**МІНІСТЕРСТВО ОСВІТИ І НАУКИ**

**УКРАЇНИ ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД**

**«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ» ІНЖЕНЕРНО-**

**ТЕХНІЧНИЙ ФАКУЛЬТЕТ**

**КАФЕДРА КОМП’ЮТЕРНИХ СИСТЕМ І МЕРЕЖ**

**КУРСОВА РОБОТА**

з дисципліни «Мультимедійні засоби» на тему:

**«Мультимедійна енциклопедія на тему: “deepfake-технології”»**

Студента 5-го курсу КІ

Усачова Олександра Андрійовича

Керівник: зав. каф. Горват П.П.

Національна шкала\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Кількість балів:\_\_\_\_Оцінка: ECTS\_\_\_\_

м. Ужгород – 2025 рік

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД

«УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ»

ІНЖЕНЕРНО -ТЕХНІЧНИЙ ФАКУЛЬТЕТ

КАФЕДРА КОМП’ЮТЕРНИХ СИСТЕМ ТА МЕРЕЖ

ТЕХНІЧНЕ ЗАВДАННЯ

на курсову роботу

з дисципліни « Мультимедійні засоби»

студенту Усачову Олександру Андрійовичу

Тема **«Мультимедійна енциклопедія на тему: “deepfake-технології”»**

1. Теоретична частина курсової роботи повинна включати реферат на задану тему та опис розробленої мультимедійної енциклопедії, обсяг реферату — 15-20 сторінок, опис розробки — 5-7 сторінок.
2. Опис розробки має включати методи та інструменти, використані при створенні МЕ, бібліотеки, формати даних, структуру та особливості розробки.
3. Мультимедійна енциклопедія повинна бути самостійним програмним продуктом, що не потребує допоміжних програм, з організацією вмісту у вигляді гіпермедіа.
4. МЕ повинна містити статичні зображення, відео-фрагменти, тривимірні моделі та аудіо-супроводження, а також обов’язково включати розділи: глосарій та список літератури.

Дата видачі „\_\_\_\_” \_\_\_\_\_\_\_ 20\_\_ р. Керівник \_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Завдання отримав \_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ВСТУП 4](#_Toc199196519)

[1. ОСНОВИ ТЕХНОЛОГІЙ DEEPFAKE 5](#_Toc199196520)

[1.1 Виникнення та історія розвитку deepfake-технологій 5](#_Toc199196521)

[1.2 Cинтетичне медіа, його класифікація та види 6](#_Toc199196522)

[1.3. Архітектури deepfake-моделей 7](#_Toc199196523)

[1.4 Найпоширеніші типи діпфейків 10](#_Toc199196524)

[1.5 Етичні та правові виклики пов’язані з використанням deepfake-технологій 13](#_Toc199196525)

[1.6. Методи виявлення синтезованого медіаконтенту 14](#_Toc199196526)

[2.1 Вибір методів та засобів розробки 19](#_Toc199196527)

[2.2 Використані бібліотеки: 19](#_Toc199196528)

[2.3 Структура проекту: 20](#_Toc199196529)

[ВИСНОВКИ 23](#_Toc199196530)

[СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ 24](#_Toc199196531)

[ДОДАТКИ 25](#_Toc199196532)

[Додаток А. Блок-схема роботи блоку керування 25](#_Toc199196533)

# ВСТУП

Метою даної курсової роботи є розробка мультимедійної енциклопедії, що надає розлогу інформацію про технології deepfake, принципи їх роботи, різновиди, методи виявлення, а також потенційний вплив на суспільство. Розроблений проєкт повинен забезпечувати структурований та інформативний огляд відповідної тематики. Формат фінальної роботи повинен мати текстові, графічні та відеоелементи, що дозволятиме йому не лише демонструвати матеріал у доступній формі, а й наочно ілюструвати механізми та результат генерації deepfake-іструментами.

У межах реалізації даного проєкту було передбачено класифікацію основних типів deepfake, таких як: заміна обличчя у фото- та відеозаписах, синхронізація артикуляції вимови з доданим аудіо, клонування голосу, генерація повної моделі тіла та створення відео з елементами deepfake у реальному часі. Також один з розділів енциклопедії було присвячено методам виявлення фальсифікованого контенту — від традиційних аналітичних алгоритмів до застосування спеціалізованих нейромереж для виявлення ознак маніпуляцій зображенням на рівні візуальних артефактів або поведінкових відхилень.

Ця курсова робота також передбачає реалізацію функціонального веб-інтерфейсу із використанням стеку Node.js та React, структурування змісту енциклопедії та реалізацію словника використаних термінів.

Виконання даної курсової роботи сприятиме не лише поглибленому вивченню основ роботи deepfake-технологій, а й формуванню практичних навичок у сфері створення цифрового контенту та розробки користувацьких інтерфейсів для інформативних ресурсів.

# 1. ОСНОВИ ТЕХНОЛОГІЙ DEEPFAKE

## 1.1 Виникнення та історія розвитку deepfake-технологій

Ідеї, що лежать в основі сучасних методів генерації зображень-підробок були сформовані ще у 2010-х роках. Однак визначальною подією стала розробка генеративних антагоністичних мереж (GAN), концепція яких була представлена у 2014 році Ієн Ґудфелоу. Незабаром після публікації наукової статті на її основі були побудовані перші моделі генеративних мереж, що почали стрімко поширюватися інтернетом, зокрема серед спільнот програмістів-ентузіастів.

