ISSN 2522-1272
online ISSN 2786-8532

Scientific Notes of the Institute of Journalism, Vol. 86 (1) 2025
Journal homepage: https://nz.knu.ua/

IN THE SYSTEM OF SOCIAL COMMUNICATIONS

https://doi.org/10.17721/2522-1272.2025.86.2
UDC 007:004.8:316.776.23

Deepfake as a Technology in the Fieldof Social and Mass Communications: Opportunities and Threats

Tetiana Ivanova, Oksana Yefremova, Artur Chulkov
Mariupol State University, Ukraine

This article analyzes the transformative role of artificial intelligence in journalism, with a particular focus on the challenges and opportunities it presents. Special attention is devoted to the threats that can contaminate the information environment, namely deepfake technologies created using AI resources. Various AI-based methods and technologies are examined, including the use of advanced software for content creation (RunwayML, HeyGen, Deepbrain AI Studios) and for the detection of deepfakes (Deepware, Resemble AI, and InVID & WeVerify), which are becoming increasingly important in contemporary media workflows. The theoretical analysis is complemented by practical examples of the application of deepfake technology in advertising and education as integral components of social communications. A comparative analysis of current deepfake generation and verification systems enables an assessment of their risks, as well as the development of strategies to identify, differentiate, and counteract them in modern journalistic practice. The findings indicate that artificial intelligence and neural networks are becoming prevalent not only for content creation but also for safeguarding its authenticity. However, further research and refinement are required to ensure that these systems meet the standards of journalistic integrity and public trust in media content.

Keywords: artificial intelligence (AI); social communications; media literacy; critical thinking; journalistic ethics; deepfake detection systems; disinformation; propaganda

Citation: Іванова, Т., Єфремова, О., Чулков, А. (2025). Deepfake як технологія сфери соціальних та масових комунікацій: можливості та загрози. *Наукові записки інституту журналістики*, 86, 21–34. https://doi.org/10.17721/2522-1272.2025.86.2

ISSN 2522-1272
online ISSN 2786-8532

Наукові записки Інституту журналістики, Том 86 (1) 2025

вебсайт журналу: https://nz.knu.ua/

У СИСТЕМІ СОЦІАЛЬНИХ КОМУНІКАЦІЙ

УДК 007:004.8:316.776.23

Deepfake як технологія сфери соціальних та масових комунікацій: можливості та загрози

Тетяна Іванова, Оксана Єфремова, Артур Чулков
Маріупольський державний університет, Україна

У статті проаналізовано трансформаційну роль штучного інтелекту (ШІ) в журналістиці, зокрема про виклики і можливості. Особливу увагу приділено загрозам, які можуть забруднювати інформаційний простір, а саме: технологіям deepfake (глибоких фейків), які створюються за допомогою ресурсів ШІ. Мета статті – визначити вплив технологій штучного інтелекту (ШІ) на журналістику, зосередившись на загрозах і наслідках розповсюдження у мережі дипфейків. Досліджуються різні методи та технології на основі ШІ: використання передового програмного забезпечення для створення (RunwayML, HeyGen, Deepbrain AI Studios) та виявлення глибоких фейків (Deepware, Resemble AI та InVID & WeVerify), які стають дедалі важливішими в сучасних робочих процесах у медіа. Теоретичний аналіз доповнено практичними прикладами застосування технології «глибоких фейків» у рекламі та освіті. Порівняльний аналіз сучасних систем створення та верифікації «глибоких фейків» дає змогу оцінити їх небезпеку, а також вміти їх ідентифікувати та розрізняти, а також протидіяти їм у роботі сучасних журналістів. Результати дослідження свідчать, що штучний інтелект і нейронні мережі використовуються не лише для створення контенту, а й для захисту його автентичності. Однак для того, щоб ці системи відповідали вимогам журналістської доброчесності та довіри до медіаконтенту, необхідні подальші дослідження та вдосконалення.

Ключові слова: штучний інтелект (ШІ); соціальні комунікації; медіаграмотність; критичне мислення, журналістська етика, системи виявлення «глибоких фейків»; дезінформація, пропаганда

Стрімкий розвиток штучного інтелекту відкриває нові можливості для соціальних комунікацій, але водночас несе значні ризики. Технологія deepfake вже використовується для створення реалістичних фальсифікацій, що можуть впливати на вибори, репутацію осіб і навіть безпеку держав. В епоху інформаційних війн та зростаючої кількості цифрових маніпуляцій необхідно досліджувати та розробляти механізми протидії загрозам ШІ, аби зберегти достовірність інформації та захистити суспільство від дезінформаційних атак.

Тетяна Іванова https://orcid.org/0000-0003-1432-4893
Оксана Єфремова https://orcid.org/0000-0002-0280-8084
Артур Чулков https://orcid.org/0009-0005-5591-2845

Цю статтю було опубліковано спочатку онлайн 24 червня 2025 року. Вона є звітом про спільне дослідження студента Артура Чулкова, заступника декана Оксани Єфремової під керівництвом доктора педагогічних наук, професора Тетяни Іванової.

Автори заявляють про відсутність конфлікту інтересів. У розробленні дослідження, зборі, аналізі чи інтерпретації даних, у написанні рукопису спонсори участі не брали.

Електронна адреса автора для листування: Оксана Єфремова **efremovaoksana7@gmail.com**

ISSN 2522-1272
online ISSN 2786-8532

Останніми роками штучний інтелект став трансформаційною силою в медіаіндустрії, адже саме він змінив способи збирання, обробки та представлення інформації. ШІ, розроблений для імітації людських когнітивних процесів, відкрив широкі можливості для працівників галузі соціальних комунікацій і медіа організацій: автоматизоване виробництво новин, інструменти для перевірки інформації й складна візуалізація даних.

