16. Deepfake video manipulates that General Zaluzhny is preparing a military coup. *Meta.mk*. URL: https://meta.mk/en/deepfake-video-manipulates-that-general-zaluzhny-is-preparing-a-military-coup/ (date of access: 12.03.2025).

17. Devi B. T., Rajkumar R. A Comprehensive Survey on Deepfake Methods: Generation, Detection, and Applications. *International Journal on Recent and Innovation Trends in Computing and Communication*. 2023. P. 656–658.
18. Digital watermarking – a meta-survey and techniques for fake news detection / A. Malanowska et al. *IEEE Access*. 2024. P. 35. URL: https://doi.org/10.1109/access.2024.3374201 (date of access: 08.05.2024).
19. Doctored Nancy Pelosi video highlights threat of "deepfake" tech. *CBS News - Breaking news, 24/7 live streaming news & top stories*. URL: https://www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25/ (date of access: 10.04.2024).
20. Eye on Tech. What is Deepfake AI? Deepfake Content and Elections, 2020. *YouTube*. URL: https://www.youtube.com/watch?v=ENdRAI-JrDc (date of access: 08.05.2024).
21. Fraud : Criminal Code no. 380(1) : as of 14 January 2024. URL: https://laws-lois.justice.gc.ca/PDF/C-46.pdf (date of access: 14.04.2024).
22. I. Weiner D., Norden L. Regulating AI Deepfakes and Synthetic Media in the Political Arena. *Brennan Center for Justice*. URL: https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena (date of access: 11.04.2024).
23. General Data Protection Regulation : of 25.05.2018 no. 6. URL: https://gdpr-text.com/read/article-6/#recital_gdpr-a-06_3e (date of access: 14.04.2024).
24. General Data Protection Regulation : of 25.05.2018 no. 7. URL: https://gdpr-text.com/read/article-6/#recital_gdpr-a-06_3e (date of access: 14.04.2024).
25. How Generative AI Is Changing Creative Work. *Harvard Business Review*. URL: https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work (date of access: 08.05.2024).

26. Jacobson N. Deepfakes and Their Impact on Society. *CPI OpenFox*. URL: https://www.openfox.com/deepfakes-and-their-impact-on-society/ (date of access: 01.04.2025).
27. Jayakumar S., Ang B., Anwar N. D. Disinformation and Fake News. Palgrave Macmillan, 2020. 114 p.
28. Journalism, fake news & disinformation: handbook for journalism education and training / ed. by C. Ireton, J. Posetti. France : UNESCO, 2018. 128 p. URL: https://unesdoc.unesco.org/ark:/48223/pf0000265552 (date of access: 08.05.2024).
29. Leading Companies Launch Consortium to Address AI's Impact on the Technology Workforce. *IBM Newsroom*. URL: https://newsroom.ibm.com/2024-04-04-Leading-Companies-Launch-Consortium-to-Address-AIs-Impact-on-the-Technology-Workforce (date of access: 08.05.2024).
30. Murphy G., Flynn E. Deepfake false memories. *Memory*. 2021. P. 1–13. URL: https://doi.org/10.1080/09658211.2021.1919715 (date of access: 24.04.2025).
31. Olaoye F., Potter K., Doris L. Generative Adversarial Networks (GANs) and their Applications. 2024. URL: https://www.researchgate.net/publication/379035298_Generative_Adversarial_Networks_GANs_and_their_Applications (date of access: 12.04.2024).
32. Ott H. AI expert says Princess Kate photo scandal shows our "sense of shared reality" being eroded. *CBS News | Breaking news, top stories & today's latest headlines*. URL: https://www.cbsnews.com/news/princess-kate-middleton-photo-scandal-ai-sense-of-shared-reality-being-eroded/ (date of access: 24.04.2025).
33. Pain A. Understanding Convolutional Neural Networks (CNNs) in Deep Learning. *LinkedIn*. URL: https://www.linkedin.com/pulse/understanding-

convolutional-neural-networks-cnns-deep-aritra-pain/ (date of access: 17.04.2024).