Перші реалізації мали досить обмежену функціональність. Вони переважно базувалися на структурно простих автокодерах і мали змогу проводити лише базову трансформацію зображення. Не зважаючи на це вже у 2017 deepfake моделі почали використовуватися загальною публікою. Однією з славнозвісних цілей було створення відео з підміною облич відомих осіб, що поширювалися на форумах Reddit та 4chan. Також у цей період у розмовний лексикон ввійшов вираз “deepfake”, ставши маркером нового технологічного явища.



Рис. 1.1. Один з перших «діпфейків», поширених у соціальній мережі Reddit у 2017 році

Однак прогрес не стоїть на місці, і впродовж наступних років моделі для створення deepfake-медіа почали архітектурно ускладнюватися та розширюватись. Також завдяки збільшення розміру мереж та використання досконаліших методів вираховувавння (inference) було досягнуто значного покращення якості синтезу деталей, таких як текстура шкіри, лицьова міміка, зміна тону та інтонації вимови, а також збереження контексту сцени при зміні освітлення або кута зйомки. Водночас з поширенням генеративних моделей, завдяки праці ряду ентузіастів почали з’являтись нові, більш зручні графічні інтерфейси для роботи з ними, що, в свою чергу, дозволило користувачам без глибоких технічних знань синтезовувати власний контент. Також завдяки стрімкому збільшенню швидкості роботи та обсягу пам’яті комерціного апаратного забезпечення значно зменшився рівень фінансових вкладень, необхідних для генерації медіа.

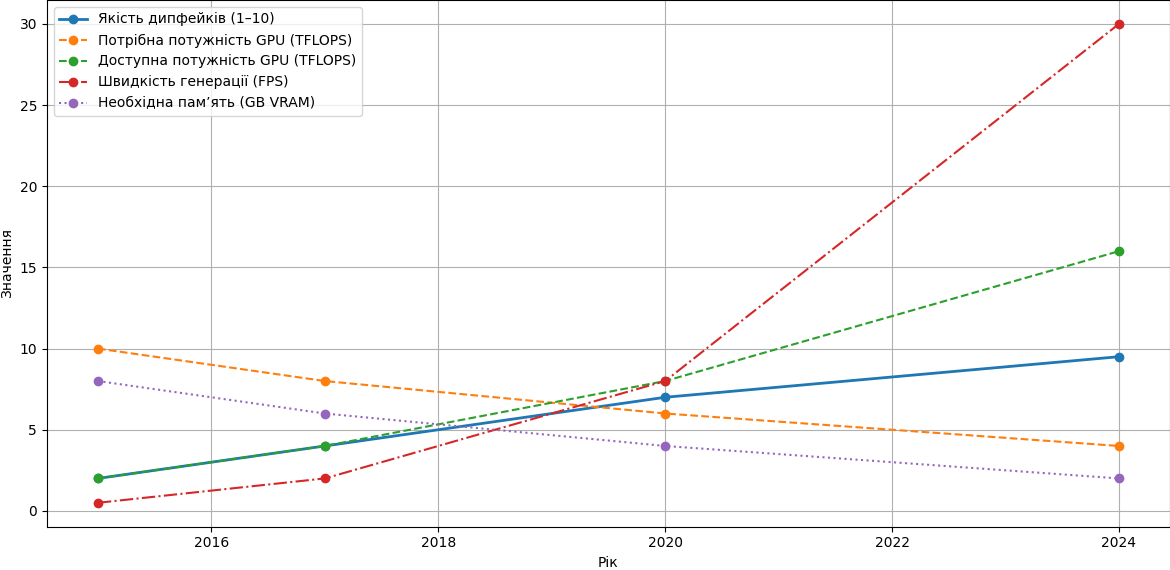


Рис. 1.2. Еволюція діпфейк-технологій

## 1.2 Cинтетичне медіа, його класифікація та види

Синтетичне медіа — узагальнений термін, позначаючий медіа-контент, що був створений за допомогою алгоритмічних моделей. Воно може бути не лише графічне, а й текстове та аудіо. Також розрізняють синтетичний мультимедіа контент, що має одразу відео, аудіо та/або текстовий вміст.

Класифікацію синтетичних медіа можна проводити за великою кількістю ознак. Однією з основних є переважаючий тип медіа: **візуальний** (переанімовані обличчя, змінена сцени/фону), **аудіо** (клонування голосу, штучна зміна інтонацій), **текстовий** (згенеровані повідомлення). Також окремо виділяють **мультимодальні** медіа, що є поєданням попередніх.

Ще одиним параметром класифікації є архітектура використаної моделі. На сьогоднішній день виділяють чотири основні типи:

* **Генеративні антагоністичні мережі (GAN)** — базуються на парі моделей “генератор-дискримінатор” і забезпечують постійне удосконалення результату;
* **Автокодери** та **варіаційні автокодери**, що дозволяють змінювати та реконструювати вхідні дані на основі їх латентних ознак;
* **Diffusion-моделі** — більш сучасний підхід, що базується на основі ітеративного "очищення" шума.
* **Трансформерні архітектури**, які показують особливу ефективність у мультимодальних сценаріях, зокрема для синхронізації синтезованого мовлення та його зображення.

Також часто класифікацію синтетичного медіа проводять, базуючись на рівні маніпуляції первинним матеріалом: від точкових змін (до прикладу корекція губ або вимови окремих слів) до повністю синтезованих сцен, що не мають жодного відношення до реальності.

