ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ В.Н. КАРАЗІНА
СОЦІОЛОГІЧНИЙ ФАКУЛЬТЕТ

Кафедра прикладної соціології та соціальних комунікацій

Пояснювальна записка

до кваліфікаційної роботи

на тему

«DEEPFAKES ЯК ІНСТРУМЕНТ МАНІПУЛЯТИВНИХ МЕДІАКОМУНІКАЦІЙ»

Виконав: студент 4 курсу групи СМК-42
першого (бакалаврського) рівня вищої освіти
спеціальності 061 Журналістика
Гармаш В.В
Керівник: кандидат соціологічних наук, доцент закладу вищої освіти кафедри соціології соціологічного факультету Зубарєв О.С

Харків – 2024

ЗМІСТ

ВСТУП

У сучасному світі, де інформація виступає ключовим ресурсом, її якість, достовірність та автентичність набувають величезного значення. Відомо, що зображення може "говорити" тисячами слів, але що, якщо зображення було спотворено або повністю синтезовано? Існує різновид штучного інтелекту, який дозволяє створювати високоякісні фальшиві аудіо та відео записи – це технологія під назвою "deepfake". Зі створенням deepfakes можливість маніпулювати реальністю досягла нового рівня, викликаючи загрози для індивідів, політики, суспільства та демократії загалом. Проблема deepfakes не просто в тому, що вони існують.

Проблема полягає в тому, як вони використовуються в медіа комунікаціях, викривлюючи правду та справжність подій. Ця проблема стає особливо актуальною в контексті політичних кампаній, рекламних стратегій або просто особистих взаємин. З цієї причини стає очевидною необхідність глибокого вивчення цієї теми.

У вузькому розумінні, дипфейки (від англ. "deep learning" – глибоке навчання та "fake" - підробка) створюються за допомогою методів, які дозволяють накласти зображення обличчя цільової особи на відео джерела, щоб створити відео, на якому цільова особа робить або говорить те саме, що й джерело. Це входить до категорії дипфейків, відомої як обмін обличчями (faceswap). У більш широкому розумінні, дипфейки - це створений за допомогою штучного інтелекту контент, який також може включати дві інші категорії, а саме: синхронізацію губ (lip-sync) та "маріонетку" (puppet-master). Дипфейки з синхронізацією губ відносяться до відео, які змінені так, щоб рухи рота відповідали аудіозапису. Дипфейки "маріонетки" включають відео цільової особи (ляльки), який оживає відповідно до міміки, рухів очей та голови іншої особи (маріонетки), що сидить перед камерою [1].

Основна сутність проблеми deepfakes виявляється у збільшенні складності відмежування реальності від вигадки в медіа комунікаціях.

Ступень наукового дослідження обраної теми

Ступінь наукового дослідження теми deepfakes в контексті маніпулятивних медіакомунікацій показує, що ця проблема є досить новою, але активно розвивається в останнє десятиліття завдяки стрімкому прогресу в галузі штучного інтелекту. В Україні ця тема практично не досліджена, і відповідних досліджень з цієї тематики не проводиться, а практично всі дослідження є перекладом найбільш значущих і актуальних досліджень зі США, Канади та країн Європейського Союзу.

На Заході проблема deepfake'ів як інструменту маніпуляцій привернула значну увагу вчених, спеціалістів із кібербезпеки та громадськості загалом [2] Проблема deepfake'ів як інструменту маніпуляцій привернула значну увагу вчених, спеціалістів із кібербезпеки та громадськості загалом, наприклад Оксфордський університет активно досліджує тему виявлення та розповсюдження дипфейків у політиці [2]. Міністерство Внутрішньої Безпеки США випустило доповідь на тему "Зростаюча загроза використання deepfakes" [3], у якій йде мова про майбутні загрози використання deepfakes та рекомендації для розробки нормативно-правової бази для порушників, які виготовляють порнографічний контент без згоди особи, яка на ньому представлена.

Є західні науковці, які спостерігають за розповсюдженням deepfakes у режимі реального часу, наприклад доктор Хані Фарід, професор Каліфорнійського університету в Берклі та експерт з цифрової криміналістики, веде сторінку Deepfakes на президентських виборах у США 2024 року, яка містить близько десятка відомих прикладів використання deepfakes у великій американській політиці [4]. Дані, отримані з одного із відкритих хабів наукових статей наприкінці 2021 року, показують значне збільшення кількості наукових статей про дипфейки в останні роки (Додаток А). Хоча отримані

цифри статей про дипфейки можуть бути менше, ніж реальна кількість, але очевидно зростає науковий інтерес до цієї теми.

Із філософської точки зору найбільш повно та ґрунтовно розкрита тема штучного інтелекту та deepfakes у видатного українського інтелектуала, професора філософії Київського Університету імені Т.Г Шевченка Андрія Баумейстера, що випустив великий цикл роликів про штучний інтелект та deepfakes та їхній вплив на масову свідомість [5, 6].

Об’єктом роботи є deepfakes як феномен медіапростору.

Предметом роботи є deepfakes як інструмент маніпулятивних медіакомунікацій.

Метою роботи є виявити вплив технології deepfakes як інструменту маніпулятивних комунікацій на масову свідомість, визначити основні технічні, соціальні та етичні аспекти їхнього використання та розробити рекомендації щодо протидії можливим негативним впливам цієї технології.

Задачі:

1) Розглянути медіа маніпуляції з deepfakes як інструмент формування та зміни громадської думки.
2) Описати термінологію, походження та технічні аспекти технології deepfakes.
3) Надати стислий огляд програмного забезпечення для створення та детекції deepfakes.
4) Вивчити приклади реального використання глибоких фейків у політиці.
5) Оцінити правовий статус глибоких фейків і пов'язані з ними законодавчі аспекти.
6) Типологізувати deepfakes із точки зору технічних ознак, інструментів створення, описати основні області застосування технології.
7) Концептуалізувати deepfakes з точки зору соціології через призму концепцій Ж. Бодріяра, М. Вебера, Т. Гейгера.
8) Провести покроковий аналіз deepfake на прикладі фальшивого відео із відомою особистістю використовуючи різні інструменти для аналізу

9) Створити власний голосовий deepfake, проаналізувати тенденції та аспекти для масового поширення, проаналізувати суспільний інтерес до цього явища.

Методи дослідження:

1) Аналітичний: аналіз даних про deepfakes у медіакомунікаціях.
2) Порівняльно-історичний: вивчення історії медіа маніпуляцій та розвитку deepfakes.
3) Аналіз вторинної соціологічної інформації: дослідження використання deepfakes в медіа.
4) Нормативно-правовий аналіз: вивчення законодавства та регулювання deepfakes.

Публікації : відображені у ДОДАТОК Д

1 МЕДІА МАНІПУЛЯЦІЇ ІЗ ВИКОРИСТАННЯМ DEEPFAKES ЯК ІНСТРУМЕНТ ФОРМУВАННЯ ТА ЗМІНИ ГРОМАДСЬКОЇ ДУМКИ

1.1 Deepfakes як загроза демократії і громадській стабільності

У контексті публічного дискурсу, deepfakes стають потужним інструментом медіа маніпуляцій, що може мати згубний вплив на демократичні процеси та соціальну стабільність. Їх можливості перевищують просте створення фальшивих новин, оскільки вони можуть відтворювати реалістичні візуальні та аудіо сценарії, які здатні вводити в оману навіть критично налаштованих споживачів інформації. Deepfakes можуть бути використані для створення матеріалів, які підривають довіру до політичних лідерів або організацій. Наприклад, можливе створення відео, де політик виглядає компрометуюче або висловлює радикальні погляди, яких він насправді не поділяє.

Що стосується аналізу впливу глибоких фейків на масову свідомість та того, як це «відгукуєтеся» у пересічних громадян, які є споживачами контенту, то тут можна навести дані дослідження 2020-го року, яке було проведене двома британськими науковцями Крістіаном Вассарі та Ендрю Чедвіком [7].
У рамках експерименту з дослідження впливу дипфейків, відомих як маніпулятивні відео, було опитано 2005 респондентів. Їх випадково розподілили на три групи: перша (653 особи) переглядала короткий, 4-секундний оманливий кліп; друга (683 особи) – більш довгий, 26-секундний оманливий кліп; а третя (669 особи) – повне відео з роз'ясненням і освітнім матеріалом. Розподіл учасників був рівномірним за демографічними характеристиками, політичними поглядами, використанням цифрових медіа, політичними обговореннями в соціальних медіа та довірою до новин у соціальних мережах, що було виміряно до експерименту. Експеримент досліджував вплив deepfakes на довіру до новин у соціальних медіа. Зокрема,

аналізувалася реакція учасників на питання, чи вважають вони, що Обама назвав Трампа непристойним словом, що могло б ввести їх в оману (H1), і наскільки вони були непевні щодо змісту відео (H2). Виявилося, що лише 50,8 % учасників не були обмануті дипфейком, що є дивним, враховуючи малоймовірність такого твердження. Група обманутих становила 16 %, тоді як 33,2 % учасників залишилися непевними. Аналіз показав статистично значущі відмінності в реакціях учасників, які переглянули різні відео.

Цікаво, що учасники, які дивилися короткі оманливі дипфейки, не були більш схильні до омани, ніж ті, хто дивився повне відео з освітнім матеріалом. Навпаки, коротке відео виявилося найменш оманливим (14,9%), за ним йшло довше відео (16,4%) , а потім вже і повне відео з освітнім матеріалом (16,9%). Однак ці відмінності були незначними та не мали статистичної значущості. В результаті гіпотеза H1, яка стверджувала, що індивіди, які дивляться дипфейк з неправдивим твердженням, більш схильні вірити цьому твердженню, була відхилена.

Автори дослідження прийшли до висновку, що перегляд дипфейків з неправдивими твердженнями, які не розкриваються як неправдиві, схильний викликати у глядачів невпевненість. Оманливі відео, незалежно від їх тривалості, збільшували рівень невпевненості серед учасників порівняно з повними відео, де було надано освітні роз'яснення. Експеримент також досліджував, як вплив дипфейків на рівень довіри до новин у соціальних медіа може бути зумовлений через зростання невпевненості. За результатами дослідження Ендрю Чедвіка, які я привів вище, випливає, що така невпевненість дійсно може знижувати довіру до інформації у соціальних мережах, що свідчить про важливість боротьби з поширенням дезінформації через дипфейки. [7, с. 5]

Хоча це дослідження і не виявило чіткої кореляції із тим, що через появу глибоких фейків рівень довіри до ЗМІ у загальному сенсі слова зменшився, проте доволі очевидним стає той факт, що точкове використання deepfake-ів може посіяти напругу і невпевненість у суспільстві, особливо, якщо їх

використання стосується політики або державного управління. Такі матеріали можуть викликати політичну дестабілізацію та впливати на результати виборів. Також можуть бути використані для розпалювання етнічних, релігійних або соціальних конфліктів. Відео, які виглядають автентично, можуть створювати ілюзію дій або висловлювань, які можуть спровокувати насильство або ненависть.

1.2 Осмислення deepfakes в українському просторі

Найпоширенішими формами висвітлення теми deepfakes в українських ЗМІ є журналістські репортажі, статті та замітки. Ці матеріали зазвичай є спробами осмислення ситуації, що склалася. Однак слід зазначити, що найчастіше deepfakes в українських ЗМІ розглядаються з погляду звичайної людини. Журналістські матеріали часто обмежуються оглядом, покликаним просто повідомити про існування проблеми. Публіцистичні тексти на кшталт "Як уберегтися від дифпейків?" або "Як використовувати deepfakes для своєї користі" також не рідкісні. Ці огляди спрямовані на підвищення обізнаності громадян про наявні ризики та можливості, пов'язані з deepfakes.

У контексті останньої політичної ситуації та війни, яку розв'язала російська федерація проти України, непоодинокими є випадки використання deepfakes як інструменту військової пропаганди, наприклад, зовсім нещодавно в мережі з'явилося відео нібито з головнокомандувачем ЗСУ Валерієм Федоровичем Залужним, який закликав до протиправних дій щодо влади.
Як виявилося пізніше, це відео було дуже якісним deepfake-ом з боку противника, покликаним підірвати моральний дух і єдність нашої країни зсередини [8]. Deepfakes наразі хоча й не мають чільної ролі в медіа маніпулюванні та військовій пропаганді, проте з кожним роком підкорюють дедалі нові й нові вершини своєї технологічної досконалості. Це наочно можна побачити навіть аналізуючи рівень тієї військової пропаганди ворога, що

постійно намагається підірвати бойовий дух ЗСУ. Наприклад, на початку російського вторгнення був розповсюджений deepfake із президентом України Володимиром Зеленським, який також пропонував здатися і капітулювати. Навіть при поверхневому аналізі видно наскільки за цей період часу покращилася якість deepfake`ів [9].

Також важливим аспектом є контекст використання deepfake`-ів для генерації оголеного фото або відео контенту з людиною без її згоди, це також викликає непідробний інтерес у читацької аудиторії українських ЗМІ, бо проблема є актуальною, і найчастіше через неї страждають дівчата і жінки [10]. Більш детально про це буде сказано у 3-му розділі роботи.

Висновки до розділу 1

У цьому розділі демонструється, як deepfake можуть впливати на масову свідомість. Вони впливають на аудиторію опосередковано через емоції, особистісний вплив та переконання, а не безпосередньо. Відповідно до Теореми Томаса з соціології, якщо люди вірять у щось, навіть якщо це не відповідає дійсності, це може мати реальні наслідки. Хоча вплив глибоких фейків не є миттєвим, їх важливість і вплив підтверджені науковими дослідженнями. Основною проблемою є низький рівень обізнаності населення щодо deepfakes та відсутність адекватної нормативно-правової бази для їхнього регулювання, що сприяє маніпуляціям, конспірології та іншим проблемам, які детально розглядаються у 4-му розділі цієї роботи. Крім того, недостатня правова регуляція також призводить до складнощів у притягненні до відповідальності тих, хто створює та поширює deepfakes, ускладнюючи боротьбу з цим феноменом.

2 DEEPFAKES: ТЕРМІНОЛОГІЯ, ПОХОДЖЕННЯ, ТЕХНІЧНІ АСПЕКТИ, ОБЛАСТІ ЗАСТОСУВАННЯ

2.1 Походження та популяризація терміну "Deepfakes"

Термін "Deepfakes" є поєднанням слів "deep learning" (глибоке навчання) та "fake" (підробка). Його популяризація почалася наприкінці 2017 року, коли на форумі Reddit з'явилася спільнота під цією назвою. У цій спільноті користувачі почали ділитися своїми роботами, в яких вони використовували алгоритми машинного навчання для заміни облич на відео.

Популяризація терміну

З моменту появи термін "Deepfakes" став широко відомим, особливо після того, як численні ЗМІ почали звертати на нього увагу через його потенційне використання для дезінформації та маніпуляцій. Популярні платформи, такі як YouTube та Twitter, були заповнені deepfake-відео, створеними як для розваги, так і з маніпулятивними цілями.

2.2 Технічні та соціокультурні аспекти Deepfakes

Різні тлумачення терміну

Технічне тлумачення - Deepfakes базуються на концепції глибокого навчання, зокрема на так званих згорткових нейронних мережах (CNN) та генеративно-заступницьких мережах (GAN). CNN використовуються для "розуміння" основних характеристик вхідного зображення або відео, тоді як GAN займається генерацією нового контенту, який непомітно відрізняється від оригіналу.

Соціокультурне тлумачення**.** Deepfake є інструментом маніпулятивних технологій, які мають значний вплив на сучасне суспільство. Під медіакомунікаціями розуміється процес передачі інформації за допомогою

різних медійних каналів, включаючи телебачення, радіо, інтернет, соціальні мережі тощо. Маніпулятивні медіакомунікації – це вид медіакомунікацій, де інформація подається таким чином, щоб впливати на сприйняття аудиторії, часто з метою вводити в оману або контролювати громадську думку. Такі комунікації можуть включати викривлення фактів, використання підмінних аргументів, емоційний вплив та інші маніпулятивні техніки. Інструменти маніпулятивних медіакомунікацій – це методи та техніки, використовувані для маніпуляції інформацією та аудиторією.

Одним із таких інструментів є deepfakes, які дозволяють створювати відео та аудіозаписи, де обличчя або голос певної особи може бути змінено або повністю підроблено. Це створює великі ризики для дезінформації та маніпуляції, адже фальшиві матеріали можуть бути настільки переконливими, що їх важко відрізнити від реальності. У соціокультурному аспекті deepfakes викликають значні проблеми: вони можуть змінювати наше розуміння правди, автентичності та довіри до медійних джерел, поширювати фальшиві новини, спотворювати факти, або навіть фальсифікувати виступи публічних осіб.

Юридичне тлумачення та контекст**.** Deepfakes може мати ряд наслідків, що стосуються порушення авторських прав, незаконного використання зображення, дифамації та інших форм незаконної діяльності. Deepfakes можуть бути використані для створення фальшивих відео, що показують осіб у компрометуючих або шкідливих ситуаціях. Це може стати основою для судових позовів про дифамацію.

Технічні аспекти

Deepfakes використовують алгоритми глибокого навчання для аналізу двох або більше зображень або відео. Ці алгоритми "навчаються" характеристикам цільового обличчя та монтують його на інше зображення або відео з максимальною точністю.