У статті досліджується негативний вплив технологій штучного інтелекту, зокрема deepfake, на інформаційний простір та соціальні комунікації. Аналізується, як генеративні алгоритми можуть бути використані для створення фейкового контенту, маніпулювання громадською думкою та підриву довіри до медіа. Розглядаються ключові загрози, пов'язані з поширенням дезінформації, політичними маніпуляціями та кіберзлочинністю, а також можливі шляхи протидії цим викликам. Показано практичне використання технології deepfake на етапах створення фейкового відео і його покрокової перевірки та верифікації.

Мета статті – визначити вплив технологій штучного інтелекту (ШІ) на журналістику, зосередившись на загрозах і наслідках розповсюдження у мережі дипфейків (deepfake), які забруднюють інформаційний простір для створення та поширення неправдивої інформації та введення в оману споживачів інформації.

Для досягнення мети необхідно вирішити такі завдання:

- описати сучасні технології використання ШІ у журналістиці, які алгоритми та методи використовуються для генерації контенту;
- з'ясувати роль штучного інтелекту (ШІ) та наслідки його застосування в журналістиці. Особливу увагу звернути на технології deepfake (глибоких фейків), які маскуючись під журналістський контент, вводять в оману споживачів інформації та шкодять довірі до професійної чесної журналістиці.
- визначити етичні межі застосування ШІ в медіа;
- запропонувати методи створення та виявлення (верифікації) deepfake; визначити, які є алгоритми та інструменти для боротьби з фейками (детектори deepfake, блокчейн, цифрові водяні знаки) і як відрізнити відповідальне використання ШІ від маніпуляцій.

Звісно, розвиток штучного інтелекту (ШІ) стимулює розвиток журналістики, відкриваючи нові можливості для збирання, аналізу та більш візуалізованого й мобільного представлення контенту. Крім того, застосування ШІ дає змогу створювати персоналізований контент, адаптований до інтересів різних аудиторій, що підвищує ефективність комунікації. Автоматизація завдань за допомогою штучного інтелекту може також впорядкувати робочі процеси, скоротити час, витрачений на створення контенту, і дати змогу швидше реагувати на тренди або новинні події. ШІ може аналізувати дані користувачів, щоб адаптувати контент до індивідуальних уподобань, підвищуючи рівень залученості та задоволеності аудиторії (Raturi, Mishra, & Maji, 2022).

Позитивна роль штучного інтелекту полягає також і в тому, що в епоху перенасичення інформацією, а через це й зростанням впливу дезінформації, маніпуляцій та фейків, саме використання ШІ стає не лише інструментом оптимізації робочих процесів, а й засобом забезпечення точності, перевірки інформації та достовірності новин. Отже, ця тема є надзвичайно актуальною, оскільки визначає майбутнє медіасфери та впливає на формування суспільної думки в цифрову епоху.

Теоретичне підґрунтя

Розвиток і застосування штучного інтелекту, зокрема технологій «глибоких фейків», перебуває у фокусі як вітчизняних, так і зарубіжних учених, таких як: А. Вальорська (2020), Т. Іванова (2024), Д. Чиж (2024), А. Кусий (2024), З. Ахтар (Akhtar, 2023), А. Шампандар, Хао Лі та ін.

ISSN 2522-1272
online ISSN 2786-8532

Т. Іванова у своєму посібнику «Медіаграмотність та критичне мислення в процесі викладання дисциплін комунікаційного циклу» наголошує, що з розвитком технологій потреба нових навичок цифрової грамотності й критичного мислення стає актуальною, бо саме технології використання ШІ як стимулюють великий прогрес, так і уможливлюють поширення в інформаційному просторі дипфейків і недостовірної інформації (Іванова, 2024).

А. Кусий зазначає, що захист від дипфейків ґрунтується на двох основних елементах: критичному мисленні та технологіях. Дипфейки апелюють до людських інстинктів та емоцій, що робить їх дуже переконливими. Законодавство багатьох розвинених країн удосконалюється, щоб протистояти цим загрозам, а такі корпорації, як Microsoft і Google, активно розробляють технологічні рішення для протидії дипфейкам (2021).

Завдяки прогресу в машинному навчанні, а також нейромережам штучний інтелект здатен аналізувати величезні обсяги даних, приймати складні рішення та створювати творчі продукти в небувалих раніше масштабах. А. Вальорська у своїй книзі «Діпфейк та дезінформація» зауважує: «Ми не повинні робити одне: піддатися спокусі заборонити діпфейки в принципі. Як і будь-яка технологія, вона теж, опріч пов'язаних з нею небезпек, відкриває безліч цікавих можливостей, зокрема для освіти, кінематографу та сатири» (2020).

Саме на цьому наголошує й закон ЄС про штучний інтелект, який зобов'язує системи, що виробляють або маніпулюють візуальним, аудіо- чи відеоконтентом, дотримуватися стандартів прозорості, вимагаючи від провайдерів розкривати, коли користувачі взаємодіють з контентом, створеним штучним інтелектом, якщо це не є очевидним (The EU's Artificial Intelligence Act, 2024). Закон про цифрові послуги (DSA) також передбачає, що платформи, які модерують користувацький контент, зокрема глибокі фейки, повинні бути прозорими щодо своїх правил модерації та механізмів їх дотримання. Для цього мають бути передбачені процедури повідомлення та усунення порушень (The EU's Digital Services Act, 2022).