## 1.3. Архітектури deepfake-моделей

Стрімкий розвиток deepfake-технологій став можливим передусім завдяки постійному вдосконаленню архітектур машинного навчання, що дозволило відтворювати дедалі складнішу структуру відео та аудіо даних. На відміну від первинних алгоритмів, що були переважно призначені для забезпечення комп’ютерного зору, архітектури, розроблені для генерації медіа мають набагото більшу кількість вхідних параметрів, що в свою черго дозволяє їм знаходити багаторівневі залежності між ознаками, що є особливо важливо при створенні зображення або звуку, наближених до реальних. Найбільш поширеними архітектурами deepfake-моделей є автокодери, генеративні антагоністичні мережі та моделі на основі дифузії або трансформерів.

**Автокодери та варіаційні автокодери**

Автокодери є одними з найпростіших генераційних архітектур, і часто використовуються для вибіркового перетворення зображень. Структурно модель автокодера складається з двох компонентів: енкодера, що перекодовує вхідні дані до латентного простору, та декодера, що виконує зворотне відображення. Автокодери часто застосовуються для виокремлення певних рис цільового зображення (до прикладу обличчя) і їх подальшої заміни на інше, але при цьому зі збереженням загальної структури зображення.

Поширеним варіантом даної архітектури є варіаційні автокодери (VAE), у яких латентний простір ймовірнісний розподіл, що дозволяє генерувати більш варіативні результати. Водночас це призводить до зменшення рівня деталізації у згенерованих зображеннях, через що використання VAE у deepfake-моделях зазвичай має допоміжний характер.

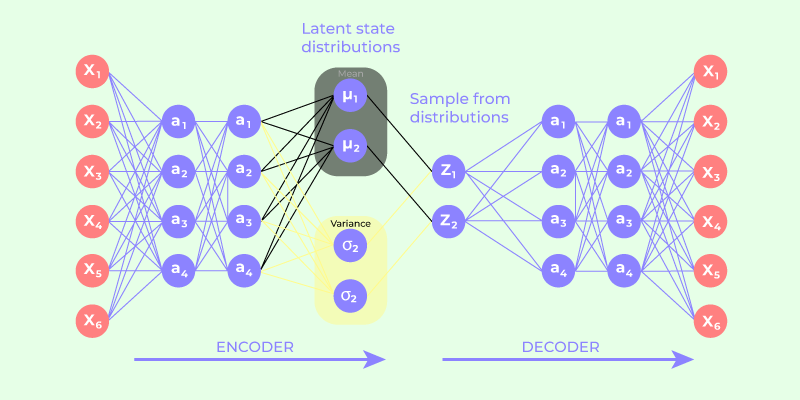


Рис. 1.3. Структура моделі на базі варіативного автокодера

**Генеративні антагоністичні мережі (GAN)**

Генеративні антагоністичні мережі (Generative Adversarial Networks) є однією з найпоширеніших архітектур deepfake моделей. Суть їх роботи полягає у протистоянні двох нейронних мереж: генератора, що створює зразки синтетичного медіа, та дискримінатора, чия задача полягає у виявленні чи вхідні дані є штучно створеними. Навчання такої моделі проводиться до моменту, коли дискримінатор не може з великою впевніностю відрізнити підробку від реального зображення.

Незважаючи свою ефективність, GAN-моделі мають кілька важливих недоліків: вони є дуже чутливими до гіперпараметрів, є порівняно складними у навчанні, а також мають ризик колапсу, після якого модель може генерувати медіа з дуже обмеженим спектром характеристик. Це обмежує використання GAN для генерації відео- та аудіо- медіа, потребуючи додаткових стабілізуючих модулів.

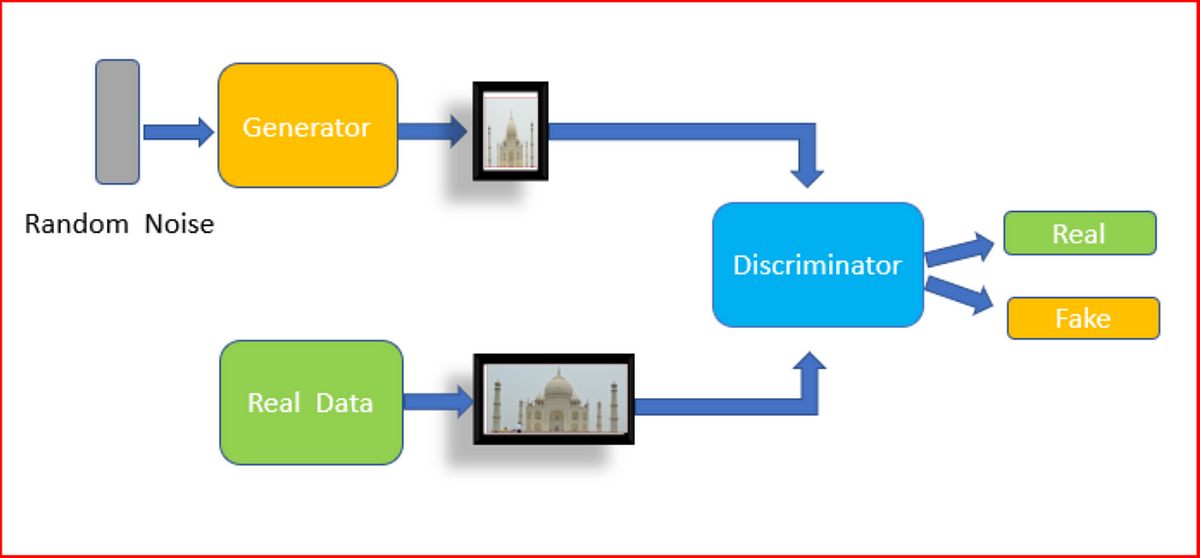


Рис. 1.4. Структура моделі на базі генеративних антагоністичних мереж

**Моделі на базі дифузії та трансформерів**

Останнім часом дедалі ширшого застосування набувають дифузійні генеративні моделі. В їх основі лежить ітеративне перетворення шуму у структуроване зображення, що дозволяє досягати високої деталізації та фотореалістичності. Однак незважаючи на те, що дифузійні моделі демонструють вищу точність у порівнянні з GAN, їх використання задля генерації відео є дуже обмеженим через великі вимоги до обчислювальних потужностей.