Навіть якщо деякі Deepfakes можуть бути створені за допомогою традиційних візуальних ефектів або методів комп'ютерної графіки, останнім часом

основний механізм створення глибоких фальшивок - це використання моделей глибокого навчання, таких як автоенкодери та генеративно-суперечливі мережі (GAN), які знайшли широке використання у галузі computer vision [11]. Ці моделі використовуються для аналізу міміки та рухів людини і створення зображень обличчя іншої людини, яка робить аналогічні вирази та рухи [12]. Методи створення глибоких фейків зазвичай вимагають великої кількості зображень і відеоданих для навчання моделей створювати фотореалістичні зображення та відео.

Оскільки в Інтернеті може бути багато відео та зображень відомих осіб, таких як знаменитості та політики, вони стають початковими цілями для створення deepfakes. Deepfakes використовувалися для заміни обличчя знаменитостей або політиків на тіла в порнозображеннях та відео. Перше відео з дипфейком з'явилося в 2017 році, де обличчя знаменитості було замінено обличчям порноактора. Це становить загрозу світовій безпеці, коли методи дипфейків можуть бути використані для створення відео з світовими лідерами, які промовляють фальшиві промови [13].

Дипфейки стали популярними завдяки високій якості відеоробіт та легкості використання їх додатків різними користувачами з різними навичками в роботі з комп'ютером, від професіоналів до початківців. Більшість цих додатків розроблені на основі технік глибокого навчання. Глибоке навчання відоме своєю здатністю представляти складні та високо-вимірювані дані. Однією з варіацій глибоких мереж з такою здатністю є глибокі автоенкодери, які широко застосовуються для зменшення розмірності та стиснення зображень [14].

Перша спроба створення дипфейку була здійснена за допомогою програми FakeApp, розробленої користувачем Reddit за допомогою структури кодера-декодера автоенкодера [15]. У цьому методі автоенкодер витягує латентні ознаки обличчя зображень, а декодер використовується для відтворення зображень обличчя. Для обміну обличчями між джерелом та цільовими зображеннями потрібно два пари кодерів-декодерів, кожна з яких

використовується для навчання на наборі зображень, і параметри кодера спільно використовуються між двома парами мереж. Іншими словами, дві пари мають однакову мережу кодера. Ця стратегія дозволяє загальному кодеру знаходити та вивчати подібність між двома наборами зображень обличчя, які, як правило, мають схожі риси, такі як очі, ніс, рот. У Додатку Б показано процес створення дипфейку, де набір ознак обличчя A з'єднаний з декодером B для відтворення обличчя B зі зображенням обличчя A. Цей підхід застосовується в таких роботах, як DeepFaceLab , DFaker [16], DeepFake tf (основані на tensorflow deepfakes) [17].

Основою для створення deepfakes є різноманітні типи нейронних мереж, які використовуються для обробки та модифікації візуальних та аудіальних даних. У даній секції роботи ми зосередимося на технологічних аспектах нейронних мереж, що використовуються для створення deepfakes.

Генеративно-заворажувальні мережі (GANs)

Цей тип нейронних мереж є найпоширенішим у сфері deepfakes. GAN складається з двох основних компонентів: генератора та дискримінатора. Генератор створює нові образи, виходячи з вхідних даних, тоді як дискримінатор намагається визначити, чи є зображення реальним чи сгенерованим. Співпраця цих двох елементів дає можливість створювати дуже переконливі візуальні образи.

Звичайна модель GAN складається з двох нейронних мереж: генератора і дискримінатора, як зображено на Додатку В вона працює наступним чином. Є дві команди: генератор (G) і дискримінатор (D). Генератор намагається створити зображення, а дискримінатор визначає, чи ці зображення виглядають як справжні, чи ні. Мета гри полягає в тому, щоб генератор створив такі зображення, як справжні, і дискримінатор не міг відрізнити їх від справжніх. Ось як ця “гра” працює :

Дискримінатор намагається правильно визначити, які зображення справжні. Він виставляє оцінки для кожного зображення. Генератор створює синтетичні зображення і намагається зробити їх настільки схожими на

справжні, щоб дискримінатор помилявся і вважав їх справжніми. Формула V(D, G) вимірює, наскільки гарно генератор та дискримінатор грають в цю гру. Ми намагаємося зробити так, щоб генератор був дуже хорошим у створенні схожих на справжні зображення, і дискримінатор не міг їх розрізнити. На Додатку Г демонструються приклади зображень, створених шляхом змішування двох латентних кодів на трьох різних масштабах, де кожна підмножина стилів керує окремими важливими високорівневими характеристиками зображення.

2.3 Області застосування deepfakes

Експерти Європолу оцінюють, що до 2026 року до 90% онлайн-контенту може бути синтетично створено [19, с. 3-4]. Синтетичні медіа означають медіа, створені або зманіпульовані за допомогою штучного інтелекту (ШІ). У більшості випадків синтетичні медіа генеруються для ігор, покращення послуг або підвищення якості життя, але збільшення синтетичних медіа та удосконалення технологій породило можливості дезінформації, включаючи deepfakes. Однією з найбільш тривожних тенденцій в області застосування deepfakes є нелегітимне використання особистих фото та відео для створення компрометуючих матеріалів. Втручання в приватність через такий канал може мати драматичні наслідки, включаючи шантаж, дискредитацію і соціальну остракізацію.

Приватні фотографії чи відео можуть бути взяті без згоди особи і оброблені за допомогою генеративно-застосовних мереж (GANs) для створення матеріалів, що можуть бути використані для шантажу. Такий контент може потім бути висланий безпосередньо жертві або поширюватися в соціальних мережах. Іншим сценарієм є створення компрометуючого контенту для дискредитації особи в професійному або особистому житті.

Створені матеріали можуть бути використані для знищення репутації, кар'єрного зростання, або навіть для її соціальної ізоляції. Дипфейки можуть бути використані для створення політичної або релігійної напруги між країнами, обману громадськості та впливу на результати виборів або створення хаосу на фінансових ринках шляхом створення фальшивих новин [20]. Вони навіть можуть бути використані для генерації фальшивих супутникових зображень Землі, які містять об'єкти, яких насправді не існує, щоб ввести в оману військових аналітиків, наприклад, створюючи фальшивий міст через річку, хоча в реальності такого моста немає. Це може збити з пантелику війська, які були направлені на перетин моста в бою [21].

Розвиток передових глибоких нейронних мереж і доступність великої кількості даних зробили підроблені зображення та відео майже не відрізними від реальних як для людини, так і для складних комп'ютерних алгоритмів. Процес створення цих маніпульованих зображень і відео також став набагато простішим сьогодні, оскільки для цього потрібно всього лише фотографія документа особи або коротке відео цільової особи. Для виробництва переконливого підробленого відеоматеріалу потрібно докладати набагато менших зусиль. Останні досягнення навіть дозволяють створювати дипфейки за однією фотографією [22].

Deepfakes несуть ризики не лише для громадських осіб, але і для звичайних громадян. Наприклад, голосовий deepfake вже використовувався для ошукання генерального директора на суму $243,000 [23]. Останній реліз програмного забезпечення, відомого як DeepNude, насторожує більше, оскільки воно може перетворювати зображення людини в еротичний контент без її згоди [24].

2.4 Створення порнографічного контенту за допомогою deepfake

Deepfakes надають їхнім творцям тривожну форму влади над іншими людьми, яка, здається, неухильно піддається порнографічному використанню. Перші відео з обміном обличчями на Reddit представляли актрис, таких як зірка "Чудо-жінка" Галь Гадот, у симульованому інцесті. Акторки не були проконсультовані і не згодились на те, щоб їхні зображення використовувались таким чином, так само як і порнографічні актори. Цей факт в кінцевому підсумку призвів Reddit закрити свій форум deepfake відповідно до своєї політики проти "неприродної порнографії". Однак відео все ще активно обговорюються в темних куточках Інтернету. Технологія deepfake втрапляє в довгострокові дискусії про порнографію і об’єтивізацію жінок. В 1970-х і 1980-х роках феміністки, такі як Андреа Дворкін і Кетрін МакКіннон, стверджували, що порнографія служить для підтвердження уявлень про жінок як іграшок для чоловічих глядачів, прості об'єкти для задоволення, а не повноцінні особи з автономною волею [25, с. 4].

Деякі феміністки стверджують, що порнографія, якщо вона виконана обережно і з повагою, може бути сумісною або навіть покращенням для визволення жінок [25, с. 6]. Такі "позитивні щодо порнографії" феміністки наголошують на добровільності окремих виконавців, а також жіночих режисерів і дистриб'юторів, які створюють порнографію, що виражає сексуальність жінок без сорому. Це залишається нерозв'язаною дискусією, яка має значну складність.

Навіть якщо Дворкін та інші правильно стверджують, що загальна соціальна функція порнографії полягає в об'єктивізації жінок, все ж може бути правдою, що місцева функція деякої порнографії, створеної феміністками, полягає в розширенні прав та можливостей та дестигматизації. Жінки загалом не згоджуються, із тим, як вони зображені в порнографії, навіть якщо деякі жінки згоджуються на їхнє особисте представлення у конкретних

порнографічних роботах. Важко зважити ці дві точки. Але deepfake анігілює будь-яку тонкість чи нюанс, оскільки ніхто не згоджується на deepfake порнографію. Журналістка Саманта Коул взяла інтерв'ю у жінок, які працювали як порнографічні акторки, щоб отримати їхню думку щодо появи deepfake порнографії. Пенсіонерка Алія Джеймс розповіла Коул: "Це дійсно тривожно... Це якраз показує, що деякі чоловіки в основному бачать жінок як об'єкти, якими вони можуть маніпулювати та змушувати робити все, що вони хочуть... Це просто демонструє повне відсутність поваги до порнографічних акторок у фільмі, а також жіночих акторок». [26, с. 12] Користувачі форуму Reddit вимагали створення індивідуальних відео з конкретними акторками, обмінюючи їх обличчями в певних сексуальних актах так само легко, як вказуючи фарбу в автомийці чи замовляючи начинку для піци. І з покращенням технології здатність трактувати зображення жінок як іграшок лише зростатиме.

Однією з нових технік, розробленою комп'ютерними вченими без злих намірів, є використання штучного інтелекту для імітації рухів всього тіла реальної людини, відображаючи їх в позах актора. Як тільки подібна техніка буде доступна для deepfake'ерів, їх обмеження вже не буде полягати в суперпозиції обличчя відомих осіб на існуючі порноролики. Замість цього вони будуть створювати нові симулякри цільових знаменитостей - гнучкі, пластичні представлення, замовлені виконувати будь-які бажання користувача.

Термін "франкенпорн" у контексті дипфейків відсилає до цифрового створення порнографічних зображень чи відео шляхом комбінування частин тіла різних людей, зазвичай без їх згоди. Це назва походить від образу Франкенштейна, літературного персонажа, який був створений із частин різних тіл. У випадку "франкенпорну", технологія дипфейків використовується для створення фальшивих порнографічних зображень, комбінуючи обличчя однієї особи з тілом іншої. Знову ж таки, це, здається,

проявом найгіршого виду об'єктивізації, яку феміністична критика завжди пов’язувала із порнографію.

2.5 Deepfakes як важіль впливу на політичні процеси та інструмент військової пропаганди

У квітні 2018 року BuzzFeed опублікував відео з deepfake, на якому був показаний президент Обама, щоб продемонструвати, наскільки легко можна вставити будь-які слова в чужі вуста. У цьому відео президент Обама говорив власним голосом , скоріш за все автор відео користувався технологією “puppet-master” (маріонетка та кукловод), де він вимовляв слова автора відео, де деякі з них були малоймовірні для реального Обами. [36]

Також, як я описував на початку цієї роботи, російська федерація, яка розпочала агресивну війну проти України, використовує свої агентурні мережі та умільців для створення глибоких фейків з метою під час війни зганьбити верховну владу, тим самим внісши розкол в українське суспільство. Ще на початку війни в соціальні мережі потрапив deepfake із президентом України Зеленським, на якому він закликав капітулювати. Пізніше було запущено глибокий фейк із головкомом ЗСУ Валерієм Залужним, який так само закликав до повстання проти чинної влади. Варто визнати, що ворожа пропаганда сильно еволюціонувала за час війни, проте й досі дипфейки можна відрізнити за допомогою натренованого зору та підручних способів.

Однак, найбільш актуальними deepfakes є під час електоральних процесів, вони здатні посіяти сумніви навіть серед найбільш затятих прихильників того чи іншого політика. Наприклад, зовсім нещодавно одним із найбільших американських медіа Newsweek було викрито дипфейк із Дональдом Трампом, на якому, він нібито, танцював із 13-ти річною дівчинкою. [37, 38]. Фотографія містила в собі не тільки зображення, а й текст, що дискредитував політика: "Фотографія Дональда Трампа на приватному

острові Епстейн, на якому він танцює з 13-ти річною дівчинкою. Трампу було 50, коли він це зробив. Як називається цей тип чоловіків".

Очевидно, що перед майбутніми електоральними процесами, які відбудуться в Америці наступного року, подібні звинувачення покликані дискредитувати і зганьбити честь і гідність політика. Як заявив передвиборчий штаб Трампа, цей дипфейк був створений командою іншого республіканця Рона Десайнтіста, проте передвиборчий штаб політика випустив заяву, яка це спростувала. Варто зауважити, що цей deepfake виконаний доволі якісно, проте містить деякі проблеми, які властиві deepfake: ми можемо бачити деформовану руку чоловіка на задньому плані, який тримає склянку з напоєм. При цьому на самій руці 6 пальців. Також deepfake видає відсутність блиску в одному оці у нібито дівчини і наявність його в іншому. Тобто, дівчина прямо дивиться в об'єктив і відображення від об'єктива має бути у двох очах. Так само інше фейкове зображення пов'язане з одіозною фігурою колишнього президента Сполучених Штатів. На ньому нібито видно затримання Трампа за звинуваченням його в штурмі Капітолію, що стався в 21-му році [39].

Однак нейронні мережі наразі можуть підробляти не тільки фото- і відеоконтент, а й аудіо. Так, відомий випадок із президентом США Джозефом Байденом і підробленим аудіозаписом, на якому він нібито розповідає про банківську кризу і крах американської економічної системи загалом. [40] Звук розповсюджувався як частина відеокліпу, у якому показано рухомий старовинний магнітофон, поки голос говорить.

На задньому плані чутні перешкоди, у той час як голос вимовляє коментарі на кшталт: «Всі гроші зникли» та «крах неминучий». Голос також пропонує використовувати «всю силу засобів масової інформації», щоб заспокоїти громадськість, і у декількох місцях вимовляє безглузді фрази. Перед аудіофрагментом у відео показаний чоловік, який припускає, що на аудіозаписі Байден розмовляє наодинці перед своєю прес-конференцією. Чоловік також говорить у відео, що не знає, чи є звук справжнім. У відео відтворюються кадри, які нібито ілюструють таємні зустрічі та заговори,

забарвлені в темні тони та з елементами старовини. Це створює атмосферу змови та таємниці. Напівтемні приміщення, де видно тільки силуети людей, що обговорюють щось важливе, додають драматизму та натякають на потаємні дії влади. У деяких моментах відео з'являються нечіткі зображення документів із нерозбірливим текстом, що можуть вказувати на конфіденційність інформації, яка обговорюється. Все це супроводжується загадковою музикою, яка підсилює відчуття невизначеності та інтриги.

2.6 Позитивні області застосування

Також deepfakes можна використовувати як в візуальних ефектах, цифрових аватарах, фільтрах Snapchat, створенні голосів для людей, які втратили свої голоси, або оновленні епізодів фільмів без повторних зйомок [7]. Дипфейки можуть мати творчий або продуктивний вплив на фотографію, відеоігри, віртуальну реальність, кінематографію та розваги, наприклад, реалістичне озвучення іноземних фільмів, освіту шляхом оживлення історичних постатей, віртуальну примірку одягу під час покупок і так далі [8]. Наприклад, лабораторія штучного інтелекту Samsung змусила Мону Лізу посміхатися та створила "живий портрет" Сальвадора Далі, Мерилін Монро та інших, використовуючи машинне навчання для створення реалістичних відеороликів з одного зображення [7]. Цей прорив у галузі штучного інтелекту демонструє, як технології можуть вдихати життя в статичні зображення, створюючи ілюзію руху та емоцій, які раніше були неможливі. Проте кількість зловживань дипфейками значно перевищує кількість позитивних використань.

Висновок до розділу 2

Таким чином, проблема глибоких фейків є відносно новим викликом, який прямопропорційно пов'язаний з розвитком обчислювальних технологій.

Сама теорія глибоких фейків була створена ще в 70-х роках минулого століття американськими та радянськими вченими, однак, обчислювальні потужності того часу не дозволяли втілити цю сміливу задумку в реальність. Хоча, доволі цікавий розумовий експеримент: уявити, що deepfakes з'явилися не в другому десятилітті 21-го століття, а років на 40 раніше, який би зараз мав вигляд світ?