34. Perfection IAS. Deepfake Technology. *IAS Intent. UPSC Monthly Current Affairs Magazine*. 2022. No. 7. P. 34–35. URL: https://www.perfectionias.com/imagebag/1675175199A5322.pdf (date of access: 12.04.2024).
35. Prothero A. Many Adults Did Not Learn Media Literacy Skills in High School. What Schools Can Do Now. *Education Week*. URL: https://www.edweek.org/teaching-learning/many-adults-did-not-learn-media-literacy-skills-in-high-school-what-schools-can-do-now/2022/09?utm_source=chatgpt.com (date of access: 13.02.2025).
36. Reardon S. The Impact of Deepfakes on Journalism | Pindrop. *Pindrop*. URL: https://www.pindrop.com/article/impact-deepfakes-journalism/ (date of access: 06.04.2025).
37. Stephen Hawking's famous voice belonged to this pioneering MIT scientist. *Big Think*. URL: https://bigthink.com/high-culture/this-mit-scientist-gave-his-voice-to-stephen-hawking/ (date of access: 05.04.2025).
38. Stupp C. Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal*. URL: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 (date of access: 02.04.2025).
39. Tassel J. V., Mary M., Schmitz J. New News: The Journalist's Guide to Producing Digital Content for Online and Mobile News. Taylor & Francis Group, 2020. 360 p.
40. The Rising Threat of Deepfakes: Detection, Challenges, and Market Growth / V. Jaikishan et al. *Liminal*. URL: https://liminal.co/articles/rising-threat-of-deepfakes/ (date of access: 09.04.2025).

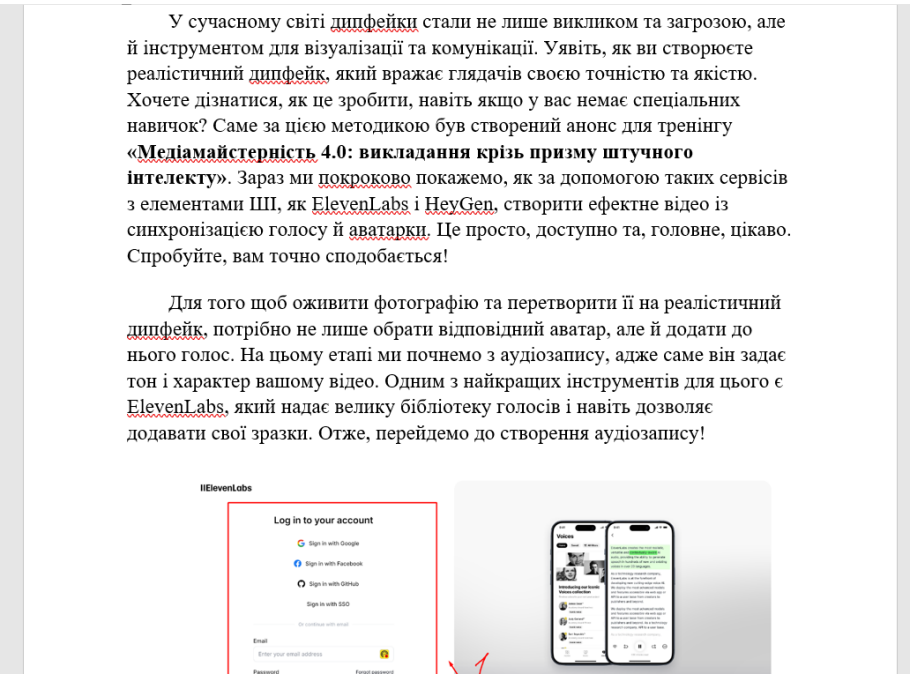
41. Top 3 Deepfake Detection Tools of 2023 | Resemble AI. *Resemble AI*. URL: https://www.resemble.ai/top-deepfake-detection-tools/ (date of access: 08.05.2024).
42. Tripathi G., Ahad M. A., Casalino G. A comprehensive review of blockchain technology: Underlying principles and historical background with future challenges. *Decision Analytics Journal*. 2023. P. 21. URL: https://doi.org/10.1016/j.dajour.2023.100344 (date of access: 08.05.2024).
43. Two New California Laws Tackle Deepfake Videos in Politics and Porn | Davis Wright Tremaine. *Davis Wright Tremaine*. URL: https://www.dwt.com/blogs/media-law-monitor/2020/02/two-new-california-laws-tackle-deepfake-videos-in (date of access: 14.04.2024).
44. Venkatesan R., Li B. Convolutional Neural Networks in Visual Computing: A Concise Guide : підручник. CRC Press, 2017.
45. Vincent J. Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news. *The Verge*. URL: https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed (date of access: 12.04.2024).
46. Unmasking the Illusion: An AI and ML Driven Approach to Face Swap Deepfake Detection / P. P et al. *International Journal for Research in Applied Science and Engineering Technology*. 2025. Vol. 13, no. 4. P. 2081–2087. URL: https://doi.org/10.22214/ijraset.2025.68647 (date of access: 24.04.2025).