У свою чергу, трансформерні моделі, що спочатку розроблялись переважно для обробки мови, починають дедалі частіше використовуватись для обробки та генерації відео. Архітектури, подібні до DALL·E, Imagen або Make-A-Video, демонструють здатність до генерації мультимедійного контенту на основі текстових описів. Головною особливістю трансформерних моделей є їх здатності зберігати логічну та просторову цілісність зображень, що є критично важливим для генерації відео.

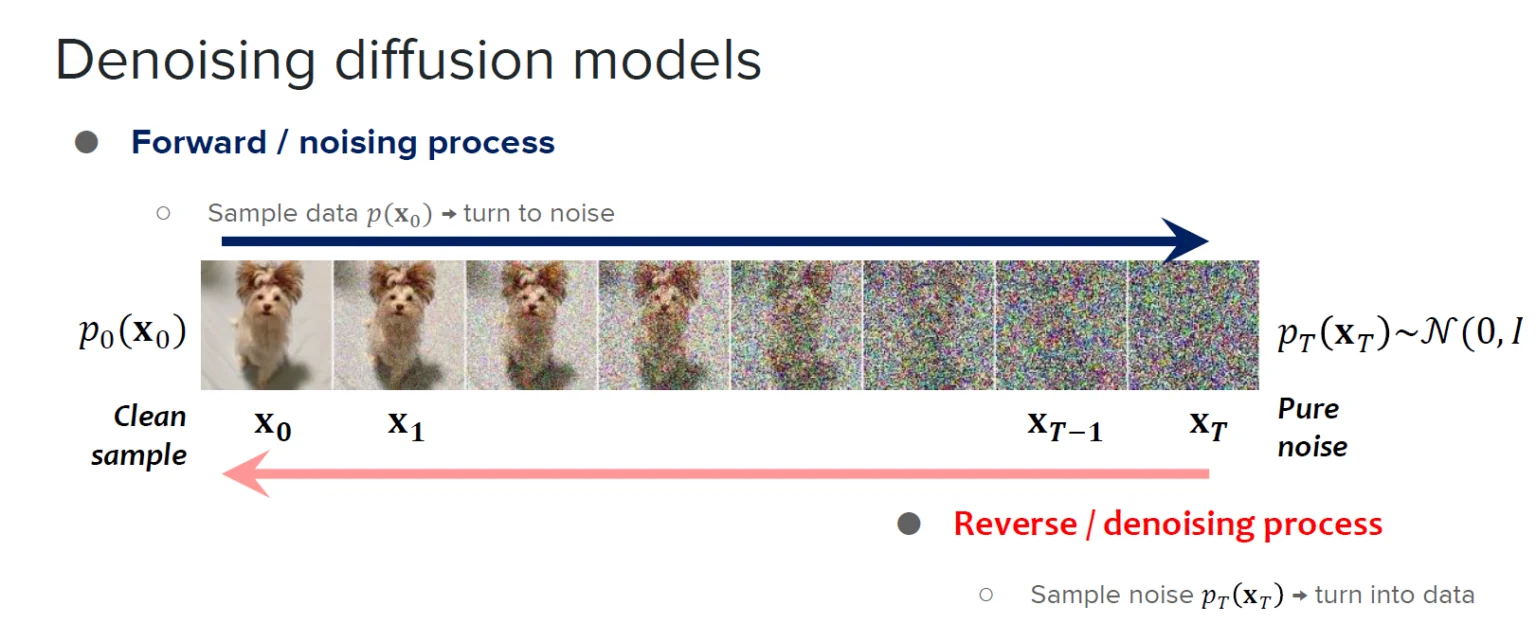


Рис. 1.5. Структура моделі на базі дифузера

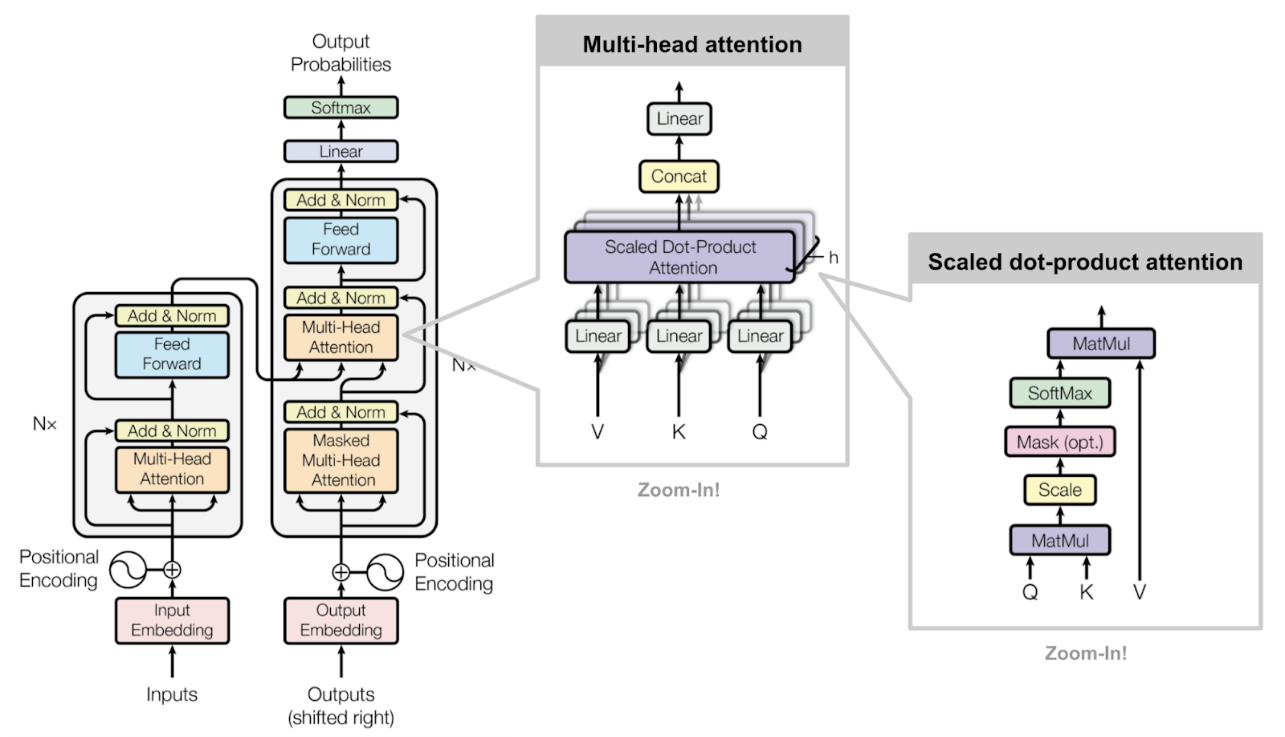


Рис. 1.6. Структура моделі на базі трансформерів

## 1.4 Найпоширеніші типи діпфейків

**Підміна обличчя (Face Swap)**

Напевно один з найвідоміших типів, підміна обличчя є дуже розповсюдженою серед deepfake-контенту. Її суть полягає у тому, що обличчя однієї людини замінюється на обличчя іншої, зазвичай схожої за рисами. Підміна обличчя набула широкого розповсюдження у розважальній індустрії, а останнім часом починає бути дедалі частіше поміченою під час проведення дезинформаційних кампаній.



Рис. 1.7. Приклад підміни обличчя

**Синхронізація руху губ (Lip Sync)**

У цьому випадку змінюється виключно артикуляція обличчя — рухи губ синхронізуються з попередньо записаною аудіодоріжкою, що дозволяє створити ілюзію вимови нових реплік, не змінюючи решту відео.

Цей інструмент набув особливого поширення у сферах цифрової реконструкції історичних персонажів, адаптації відеоконтенту для різних мов, а також для створенні навчальних або рекламних матеріалів.

**Клонування голосу (Voice Cloning)**

Клонування голосу полягає в моделюванні мовлення конкретної особи з відтворенням її тембру, інтонаційних характеристик, ритміки та фонетичних особливостей. Моделі, що застосовуються для цього, можуть функціонувати як на основі класичних автокоригувальних архітектур, так і з використанням дифузійних підходів до синтезу мовлення.