Однак, варто зауважити, що сучасні обчислювальні потужності дають можливість створювати подібні фальшивки "промисловим" способом, про це детально йтиметься в 5-му розділі цієї роботи. Наразі пересічному користувачеві мережі, який захоче поекспериментувати з глибокими фейками, достатньо просто знайти кілька застосунків і вибрати той, який йому здається найкращим. Подібних варіантів в інтернеті існує маса, аж до того, що на хостингу відкритого програмного забезпечення github можна знайти вихідні коди подібних моделей і, використовуючи базові навички програмування, переробити цю модель під себе. Чим і займаються окремі заповзятливі громадяни, але, як правило, подібні дії порушують закон - найпопулярнішими є додатки, що представляють послуги з "роздягання" людини, і підписка на такі сервіси коливається від двадцяти до тридцяти доларів на місяць.

Таким чином, підбиваючи підсумок цього розділу, виходячи з проаналізованих мною даних, проблема глибоких фейків є відносно новим викликом, який прямопропорційно пов'язаний з розвитком обчислювальних технологій. Тож практично будь-яка людина може бути жертвою deepfake, який буде створений для вимагання, викрадення особистих даних, та автентифікації, наприклад, у банківському застосунку, або ж взагалі стати жертвою "роздягання", що найпроблематичніше для жінок. Також варто згадати те, що з погляду моралі найбільш популярне застосування глибоких фейків - створення порнографії, є неетичним і прямо сприяє об'єктивізації жінок і переходу цілісної жіночої особистості у статус простої картинки, яку можна обирати, як фарбу для автомобіля.

Це сприяє моральному розкладанню суспільства і зсуває норми моралі, роблячи наше суспільство ще більш жорстоким. Однак, це може сприяти і

розвитку психічних захворювань, наприклад, людина з певними відхиленнями може спочатку використовувати подібні програми у власних цілях, а потім переключиться не тільки на оголені зображення. Дана тема є доволі цікавою з точки зору наукового дослідження, однак я не зустрічав жодного серйозного дослідження, присвяченого цій проблемі. Проте, глубокі фейки можуть також використовуватися для анімації вже існуючих статичних зображень, наприклад, витворів мистецтва, «оживляти» історичних персонажів . Так, наприклад, один із авторів на платформі Youtube створив відео, у якому він за допомогою Photoshop та ШІ зміг оживити римських імператорів і перетворити білі статуї у справжніх людей [45]. Проаналізувавши сучасні випадки використання глибоких фейків не лише під час електоральних процесів, а й з метою дискредитації уряду, можна підсумувати тим, що сучасна політика, а зокрема й американська, як найбільш прогресивна за своїми інструментами та взаємодією з аудиторією спромоглась пристосувати deepfakes для так званого "чорного піару", а на колишньому президенті США Дональді Трампі було використано метод "кидка тухлої риби", тобто яким би чином політик не намагався спростовувати ці заяви, сам публічний дискурс, який направлений на актуалізацію даної проблеми (також необхідно враховувати "розмір і статус політика" та його репутацію, а також "репутацію"), є неабиякою проблемою.

Суто гіпотетично, дане зображення підпадає під статтю про наклеп однак, існуюча юридична практика США ще не знає прикладів з політиками такої величини, щоб за це будь-хто поніс покарання.

3 СТИСЛИЙ ОГЛЯД ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ СТВОРЕННЯ DEEPFAKES ТА ЇХНЬОЇ ДЕТЕКЦІЇ

3.1 Огляд найпопулярнішого ПО

MidJourney є однією з платформ, яка зосереджена на створенні фотографічних deepfakes з високим ступенем реалізму [28]. Основуючись на нейронних мережах та алгоритмах машинного навчання, MidJourney надає користувачам можливість генерувати фото-ілюзії з вишуканими деталями та узгодженістю. Серед ключових особливостей MidJourney можна відзначити використання варіативних автоенкодерів для забезпечення глибокої семантичної адаптації фотографій. Це дозволяє не просто "перенести" обличчя однієї особи на фотографію іншої, але й зробити це так, що взаємодія між освітленням, текстурою та анатомічними характеристиками обличчя зберігається в гармонійний спосіб. Можливості цієї платформи виявляються особливо актуальними в контексті медіаманіпуляцій. Наприклад, зображення політичного лідера може бути сфальшоване так, що він або вона були зображені в контексті, який може викликати суспільний резонанс, таким чином впливаючи на громадську думку. Так, дуже великого розголосу зазнало використання MidJourney для створення глибоких фейків за участю Римського Папи Франциська, на якому він зображений у великій білій куртці, і зовнішнє схожий більше на реп виконавця, а не католицького понтифіка [26].

Користувачі Інтернету сприйняли зображення за чисту правду – і вихваляли Папу за його модний вибір. "Хлопці з Брукліна могли б тільки мріяти про такий рівень стилю," один із найпопулярніших коментарів під цим постом про верховного понтифіка на Twitter. Воно (зображення) має більше 20,3 мільйонів переглядів і, наразі, позначку, що воно фейкове. Наступні портрети зображували Франциска у стильних білих рукавичках і чистих білих кросівках — можливо, не дуже далеко від звичайної вишуканої форми Папи,

але все одно повністю фейкові. Ці зображення створив звичайний 31-річний будівельник з Чикаго, Пабло Ксав'єр із метою розваги, проте, як часто буває, автор виклав цей фейк у соціальні мережі та почалась «вірусіння» цього зображення і за декілька тижнів воно зібрало більше 3 мільйонів переглядів.

Через подібні скандали MidJourney вирішила обмежити доступ до свого боту, який створює зображення. Зараз необхідно оформлювати підписку, вказавши свої контактні данні та банківську карту. Натомість, MidJourney також пропонує набір інструментів для виявлення deepfakes, що базується на аналізі текстурних аномалій та інших маркерів, які зазвичай присутні в згенерованих зображеннях. Це робить платформу унікальною, оскільки вона є не тільки інструментом для створення медіаманіпуляцій, але й інструментом для їх виявлення та аналізу.

DeepFaceLab [27] представляє собою відкритий проект, доступний на платформах зберігання вихідного коду, таких як GitHub. Ця платформа надає користувачам велику гнучкість у створенні deepfake відео та інших типів медіаконтенту. Вона включає в себе модульну архітектуру, що дозволяє з легкістю інтегрувати нові алгоритми машинного навчання. DeepFaceLab пропонує широкий арсенал інструментів для тренування нейронних мереж: можливість вибору датасетів, налаштування гіперпараметрів та візуалізація процесу навчання. Сфера використання DeepFaceLab надзвичайно широка.

Окрім очевидних сценаріїв з медіаманіпуляцій, платформу використовують для наукових досліджень, кінематографії та реклами. Однак її доступність і потужність зробили її зручним інструментом для створення політичних deepfakes, які можуть змінити динаміку виборчих кампаній або дискредитувати публічних діячів.

Runway [29] є комерційною платформою, яка використовує хмарні обчислення для генерації deepfakes. Завдяки цьому користувач не обмежений потужностями власного обладнання і може генерувати досить складні deepfakes в реальному часі. Платформа має інтуїтивно зрозумілий інтерфейс і набір інструментів, який значно спрощує процес створення deepfake контенту.

Важливим аспектом є і те, що Runway пропонує користувачам API для інтеграції з іншими додатками та сервісами. Це робить Runway зручною платформою не тільки для окремих користувачів, але і для організацій, які можуть вбудовувати технологію deepfakes в свої продукти або дослідницькі проекти. Обидві платформи мають власні ризики та потенціал для зловживань, особливо в контексті медіаманіпуляцій. Зокрема, можливість створення високоякісних deepfakes на персональних комп'ютерах або навіть мобільних пристроях значно знижує поріг входу для потенційних зловмисників.

3.2 Огляд ПО для детекції deepfakes

Виявлення дипфейків зазвичай вважається завданням бінарної класифікації, де використовують класифікатори для розпізнавання між аутентичними та підробленими відео. Такий тип методів вимагає великої бази даних реальних і фальшивих відео для навчання класифікаційних моделей. Кількість фальшивих відео постійно зростає, але все ще обмежена з точки зору створення бенчмарків для перевірки різних методів виявлення. Для вирішення цієї проблеми Коршунов і Марсель [30] створили відомий набір даних з deepfakes, який складається з 620 відео на основі моделі GAN за допомогою відкритого вихідного коду Faceswap-GAN [31]. Для генерації низькоякісних і високоякісних дипфейк-відео використовувалися відео з загальнодоступної бази даних VidTIMIT [32], які ефективно імітували вирази обличчя, рухи рота та плескання очима. За допомогою цих відео були протестовані різні методи виявлення дипфейків. Результати тестів показали, що популярні системи розпізнавання облич VGG [33] та Facenet [34] не в змозі ефективно виявляти дипфейки.

Інші методи, такі як підходи з синхронізацією губ [35] та метрики якості зображення зі згортковою машиною підтримки (SVM) [36] показали дуже високий рівень помилок при спробі виявити deepfake з цього новоствореного

набору даних. Це викликає занепокоєння щодо критичної потреби у розвитку більш надійних методів, які зможуть виявляти дипфейки серед справжніх відео.

Висновок до розділу 3

Отже, можна відзначити, що програмне забезпечення для створення та детекції глибоких фейків розвивається з великою швидкістю, пропонуючи як потужні інструменти для їх створення, так і методи для виявлення. Платформи як MidJourney, DeepFaceLab, та Runway демонструють різноманітність підходів та можливостей у цій сфері, від створення фотографічних ілюзій з високим ступенем реалізму до генерації deepfake відео та інших типів медіаконтенту. Проте, поряд з технологічним прогресом зростає і потенційна загроза медіаманіпуляцій, що вимагає розвитку ефективних методів детекції та протидії. Використання інноваційних методів, таких як аналіз текстурних аномалій та розробка баз даних для навчання класифікаційних моделей, є критично важливим у боротьбі з поширенням глибоких фейків. Таким чином, розуміння потенціалу та обмежень цих технологій є ключовим для забезпечення безпеки інформаційного простору.

4 ПРАВОВИЙ СТАТУС DEEPFAKES

4.1 Юридична практика у США

У Сполучених Штатах Америки не існує конкретного федерального закону, який би у всіх випадках криміналізував створення або розповсюдження deepfakes. Однак, декілька законопроектів та законів штатів займаються питанням deepfakes, особливо тих, що використовуються зі зловмисною метою, наприклад, для переслідування, порнографії без згоди або втручання в вибори. Ось деякі помітні законодавчі ініціативи та закони штатів:
Законопроєкт про відповідальність за deepfakes (DEEPFAKES Accountability Act). Внесений до Палати представників США, цей законопроект мав на меті криміналізувати зловмисні deepfakes. Він пропонував вимоги щодо маркування контенту deepfake та зобов'язував творців нести відповідальність за шкоду, спричинену немаркованими deepfakes [41]. Станом на листопад 2023 року цей законопоєкт не був прийнятий.
Акт про заборону зловмисних deepfakes (Malicious Deep Fake Prohibition Act). Цей акт був представлений у Сенаті США для криміналізації зловмисного створення та розповсюдження deepfake. Конкретні штрафи мали бути визначені після прийняття і введення закону в дію, станом на зараз він так само знаходиться на розгляді в Сенаті [42].
Закони штатів.
Каліфорнія ухвалила кілька законів, що стосуються deepfake. Законопроект Асамблеї № 602 надає право на подання позову особам, які стали мішенями сексуально експліцитного контенту deepfake, створеного без їхньої згоди. Законопроект Асамблеї № 730 забороняє розповсюдження зловмисних deepfake аудіо або візуальних матеріалів, спрямованих на кандидата, який балотується на громадську посаду, за 60 днів до виборів [43].

Вірджинія ухвалила закон, який робить незаконним поширення порнографії deepfake без згоди, що карається до року ув'язнення та штрафом до 2,500 доларів [43].

Техас у 2019 році ввів закон, який робить кримінальним злочином створення відео deepfake з метою завдати шкоди репутації кандидата та вплинути на вибори, із штрафами, які варіюються залежно від конкретних обставин та заподіяної шкоди [43].

4.2 Юридична практика у Євросоюзі

Що стосується старого світу, то наразі немає жодного нормативно-правового акту, який би регулював створення, розповсюдження deepfake. Згідно звіту Європолу та Європейського парламенту «Боротьба із глибокими фейками у європейській політиці» «регуляторний ландшафт» у Європейському Союзі, пов'язаний з deepfake, «включає складну мережу конституційних норм, а також жорсткого та м'якого регулювання як на рівні ЄС, так і на рівні держав- членів» [19, с. 9]. Мінімальні вимоги, які рекомендує погодити Європол включають маркування контенту як deepfake, щоб було ясно, що користувачі мають справу з маніпулятивними записами, а програмне забезпечення для створення deepfake, буде вважатись правоохоронними органами таким, що підпадає під категорію "високого ризику", оскільки вважається, що воно становить загрозу правам та свободам особистості. [19, с. 11].

4.3 Юридична практика в Україні

Як я писав вже раніше, питання deepfake взагалі не існує для українського законодавства, за законом немає жодного нормативно-правового

акту який би легалізував або криміналізував створення такого контенту. Як пише кандидат юридичних наук, доцент Харківського національного університет внутрішніх справ Юртаєва Ксенія Володимирівна, це проблема потребує всебічного розгляду, дослідження та вивчення соціальної і законодавчої проблематики із регулюванням цього нового інструменту [44].

Висновок до розділу 4

У цьому розділі розглянуто правовий статус deepfake з акцентом на ситуацію в Сполучених Штатах, Євросоюзі та Україні. У США, хоча федеральний закон, який би в усіх випадках криміналізував створення чи розповсюдження deepfake, відсутній, існують різні законодавчі ініціативи та закони на рівні штатів. Ці закони зосереджені на протидії зловмисному використанні deepfake, до якого можна віднести переслідування, порнографію без згоди, чи втручання в вибори. В Євросоюзі поки що немає специфічного нормативно-правового акту, але регулювання deepfake здійснюється в рамках більш широкої регуляторної рамки для штучного інтелекту. В Україні ж питання deepfake поки що не регулюється законодавчо, що вимагає додаткового вивчення та розробки відповідного законодавства. Загалом, правовий deepfake залишається складним і різниться в залежності від регіону. Розвиток технологій вимагає від законодавців активної реакції та адаптації до нових викликів, пов'язаних із цифровою ерою.

5 DEEPFAKES ЯК ІНСТРУМЕНТ МАНІПУЛЯТИВНИХ МЕДІАКОМУНІКАЦІЙ. ПРАКТИЧНА ЧАСТИНА

Цей розділ присвячений практичній частині роботи. Я хотів би почати із соціологічної концептуалізації цієї теми. Необхідно зауважити, що тема впливу deepfake-ів доволі нова, і людство наразі перебуває лише на етапі усвідомлення та інтеграції можливостей у своє повсякденне життя в усіх сферах, і це стосується не лише deepfake-ів, а й штучного інтелекту зокрема.

Наприклад, відоме видання The New York Times [61] випустила цілу статтю, в якій колумністка ділитися власним досвідом взаємодії зі штучним інтелектом для домашніх справ, а саме сервісом на штучному інтелекті "Йогана", який, як заявляється, допомагає молодим жінкам автоматизувати більшу частину домашніх справ: наприклад, написання списку покупок, виставлення таймерів для догляду за дитиною, домашнім улюбленцем і так далі. Також чат-бот може давати поради з облаштування сімейного життя, може радити покупки, шукати їх в інтернеті, викликати клінінгову службу тощо. Варто зауважити, що людству і вченим тільки-но належить усвідомити, яким чином праця і побут людини змінитися під впливом штучного інтелекту і нейронних мереж. Однак, варто зауважити, що нейронні мережі і deepfakes можуть використовуватися і в якості розваги, про що більш детально було написано в теоретичній частині роботи в РОЗДІЛ 2, 2.6 Позитивні сфери використання.

У моєму розумінні концептуалізація з погляду класичної соціологічної науки тут необхідна з огляду на те, що цей феномен використання нейронних мереж, штучного інтелекту та алгоритмів для створення deepfake-ів уже давно стали частиною політичних, економічних і соціальних взаємовідносин. Більш детально це описано в Розділі 2, 2.5 Deepfakes як важіль впливу на політичні процеси та інструмент військової пропаганди. Хоча масове застосування deepfakes є новим явищем, однак, воно вкладається в загальну канву

соціологічної науки: прийоми впливу та впливу описані класиками соціологічної науки (Бодрійяром, Вебером та іншими вченими-соціологами), адже за своєю концептуальною сутністю, deepfake з погляду бодріярівської соціології є симулякром, однак, варто зауважити, що хоч він і створює конструкти, що видаються реальними, це не відображає всієї суті явища глибоких fakes-ів. Абсолютно нова більш необхідна через те, що окрім створення "замінника" симулякра, deepfake може створювати й унікальні твори.

Наприклад, на конференції World Internet Conference було представлено унікального робота-телеведучого зі штучним інтелектом [62] , створеного новинним агентством Сіньхуа та технологічною компанією Sogou Inc. Робот може копіювати людські міміку та манери, читаючи новини. Цей прототип, який одягнений у чорний костюм і червоний краватоk, створений на основі китайського диктора Цю Хао і є частиною китайської стратегії розвитку в галузі технологій штучного інтелекту. Наразі невідомо, коли нова технологія почне використовуватися на практиці.