Додатки

1. Посилання на вебплатформу Canva для відеомонтажу, додавання титрів, логотипів, інтро, завершення та візуальної стилізації — https://www.canva.com
2. Посилання на онлайн-платформу HeyGen для створення talking head-анімації: синхронізація зображення й озвучення, генерація відео — https://www.heygen.com
3. Посилання на онлайн-середовище Figma для розробки макетів інтерфейсу сайту медіапродукту «DeFake It!», створення графіки та дизайну користувацьких елементів — https://www.figma.com
4. Посилання на платформу ChatGPT для створення сценарію, стилізації мовлення історичних діячів, генерації текстових запитів (промптів) — https://chat.openai.com
5. Посилання на платформу Sora (OpenAI) для генерації фотореалістичних зображень історичних постатей та персонажів дипфейку — https://openai.com/sora
6. Посилання на платформу для генерації відео за допомогою ШІ з можливістю автоматичного редагування сцен, стилів і ефектів — https://runwayml.com/
7. Посилання на інтерактивний освітній сайт «DeFake It!», створений у межах кваліфікаційної роботи для навчання медіаграмотності у сфері deepfake-технологій — https://defakeit.framer.website/
8. Посилання на сервіс DeepL для автоматичного перекладу сценарію з англійської на українську мову з дотриманням граматики та стилістики — https://www.deepl.com/translator
9. Посилання на сервіс ElevenLabs для створення голосів історичних діячів із заданими параметрами, акцентом і тембром — https://elevenlabs.io

10. Посилання на сервіс Suno AI для генерації авторського музичного супроводу на основі текстового промпту — https://app.suno.ai

11. Приклад матеріалу на освітній інтерактивний сайт DeFake It! під назвою «AI говорить за нас: як створити реалістичне дипфейк-відео з голосом і фото»:



У сучасному світі дипфейки стали не лише викликом та загрозою, але й інструментом для візуалізації та комунікації. Уявіть, як ви створюєте реалістичний дипфейк, який вражає глядачів своєю точністю та якістю. Хочете дізнатися, як це зробити, навіть якщо у вас немає спеціальних навичок? Саме за цією методикою був створений анонс для тренінгу **«Медіамайстерність 4.0: викладання крізь призму штучного інтелекту»**. Зараз ми покроково покажемо, як за допомогою таких сервісів з елементами ШІ, як ElevenLabs і HeyGen, створити ефектне відео із синхронізацією голосу й аватарки. Це просто, доступно та, головне, цікаво. Спробуйте, вам точно сподобається!

Для того щоб оживити фотографію та перетворити її на реалістичний дипфейк, потрібно не лише обрати відповідний аватар, але й додати до нього голос. На цьому етапі ми почнемо з аудіозапису, адже саме він задає тон і характер вашому відео. Одним з найкращих інструментів для цього є ElevenLabs, який надає велику бібліотеку голосів і навіть дозволяє додавати свої зразки. Отже, перейдемо до створення аудіозапису!