Говорячи про те, що за допомогою штучного інтелекту в медіапросторі з'являється чимало фейків, один з дослідників штучного інтелекту А. Шампандар наголошує на необхідності підвищення обізнаності громадськості про те, як швидко технологія «глибоких фейків» може спотворювати інформацію, розглядаючи цю проблему не як суто технічну, а як таку, що сприяє зміцненню довіри до медіа (Bezmalinovic, 2018). Хао Лі, доцент кафедри комп'ютерних наук Університету Південної Каліфорнії, також застерігає, що «глибокі фейки», створені зі шкідливими цілями (наприклад, поширення неправдивих новин), можуть стати ще більш шкідливими без підвищення рівня обізнаності та освіти щодо цієї технології (WBUR, 2019).

Технологія deepfake стала поворотним моментом в історії ШІ в медіаіндустрії. Це явище зародилося в ранніх дослідженнях штучного інтелекту та машинного навчання наприкінці 20 століття, а широкого визнання набуло лише у 2017 р., коли доступні інструменти ШІ почали дозволяти користувачам безперешкодно маніпулювати обличчями, голосами та жестами людей. Часто це відбувається саме з публічними особами, щоб ввести аудиторію в оману, вплинути на громадську думку або згенерувати вірусний контент завдяки високому рівню залученості.

Отже, спочатку ця технологія була розроблена як засіб для розвитку цифрової творчості та імітації реалістичного відео, аудіо та зображень. Однак зараз вона стрімко розвинулась, і це уможливило створення дуже переконливих, але повністю сфабрикованих відео/аудіо повідомлень. Це несе величезні ризики для споживачів інформації, але водночас створює унікальні можливості для журналістської діяльності.

Технологію «глибоких фейків» можна застосовувати в різних сферах соціальних ко-

ISSN 2522-1272
online ISSN 2786-8532

мунікацій. Наприклад, дипфейки можна використовувати в рекламних кампаніях, де будь-який бренд, наприклад бренд одягу, може бути представлений у вигляді реалістичних віртуальних примірок. Це дає змогу клієнтам побачити, як конкретний одяг на них виглядатиме і чи пасуватиме їм.

В освіті «глибокі фейки» можна використовувати для створення захопливого навчального процесу, наприклад, для візуалізації історичних подій за допомогою реалістичних зображень ключових суспільно-політичних діячів тощо.

Проте водночас працівники галузі соціальних комунікацій мають бути обережними у використанні та споживанні дипфейків, бо саме вони досить ефективно створюють ілюзію реальності та вводять споживачів інформації в оману. Треба ретельно перевіряти сумнівні відео- та аудіозаписи із застосуванням як людської верифікації, так і автоматизованих інструментів для виявлення «глибоких фейків». Необхідно встановити чіткі принципи й правила етичного використання технології «глибоких фейків», щоб зменшити кількість потенційних випадків зловживання.

Тому автори публікацій про штучний інтелект наголошують на нагальній потребі в підвищенні цифрової грамотності, нормативно-правової базі та технологічному захисті для зменшення ризиків, пов'язаних з технологією дипфейків. Визнаючи значний прогрес, який приносить штучний інтелект, вчені висловлюють занепокоєння щодо його потенціалу для зловживань, підкреслюючи, що обізнаність, критичне мислення та довіра до журналістики важливі для боротьби з поширенням дезінформації.

Метод

Для досягнення мети дослідження використано комплекс методологічних підходів: а) теоретичний аналіз для вивчення наукових праць та європейських нормативно-правових документів, щоб визначити основні тенденції та регуляторні підходив до технологій deepfake; б) порівняльний аналіз систем створення (RunwayML, HeyGen, Deepbrain AI Studios) та виявлення deepfake (Deepware, Resemble AI, InVID & WeVerify) дає змогу оцінити їх точність, швидкість, доступність та інтеграцію в робочі процеси журналістів; в) метод синтезу та узагальнення дав змогу створити цілісний огляд переваг і ризиків технологій deepfake в журналістиці, рекламі та освіті; г) метод case-study охопив приклади використання deepfake у рекламі (State Farm, Суперкубок 2021) та освіті (навчальне відео «Ксена-українка»); д) емпіричний аналіз відбувся завдяки практичному тестуванню систем створення й виявлення deepfake, спрямованому на оцінювання їх ефективності в реальних кейсах у медіаіндустрії; е) експеримент був проведений у межах дослідження: за допомогою RunwayML створено авторський deepfake із зображенням Авраама Лінкольна. Цей deepfake перевірено через сучасні системи виявлення дипфейків (Deepware, Resemble AI, InVID & WeVerify), щоб оцінити їхню ефективність в ідентифікації штучно створеного контенту. Аналіз результатів експерименту дав змогу не лише підтвердити практичну функціональність цих інструментів, а й визначити їхні сильні та слабкі сторони.

Результати та обговорення

Спробуємо з'ясувати, як журналістика використовує технології ШІ для вирішення своїх завдань і які є мовні моделі, що допомагають у цьому. Перш за все треба зрозуміти, що ШІ не можуть взяти на себе повністю написання матеріалів, тому що основна функція журналіста – надавати споживачеві саме авторський, ексклюзивний контент та професійний аналіз подій, але є сфери, де технології штучного інтелекту можуть допомогти. Це дає

ISSN 2522-1272
online ISSN 2786-8532

змогу журналістам, а також іншим працівникам галузі масових комунікацій, автоматично озвучувати статті та новини, створювати гібридні медіапродукти (подкасти, аудіокниги).

Для допомоги у створенні контенту (пошук ідей, стилю подання, систематизації тощо) ШІ розв'язує такі завдання:

- автоматичне написання новин, прес-релізів, аналітичних записок тощо (наприклад, про спортивні події, фінанси, погоду);
- генерація аналітичних звітів та резюме новин;
- переклад новинних матеріалів.