Однак, у деяких випадках класичне визначення Бодріяровського симулякра спрацьовує вірно, наприклад, на одному популярному еротичному сатйе модель самостійно створила свого двійника зі штучним інтелектом для того, щоб "автоматизувати" (https://nypost.com/2023/06/24/my-man-made-an-ai-clone-of-me-so-others-can-enjoy-my-company/) спілкування зі своїми шанувальниками. Із соціологічного погляду доволі важко охарактеризувати цей феномен: подібне перебуває на стику технологій штучного інтелекту та deepfakes, цей "клон" може жартувати, фліртувати та надсилати повідомлення інтимного змісту своїм користувачам за підписку 10 доларів на місяць.

Підсумовуючи, у даному розділі роботи розглядається практичне використання deepfake та штучного інтелекту в різних сферах суспільного життя, демонструючи широкий спектр можливостей і викликів, які вони представляють. Сучасне використання штучного інтелекту в повсякденному

житті і політичних процесах починає переосмислюватись з точки зору соціології, вказуючи на необхідність глибшого аналізу впливу цих технологій. Важливо, що deepfake і симулякри, які вони створюють, можуть служити як інструмент впливу та розваг, так і представляти собою унікальні новаторські рішення у технологічній галузі, як це видно з прикладу робота-телеведучого. Однак, з соціологічної точки зору це викликає нові етичні та моральні питання, що вимагають більшої уваги вчених і громадськості до наслідків використання таких технологій в майбутньому.

5.1 Соціологічна концептуалізація deepfakes (за Ж. Бодріяром, М. Вебером, Т. Гейгером)

Концептуалізація за Жаном Бодріяром

Жан Бодріяр у своїй праці "Симулякри і симуляція" розвиває концепцію гіперреальності, наголошуючи на тому, що в сучасному світі відмінності між реальністю та її представленням стають дедалі більш розмитими. Бодріяр арґументує, що сучасні медіа створюють симулякри - копії, які замінюють і спотворюють реальність, перетворюючи її на серію знаків, що не мають безпосереднього стосунку до оригінальних об'єктів або подій. Він стверджує, що ці симулякри не просто відтворюють або спотворюють реальність, а й замінюють її, створюючи "гіперреальність", де штучне стає сприйманим як більш реальне, ніж сама реальність [56].

Бодрійяр описує симулякр не просто як копію реальності, але як щось, що стає більш реальним, ніж сама реальність. У цьому контексті, симулякри поводяться як замінники реальності, які не просто імітують або відтворюють оригінал, а й маскують і спотворюють основну суть того, що вони представляють. Симулякр не тільки імітує реальність, а й пропонує альтернативу їй, іноді навіть кращу за саму реальність. Deepfakes порушують

питання, які Бодрійяр вважав би надзвичайно важливими для розуміння сучасної культури. Ці штучно створені відео та аудіозаписи, які використовують машинне навчання для створення відео, де люди кажуть або роблять речі, яких вони насправді не казали або не робили, є прикладами симулякрів, що створюють гіперреальність. Вони не тільки маскують брехню під правду, а й можуть бути використані для маніпуляції громадською думкою або навіть для зміни історії. Концепція гіперреальності Бодрійяра особливо актуальна при аналізі deepfakes. Гіперреальність описує стан, у якому відмінність між реальністю і сценаріями, створеними медіа, стає невиразною [57, с. 3]

В епоху deepfakes, коли технологічні засоби дають змогу з легкістю створювати переконливі фальсифікації, аудиторія може більше не розрізняти, що реально, а що ні. Це підриває довіру до медіа та інформаційних джерел, поглиблюючи культурну кризу уявлення. Беручи до уваги думки Бодрійяра, deepfakes виступають як потужний інструмент маніпуляції. Вони ілюструють, як технології можуть використовуватися не лише для розваги чи освіти, а й для створення потужних інструментів маніпуляції, здатних формувати політичні порядки денні, маніпулювати індивідуальним і колективним сприйняттям. Це створює виклики для юридичних та етичних норм, що регулюють використання таких технологій.

Концептуалізація за Максом Вебером

Макс Вебер, один із засновників соціології як науки, вивчав зміни в структурі суспільства, особливо процес раціоналізації в західних суспільствах. Його теорії раціоналізації та в'янення харизми особливо актуальні для аналізу феномена deepfakes.

Раціоналізація для Вебера - це процес, під час якого традиційні та афективні дії (засновані на почуттях, емоціях і вірі) поступаються місцем діям, заснованим на розумі та ефективності. Це передбачає зміцнення

бюрократичних структур і повсюдне застосування технічних засобів для досягнення контролю та ефективності в управлінні суспільством. [58, с. 16] .

Deepfakes можна розглядати як продукт і підсилювач процесу раціоналізації в сучасному медіапросторі. Технології, що стоять за створенням deepfakes, ґрунтуються на алгоритмах штучного інтелекту, які дають змогу досягати високого ступеня точності в підробці відео та аудіо записів. Це перетворює мистецтво створення медіаконтенту на технічний процес, де раціональний підхід до створення і поширення інформації стає домінуючим. Вебер також обговорював зів'янення харизми, коли унікальні особисті якості лідера, здатні надихати і вести за собою людей, заміщуються рутинними і бюрократичними функціями. Deepfakes підсилюють цей процес, даючи змогу створювати зображення і записи, в яких лідери здаються такими, що говорять або діють у певний спосіб, що може бути повністю штучним. Це спотворює сприйняття харизми, роблячи її продуктом технологічного маніпулювання, а не особистої харизматичності.

Використання deepfakes у політиці може призвести до "інженерії згоди", де сприйняття політичних фігур та їхніх рішень формується не через реальні дії та переконання, а через маніпульований медіапростір. Це перетворює політичну взаємодію на обачливий процес, мета якого - не інформування чи натхнення суспільства, а контроль над громадською думкою. З точки зору Вебера, deepfakes є яскравим прикладом того, як сучасні технології можуть сприяти подальшій раціоналізації суспільства, підриваючи традиційні механізми соціальної взаємодії та в'янучи харизму як ключовий елемент лідерства. Це створює нові виклики для суспільства, яке повинно навчитися розпізнавати і критично оцінювати джерела інформації в епоху гіперреальності.

Концептуалізація за Теодором Гейгером

Теодор Ґейґер, видатний німецький соціолог, зробив значний внесок у вивчення структури суспільства, особливо в контексті статусу та ролей. Його ідеї можуть бути застосовані для аналізу феномена deepfakes, які впливають на сприйняття соціальних ролей і статусів.

Основи теорії Гейгера

Ґейґер досліджував, як соціальні структури формуються і підтримуються через певні статуси та ролі, які люди займають у суспільстві. Статус пов'язаний із престижем і владою, які суспільство приписує певним позиціям, тоді як роль визначає очікування поведінки, пов'язані з цими позиціями. Цей поділ допомагає підтримувати порядок і передбачуваність у соціальних взаємодіях [59 с. 5]

Deepfakes можуть серйозно впливати на сприйняття статусів і ролей у суспільстві: Маніпуляція сприйняттям: Використання deepfakes для створення контенту, в якому відомі особистості або громадські діячі здаються такими, що говорять або роблять щось, чого вони насправді не робили, може кардинально змінити сприйняття цих людей. Це впливає на їхній соціальний статус і може порушувати очікувані рольові поведінки, запропоновані їхнім соціальним становищем. Оскільки статус значною мірою залежить від репутації, deepfakes можуть швидко і незворотно зашкодити репутації людини, знижуючи її статус або змінюючи її роль у суспільстві. Використання deepfakes викликає серйозні етичні питання, особливо пов'язані з довірою і достовірністю в соціальних відносинах. Гейгер наголошував на значенні соціального порядку й очікувань, пов'язаних із різними ролями, а deepfakes можуть підірвати ці основи, призводячи до соціальної нестабільності.

Приклади можуть включати використання deepfakes у політичній агітації, де створюються відео, що показують політичних опонентів у негативному світлі, що може несправедливо змінити їхній статус і вплив у суспільстві. Також, в індустрії розваг, де акторів можуть показати в

контекстах, несумісних з їхнім суспільним образом, що впливає на їхню кар'єру та особисте сприйняття. Вивчення впливу deepfakes через призму теорії Гейгера про статус і ролі в суспільстві підкреслює, як технологічні інновації можуть мати глибокі соціальні наслідки. Важливо, щоб суспільство розробляло методи виявлення та заходи протидії для мінімізації потенційної шкоди від цього потужного, але потенційно небезпечного інструменту. Соціологічний підхід Гейгера нагадує нам про важливість розуміння соціальних структур і динаміки, які технології можуть ненавмисно порушувати або змінювати.

5.2 Власна типологія deepfakes за технічним інструментарієм та областями застосування

Переходячи до практичної частини цієї роботи, під час переосмислення і детальнішого занурення в матеріал, я склав типологію нейромереж за двома критеріями: типологія за технічними інструментами, та типологія за областями застосування. Нижче представлена типологія за технологічними інструментами, себто, під екземплярами типології маються на увазі окремі екземпляри алгоритму для створення deepfake : усі вони у головному схожі – усі працюють із даними, на яких були треновані (Більш детально це описано у теоретичній частині роботи, а саме «Технічні та соціокультурні аспекти deepfakes).

Більше того, будь-яка програма, будь-то сайт банку або застосунок для створення відео з котиками, або готовий для використання сервіс deepfakes поєднує одне – в усіх випадках йдеться про алгоритм, який працює із даними. У нашому випадку, ця типологія має узагальнити дуже обширний перелік доволі різних алгоритмів, які здібні виконувати генерацію голосу, заміну облич із існуючим відео, або навпаки, розпізнавати людський голос і працювати як голосовий асистент (Наприклад застосунок Siri у iPhone)

Типологія за технологічними інструментами - типологія, яка покликана розмежувати різні інструменти та підходи для створення deepfakes,

класифікуючи їх на основі методів і алгоритмів, що використовуються для генерації або модифікації аудіовізуального контенту. Це включає в себе оцінку генеративно-змагальних мереж, автоенкодерів, нейронних мереж та інших технологічних рішень, кожне з яких має свої особливості, можливості та обмеження. Такий підхід дає змогу не тільки зрозуміти технічні аспекти кожного інструменту, а й оцінити потенційні ризики та переваги їхнього використання в різних сферах, від розваг до політичної агітації та освіти.

1) Генеративно-змагальні мережі (GANs)

Опис: GANs складаються з двох нейронних мереж, що змагаються одна з одною: генератора, який створює зображення, і дискримінатора, який оцінює, наскільки добре вони імітують реальні дані. Ця методика широко використовується для створення високоякісних зображень і відео.

Приклади застосування: Створення реалістичних людських облич, які не існують у реальності, або зміна атрибутів обличчя у відео.

Характерні риси GANs : являють собою унікальний клас алгоритмів у машинному навчанні, який використовується для створення deepfakes.

Ось основні характеристики, які відрізняють GANs від інших методів:

- Подвійна структура мережі

  GANs унікально працюють із двома нейронними мережами - генератором і дискримінатором, які працюють у змагальному режимі. Генератор прагне створити якомога реалістичніші дані, тоді як дискримінатор намагається відрізнити підроблені дані від справжніх. Це змагання сприяє поліпшенню якості створюваних зображень або відео.

- Здатність до інновацій

  Однією із сильних сторін GANs є їхня здатність генерувати нові дані, які ніколи не існували, що робить їх ідеальними для створення унікальних

облич або сцен. Вони не просто модифікують наявні дані, а створюють нові з "нічого", ґрунтуючись на вивчених розподілах даних.

- Висока якість синтезу

  Завдяки динаміці навчання між генератором і дискримінатором, GANs здатні виробляти високоякісні та реалістичні результати. Це робить їх особливо цінними для завдань, де потрібен високий ступінь нерозрізнення, наприклад, у створенні фотореалістичних зображень людей або зміні вихідних відеоматеріалів.

  - Труднощі в навчанні

  Навчання GANs може бути складним і нестабільним процесом через їхню змагальну природу. Налаштування параметрів вимагає уважного підходу, щоб уникнути проблем, таких як колапс моди, коли генератор починає виробляти обмежену кількість типів вихідних даних, або нездатність дискримінатора адекватно оцінювати якість генерованих даних.

  - Застосовність у різних доменах:

  Незважаючи на своє широке застосування у створенні зображень, GANs також адаптовані для роботи з текстом, аудіо та іншими типами даних, що дає змогу їх використовувати в різноманітних галузях від медіа-розваг до академічних досліджень.

  2) Автоенкодери - це тип нейронних мереж, що використовуються для несупервайздного (неконтрольованого) навчання. Вони працюють, навчаючись ефективно стискати (кодувати) дані, а потім відновлювати (декодувати) їх зі стисненого подання.

  Приклади застосування: Заміна облич у відео, де обличчя однієї людини накладається на обличчя іншої в цільовому відео. Автоенкодери посідають особливе місце серед інструментів для створення deepfakes, завдяки їхній здатності до навчання за принципом стиснення і

відновлення даних. Ось кілька ключових особливостей, які відрізняють автоенкодери від інших технологій:

- Принцип роботи:

Автоенкодери складаються з двох основних компонентів: кодувальника, який перетворює вхідні дані на компактне, стиснене подання, і декодера, який використовує це подання для відновлення вихідних даних. Це відрізняє їх від GANs, де дві мережі працюють у протистоянні.

- Мета навчання:

На відміну від інших підходів, автоенкодери навчаються мінімізувати втрати між вхідними і відновленими даними, що робить їх ефективними для завдань, де необхідне точне відтворення вихідних характеристик даних.

- Здатність до деталізації:

Автоенкодери ідеально підходять для завдань, що вимагають детального відновлення даних, оскільки вони здатні вчитися на малих деталях вхідного сигналу. Це робить їх особливо корисними для таких додатків, як реставрація зображень або зміна облич у відео.

- Втрата даних:

Одним із недоліків автоенкодерів є потенційна втрата інформації під час кодування даних у більш компактну форму. Це може призвести до того, що відновлені дані будуть дещо відрізнятися від оригінальних, особливо якщо дані складні або їх багато.

- Застосування:

Автоенкодери найефективніші в завданнях, де необхідно аналізувати й відновлювати дані з високим рівнем точності та деталізації, як-от заміна облич у відео або видалення шуму із зображень.

- Незалежність від міток:

На відміну від супервайзд (контрольованого) навчання, автоенкодери не вимагають розмічених даних, що робить їх придатними для сценаріїв, де розмічені дані важкодоступні або їх отримання дороге.

3) Нейронні мережі для синтезу мови (TTS і Voice Cloning)

Опис: Технології, засновані на нейронних мережах, дають змогу синтезувати мовлення, яке звучить як людський голос, імітуючи його тембр, інтонацію та ритм.

Приклади застосування: Створення голосових повідомлень від відомих особистостей або генерація синтетичної мови для віртуальних помічників. Характерні риси нейронних мереж для синтезу мови (TTS і Voice Cloning) . Нейронні мережі для синтезу мовлення, включно з технологіями Text-to-Speech (TTS) і Voice Cloning, мають унікальні характеристики, що вирізняють їх з-поміж інших методів створення і маніпуляції аудіоданими.

Ось основні риси цих технологій:

- Висока реалістичність:

Сучасні системи TTS і Voice Cloning здатні синтезувати мову, яку важко відрізнити від справжнього людського голосу. Вони можуть імітувати тембр, інтонацію, емоції та ритм мови, що робить їх ідеальними для створення персоналізованих голосових повідомлень.

- Адаптація до контексту:

Ці системи можуть адаптуватися до різних мов і діалектів, забезпечуючи більш природне звучання для конкретних культурних і лінгвістичних контекстів. Це досягається завдяки навчанню на великих наборах даних із різноманітних джерел.

- Здатність до навчання та адаптації:

Нейронні мережі можуть навчатися на обмеженому наборі голосових даних і потім клонувати будь-який голос, для якого надано аудіозразки.

Це дає змогу створювати кастомізовані голосові профілі для використання в різних застосунках.

- Широкий спектр застосування:

Від віртуальних помічників до аудіокниг, від інтерактивних ігор до систем попередження в автомобілях - можливості застосування нейронних мереж для синтезу мови величезні. Це робить їх затребуваними в багатьох секторах.

- Мінливість голосу:

На відміну від традиційних методів синтезу мови, нейронні мережі можуть генерувати безліч варіацій одного й того самого голосу, змінюючи емоційне забарвлення, швидкість та інтонацію залежно від контексту розмови.

- Вимоги до навчальних даних:

Хоча нейронні мережі можуть демонструвати видатні результати в синтезі мови, вони вимагають великих обсягів навчальних даних і значних обчислювальних ресурсів для ефективного навчання. Це може бути обмеженням для деяких проєктів.

Ці характеристики роблять нейронні мережі для синтезу мовлення особливо цінними у створенні реалістичних і адаптованих голосових deepfakes, які можуть бути використані в широкому спектрі застосунків, від поліпшення користувацького досвіду до створення нового контенту.

4) Deep Video Portraits

Опис: Технологія, що дає змогу маніпулювати відео портретами в реальному часі, змінюючи напрямок погляду, вирази обличчя і рухи голови. Приклади застосування: Використання в кіноіндустрії для адаптації виразів облич акторів або заміни облич у вже знятих сценах.