Рис. 1

12. Приклад матеріалу на освітній інтерактивний сайт DeFake It! під назвою «Навчати через маніпуляцію? Як дипфейк може стати інструментом медіаграмотності»:

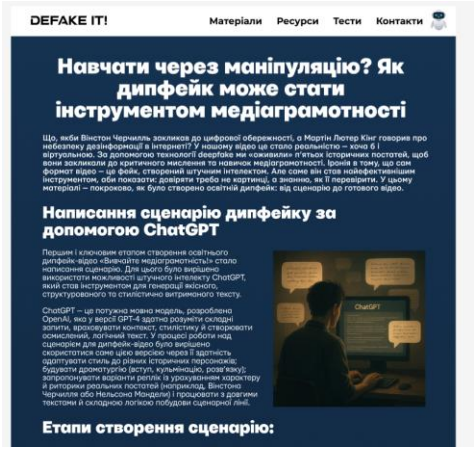


Рис. 2

13. Прототип освітнього інтерактивного сайту DeFake It! у Figma:

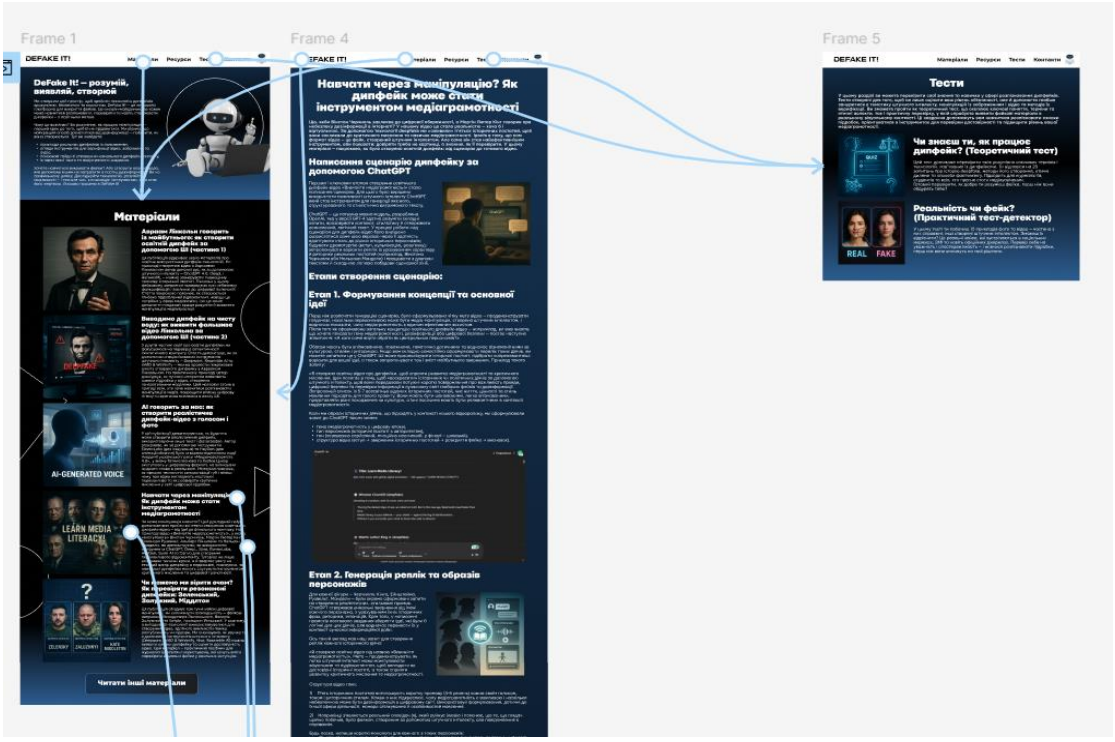


Рис. 3