Потенційні ризики, пов’язані із застосуванням таких моделей, є значними — від імітації телефонних дзвінків у шахрайських схемах до створення фальсифікованих аудіозаписів для маніпуляцій у публічному просторі. А при наявності оригінальних записів можливе створення копій, які практично неможливо відрізнити від автентичних.



Рис. 1.8. Реклама сервісу з синтезу мовлення та її синхронізації з відео

**Повна генерація людського тіла**

Цей напрям, попри меншу поширеність, вважається перспективним з огляду на створення повністю цифрових аватарів. Для його реалізії використовуються складні мультикомпонентні моделі, до складу яких входять алгоритми трекінгу скелетної структури, генерації поз та матеріалів одягу, адаптації тіла до симульованих фізичних умов, та інші. Основні труднощі виникають під час побудови сцен із кількома об'єктами, або коли необхідна детальна міжоб’єктна координація.



Рис. 1.9. Відеосюжет, повністю згенерований за допомогою Google Veo 3

**Зміна медіа в реальному часі**

Цей клас deepfake-технологій передбачає миттєве перетворення візуального чи голосового сигналу під час його трансляції. Системи такого типу повинні функціонувати з мінімальною затримкою, що вимагає більш оптимізованих моделей, а також збільшує рівень вимог до обчислювальних потужностей.

Використовується як у розважальних трансляціях, так і для потенційно протиправних діях — до прикладу для імітації присутності іншої особи у відеоконференції, що становить значну загрозу для безпеки в цифровому середовищі.

## 1.5 Етичні та правові виклики пов’язані з використанням deepfake-технологій

**Використання біометричних ознак без згоди особи**

Проблема несанкціонованого використання обличчя або голосу людини залишається однією із найбільш чутливих аспектів синтетичних медіа. Застосування deepfake-технологій для створення контенту, у якому зображується конкретна особа без її відома й згоди суперечить основоположним принципам цифрової етики та може підпадати під дію законів про захист персональних даних.