Характерні риси Deep Video Portraits :

Deep Video Portraits - це передова технологія в галузі маніпулювання відео, яка використовує складні алгоритми для зміни виразів облич,

напрямку погляду та інших аспектів портретних відео в реальному часі. Ось кілька ключових особливостей, які відрізняють Deep Video Portraits від інших технологій:

- Висока точність маніпуляції:

Deep Video Portraits дають змогу проводити тонку і точну маніпуляцію портретними відео, що передбачає зміну виразів обличчя, рухів голови і погляду. Це досягається завдяки використанню складних моделей навчання, які детально аналізують і відтворюють мікрорухи обличчя.

- Реалізм: одна з головних переваг цієї технології - здатність створювати надзвичайно реалістичні відео. Це забезпечується за рахунок точної симуляції освітлення, текстур шкіри і природних рухів, що робить відео практично не відрізняються від реальних.
- Застосування в реальному часі: технологія здатна працювати в реальному часі, що робить її ідеальною для використання в кіно і телебаченні, де необхідно швидко адаптувати або змінити сцени без перезйомок.
- Доступність для постпродакшн змін:

На відміну від традиційних методів, які вимагають тривалого часу на підготовку і реалізацію, Deep Video Portraits дають змогу вносити зміни після закінчення основної зйомки, забезпечуючи тим самим більшу гнучкість у процесі постпродакшн.

- Етичні та юридичні питання:

Як і інші методи створення deepfakes, Deep Video Portraits стикаються з етичними та юридичними викликами, пов'язаними з правдоподібністю і використанням чужого образу без згоди. Ці особливості роблять Deep Video Portraits унікальним і потужним інструментом в індустрії розваг, особливо в кіновиробництві та телевізійних проєктах, де потрібен високий ступінь адаптивності та реалізму під час зміни відеоконтенту.

5) Технології заміни аудіо (Audio Replacement і Audio Synthesis)

Опис: Використовуючи просунуті методи обробки звуку, можна не тільки клонувати чиїсь голоси, а й створювати повністю нові звукові доріжки, які звучать природно. Приклади застосування: Заміна оригінального аудіо у відео, щоб змінити сказане слова або створити новий контекст. Характерні риси технологій заміни аудіо (Audio Replacement та Audio Synthesis) . Технології заміни аудіо, включно з Audio Replacement та Audio Synthesis, являють собою сучасні підходи до маніпуляції та створення звукових доріжок. Ці методи мають кілька ключових особливостей, які відрізняють їх від інших аудіотехнологій:

- Гнучкість у створенні контенту:

Технології заміни аудіо дають змогу не тільки модифікувати наявні аудіозаписи, а й створювати з нуля нові звукові доріжки, що ідеально підходить для ситуацій, які потребують високого ступеня кастомізації аудіоконтенту.

- Реалістичність звучання:

Сучасні методи синтезу аудіо можуть досягти високого ступеня реалістичності, що робить нові аудіодоріжки практично не відрізняються від оригінальних записів. Це досягається завдяки використанню складних алгоритмів обробки та синтезу звуку.

Застосування в різноманітних галузях:

Від кіновиробництва до ігрової індустрії та від віртуальних помічників до освітніх платформ, технології заміни аудіо знаходять застосування в широкому спектрі індустрій, що вимагають точної та переконливої аудіальної комунікації.

- Маніпуляція контекстом:

Ці технології дають змогу змінювати сказане в аудіозаписах, додаючи або змінюючи слова і фрази для створення нового контексту або для виправлення помилок в оригінальному записі.

- Етичні та юридичні питання:

Як і інші форми deepfake технологій, заміна аудіо стикається з етичними та юридичними проблемами, особливо коли йдеться про маніпулювання записами, які можуть вводити в оману або бути використані в шахрайських цілях.

- Доступ до складного обладнання та програмного забезпечення:

Хоча технології стають доступнішими, високоякісний аудіосинтез часто вимагає просунутих програмних рішень і, в деяких випадках, спеціалізованого обладнання для обробки і синтезу звуку. Ці характеристики роблять технології заміни аудіо цінним інструментом для творців контенту, які шукають способи поліпшити або повністю змінити аудіокомпонент своїх проєктів, забезпечуючи водночас високий ступінь реалізму й адаптивності.

Типізація областей застосування deepfakes

Також, необхідно зазначити, що крім технічного аспекту важливо не випускати з уваги і сфери застосування deepfakes. Сам інструмент практично повністю марний без грамотного використання в контексті, який приносить цінність для користувачів або служить певним цілям. У кожній сфері застосування deepfakes можуть нести як позитивний, так і негативний потенціал. Наприклад, у медіа та розвагах вони можуть покращувати візуальну якість контенту, тоді як у політиці можуть використовуватися для поширення дезінформації. Тому усвідомлення етичних і соціальних наслідків їх застосування стає критично важливим.

Розробникам і користувачам слід ретельно зважувати потенційні ризики та вигоди, а також розробляти стратегії мінімізації шкоди, пов'язаної з використанням цієї технології.

Аналіз типології за областями застосування deepfakes і відповідні технологічні інструменти

1) Телекомунікації:

Застосування: Використання для підробки голосів у телефонних дзвінках для шахрайства або шантажу.

Технічні інструменти: Нейронні мережі для синтезу мови (TTS і Voice Cloning) є ключовими технологіями в цій галузі, даючи змогу створювати переконливі та реалістичні копії людських голосів. Ці технології можуть бути використані для імітації голосу знайомої людини, викликаючи довіру у жертви.

2) Розваги та медіа:

Застосування: Створення аудіо- та відеоконтенту за участю відомих особистостей без їхньої фактичної участі.

Технічні інструменти:

Аудіоконтент: Нейронні мережі для синтезу мови (TTS і Voice Cloning) дають змогу створювати мову, яка ідеально імітує знаменитостей.

Відеоконтент: Генеративно-змагальні мережі (GANs) і Deep Video Portraits використовуються для створення або модифікації відеоматеріалів, даючи змогу, наприклад, змінювати вирази облич акторів або відтворювати молоді версії акторів у кіно.

3) Політична арена:

Застосування: Створення фальсифікованих аудіо- та відеозаписів для введення в оману виборців або дискредитації суперників.

Технічні інструменти:

Аудіозаписи: Технології заміни аудіо та нейронні мережі для синтезу мови часто використовуються для підробки промов і заяв політиків.

Відеозаписи: GANs і Deep Video Portraits можуть бути використані для створення відеороликів, у яких політики здаються тими, хто говорить або робить щось компрометуюче.

4) Освіта:

Застосування: Використання для створення освітніх матеріалів за участю голосів відомих учених або історичних особистостей. Технічні інструменти: Нейронні мережі для синтезу мови можуть відтворювати лекції в голоси історичних постатей, роблячи освітній процес більш інтерактивним і зануреним.

5) Психологічне консультування:

Застосування: Створення віртуальних терапевтів, які можуть запропонувати персоналізовану підтримку на основі імітації реальних терапевтичних голосів. Технічні інструменти: Технології TTS і Voice Cloning можуть створювати втішні та реалістичні голоси, що допомагають у встановленні довірчих відносин із клієнтами.

6) Правозастосування:

Застосування: Відтворення голосів або відеозображень злочинців для реконструкції злочинів або тренувань правоохоронних органів.

Технічні інструменти: GANs і Deep Video Portraits можна використовувати для візуалізації сцен злочинів або підозрюваних, покращуючи тренувальні програми та реконструкції.

5.1 Детальний аналіз та детекція deepfake (на прикладі deepfake із Валерієм Залужним).

Цей розділ роботи є практичним, саме тут я наочно продемонструю методи аналізу та створення deepfakes. Варто зауважити, що актуальність цієї теми складно переоцінити. Ворог в особі так званої РФ активно користується політичною ситуацією всередині країни, намагаючись отримати для себе з цього вигоду. У цьому ж випадку, крім базових інструментів військової пропаганди, було використано також deepfakes. Так, російською стороною з метою дестабілізації внутрішньої ситуації було створено deepfake з екс-головнокомандувачем Валерієм Федоровичем Залужним, на якому він

закликав до військового метежу і державного перевороту. Варто зауважити, що росіяни з початку війни неодноразово вдавалися до подібних технологій для дестабілізації ситуації. Однак, в даному випадку, внутрішньополітична ситуація, що склалася, і відхід одного з найбільш поважних та рейтингових воєначальників і без того спричинив деякі заворушення в суспільстві, проте ворог вирішив використати цю ситуацію на свою користь.

Відео, яке буде докладено до цієї роботи, є deepfak-ом, що був створений російською стороною та розповсюджувався серед російський засобів масової інформації. Перші кадри відео - заставка, яка повідомляє про, нібито, звернення генерального штабу України. Після, на відео ми бачимо ляльку маріонетку (puppet), яка намагається зобразити Валерія Залужного. Одразу необхідно зауважити, що персонаж на відео (нібито Залужний) розпочинає свою промову без привітання, з фрази "у мене є точна інформація", - схоже на те, що первісний шматок відео було вирізано через те, що він погано вийшов, проте з перших кадрів можна помітити перший та головний доказ 100% deepfake: губи персонажа видніються набагато чіткіше, ніж все інше обличчя. Цей прийом був описаний мною в ДОДАТОК Б, на якому наочно демонструється процес створення такого відео. Конкретно на ньому, були використані латентні риси обличчя А (справжнього Залужного) та риси іншого обличчя, скоріш за все, людини, яка це озвучувала, або ж обличчя, яке також було згенероване нейромережею, після чого відбулося змішування зображень.

У цьому описі максимально наочно відображається логіка і принцип функціонування прегенеративних нейронних мереж. Детальне зображення невідповідності частини обличчя, а саме рота, представлено на ДОДАТОК Ж. Також необхідно зауважити, що на відео, людина, яка нібито є Залужним, здійснює різкі кивки головою в різні боки, що є неприродними для поведінки людини. Найбільш це помітно на проміжку відео 0:25-0:30, також необхідно зауважити, що в проміжку між 0:05 і 0:06 секундами ролика спостерігаються так звані "артефакти" - через недосконаліть нейронних мереж кодер і декодер

роблять тільки поверхневу обробку відео, при цьому не роблячи детального аналізу тіней і якості зображення, таким чином, в одному кадрі одне досить погане (480p), а в іншому (600p), при чому з'являється цей артефакт лише на частку секунди. При чому, ті складки шкіри і міміка "згладжується" - декодер визначає невідповідність, надсилає запит до енкодера, який, своєю чергою, застосовує згладжування зображення.

Варто зазначити, що наразі динамічні об'єкти в просторі досить важко змоделювати до якості, яку не відрізнити від оригіналу, і цей ролик не є винятком. Крім того, що фігура на відео неприродно пересуває головою, водночас торс практично залишається нерухомим, що вже зі 100% ймовірністю свідчить про фейковість цього ролика. Подібні артефакти з досить сильним згладжуванням і різницею якості обличчя також спостерігаються на 0:51-0:55 секундах відеоролика. Однак, окрім простої візуальної невідповідності, також необхідно зауважити те, що на цьому відеоролику не збігається аудіодоріжка. Через недосконаліть генеративних нейромереж, досі досить складним завданням залишається зіставлення міміки губ і голосу. Також варто відзначити й те, що на записі не чути жодних сторонніх шумів, він є дуже "камерним" для просторої зали в якій зазвичай відбуваються брифінги міністерства оборони. Особливо це помітно на 0:25-0:30 секундах, коли маріонетка повертає голову і голос мав би бути більш віддаленим і приглушеним, проте ми цього не спостерігаємо. Ба більше, також важливий і тембр та стиль подачі.

Хоча й копіювання голосу Залужного зроблено без помилок, варто зауважити, що темп мови однаковий. Наприклад між 0:11-0:25 секундами маріонетка говорить без зупинки. Що досить важко зробити людині, їй банально може забракнути повітря, і необхідно робити паузи для вдиху та видиху. Це навіть не схоже на читання з аркуша, що вже казати про те, що це відео подається як "заява" у прямому ефірі. Текст озвучений монотонно, без пауз.

Також необхідно розібрати і смислову частину відео: на ньому лялька використовує класичні тейки російської пропаганди на кшталт війни до останнього українця, геноциду українського населення своїм же урядом, водночас варто відзначити і те, що текст складний доволі просто, особливо впадає в око фраза про "здачу держави", - військове керівництво ніколи не аперує подібними термінами, також у негативному ключі згадуються й військові успіхи, які, як озвучено, були провальними.

Варто зауважити, що подібні заяви неможливо обговорювати поза контекстом оригінального, реального Залужного, який ані за своєю риторикою, ані за стилем поведінки не мав ані найменшого шансу сказати щось подібне. Тут необхідно згадати дослідження Крістіана Вассарі та Ендрю Чеддвіка, яке я наводив у першій главі цієї роботи, - очевидно, що у ворога, найімовірніше, не було на меті переконати будь-кого в правдивості даних висловлювань, скоріше, посіяти сум'яття й невпевненість, тим самими похитнув внутрішню стабільність держави, що воює.

5.2 Аналіз розповсюдження deepfake : основні способи розповсюдження, специфіка розповсюдження, реакція, протидія.

За даними порталу stopfake [53], цей відеоролик спочатку опублікували у фейсбуці на підконтрольних РФ акаунтах, після чого на них почали посилатися російські телеграм-канали, а потім розпочалася "ланцюгова реакція" переписів і пересилань багатьма пов'язаними з пропагандистською машиною РФ телеграм-каналами і "новинними" виданнями. Необхідно відзначити оперативну реакцію державного центру протидії дезінформації, який за 2 години (7 листопада, 23:08) [54] , опублікував викриття цього відеоролика, після чого ця інформація аналогічним чином спричинила "ланцюгову реакцію", і цей пост зібрав значну кількість переглядів (41.500 переглядів і 270 перепостів на момент 13.04.2024). Що свідчить не тільки про оперативну роботу з населенням, а й про те, що під час турбулентності ціна часу дуже висока.

Як я вже писав вище, у даному кейсі російські пропагандисти не планували зробити так, щоб українські глядачі повірили в цей обман, а скоріше "розмити" правду, зробити офіційні джерела інформації менш цінними в очах населення та використати безпрецедентний рівень довіри в очах населення до Валерія Федоровича для критики влади і підриву її легітимності. Адже, з точки зору риторики, яку транслює маріонетка на відео, можна зробити висновок про те, що ворог використовує тактику сепарації військового і політичного керівництва, тим самим, не намагаючись підірвати авторитет самого головнокомандувача, бо в поточних реаліях із рівнем підтримки Залужного це не можливо, а здійснити удар по верховному керівництву, яке, аналогічно, проросійські пропагандисти часто звинувачують у непопулярних рішеннях. Необхідно зауважити, що ціна часу під час протидії досі є ключовою під час розповсюдження чутливої інформації для воюючої країни. Подібним питанням переймаються і японські дослідники, так група вчених з Токійського Інституту Технологій у своїй роботі "The Emergence of Deepfakes and its Societal Implications: A Systematic Review" аналізує останні наукові праці на дану тему [55]

Дослідження було спрямоване на вивчення соціальних наслідків технологій створення deepfakes і штучного інтелекту. Методика включала пошук статей у популярних наукових базах даних, таких як Springer, IEEE Xplore, ACM, Web of Science і Scopus, а також у Google Scholar для включення попередніх і не рецензованих публікацій. Спочатку було витягнуто 787 статей, з яких після ручної фільтрації та аналізу було відібрано 88 статей для подальшого аналізу.

Дослідження розділилося на два основних питання:

1) Які види досліджень проводять у рамках теми deepfakes?
2) Як розподілені дослідження щодо психологічних і соціальних наслідків deepfakes ?

Було виявлено такі типи досліджень: систематичні огляди, літературні огляди, філософські дослідження, експериментальні дослідження, аналіз мереж, аналіз змісту, концептуальні пропозиції та коментарі. Більшість статей (30 із 88) зосереджені на критичних оглядах попередньої літератури, а 21 стаття містила активні експерименти з реальними користувачами для вивчення соціальних і психологічних аспектів сприйняття deepfakes. Основні категорії досліджень ґрунтувалися на методологічних підходах і областях фокусування, таких як безпека, правові питання, психологічні аспекти, медійні та політичні перспективи. Особлива увага приділялася соціальним наслідкам deepfakes, включно із загрозами безпеці та етичними дилемами. Використання програмного забезпечення Gephi для аналізу мереж дало змогу оцінити зв'язки між авторами та статтями, підкреслюючи малу зв'язність дослідників у цій галузі, що вказує на необхідність більш інтегрованого та міждисциплінарного підходу до дослідження соціальних наслідків дипфейків. Насамкінець, результати показують, що соціальна наука про дипфейки перебуває на стадії формування і потребує глибшого аналізу та співпраці в рамках різних наукових дисциплін для повного розуміння можливих загроз і рішень.

Насамкінець, результати показують, що соціальна наука про deepfakes перебуває на стадії формування і потребує глибшого аналізу та співпраці в межах різних наукових дисциплін для повного розуміння можливих загроз і рішень.