ШІ також допомагають журналістам у генерації аудіо та відеозображень. До одних з таких технологій можна віднести і deepfake. Ці технології можуть застосовуватись у сфері масових комунікацій для:

- реконструкції історичних подій у відеоформаті;
- створення персоналізованого контенту (наприклад, автоматичне озвучення статей);
- використання вокодерів та мовних синтезаторів для того, щоб імітувати голоси реальних людей або створювати повністю синтетичний голос.

Зазначимо, що технологія deepfake, на жаль, також може використовуватися у створенні фейкових новин та маніпуляціях. І в цьому випадку, звісно, мова вже йде не про журналістику, а про пропаганду. На щастя, ШІ також надає можливості для аналізу даних та фактчекінгу. Штучний інтелект не тільки створює контент, а й допомагає журналісту боротися з фейками та їх верифікувати. Про це якраз і буде йти мова у нашому дослідженні. Прикладами таких технологій можуть бути:

- AI Fact-Checking (від Google, Meta, Full Fact) – аналіз фактів у реальному часі;
- Adversarial Neural Networks for Fake News Detection – ШІ-методи виявлення підробленого контенту;
- Forensic AI (Amnesty International, DARPA) – аналіз відео на наявність маніпуляцій. Їх використання уможливлює автоматичну перевірку джерел новин, виявлення фейкових фото та відео, а також аналіз контенту на предмет упередженості.

Отже, ми бачимо, що штучний інтелект вже кардинально змінює журналістику. Завдяки генеративним моделям ШІ автоматизує процеси, прискорює створення контенту та відкриває нові можливості для медіа. Проте разом із цим виникають ризики маніпуляцій та дезінформації, особливо через технології deepfake. Саме тому важливо поєднувати розвиток генеративного ШІ з надійними методами фактчекінгу, щоб зберегти довіру до журналістики.

Технології deepfake, які базуються на штучному інтелекті та алгоритмах глибокого навчання, стали потужним інструментом у сфері медіа. Вони відкривають нові можливості, але водночас несуть серйозні загрози для журналістики та суспільства загалом. Спробуємо окреслити можливості використання deepfake для журналістики. Технологія deepfake допомагає реконструювати історичні події. Цю можливість можна використати для відтворювання історичних особистостей, розуміння їх творчості, життєвого шляху та мотивів вчинків. Журналісти можуть використовувати цю технологію для створення документальних фільмів, а викладачі журналістики – для підготовки навчальних матеріалів для студентів. Звичайно, застосування цієї технології робить історію більш наочною та інтерактивною, а також допомагає відтворювати голоси та зображення загиблих (померлих) журналістів чи очевидців подій.

Серед навчальних проектів можна навести приклад «Dalí Lives», який за допомогою deepfake відтворили постать Сальвадора Далі у відеоформаті. Ще одним прикладом може бути проект «Такими ви їх ще не бачили. Історичні особистості оживають...» У цьому відео за допомогою передових технологій штучного інтелекту відтворено образи видатних особистостей минулого. Звичайно, це занурює студентів у процес навчання, робить його

ISSN 2522-1272
online ISSN 2786-8532

більш цікавим та наочним. Також є приклади, де за допомогою штучного інтелекту відтворюються літературні твори українських письменників та поетів, творчість яких вивчається майбутніми журналістами.

Ще одним прикладом є проект бельгійського художника Ксандера Стінбрюгге, який за допомогою нейромережі StableDiffusion створив трихвилинний ролик про історію Землі та її можливе майбутнє[3]. А у 2023 р. у Житомирі відбулася виставка українського письменника Мартина Якуба, який представив світлини діячів «Розстріляного відродження», створені штучним інтелектом[4].

Deepfake може адаптувати контент під аудиторію, створюючи динамічні відеоновини на основі текстових матеріалів. Наприклад, ШІ-модель BloombergGPT, створена для аналізу фінансових новин, потенційно може автоматично генерувати відеоновини. Це дає змогу прискорювати створення відеоматеріалів й адаптувати інформацію під конкретного глядача.

Технології deepfake можуть покращити якість журналістських розслідувань, а також забезпечити безпеку інсайдерів. Їх можна використовувати для відновлення спотворених відео або захисту анонімності свідків, що допомагає приховувати особисті дані джерел інформації та використовується в боротьбі з цензурою в авторитарних країнах.

Попри позитивні моменти, перелічені вище, технологія deepfake може створювати справжні загрози для журналістики, шкодити іміджу чесної та етичної журналістської діяльності, паплюжити головну місію журналіста – надавати споживачеві достовірну інформацію. Наприклад, нещодавно папа Римський Франциск I назвав себе «жертвою цифрових технологій». Він обґрунтовував це тим, що у соціальних мережах «завірусилося» нібито його фото у стильному білому пуховику. Насправді, це був фейк, але багато людей йому повірили, а деякі поважні видання опублікували цей знімок як достовірний. Саме з цієї причини папа Римський закликав журналістську спільноту до всесвітнього регулювання сфери штучного інтелекту[5].

Тож основна загроза використання deepfake для журналістики полягає в тому, що ця технологія розповсюджує дезінформацію та маніпуляції. Deepfake дозволяє створювати фальшиві відео, які виглядають дуже реалістично, що загрожує довірі до медіа та може використовуватися для пропаганди. Як приклади можна навести:

- фейкове відео Зеленського у 2022 р., де він, начебто, закликав скласти зброю[6];
- дипфейк Залужного, який закликає українців до держперевороту[7];
- діпфейк з Кейт Уілтон, де вона, начебто, у своїй «передсмертній промові» пропонує молоді скористатися «можливостями фінансової піраміди та почати заробляти легкі гроші».

Усі ці приклади підривають довіру до журналістики, що може призвести до того, що люди сумніватимуться навіть у правдивих новинах. Також не можна не зазначити, що існують певні труднощі в боротьбі з deepfake, адже попри розробку алгоритмів розпізнавання, технології вдосконалюються дуже швидко.