Особливу небезпеку складають випадки несанкціонованого використання технологій у створенні матеріалів порнографічного характеру, що спрямовані на дискредитацію осіб або втручання в їхнє особисте життя. Така практика фактично створює новий різновид віртуального насильства, який складно виявити, і ще складніше довести в юридичній площині.

**Регуляторна невизначеність і відсутність уніфікованого правового підходу**

Правова база у більшості держав не здатна оперативно реагувати на інновації у сфері штучного інтелекту, зокрема генеративних моделей. Правове поле у цій галузі залишається розмитим: існують окремі ініціативи та акти, однак вони не становлять цілісної системи, що могла б регулювати створення й розповсюдження синтетичного контенту.

Зусилля окремих країн, таких як США, Китай та ряду держав Європейського Союзу мають переважно фрагментний характер і зосереджуються на обмеженні найбільш деструктивних видів використання deepfake-технологій. Водночас відсутність координації на міждержавному рівні унеможливлює ефективну протидію транснаціональним загрозам та підриває механізми правового реагування.

**Втрата довіри до цифрового контенту та «дивіденд брехуна»**

Поширення синтетичного контенту з високим рівнем правдоподібності спричиняє поступову втрату довіри до візуальних і звукових свідчень, що раніше сприймалися як об'єктивні форми фіксації подій. Унаслідок цього виникає явище, яке в академічній літературі отримало назву *liar’s dividend* (у перекладі — «дивіденд брехуна»): можливість заперечення навіть справжніх відео- та аудіоматеріалів, аргументуючи це їх потенційною підробкою.

Таке середовище сприяє поширенню дезінформації, унеможливлює об’єктивне розслідування фактів, підриває можливість публічного контролю та завдає шкоди демократичним інституціям. Втрата довіри до цифрових доказів, зокрема в журналістиці, судочинстві та громадському активізмі може мати довгострокові наслідки, які поки що складно оцінити повною мірою.

## 1.6. Методи виявлення синтезованого медіаконтенту

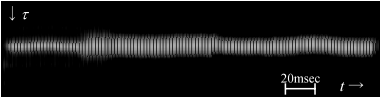
Для виявлення «штучного» медіаконтенту загалом використовуються два основні підходи. Перший з них застосовує класичні алгоритмічні методи, що ґрунтуються на чітких правилах та евристичних підходах, а другий — використовує інструменти на основі машинного навчання.



Рис. 1.10. Приклад артефакту з спотвореними очима

**Класичні алгоритмічні методи**

Класичні методи ґрунтуються на певних правилах та евристичних принципах, таких як аналіз частотних характеристик відео, виявлення статистичних аномалій у відеопотоках або оцінка стиснення відеофайлів. Особливістю таких методів є їх здатність виявляти аномалії в контенті навіть при низькій якості або явних змінах у відеофрагментах. Наприклад, алгоритми можуть виявляти аномалії в частотному спектрі або зчитувати ефективність стиснення відео для пошуку нетипових маніпуляцій. Однак їх ефективність почала значно зменшуватись при скануванні матеріалів, створених за допомогою сучасних моделей генерації медіа через їхню кращу якість та здатність до маскування.



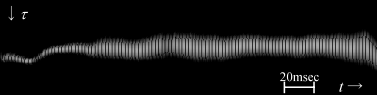


Рис. 1.11. Зміна тембру справжньої (верхній ряд), та синтезованої (нижній ряд) вимов