Однак, необхідно зазначити, що на тему протидії deepfakes у соціальній сфері практично немає інформації через новизну теми. Науковому співтовариству ще необхідно буде детально дослідити й описати механізми впливу deepfakes, так само як і механізми протидії. Так, наприклад, Міжнародний журнал досліджень у галузі прикладних наук та інженерних технологій на своїй щорічній конференції поставив собі це питання, однак, окрім дослідження впливу на масову аудиторію, він також задався питанням наявності конкретних стратегій і тактик для боротьби з deepfakes в масовому просторі.

Однак, аналогічним чином автор дослідження підсумував вищесказане практично повною відсутністю тактик і стратегій протидії цій загрозі в медійній площині[60, с. 10]

5.3 Розробка покрокового алгоритму для детекції deepfake

Таким чином, було наочно продемонстровано можливість детекції дипефейків без використання спеціалізованих інструментів. Хоча, необхідно зауважити, що подібних інструментів зараз існує безліч, як і нейромережі, які можуть їх створювати спеціалізовані платформи, а також великі видавництва, які створили власні відділи по боротьбі з дезінформацією та створили власні алгоритми, про що було написано в теоретичній частині роботи в главі "Детекція deepfakes". Тут необхідно пояснити, що хоча й існує безліч платформ для детекції, однак, найімовірніше, пересічний користувач, який побачив подібну фальшивку десь у телеграм-каналі, навряд чи використовуватиме спеціалізовані платформи для детекції. Це становить велику небезпеку для недосвідчених користувачів із високим рівнем довіри та відсутністю критичного мислення. Однак, як було продемонстровано вище, провести детекцію deepfake можна і "на око", без спеціального обладнання. Для цього, нижче, я описав власний алгоритм детекції deepfake, яким може скористатися кожен охочий :

1) Невідповідність губ і мови:

Зверніть увагу на синхронізацію руху губ із вимовленими словами. У deepfake відео часто присутня затримка або невідповідність між видимим рухом губ і аудіодоріжкою, оскільки точно синхронізувати фонеми і міміку складно.

2) Артефакти зображення і якість відео:

Шукайте раптові "артефакти" або незвичайні зміни в якості зображення, особливо навколо очей, рота і країв обличчя. Це можуть бути розмиті ділянки, дивні лінії або безперервні патерни. Також зверніть увагу на різкі перепади

якості відео, коли одні ділянки відео мають гірший або кращий вигляд, ніж інші.

3) Нестандартні рухи та міміка:

Підозріло виглядають неприродні або перебільшені рухи, особливо голови та обличчя. Наприклад, якщо голова персонажа робить різкі або механічні рухи, які не відповідають нормальній поведінці людини.

4) Невідповідність фону та освітлення:

Фон і освітлення можуть виглядати неприродно, якщо вони не збігаються з освітленням або перспективою суб'єкта. Наприклад, тіні на обличчі можуть не відповідати іншим об'єктам у кадрі.

5) Звукові аномалії:

Слухайте аномалії в звуці, як-от відсутність фонових шумів, які мають бути чутні, або неприродно чистий звук в умовах, де очікуються перешкоди або відлуння.

6) Зміст і контекст мовлення:

Аналізуйте логіку та зміст висловлювань. Deepfake часто використовують для поширення дезінформації або створення провокаційних висловлювань, які можуть бути нехарактерними для даної людини. Нестиковки в мові, незвичайний тембр або стиль подачі також можуть бути індикаторами підробки.

Висновок до розділу 5

Висновок з аналізу deepfake з Валерієм Залужним підкреслює серйозні виклики та загрози, з якими стикається сучасне інформаційне суспільство в епоху розвитку технологій штучного інтелекту. Відеоролик, створений з використанням технології deepfake і приписаний головнокомандувачу ЗСУ Валерію Залужному, є яскравим прикладом маніпулятивного впливу на громадську думку і спроби дискредитації українського військового і політичного керівництва з боку проросійських сил.

Технічний аналіз цього відео показує характерні ознаки використання генеративних нейронних мереж, такі як неприродна міміка, порушення в синхронізації руху губ і голосу, а також відсутність природних шумів і фонових звуків, що свідчить про високу якість імітації, але водночас і про її фальшивість. Ці технічні особливості, виявлені під час аналізу, наочно демонструють здібності сучасних технологій до створення переконливого, але неправдивого контенту. Зміст відео, що містить типові ознаки російської пропаганди, також вказує на мету його творців - не просто ввести в оману глядачів, а скоріше "розмити" межі між правдою і брехнею, знижуючи довіру до офіційних джерел інформації та посилюючи внутрішні протиріччя в українському суспільстві. Це відображає стратегію ворожих держав використовувати інформаційні операції для підриву єдності та стабільності національної безпеки України. Заходи у відповідь, включно з оперативним оприлюдненням інформації про фальсифікацію та широким інформуванням через центри протидії дезінформації, показують ефективність українських механізмів реагування на інформаційні загрози.

Однак цей випадок підкреслює необхідність подальшого посилення міжнародного співробітництва та розробки більш просунутих технологічних і правових інструментів для протидії подібного роду загрозам. Таким чином, deepfake з Валерієм Залужним стає важливим попередженням про те, як

технології можуть бути використані для ведення гібридної війни, підкреслюючи необхідність комплексного підходу до забезпечення інформаційної безпеки на національному та міжнародному рівнях.

6 СТВОРЕННЯ ВЛАСНОГО ГОЛОСОВОГО DEEPFAKE У МУЗИЧНОМУ ЖАНРІ НЕЙРО-КАВЕРУ

Цей практичний розділ присвячений другій частині мого практичного завдання, в цьому розділі буде описана технологія створення голосового deepfake з використанням одного з популярних сервісів.

Штучний інтелект значно прискорює створення аудіо-deepfake, що викликає тривогу у сферах від політики до фінансових шахрайств. Федеральний уряд США заборонив використання рободзвінків із голосами, що генеруються штучним інтелектом, і пропонує грошову винагороду за рішення, що зменшують шкоду від шахрайства з клонуванням голосу. Науковці та приватний сектор активно працюють над розробкою програмного забезпечення для виявлення клонованих голосів, яке часто маркетують як інструменти для виявлення шахрайства. Створення deepfake може коштувати всього кілька доларів і зайняти 8 хвилин [47] . Однак програмне забезпечення для виявлення часто помиляється, що може мати серйозні наслідки.

Наприклад, якщо реальний аудіозапис буде помічений як фальшивий у політичному контексті, це може призвести до втрати довіри до всього, що ми чуємо. Технологічні рішення для виявлення AI-генерованих голосів не є ідеальними. На основі експерименту NPR, було виявлено, що програмне забезпечення часто не може ідентифікувати AI-генеровані кліпи або помилково визначає реальні голоси як такі, що генеруються AI. Наприклад, інструмент компанії Pindrop Security [47] правильно ідентифікував майже всі зразки, тоді як інструмент компанії AI or Not помилився приблизно в половині випадків.

Технології виявлення зазвичай включають навчання моделей машинного навчання, які аналізують зразки реального і фальшивого аудіо, перетворюючи їх на цифрові дані для комп'ютерного аналізу. Ці моделі можуть виявляти різниці, непомітні для людського вуха.

За словами Сари Баррінгтон, дослідниці з AI та форенсики з Університету Каліфорнії в Берклі, важливо, щоб розробники deepfakes і детекторів працювали разом над поліпшенням технологій виявлення [47], але наразі технологія для генерації голосових deepfakes значно випереджає технологію для детекції. Це пояснюється також і фінансовим інтересом: більша кількість ресурсів іде на створення сервісів, які згодом будуть монетизовані споживачем, ніж ПЗ для детекції подібних deepfakes. Однак, деякі компанії вже почали маркувати аудіотвори, що створені за допомогою їхніх сервісів спеціальними позначками, але, разом з цим, ніхто не заважає простому користувачеві використати інший сервіс або створити його самостійно.

6.1 Загальні відомості про технологію Audio Synthetics

З моменту написання основної частини роботи минуло доволі багато часу в рамках розвитку сучасних технологій, за цей час встиг з'явитися і розвинутися жанр так званих нейро-каверів (ai cover), тобто, відомі виконавці виконують пісні у своєму фірмовому стилі, але з творами, які вони ніколи не виконували. Наприклад, відомі американський виконавець Френк Сінатра виконав пісню гурту Nirvana 'Smells like a teen spirit' [48], при цьому зібравши значні перегляди на відеоролику (понад 700 тисяч переглядів на момент написання), що може свідчити про неабиякий інтерес до даної тематики у користувачів. Хіба ніхто ніколи не мріяв, щоб його улюблену пісню хтось виконав в іншому жанрі, при цьому зробив це професійно і це можна було б послухати зі своєї оселі не відвідуючи концерти. Саме цей запит і намагаються задовольнити нейромережі, які спеціально були створені для подібних цілей, наприклад, одна з них suno.ai . Базуючись на останніх досягненнях у галузі машинного навчання та штучного інтелекту, ця технологія пропонує унікальний спосіб створення музичного контенту. Suno.ai аналізує безліч музичних жанрів і стилів, щоб створювати пісні, які не тільки звучать природно, а й здатні передати певні емоції та настрої. Використовуючи складні алгоритми обробки природної мови, Suno.ai може генерувати як

мелодію, так і текст пісні, роблячи кожне творіння унікальним. Це відкриває нові можливості для музикантів, продюсерів і всіх, хто зацікавлений у музичній творчості, надаючи інструмент для експериментів і досліджень у музичній індустрії.

Suno.ai також може слугувати освітнім ресурсом, допомагаючи вивчати музичне композиційне мистецтво. Завдяки можливості миттєвого зворотного зв'язку, користувачі можуть покращувати свої навички в написанні пісень, експериментуючи з різними музичними формами і структурами. Нейромережа не тільки прискорює процес створення музики, а й пропонує нові способи натхнення для творчості в музичному світі. Так, наприклад британський таблод 'The Guardian' [49], описує suno.ai як аналог ChatGPT для музики, що є новітнім розвитком у галузі генеративного штучного інтелекту.

Ця система дозволяє користувачам ввести музичний стиль, жанр і текст для генерації повної пісні за лічені секунди. Розроблена групою фахівців з машинного навчання з Кембриджа, вона вже два роки радує користувачів, хоча тексти пісень іноді можуть здаватися поверхневими і позбавленими душі. Suno AI дає змогу створювати до 10 пісень на день на безоплатній основі, а за 10 доларів на місяць користувачі можуть генерувати до 500 пісень і навіть монетизувати їх. Компанія заявляє, що мета Suno не в тому, щоб замінити артистів, а в тому, щоб зробити створення музики доступнішим і цікавішим.

6.2 Технологія Audio Synthetics: інновація, яка поки не регулюється авторським правом (На прикладі Suno.ai)

Однак виникають питання щодо використання авторських матеріалів і впливу AI на музичну індустрію. Відомі артисти висловили занепокоєння з приводу потенційної шкоди від AI-твореної музики для індустрії. Suno AI включає непомітні водяні знаки в кожну пісню, щоб можна було

ідентифікувати AI-генеровану музику. Через новизну цього жанру досить важко дати точну оцінку цьому явищу, а саме яким чином оцінити соціальний ефект і виробити правові норми для взаємодії з подібним контентом на великих майданчиках, наприклад, відомий майданчик Spotify схиляється до дозволу такого типу контенту, його точно так само можна буде слухати, під це буде окремий жанр (ai cover), однак досі відбуваються запеклі обговорення цього питання, деякі автори, все-таки, серйозно вважають, що це, швидше за все, негативно позначиться на індустрії [50]. В інтерв'ю генеральний директор компанії Даніель Ек повідомив, що Spotify не має наміру видаляти всі AI-генеровані матеріали, а тільки ті пісні, які імітують реальних виконавців без їхньої згоди. Проте на платформі збережеться місце для пісень, натхненних артистами або повністю створених за допомогою AI, за умови, що вони не стверджують, що є творами конкретних виконавців. Крім того, Spotify запустив функцію для подкастів, що дає змогу перекладати записи англійською мовою іншими мовами, зберігаючи при цьому оригінальний голос ведучого. Цю функцію було реалізовано з використанням новітніх розробок OpenAI у сфері генерації голосу. Поки що подкасти кількох відомих ведучих доступні іспанською мовою, а незабаром очікується додавання епізодів французькою та німецькою. Spotify вважає, що продуманий підхід до використання штучного інтелекту може поглибити зв'язок між слухачами і творцями контенту, що є ключовим аспектом місії компанії.

Однак, як я писав вище, юридичні наслідки і статус подібних творів залишаються незрозумілими, наприклад, як пише про це видання daily.dev [51], говорить про те, що подібні інструменти сприятимуть демократизації індустрії, позитивно позначаться на розмаїтті музичного наповнення в кіно-мистецтві, проте юридичні питання зависають у повітрі, бо незрозуміло, на яких творах нейромережа "навчалася", і як саме можна формалізувати авторське право в цій царині.

Suno дає змогу людям без спеціальних музичних знань створювати власні композиції, використовуючи лише комп'ютер та інтернет. Це відкриває музичну творчість для широкої аудиторії і дає змогу ділитися унікальними звуками та ідеями. Професійні музиканти також можуть використовувати Suno для нових форм співробітництва та створення музики, проте дехто побоюється, що Suno може скоротити кількість робочих місць для композиторів і музикантів. Однак інші вважають, що Suno може доповнювати роботу музичних професіоналів, а не замінювати їх. Вартість музики для реклами і кіно може знизитися, якщо AI-музика виявиться дешевшою. Однак доходи від концертів та інших живих виступів, ймовірно, не зміняться значно. Suno планує поліпшити інтерактивність створення музики, розширити музичний діапазон і стилі, а також запропонувати інтуїтивно зрозумілі інтерфейси для користувачів без музичної освіти. Крім того,

Suno працює над можливістю комерційного ліцензування музики, створеної за допомогою AI, забезпечуючи дотримання нових правил щодо контенту, створеного штучним інтелектом. Suno прагне зробити створення AI-музики доступним і корисним для всіх, вбачаючи в AI не тільки інструмент, а й партнера в музичній творчості.

6.3 Питання авторства та захисту контенту при застосуванні інструменту Audio Synthetics (на прикладі Suno.ai)

Питання авторства в контексті використання штучного інтелекту (ШІ) в музичній творчості посідає особливе місце у філософських та етичних дискусіях. Традиційне розуміння авторства передбачає наявність людського внеску, оригінальності ідей та індивідуального творчого процесу. Однак, коли музику створює ШІ, ці аспекти ставляться під сумнів, оскільки ШІ не володіє свідомістю, емоціями або здатністю до самовираження у звичному для нас сенсі. З одного боку, можна стверджувати, що справжні творці в разі

використання ШІ - це розробники програмного забезпечення та алгоритмів. Ці люди створюють умови, в рамках яких ШІ генерує музику, тому їх можна розглядати як авторів. Однак у такому разі авторство стає результатом колективної розумової праці, а не індивідуального творчого акту.

З іншого боку, порушується питання про переосмислення самої природи авторства і творчості в еру цифрових технологій. Можливо, слід запровадити нову категорію "авторства", яка враховуватиме внесок як людини, так і машини. Така зміна може включати визнання інструментальної ролі ШІ як "ко-автора", який, хоч і не має власної свідомості, відіграє ключову роль у творчому процесі. Це також передбачає новий погляд на правові та етичні аспекти інтелектуальної власності та розподілу визнання і доходів від створених творів.

Проблема автентичності проявляється особливо гостро, коли ШІ створює твори, що стилістично або вокально імітують відомих виконавців. Це створює ризик введення в оману аудиторії, яка може вважати, що твір створено людиною, а не машиною. Етичні дилеми такого підходу полягають у можливій втраті довіри до мистецтва, коли незрозуміло, чия творчість - людини чи ШІ - стоїть за конкретним твором. З точки зору етики, важливо забезпечити чесність і прозорість у використанні ШІ в мистецтві. Якщо використання ШІ буде ясно позначено, то такий підхід може бути визнаний не обманом, а новою формою мистецтва. Це дасть змогу аудиторії усвідомлено оцінювати і сприймати ШІ-генеровану творчість, відкриваючи нові кордони для експериментів та інновацій у мистецтві.

Наприклад, деякі дослідники серйозно займаються створенням захисту від нейромереж, як-от робота Нані Ванг, Х'ю Ву, Фенг Ху, Леї Ченг у своїй роботі [52], яка описує взаємодію штучного інтелекту та музичної творчості, а також зачіпає важливі аспекти захисту авторських прав у цифрову епоху. Основний фокус робиться на використанні комп'ютерних алгоритмів для створення музики та методах захисту музичних творів за допомогою

блокчейн-технологій. ШІ та машинне навчання застосовуються для автоматизації процесу створення музики. Описується, як ШІ аналізує і створює музику на основі наявних даних, не вимагаючи при цьому втручання людини. Захист авторських прав з використанням блокчейн: Запропоновано методи захисту музичних творів через блокчейн, включно зі створенням незмінюваного запису оригінального володіння та використанням "розумних" контрактів для ліцензування музики. Це дає змогу забезпечити безпечне і законне використання музичних файлів. Автори також наголошують на наявних проблемах у системі управління цифровими музичними правами, таких як централізоване управління і незбалансований розподіл вигод, що веде до частих суперечок щодо прав і порушень.