Отже, deepfake – це потужний інструмент, який можна використовувати як на користь журналістики, так і для її руйнування. Тож необхідно, щоб журналісти завжди вміли використовувати ШІ для перевірки контенту за допомогою детекторів deepfake. Споживачі

[3] URL: https://www.youtube.com/watch?v=6-97Yxkfd2A
[4] URL: https://mezha.net/ua/bukvy/shtuchnyi-intelekt-zheneruvav-portrety-diiachiv-rozstrilianoho-vidrodzhennia-foto/
[5] URL: https://www.youtube.com/shorts/emgbzI-LS0g
[6] URL: https://www.048.ua/news/3351641/uvaga-fejk-v-merezi-zavilos-video-na-akomu-nibito-volodimir-zelenskij-proponue-ukraincam-sklasti-zbrou-video
[7] URL: https://espreso.tv/merezheyu-shiritsya-dipfeyk-iz-nibito-zaluzhnim-i-zaklikom-do-zbroynogo-povstannya

ISSN 2522-1272
online ISSN 2786-8532

також мають розвивати якості критичного мислення та медіа грамотності. Необхідно також посилювати регулювання deepfake на законодавчому рівні.

Одним з наших завдань було запропонувати працівникам комунікаційної сфери практичні методи виявлення deepfake. Зауважимо, що ми використовували принцип десакралізації, який реалізується через пізнання, практичну дію та усвідомлення. Саме з цією метою ми пропонуємо комунікаційникам спочатку самим навчитися, як створюються діпфейки, а потім поетапно їх верифікувати.

Ми вже зазначили, що технологія deepfake передбачає використання штучного інтелекту та алгоритмів глибокого навчання для створення надреалістичних маніпуляцій із зображеннями, відео або аудіозаписами. Алгоритми глибокої підробки, обробляючи величезні масиви даних (відеоматеріали або голосові записи), можуть генерувати цифровий контент, який виглядає достовірним, навіть якщо він повністю сфабрикований. Згенеровані штучним інтелектом обличчя справляють враження дуже переконливих, набуваючи реальних рис. Згідно з дослідженням, опублікованим MDPI, є чотири основні типи таких маніпуляцій: підміна особистості (заміна обличчя однієї людини обличчям іншої), реконструкція обличчя (перенесення міміки й рухів однієї людини на іншу), маніпуляція атрибутами (зміна атрибутів обличчя, таких як вік, стать, колір волосся або волосся на обличчі) і синтез обличчя (використання генеративного ШІ для створення абсолютно нового обличчя, у результаті чого виходить унікальна особистість людини, якої не існує) (Akhtar, 2023).

Упровадження штучного інтелекту, зокрема дипфейк-технологій, створює як переваги, так і виклики, які значно впливають на журналістську діяльність і на сприйняття медіаконтенту авдиторією. З одного боку, ШІ може генерувати новий контент, зокрема відео, аудіо та візуальні матеріали, які можуть покращити сторітелінг у журналістиці, рекламі та розважальних програмах. З іншого боку, дипфейки, як одна з технологій штучного інтелекту, може використовуватися для створення дезінформації, маніпуляції думками та поведінкою людей, особливо під час виборів або кризових ситуацій. Також дипфейки можуть порушувати приватне життя людей, особливо коли особисті зображення або аудіозаписи використовуються без їхньої згоди. «Глибокі фейки» можуть підірвати довіру до журналістики, дозволяючи особам, які перебувають у невигідній або компрометуючій ситуації, стверджувати, що відео є сфабрикованим. Це явище називають «дивідендом брехуна», і воно створює додаткові проблеми для журналістів, які прагнуть притягнути людей до відповідальності за їхню поведінку. Крім того, зростаюча складність у розрізненні автентичного контенту від маніпулятивного може призвести до «апатії до реальності», коли люди, піддаючись загрозі дезінформації, починають неохоче довіряти навіть легітимним джерелам новин (Schiff, Schiff, & Bueno, 2021).

Говорячи про трансформаційні можливості використання «глибоких фейків», наведемо цікавий приклад використання штучного інтелекту для створення реклами, що запам’ятовується. Компанія State Farm застосувала технологію deepfake у своїй рекламі до Суперкубка 2021 року, щоб повернути до життя легендарного тренера Національної футбольної ліги (NFL) Вінса Ломбарді[8].

У цьому ролику риси обличчя Ломбарді були відтворені з надзвичайною точністю, бездоганно передаються його характерні манери. Він навіть з’являється у своїх культових окулярах, капелюсі та пальто, що робить його імідж миттєво пізнаваним. Виділяється й голос Ломбарді, коли він виголошує мотиваційну промову на теми єдності, мужності та наполегливості. Цей голос був створений компанією Respeecher, яка займається реплікацією голосу та яка зробила програму Comcast NBC Universal LIFT Labs Accelerator на Techstars у 2019 р. (Comcast NBC Universal LIFT Labs, 2021). Отже, неочікувана поява

[8] URL: https://x.com/NFL/status/1355973428423979013

ISSN 2522-1272
online ISSN 2786-8532

такої впливової постаті минулого створила емоційний резонанс серед футбольних фанатів, викликавши в них почуття ностальгії та мотивації.

Використання технології «глибоких фейків» стає потужним інструментом і у сфері навчання медіаграмотності. Демонструючи під час навчання, як за допомогою технології deepfake створюються реалістичні маніпуляції, викладачі переконують майбутніх журналістів у небезпеці цього явища, подібно до «шкідливих порад» Остермана. Такий навчальний метод сприяє глибшому розумінню того, як швидко та легко створюється дезінформація. На тренінгах з медіаграмотності процес створення та верифікації дипфейків дає студентам практичний досвід, що дає змогу проілюструвати етичні дилеми, з якими можуть стикатися як журналісти, так і споживачі інформації.