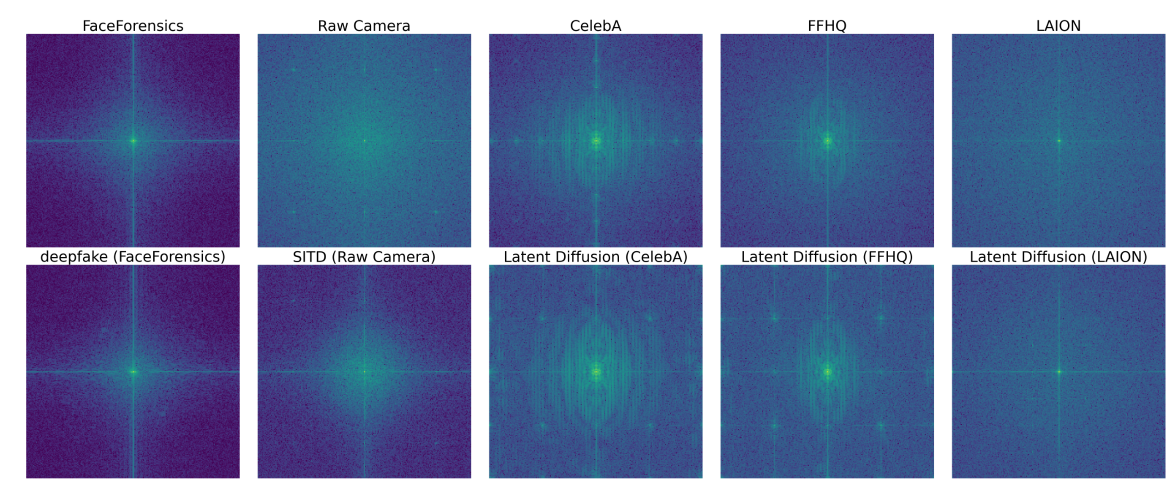


Рис. 1.12. Спектральний розподіл шумів у реальних та синтезованих зображеннях

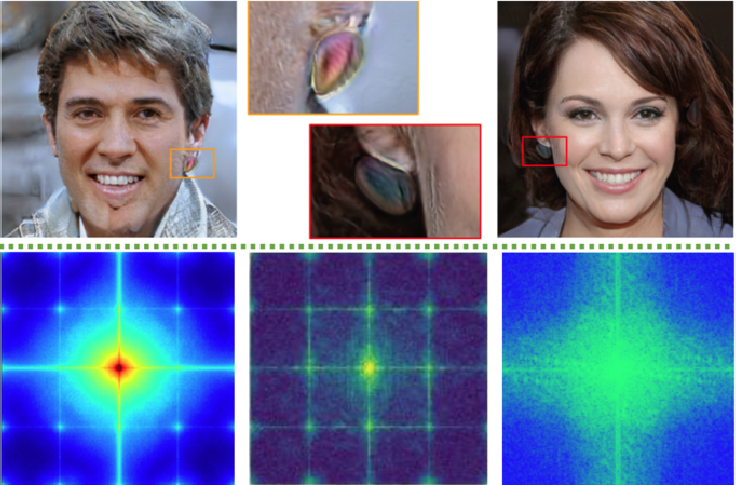


Рис. 1.13. Видимі (верхній ряд, зображення), та невидимі (нижній ряд, спектр Фур’є, частотний аналіз)

**Алгоритми на основі машинного навчання**

Натомість алгоритми, засновані на машинному навчанні, значно збільшують шанс виявляти ознаки фальсифікацій. Вони часто є натренерованими на великих наборах реальних і синтезованих медіа і можуть виявляти закономірності, невидимимі для традиційних методів. Серед найбільш популярних типів нейронних мереж, що використовуються для детекції deepfake, є convolutional neural networks (CNN), які ефективно працюють з просторовими ознаками зображень, а також recurrent neural networks (RNN), що застосовуються для аналізу часових залежностей між кадрами відео. Окрім того, архітектури на основі трансформерів також демонструють високу ефективність у виявленні фальсифікацій завдяки їх здатності працювати з часово-об’єднаними послідовностями кадрів.

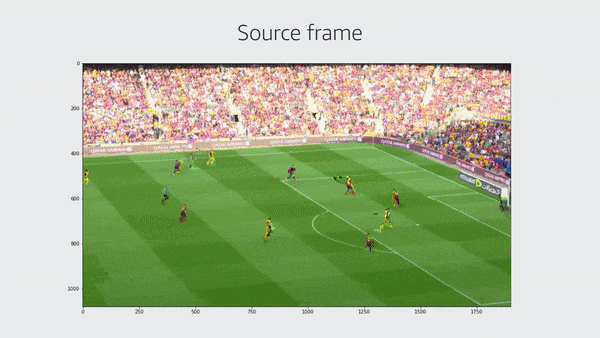


Рис. 1.14 Модель, розроблена Amazon для пошуку пошкоджених кадрів, пізніше адаптована для пошуку deepfake артефактів

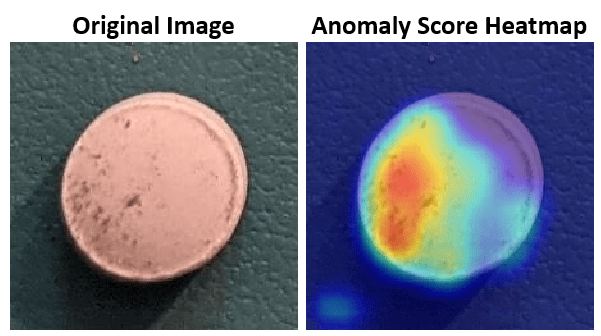


Рис. 1.15 Хітмап вірогідності артефакту у зображенні

2. РОЗРОБКА МУЛЬТИМЕДІЙНОЇ ЕНЦИКЛОПЕДІЇ

## 2.1 Вибір методів та засобів розробки

Для створення мультимедійної енциклопедії було обрано сучасний стек веб-технологій, що забезпечує кросплатформеність, масштабованість та зручність у розробці й підтримці. Основною мовою програмування є JavaScript, що  використовується як на клієнтській, так і на серверній частині. Для організації структури проекту застосовано Node.js, що дозволяє виконувати JavaScript-код на сервері, а також забезпечує роботу з файловою системою та обробку запитів клієнта. Інтерфейс користувача реалізовано з використанням бібліотеки React, що дозволяє створювати динамічні односторінкові додатки (SPA) з розвиненою структурою компонентів. Це забезпечує гнучкість у розробці, повторне використання коду та легкість масштабування проєкту. Для стилізації застосовано CSS-модулі, які дозволяють ізолювати стилі окремих компонентів і уникати конфліктів між ними.

## 2.2 Використані бібліотеки:

* **React** — бібліотека для побудови інтерфейсу користувача, що забезпечує компонентний підхід до розробки.
* **React Router** — бібліотека для організації маршрутизації між сторінками енциклопедії, що дозволяє реалізувати переходи між розділами без перезавантаження сторінки.
* **Express** — серверний фреймворк для Node.js, що забезпечує обробку запитів, маршрутизацію, роботу з файлами та статичними ресурсами.
* **classnames** — допоміжна бібліотека для роботи з класами CSS у React-компонентах.