6.4 Створення власного голосового deepfake за допомогою Audio Syntetics (на прикладі Suno.ai)

В даному підрозділі я вирішив використати нейромережу suno.ai, яка зараз перебуває на хвилі популярності та спробувати створити нейро-кавер на відому пісню чи твір. Переді мною одразу став вибір між популярними українськими композиторами та поетами, проте я вирішив створити кавер у стилі джазу на, напевно відомий кожному українцю твір Тараса Григоровича Шевченка "Реве та стогне Дніпр широкий" , для того, щоб створити якийсь зручний результат, мені не знадобилося здійснювати велику кількість дій, всього-то :

1. Зайти на вебсайт платформи suno.ai

2. авторизуватися за допомогою облікового запису Google

3. Після цих дій я мав повний доступ до створення нейрокаверів або до створення повноцінних творів, які були згенеровані нейромережею. Варто зауважити, що базовий функціонал надається безоплатно, користувачеві

доступне створення 5-ти творів щоденно на один акаунт, тобто, якщо ви маєте кілька акаунтів, то це число подвоюється.

Інтерфейс додатка зображено на ДОДАТОК З

Стартова сторінка додатка - домашня, з неї можна переглянути твори, що зараз перебувають у тренді (їх аналогічно створено або змінено нейромережею), після чого видно розбиття на різні категорії популярних уже створених нейро-каверів або повноцінних творів (блюз, фонк, реп, джазз, опера, класика та інше).

Перейшовши на вкладку Create (створення) користувач може створити власну композицію, при цьому використовувати власний придуманий або існуючий текст пісні, також можна завантажити абсолютно довільний текст.

Після чого необхідно написати стиль музики. Можна використовувати оперний, класичний, реп, поп та інші стандартні стилі. Однак можна і використовувати більш нестандартні жанри на кшталт євродиско, електронної музики та інше. Однак, важливо зауважити, що творці нейромережі убезпечили себе від прямого копіювання авторів, бо більшість сучасних і трендових пісень оподатковуються авторським правом, неможливо створити стиль повністю аналогічний конкретному автору, бо нейромережа відмовиться його відтворювати і виведе помилку. Музичний твір може лише бути створено в якомусь стилі, що не можна назвати прямим копіюванням з точки зору законодавства. Наприклад, твір, який було прикріплено до цієї роботи, створено за стилем Френка Сінатри, американський джазз 40-50-х років. Після чого необхідно дочекатися закінчення генерації твору і прослухати вподобаний. Цей твір прикріплено до роботи ДОДАТОК І.

Висновок до розділу 6

Таким чином наочно було показано можливість генерації твору в стилі нейро-каверу, здібності, переваги та недоліки сучасних алгоритмів створення

голосових deepfakes. Необхідно зауважити, що подібна технологія має неабияку перспективу і в наступні кілька років трансформується ще більш сильно. Однак, необхідно уточнити, що разом із розважальною (рекреаційною) функцією також є варіанти використання, що виходитимуть за рамки моралі та етичних норм. Маючи технологію - алгоритм, здатний генерувати людський голос, у такій якості його можна використовувати для створення фальшивок з політичними діячами, чиновниками, популярними артистами, ведучими, що, звісно, зі свого боку не може не викликати побоювання і настороженості. Однак, людству все ще належить усвідомити наслідки такого швидкого ривка до масового використання штучного інтелекту в спільному гуртожитку і в житті кожного з нас.

Після прослуховування твору, створеного штучним інтелектом. Відтворення стилю Френка Сінатри в контексті джазової манери 40-50-х років було виконане з вражаючою точністю, при цьому уникаючи прямого копіювання його робіт, що могло б становити порушення авторських прав. Цей досвід виявив значні інноваційні та творчі можливості технологій у музичній сфері.

Застосування платформи для створення музики виявилося інтуїтивно зрозумілим і доступним, що робить її прекрасним інструментом для музичних ентузіастів, зацікавлених у експериментуванні та створенні унікальних композицій без необхідності глибоких знань у сфері музичної композиції. Відкриття нових горизонтів для творчої самореалізації може істотно змінити уявлення про процес створення музики.

Однак, слід відзначити, що використання штучного інтелекту для створення музики викликає серйозні питання етичного та правового характеру. Зокрема, платформа, з якою проводився експеримент, забезпечує дотримання авторських прав, запобігаючи створенню творів, які могли б порушити виключні права оригінальних творців. Це підкреслює необхідність

знаходження балансу між інноваційним використанням технологій і повагою до існуючих творчих традицій та їх захистом.

Експеримент із платформою suno.ai демонструє потенціал сучасних технологій у музичній індустрії для розширення творчих кордонів. Однак, важливо продовжувати діалог про те, як ці інструменти слід використовувати для забезпечення справедливого та етичного використання музичного контенту в цифрову епоху, забезпечуючи, щоб інновації сприяли розширенню мистецьких можливостей, не ставлячи під загрозу моральні та правові норми.

ВИСНОВКИ ДО РОБОТИ

Ця робота стане, хоч і доволі маленькою, але все одно підмогою в подальшому вивченні глибоких фейків та їхнього впливу на навколишню соціальну дійсність. Проаналізувавши матеріали до цієї роботи, я дійшов чіткого висновку, що проблему deepfake треба виносити не тільки на рівень громадського обговорення, а й юридичного та глибшого вивчення, а саме того, як вони впливають на соціальну сферу. Не можна сказати, що deepfake є чимось принципово новим, з подібними викликами людство стикалося впродовж усієї історії: від цензури та переписування історії, до банального замовчування, ретушування фотографій, як це, наприклад, було за часів репресій у 30-х роках минулого століття. Однак, deepfakes страшні зовсім не цим, а тим, що дана проблема рефлектується і продовжить загострюватися.

Навіть зараз, приклади, які я навів у своїй роботі, можуть дати чітке розуміння того, що хоча зараз deepfakes не мають гнітючого або бодай значного впливу, досі, зокрема і під час воєнних конфліктів, використовується вже давно відома пропаганда і класичні прийоми, що були описані ще Едвардом Бернсом у своєму magnum opus – "Пропаганда". Однак, буквально за кілька років deepfakes можуть похитнути першість серед інструментів пропаганди. Наприклад, у США вже активно в передвиборчій кампанії різних політичних сил, про що йшлося в розділі 4, використовуються deepfakes. Найчастіше це відбувається з метою дискредитації опонента, наприклад, для того, щоб звинуватити його в педофілії або в державній змові, як це було з Трампом у першому випадку і Байденом у другому. Якщо раніше для того, щоб створити фальшивку, наприклад, використовуючи photoshop, були потрібні навички графічного дизайну та роботи в застосунку, то тепер подібний фейк може створити практично кожен. І та якість, яку можуть надати сучасне програмне забезпечення для їхнього створення, справді викликає занепокоєння. Масове використання глибоких фейків може призвести до проблем довіри в суспільстві, хоча, як видно з дослідження Чедвіка, що було

описано в Розділі 1.Deepfakes, хоча й не викликають довіри у більшості респондентів, але все одно залишають певний відбиток на підсвідомій сфері. Звісно, констатувати те, що deepfakes працюють за принципом "чарівної кулі" Лассуела - неправда, однак, так само не можна заперечувати, що постійний перегляд таких підробок, тим паче, коли їх надають без жодних пояснень аудиторії, слабкій в інтелектуальному плані, може мати вибуховий ефект і може вплинути не лише на вибори й електоральні процеси, а й сприяти розпалюванню міжнаціональної, релігійної та расової ворожнечі й дискримінації. Однак, також варто зауважити, що deepfakes так само дають неоціненні можливості для різних галузей, таких як кінематографія, медицина, освіта та інші.

Далі я хотів би додати наступні пункти висновків до моєї роботи:

- Юридичний аспект: Глибокі фейки вимагають розвитку нових юридичних норм та регуляцій, які могли б адекватно реагувати на виклики, що виникають із їх використанням. Необхідно розробити законодавчі акти, які визначатимуть відповідальність за створення та розповсюдження deepfakes, що мають на меті введення в оману або завдають шкоди окремим особам чи групам населення.
- Соціальний вплив: Важливо зрозуміти, як deepfakes впливають на соціальну довіру та міжособистісні відносини. Зростання кількості та якості таких фейків може призвести до загальної недовіри до відео- та аудіоматеріалів, що в свою чергу може послабити соціальні зв'язки та ускладнити комунікацію.
- Етичний аспект: Використання deepfakes у розважальних цілях, наприклад, у кіно чи рекламі, ставить питання про етичні межі їх застосування. Необхідно визначити, де проходить межа між креативним використанням технологій та маніпуляцією, яка може завдати шкоди.
- Освітні програми: Розробка освітніх програм, спрямованих на підвищення цифрової грамотності населення, є необхідним кроком для

протидії маніпулятивним впливам deepfakes. Це включає навчання розпізнаванню фейкових матеріалів та розвиток критичного мислення.

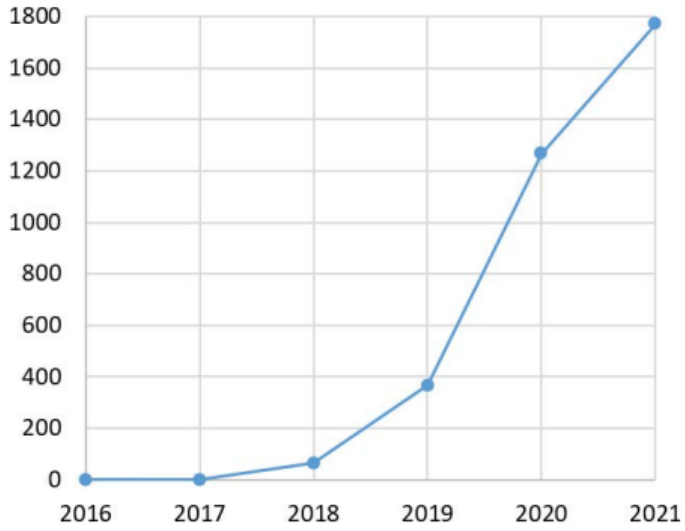
- Технологічний розвиток: Паралельно з розвитком технологій для створення deepfakes повинні розвиватися і технології для їх виявлення. Інвестиції в дослідження та розробку таких технологій є критично важливими для забезпечення інформаційної безпеки.

Для подальших досліджень:

Ця робота може стати основою для подальших досліджень у галузі впливу deepfakes на суспільство. Зокрема, варто поглиблено досліджувати юридичні аспекти їх використання та розробляти ефективні стратегії захисту від маніпуляцій, а також удосконалювати технології для автоматичного виявлення deepfakes. Такий підхід дозволить створити комплексну систему протидії інформаційним загрозам, пов'язаним із використанням глибоких фейків, і підвищить загальний рівень інформаційної безпеки в суспільстві. Але технологія deepfakes зараз знаходиться лише у стадії осмислення та концептуалізації як у науковому світі, так і не до кінця зрозумілі наслідки їх використання і розповсюдження, а засоби контролю та протидії ще потрібно розроблювати та систематизувати.

ДОДАТОК А

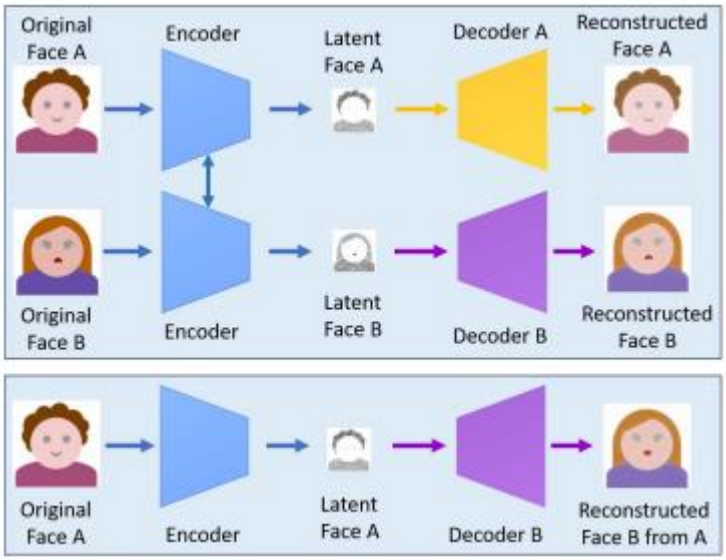
Графік , що відображає кількість публікацій по темі «deepfakes»



Зображення зверху демонструє кількість публікацій, пов'язаних із "глибокими фейками", за період з 2016 по 2021 роки, отримана з сайту https://app.dimensions.ai наприкінці 2021 року за ключовим словом "deepfake", застосованим до повних текстів наукових статей.

ДОДАТОК Б

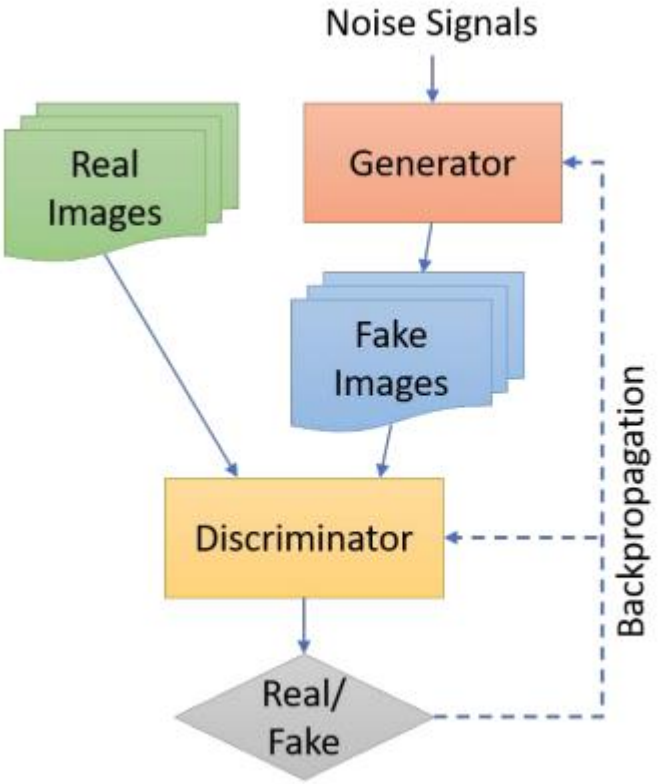
Демонстрація моделі для створення deepfakes



Зображення зверху демонструє модель створення глибоких підробок з використанням двох пар кодер-декодер. Дві мережі використовують один і той самий кодер, але різні декодери для навчання (вгорі). Зображення обличчя А кодується загальним кодеру і декодується декодером В для створення глибокої підробки (внизу). Реконструйоване зображення (внизу) - це обличчя В з ротом А. Обличчя В спочатку має рот у вигляді перевернутого серця, а реконструйоване зображення серця, тоді як реконструйоване обличчя Б має рот у формі звичайного серця.

ДОДАТОК В

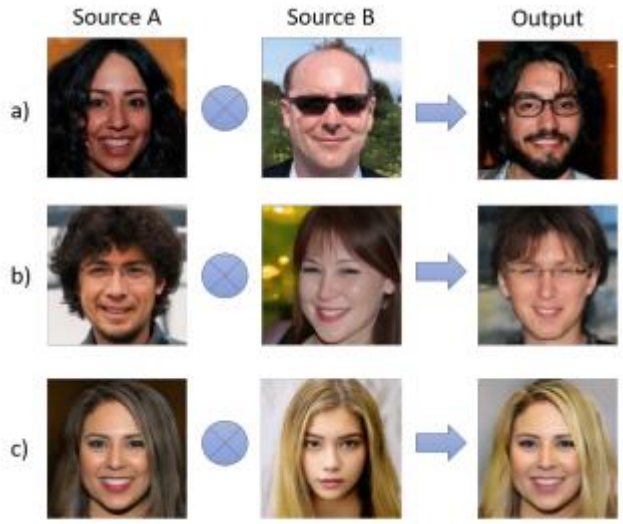
Зображення архітектури GAN нейромережі



Зображення зверху демонструє архітектуру GAN, яка складається з генератора і дискримінатора, кожен з яких може бути реалізований нейронною мережею.