Одним з прикладів такого використання для навчання медіаграмотності є дипфейк «Ксена-українка», у якому професорка Тетяна Іванова представила актуальність вивчення можливостей та загроз штучного інтелекту. Цей ролик був використаний на тренінгу «Медіаграмотність під час війни» від Академії Української преси[9].

Зауважимо, що образ «Ксени-українки» був створений за допомогою програмного забезпечення Cartoon Animator, який інтегрує штучний інтелект для анімації персонажів. У підробленому відео аватарка Тетяни Іванової тепло запрошує учасників тренінгу вирушити у святкову подорож до «Країномедії», вигаданої країни, що символізує медійний простір, і наголошує, як у ньому потрібно бути обережним. Цей метафоричний навчальний кейс демонструє, як завдяки креативним програмам штучного інтелекту можна не тільки нестандартно та інтерактивно почати тренінг (лекцію), а й зацікавити та мотивувати студентів на вивчення теми, присвяченої штучному інтелекту і його медіаграмотному використанню.

Для створення дипфейків у навчальних, творчих або професійних цілях, що відповідають стандартам журналістської етики, важливо вміти використовувати сучасні програмні забезпечення. Ці інструменти не лише розширюють можливості в галузі соціальних комунікацій, а й допомагають зрозуміти механізми роботи дипфейків, що є важливим для протидії дезінформації та підвищення медіаграмотності. Проаналізуємо деякі з них, адже використання таких програм дозволяє студентам і фахівцям на практиці зрозуміти, як створюються й викриваються дипфейки, що формує критичний підхід до аналізу контенту:

- **RunwayML.** Це багатофункціональна платформа для створення дипфейків, зокрема для роботи з відео, аудіо та зображеннями. Інструмент базується на штучному інтелекті й дає змогу користувачам генерувати високоякісні відео із заміною облич або накладенням стилів.
- **HeyGen.** Платформа для створення персоналізованих відео, яка дозволяє користувачам генерувати відеоконтент на основі текстових команд з можливістю додавання голосу та анімації. HeyGen використовується в маркетингу, рекламі та освітніх проєктах завдяки своїм функціональним можливостям.
- **DeepBrain AI Studios.** Це платформа для створення реалістичних аватарів, які можуть вимовляти заданий текст на різних мовах. Інструмент дає змогу створювати навчальні відео, презентації або рекламний контент, подібний до реального. Особливістю є можливість інтеграції штучного інтелекту для природного мовлення та міміки.

Представляємо результати порівняльного аналізу, де детально розглянуто особливості систем для створення дипфейків RunwayML, HeyGen та DeepBrain AI Studios, зокрема їхню функціональність, простоту використання, швидкість обробки та можливості адаптації під різні потреби (див. Табл. 1):

[9] URL: https://www.youtube.com/watch?v=Ixtt64zKdbM

ISSN 2522-1272
online ISSN 2786-8532

Таблиця 1.
Оцінка ефективності систем створення дипфейків за критеріями

Критерій	RunwayML	HeyGen	DeepBrain AI Studios
Точність створення дипфейків	Висока для відео та зображень	Висока для відео та анімацій	Дуже висока, особливо для аватарів
Підтримка форматів медіа	Відео (mp4), зображення (png, jpg), аудіо (mp3)	Відео, аудіо	Відео, зображення, текст
Швидкість обробки	Швидка	Швидка	Помірна
Інтерфейс користувача	Інтуїтивно зрозумілий, для професіоналів	Простий, орієнтований на кінцевих користувачів	Інтуїтивно зрозумілий, для бізнесу та навчання
Тип контенту	Відео, зображення, аудіо	Персоналізовані відео	Реалістичні аватари, текстові відео
Алгоритми ШІ	Глибоке навчання, генерація стилів	Генеративний ШІ для відео та анімації	ШІ для синтезу мови та міміки
Можливість інтеграції	API доступне	API доступне	Відсутнє
Мультиплатформність	Вебсайт, застосунок для Windows/Mac	Вебсайт, застосунок для iOS/Android	Вебсайт, не підтримує мобільні додатки
Захист даних користувача	Висока, зберігаються тільки необхідні дані	Висока	Висока
Додаткові функції	Генерація стилів, автоматизація роботи з контентом	Генерація голосу, створення анімацій	Підтримка створення реалістичних персонажів
Мова інтерфейсу	Англійська, доступність інших мов	Англійська, доступність інших мов	Англійська, доступність інших мов
Ціна/умови ліцензії	Платна підписка з безкоштовним планом	Платна підписка з безкоштовним планом	Платна підписка з безкоштовним планом
Підтримка користувачів	Обмежена, основна через онлайн-ресурси	Активна, доступна через чат та e-mail	Активна, цілодобова підтримка
Репутація та відгуки	Висока, користується популярністю в індустрії	Висока, отримала позитивні відгуки користувачів	Висока, використовується в академічних та бізнес-цілях
Актуальність алгоритмів	Висока, регулярно оновлюється	Висока, використовує новітні технології	Висока, активно інтегрує нові розробки в штучному інтелекті

З таблиці порівняльного аналізу ми можемо побачити, що всі три системи створення дипфейків – RunwayML, HeyGen та DeepBrain AI Studios – демонструють високу точність і реалістичність результатів, але відрізняються за функціональністю, швидкістю обробки та можливостями інтеграції. RunwayML – універсальний інструмент, який підходить для створення відео, аудіо та зображень. Відзначається високою швидкістю обробки та наявністю API для інтеграції в робочі процеси, проте потребує технічної підготовки для використання повного функціоналу. HeyGen – ідеальний система для створення персоналізованих відео. Вона проста у використанні, швидка й зручна, орієнтована на маркетинг, рекламу та навчання. Інтеграція через API робить її особливо привабливою для бізнесу.