## 2.3 Структура проекту:

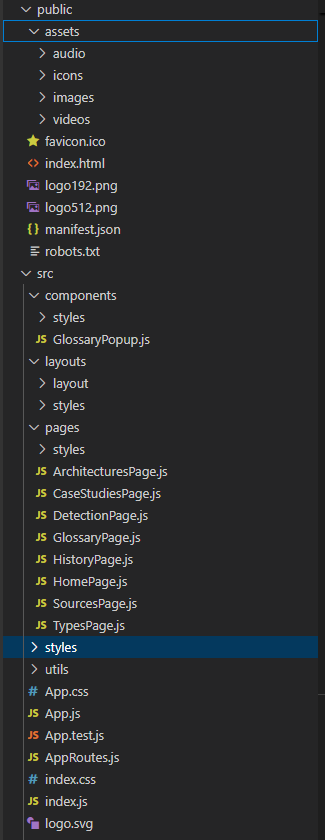
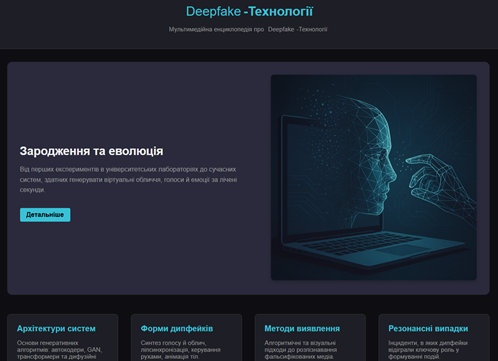
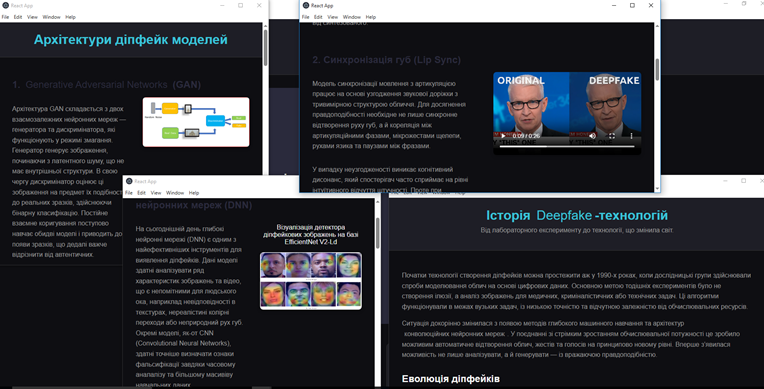
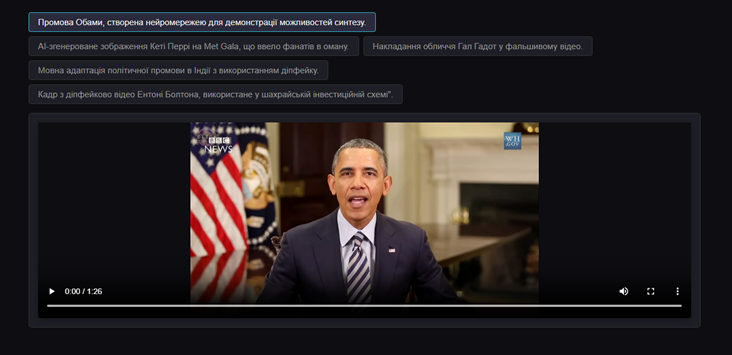
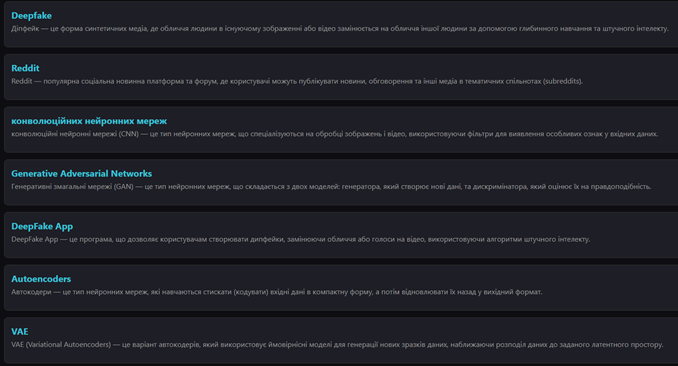


Рис. 2.1 Структура проекту

Рис. 2.2 Головна сторінка

Рис. 2.3 Статті

Рис. 2.4 Галерея

Рис. 2.5 Глосарій

# ВИСНОВКИ

# 

У результаті виконання курсової роботи було реалізовано мультимедійну енциклопедію, що окреслює головні аспекти deepfake -технологій, а саме їх різновиди, принципи функціонування, методи виявлення та суспільні виклики, пов’язані з їх застосуванням. Проєкт поєднує текстову, графічну й відеоінформацію у структурованому веб-інтерфейсі, створеному з використанням технологій Node.js і React. Особливу увагу було приділено класифікації основних типів deepfake-контенту та методам його детекції — як традиційним, так і заснованим на штучному інтелекті. Реалізація проєкту дозволила не лише поглибити теоретичне розуміння тематики, а й здобути практичний досвід у створенні інтерактивних інформаційних ресурсів.

# СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

1. *Chende Zheng, Chenhao Lin, Zhengyu Zhao, Hang Wang, Xu Guo, Shuai Liu, Chao Shen* Breaking Semantic Artifacts for Generalized AI-generated Image Detection. – NeurIPS 2024. – 4 с.
2. https://freedium.cfd/https://pub.towardsai.net/top-important-computer-vision-papers-for-the-week-from-27-05-to-02-06-f3f5625b9166
3. <https://www.researchgate.net/publication/374313999_GAN-Generated_Faces_Detection_A_Survey_and_New_Perspectives>
4. https://freedium.cfd/https://medium.com/htx-s-s-coe/uncovering-the-real-voice-how-to-detect-and-verify-audio-deepfakes-42e480d3f431