ISSN 2522-1272
online ISSN 2786-8532

DeepBrain AI Studios – спеціалізується на створенні реалістичних аватарів із синтезованим мовленням. Вона чудово підходить для освітніх і корпоративних проєктів завдяки високому рівню реалістичності, проте має обмежені можливості інтеграції й вимагає більше часу на обробку.

Зазначені критерії допоможуть об'єктивно оцінити кожну із систем та вибрати оптимальний інструмент залежно від ваших цілей – чи то створення навчальних матеріалів, чи бізнес-презентацій, чи маркетингового контенту. За потреби можна додати критерії для більш точного порівняння в межах конкретного проєкту.

Уміння виявляти та розпізнавати такі «глибокі підробки» також є вкрай важливим, щоб зберегти цілісність медіаконтенту в епоху, коли штучний інтелект може виробляти переконливі маніпулятивні медіаматеріали. Майбутні журналісти мають знати, як за допомогою різних методів (криміналістичний аналіз, біометричне порівняння, моделі машинного навчання та спеціалізовані інструменти), можна навчитися розпізнавати дипфейк. Саме це допоможе мінімізувати ризики, пов'язані з дезінформацією, порушенням приватності та маніпулюванням громадською думкою. Зрештою, всі ці навички допоможуть підтримувати довіру до цифрового світу.

Зазвичай для виявлення дипфейків використовують передові алгоритми машинного навчання, які ретельно аналізують риси обличчя, жести та інші ключові елементи у відеозапису. Ці алгоритми навчаються на великих масивах даних, що містять як реальні, так і синтетичні (глибоко підроблені) відео, і це дає їм змогу розпізнавати чіткі патерни та аномалії. Наприклад, вони можуть виявляти розбіжності в рухах обличчя, які можуть відрізняти справжню взаємодію від згенерованої штучним інтелектом. Крім того, системи розпізнавання дипфейків часто досліджують невідповідності й у освітленні, тінях і відблисках, які можуть бути неправильно вирівняні в зманіпульованих відео. Такі аномалії можуть бути характерними ознаками цифрової фальсифікації (Бірюк, 2022).

Сучасні системи виявлення «глибоких підробок» є найкращим способом боротьби зі шкідливим медіаконтентом, створеним штучним інтелектом. Проте важливо зазначити, що, хоча ці системи значно розширюють можливості виявлення, вони не є безпомилковими. Щоб забезпечити точність та надійність оцінок, зроблених цими інструментами, необхідна додаткова перевірка людиною. Такий комбінований підхід дозволяє авдиторії взаємодіяти з медіа на більш компетентному та професійному рівні (Garriga, Ruiz-Incertis, & Magallón-Rosa, 2024).

Проаналізуємо сучасні системи, що дають змогу виявляти «глибокі підробки» у відеочи аудіозаписах. Ці програмні забезпечення доступні безкоштовно й вимагають від користувачів лише реєстрації через обліковий запис Google або підтвердження номера телефону:

- **Deepware.** Він визначає, чи є відео підробкою, оцінюючи кілька критеріїв, зокрема рухи обличчя, вираз обличчя та синхронізацію зі звуком, а також аналізуючи невідповідності в освітленні, тінях та елементах фону. Система використовує передові алгоритми машинного навчання для виявлення патернів, характерних для глибоко підроблених відео, та ідентифікує аномалії, які можуть свідчити про маніпуляції. Зручний інтерфейс дозволяє користувачам швидко завантажувати відео для аналізу, додавши його безпосереднього з комп`ютера або за допомогою посилання на YouTube.
- **Resemble AI.** Він містить інструменти для виявлення голосових підробок, що дає користувачам змогу перевіряти автентичність аудіоконтенту та оцінювати потенційні маніпуляції. Система порівнює характеристики згенерованого голосу з автентичними записами, шукає розбіжності в тоні, висоті та мовленнєвих патернах, щоб визначити ймовірність підробки голосу.
- **InVID & WeVerify.** Цей інструмент пропонує аналіз відео, зворотний пошук зобра-

ISSN 2522-1272
online ISSN 2786-8532

жень і перевірку метаданих. Система може ідентифікувати дипфейк, якщо користувач зробить скриншот підозрілого об'єкта у відео та завантажить його в програму. Платформа використовує метод глибокого навчання Mantranet для візуального окреслення ділянок, які, за його алгоритмами, підозрюються в маніпуляціях.

Пропонуємо вашій увазі таблицю порівняльного аналізу, де ми оцінили вищезазначені системи виявлення фейків Deepware, Resemble AI та InVID & WeVerify за наведеними критеріями (див. Табл. 2).

Таблиця 2.
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
Мультиплатформність	Вебсайт	Вебсайт та застосунок для iOS/Android	Розширення для браузера
Захист даних користувача	Обмежена політика конфіденційності	Висока	Висока
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

З таблиці порівняльного аналізу ми можемо побачити, що всі три системи виявлення підробок – Deepware, Resemble AI та InVID & WeVerify – демонструють високу точність і доступність, проте вони відрізняються за швидкістю роботи та можливостями інтеграції. Deepware – спеціалізується на аналізі відео, швидка та проста у використанні. Підходить для швидкої перевірки відеофайлів. Проте бракує інтеграції та глибоких додаткових функцій. Resemble AI – найкраще працює з аудіоконтентом, уможливлюючи розпізнавання дипфейків голосу. Має інтеграцію через API, але потребує підписки для використання повного функціоналу. InVID & WeVerify – універсальний інструмент для журналістів. Відзначається підтримкою відео, зображень, аналізом метаданих і розширенням для браузера. Найкраще підходить для комплексної перевірки контенту.

ISSN 2522-1272
online ISSN 2786-8532

Указані критерії допоможуть об'єктивно оцінити кожну з програм та вибрати оптимальний інструмент для ваших цілей. Якщо потрібно, можна додати специфічні критерії для вашого проєкту.

Висновки

Отже, технології створення та декодування дипфейків стають ключовим викликом для професійної комунікації в сучасному світі. З одного боку, дипфейки відкривають нові можливості для творчих проєктів і маркетингових кампаній, створюючи інноваційні формати контенту. З іншого боку, вони становлять серйозну загрозу для довіри до інформації, підсилюючи ризики дезінформації та маніпуляцій. Дослідження також демонструє, що використання ШІ в боротьбі з дезінформацією, яка створена за допомогою глибоких фейків, не лише сприяє підвищенню журналістської доброчесності, а й заохочує критичне мислення, адаптивність і відповідальне використання ШІ в професійній діяльності журналіста. Необхідно також зазначити, що адаптація технологій штучного інтелекту для широкого використання в журналістиці досі потребує постійних досліджень для вдосконалення методів виявлення, а також створення нормативно-правової бази, яка зможе ефективно протистояти майбутнім викликам.

Для комунікаційників, особливо для журналістів, важливо не лише розуміти принципи роботи цих технологій, а й бути готовими до оперативного виявлення фейків, розробляючи стратегії захисту брендів і аудиторії. Інвестиції у навчання медіаграмотності та критичному мисленню, розробка етичних стандартів використання ШІ та співпраця з технічними експертами стануть важливими кроками для адаптації до цієї нової реальності.

Внесок авторів: Оксана Єфремова – концептуалізація, обговорення проблеми; Тетяна Іванова – огляд літератури, написання основного тексту; Артур Чулков – огляд літератури, підготовка англомовного резюме

Декларація про генеративний штучний інтелект та технології, що використовують штучний інтелект у процесі написання.

Під час підготовки цієї статті автори не використовували інструменти штучного інтелекту. Автори статті несуть повну відповідальність за правильне використання та цитування джерел.

Посилання

Akhtar, Z. (2023). Deepfakes generation and detection: A short survey. *Journal of Imaging, 9*(1), 18. https://doi.org/10.3390/jimaging9010018

Bezmalinovic T. Wenn Merkel plötzlich Trumps Gesicht trägt: Die gefährliche Manipulation von Bildern und Videos. (n.d.). *Aargauer Zeitung*. Retrieved October 31, 2024, from https://www.aargauerzeitung.ch/leben/digital/wenn-merkel-plotzlich-trumps-gesicht-tragt-die-gefahrliche-manipulation-von-bildern-und-videos-ld.1481742 [in German].

Comcast NBC Universal LIFT Labs. (2021). Lombardi Super Bowl Ad Features Voice Technology by LIFT Labs Alum Respeecher. URL: https://lift.comcast.com/2021/03/15/lombardi-super-bowl-ad-features-voice-technology-by-lift-labs-alum-respeecher/ (дата звернення: 01.11.2024).

Garriga, M., Ruiz-Incertis, R., & Magallón-Rosa, R. (2024). Artificial intelligence, disinformation and media literacy proposals around deepfakes. *Observatorio (OBS)**. https://doi.org/10.15847/obsobs18520242445

Raturi, S., Mishra, A. K., & Maji, S. (2022). Fake news detection using machine learning. *DeepFakes* (pp. 121–133). New York. https://doi.org/10.1201/9781003231493-10

Schiff, K. J., Schiff, D., & Bueno, N. S. (2021). The liar's dividend: How deepfakes and fake news affect politician support and trust in media. 50 p. URL: https://osf.io/preprints/socarxiv/x43ph/download

The EU's Artificial Intelligence Act. (2024, March 13). URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf.

The EU's Digital Services Act. (2022, October 19). URL: https://www.eu-digital-services-act.com/.

ISSN 2522-1272
online ISSN 2786-8532

WBUR. (2019). Perfect deepfake tech could arrive sooner than expected. URL: https://www.wbur.org/hereandnow/2019/10/02/deepfake-technology.

Бірюк, Я. (2022). Не вір очам своїм. Як deepfake змінює нашу реальність [Don't believe your eyes. How deepfake changes our reality]. *Ximera*. URL: https://ximera.com.ua/2022/07/deepfake/.

Вальорська, А. (2020). *Діпфейк та дезінформація* (переклад з німецької мови В. Олійника) [*Deepfake and disinformation* (V. Oliynyk, Trans.)]. Київ: Академія української преси; Центр вільної преси.

Іванова, Т. (2024). Медіаграмотність та критичне мислення в процесі викладання дисциплін комунікаційного циклу: навч. Посібник [Media literacy and critical thinking in the process of teaching disciplines of the communication cycle: A textbook]. Київ: Академія української преси; Центр вільної преси. URL: https://www.aup.com.ua/uploads/Mediagramotnist_ta_kritichne_AUP24.pdf.

Кусий, А. (2024). Діпфейки: як розпізнати та захиститися [Deepfakes: How to recognize and protect yourself]? *Інтернет Свобода*. URL: https://netfreedom.org.ua/article/dipfejki-yak-rozpiznati-ta-zahistitisya.

Чиж, Д. (2024). Що таке діпфейки і чому український додаток Reface створює інакший контент [What are dipfakes and why the Ukrainian app Reface creates different content]. *Bazilik Media*. URL: https://bazilik.media/shcho-take-dypfejky-i-chomu-ukrainskyj-dodatok-reface-stvoriuie-inakshyj-kontent/.

Надійшла до редакції 31.12.2024
Прийнято до друку 09.03.2025
Оприлюднено 24.06.2025