ДОДАТОК Г

Демонстрація того, яким чином нейронна мережа «навчається»



Зображення зверху демонструє , що нейромережа може навчатися за допомогою зворотного поширення, що дозволяє обом мережам покращувати свої можливості. Приклади змішування стилів за допомогою StyleGAN: вихідні зображення згенеровані шляхом копіювання заданої підмножини стилів з джерела B, а решту - з джерела A.

a) Копіювання грубих стилів з джерела B призведе до створення зображень, які матимуть високорівневі аспекти з джерела B, а всі кольори та дрібні риси обличчя - з джерела A;

b) якщо скопіювати стилі середньої роздільної здатності з джерела B, вихідні зображення матимуть дрібніші риси обличчя з B і збережуть позу, загальну форму обличчя та окуляри з A; c) якщо скопіювати стилі дрібної роздільної здатності з джерела B, згенеровані зображення матимуть кольорову гаму та мікроструктуру джерела B

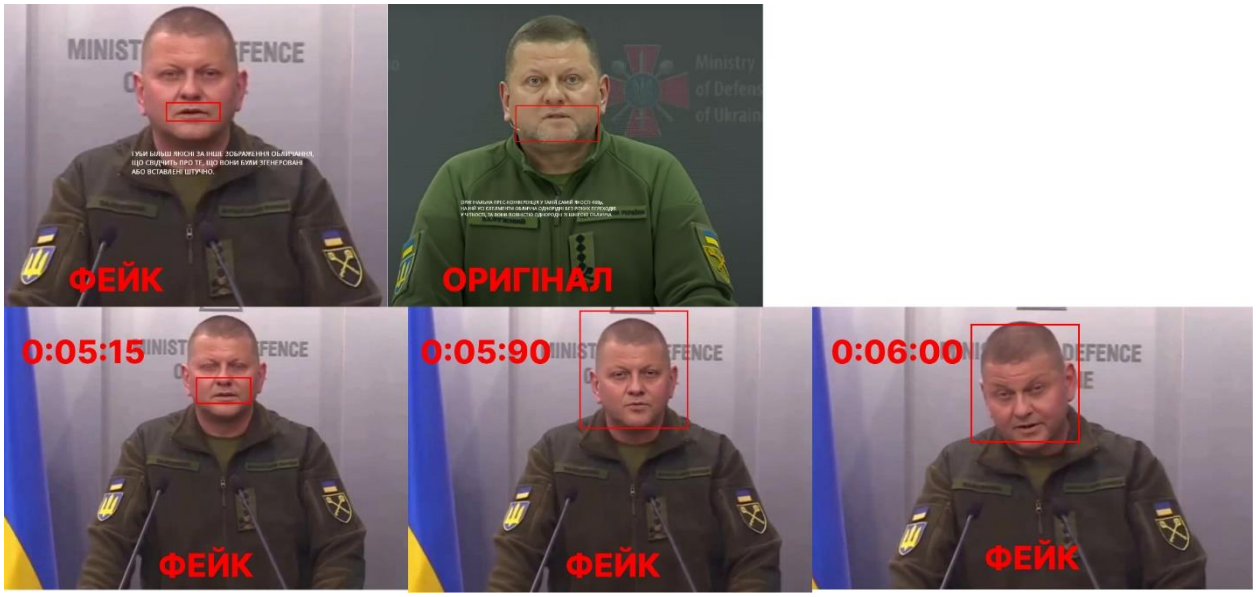
ДОДАТОК Д

Тези до наукової конференції у збірнику доступні за цим посиланням :

https://archive.liga.science/index.php/conference-proceedings/issue/view/ukr-15.12.2023/57 , сторінка 440.

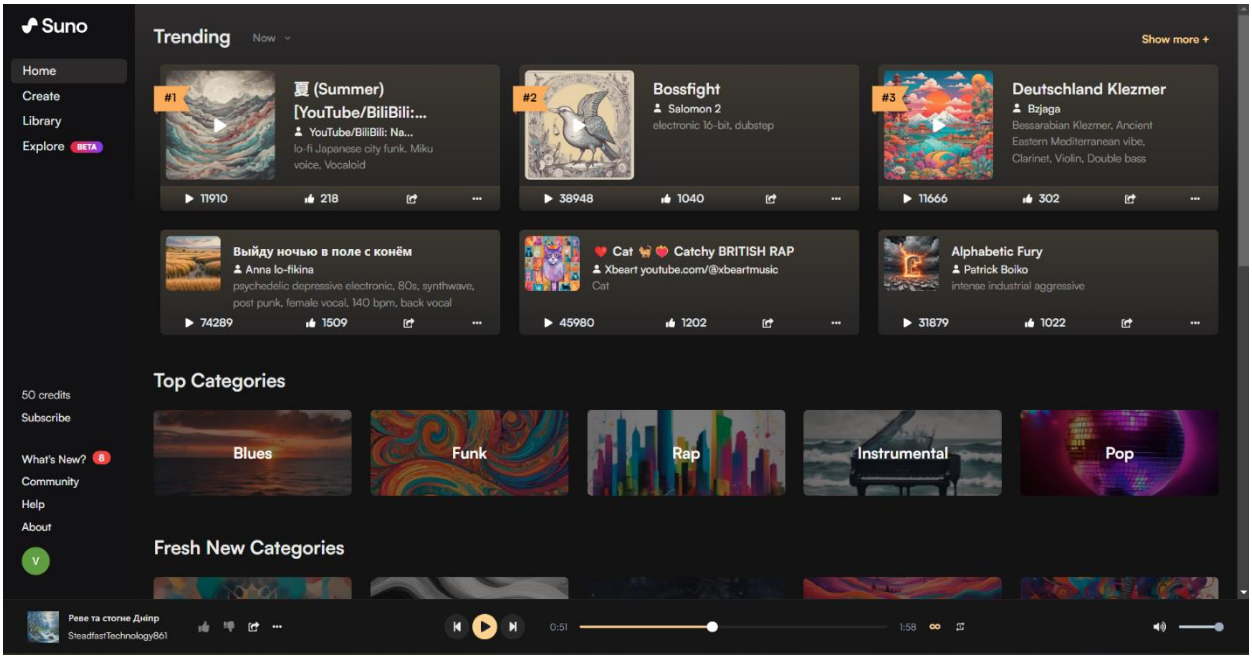
ДОДАТОК Ж

Фрагменти відеоролику, що демонструють невідповідність рис обличчя.



ДОДАТОК З

Зображення користувацького інтерфейсу застосунку suno.ai



ДОДАТОК І

Музичний твір, що був створений завдяки застосунку suno.ai

Реве та стогне
Дніпр широкий,.mp

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1) Agarwal S., Farid H. Protecting World Leaders Against Deep Fakes. *CVPR Workshop Paper*. 2019. С. 5. URL: https://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf (дата звернення: 05.05.2024) .

2) Lyu S. Detecting ‘deepfake’ videos in the blink of an eye. *The Conversation*. URL: https://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072 (дата звернення: 05.05.2024).

3) Homeland Security of USA. Increasing Threat of deepfake Identities. *https://www.dhs.gov*. URL: https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf (дата звернення: 05.05.2024).

4) Deepfakes in the 2024 US Presidential Election. *Hany Farid*. URL: https://farid.berkeley.edu/deepfakes2024election/ (дата звернення: 05.05.2024).

5) Andrii Baumeister. Почему ИИ не сможет хакнуть человека?, 2023. *YouTube*. URL: https://www.youtube.com/watch?v=XKT0_QfUays (дата звернення: 05.05.2024).

6) Andrii Baumeister. Имеет ли машина душу? Как искусственному интеллекту пробудиться к жизни?, 2023. *YouTube*. URL: https://www.youtube.com/watch?v=iobc_I3nmj4 (дата звернення: 05.05.2024).

7) Vaccari C., Chadwick A. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*. 2020. Т. 6, № 1. С. 205630512090340. URL: https://doi.org/10.1177/2056305120903408 (дата звернення: 05.05.2024).

8) «Залужний заявив про держпереворот і пригрозив Зеленському». ЩО? Та звісно, ні! Просто росіяни запустили два діпфейки з головкомом, а ми їх розібрали. Штучний і власний інтелект у поміч. *Бабель | Розповідаємо про*

політику, культуру і суспільство в Україні. Останні новини детально і неупереджено. URL: https://babel.ua/texts/100408-zaluzhniy-zayaviv-pro-derzhperevorot-i-prigroziv-zelenskomu-shcho-ta-zvisno-ni-prosto-rosiyani-zapustili-dva-dipfeyki-z-golovkomom-a-mi-jih-rozibrali-shtuchniy-i-vlasniy-intelekt-nam-u-pomich (дата звернення: 05.05.2024).

9) Мирончук Р. Росія готує фейкове "відео з Зеленським". Розвідка попереджає українців. РБК-Украина. URL: https://www.rbc.ua/ukr/news/rossiya-gotovit-feykovoe-video-zelenskim-1646278231.html (дата звернення: 05.05.2024).

10) ШІ почав створювати нюдси та діпфейки: що робити, якщо ви постраждали від цього. *Рубрика*. URL: https://rubryka.com/article/shi-pochav-stvoryuvaty-nyudsy/ (дата звернення: 05.05.2024).

11) Punnappurath A., Brown M. S. Learning Raw Image Reconstruction-Aware Deep Image Compressors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020. Vol. 42, no. 4. P. 1013–1019. URL: https://doi.org/10.1109/tpami.2019.2903062 (дата звернення : 05.05.2024).

12) Guardian staff reporter. Chinese deepfake app Zao sparks privacy row after going viral. *the Guardian*. URL: https://www.theguardian.com/technology/2019/sep/02/chinese-face-swap-app-zao-triggers-privacy-fears-viral (дата звернення: 05.05.2024).

13) Lyu S. Detecting deepfakes: how can we ensure that generative AI is used for good?. *Futurum Careers*. 2024. URL: https://doi.org/10.33424/futurum481 (date of access: 05.05.2024).

14) Punnappurath A., Brown M. S. Learning Raw Image Reconstruction-Aware Deep Image Compressors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020. Vol. 42, no. 4. P. 1013–1019. URL: https://doi.org/10.1109/tpami.2019.2903062 (дата звернення : 05.05.2024).

15) GitHub - deepfakes/faceswap: Deepfakes Software For All. *GitHub*. URL: https://github.com/deepfakes/faceswap (дата звернення: 05.05.2024).

16) GitHub - dfaker/df: larger resolution face masked, weirdly warped, deepfake,. *GitHub*. URL: https://github.com/dfaker/df (дата звернення: 05.05.2024).

17 GitHub - StromWine/DeepFake_tf: Deepfake based on tensorflow. *GitHub*. URL: https://github.com/StromWine/DeepFake_tf (дата звернення: 05.05.2024).

18) Generative adversarial networks / I. Goodfellow та ін. *Communications of the ACM*. 2020. Т. 63, № 11. С. 139–144. URL: https://doi.org/10.1145/3422622 (дата звернення: 05.05.2024).

19 Facing reality? Law enforcement and the challenge of deepfakes | Europol. *Europol*. URL: https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes (дата звернення: 05.05.2024).

20) A Survey of Speaker Recognition: Fundamental Theories, Recognition Methods and Opportunities / M. M. Kabir та ін. *IEEE Access*. 2021. Т. 9. С. 79236–79263. URL: https://doi.org/10.1109/access.2021.3084299 (дата звернення: 05.05.2024).

21) Tucker P. The Newest AI-Enabled Weapon: ‘Deep-Faking’ Photos of the Earth. *Defense One*. URL: https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/ (дата звернення: 05.05.2024).

22) Few-Shot Adversarial Learning of Realistic Neural Talking Head Models. *arXiv.org*. URL: https://arxiv.org/abs/1905.08233 (date of access: 05.05.2024).

23) Damiani J. A Voice Deepfake Was Used To Scam A CEO Out Of $243,000. *Forbes*. URL: https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=2cda65782241 (дата звернення: 05.05.2024).

24) Samuel S. A guy made a deepfake app to turn photos of women into nudes. It didn’t go well. *Vox*. URL: https://www.vox.com/2019/6/27/18761639/ai-deepfake-deepnude-app-nude-women-porn (дата звернення: 05.05.2024).

25) Rini R., Cohen L. Deepfakes, Deep Harms. *Journal of Ethics and Social Philosophy*. 2022. Т. 22, № 2. URL: https://doi.org/10.26556/jesp.v22i2.1628 (дата звернення: 05.05.2024).

26) Keller E. Pope Francis in Balenciaga deepfake fools millions: ‘Definitely scary’. *New York Post*. URL: https://nypost.com/2023/03/27/pope-francis-in-balenciaga-deepfake-fools-millions-definitely-scary/ (дата звернення: 05.05.2024).

27) GitHub - iperov/DeepFaceLab: DeepFaceLab is the leading software for creating deepfakes. *GitHub*. URL: https://github.com/iperov/DeepFaceLab (дата звернення: 05.05.2024).

28) Midjourney AI - Free Image Generator. *Midjourney AI - Free Image Generator*. URL: https://midjourney.co/ (дата звернення: 05.05.2024).

29) Runway - Advancing creativity with artificial intelligence. *Runway - Advancing creativity with artificial intelligence.* URL: https://runwayml.com/ (дата звернення: 05.05.2024).

30) P. Korshunov and S. Marcel, "Vulnerability assessment and detection of Deepfake videos," 2019 International Conference on Biometrics (ICB), Crete, Greece, 2019, pp. 1-6, doi: 10.1109/ICB45273.2019.8987375.

31) GitHub - shaoanlu/faceswap-GAN: A denoising autoencoder + adversarial losses and attention mechanisms for face swapping. *GitHub*. URL: https://github.com/shaoanlu/faceswap-GAN (дата звернення: 05.05.2024).

32) Ron DeSantis ad uses AI-generated photos of Trump, Fauci. *Fact Check*. URL: https://factcheck.afp.com/doc.afp.com.33H928Z (дата звернення: 05.05.2024).

33. Conrad Sanderson - VidTIMIT dataset. *Conrad Sanderson - home page*. URL: https://conradsanderson.id.au/vidtimit/ (дата звернення: 05.05.2024).

34) GitHub - davidsandberg/facenet: Face recognition using Tensorflow. *GitHub*. URL: https://github.com/davidsandberg/facenet (дата звернення: 05.05.2024).

35) Chung J. S., Zisserman A. Learning to lip read words by watching videos. *Computer Vision and Image Understanding*. 2018. T. 173. C. 76–85. URL: https://doi.org/10.1016/j.cviu.2018.02.001 (дата звернення: 05.05.2024).

36) BuzzFeedVideo. You Won’t Believe What Obama Says In This Video! 😉, 2018. *YouTube*. URL: https://www.youtube.com/watch?v=cQ54GDm1eL0 (дата звернення: 05.05.2024).

37) Silverman C. How To Spot A DeepFake Like The Barack Obama-Jordan Peele Video. *BuzzFeed*. URL: https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed (дата звернення: 05.05.2024).

38) Kuklychev Y. Is photo of Donald Trump "dancing with 13-year-old" girl real?. *Newsweek*. URL: https://www.newsweek.com/photo-donald-trump-dancing-girl-real-ai-1806655 (дата звернення: 05.05.2024).

39) Сторінка професора Хані Фаріда, університету Берклі, Каліфорнія, на якому є зображення фейку із Д.Трампом ,URL: https://farid.berkeley.edu/deepfakes2024election/assets/2023-06-14-Trump-13year-old-girl.jpeg (дата звернення 28.11.23).

40 By Kayleen Devlin and Joshua Cheetham. Fake Trump arrest photos: How to spot an AI-generated image. *BBC Home - Breaking News, World News, US News, Sports, Business, Innovation, Climate, Culture, Travel, Video & Audio*. URL: https://www.bbc.com/news/world-us-canada-65069316 (дата звернення: 05.05.2024).

41) Fake audio falsely claims to reveal private Biden comments. *AP News*. URL: https://apnews.com/article/fact-check-biden-audio-banking-fake-746021122607 (дата звернення: 05.05.2024).

42) Manipulated Reality, Menaced Democracy: An Assessment of the DEEP FAKES Accountability Act of 2019 – N.Y.U. Journal of Legislation & Public Policy. *N.Y.U. Journal of Legislation & Public Policy – A nonpartisan periodical specializing in the analysis of local, state, and federal legislation and policy*. URL: https://nyujlpp.org/quorum/lipkowitz-manipulated-reality-menaced-democracy-deepfakes-accountability-act/ (дата звернення: 05.05.2024).

43) GovInfo. *GovInfo | U.S. Government Publishing Office*. URL: https://www.govinfo.gov/app/details/BILLS-115s3805is (дата звернення: 05.05.2024).

44) Brown N. I. Congress Wants to Solve Deepfakes by 2020. That Should Worry Us. *Slate Magazine*. URL: https://slate.com/technology/2019/07/congress-deepfake-regulation-230-2020.html (дата звернення: 05.05.2024).

45) КРИМІНОЛОГІЧНИЙ АНАЛІЗ ВИКОРИСТАННЯ ТЕХНОЛОГІЇ DEEPFAKE: КОЛИ ФЕЙК СТАЄ ЗЛОЧИНОМ : дис. канд. юр. наук : ISSN 2304-4756 / . – Харків, 2021. – 12 с.

46) Mystery Scoop. 2nd Century Roman Emperors | Realistic Face Reconstruction Using AI and Photoshop, 2021. *YouTube*. URL: https://www.youtube.com/watch?v=IdK_FxC2Zog (дата звернення: 05.05.2024).

47) Jingnan H. Using AI to detect AI-generated deepfakes can work for audio – but not always. *NPR*. URL: https://www.npr.org/2024/04/05/1241446778/deepfake-audio-detection (дата звернення: 05.05.2024).

48) AI MUSIC WORLD. Frank Sinatra - Smells Like Teen Spirit (AI Cover), 2023. *YouTube*. URL: https://www.youtube.com/watch?v=Num0q-l-ldc (дата звернення: 05.05.2024).

49) Taylor J. Suno AI can generate power ballads about coffee – and jingles for the Guardian. But will it hurt musicians?. *the Guardian*. URL: https://www.theguardian.com/technology/2024/apr/13/ai-generated-music-app-suno-ai-impact-musicians-music-rights (дата звернення: 05.05.2024).

50) Spotify won't remove all AI-generated content, as it rolls out some of its own. *ZDNET*. URL: https://www.zdnet.com/article/spotify-wont-remove-all-ai-generated-content-as-it-rolls-out-some-of-its-own/ (дата звернення: 05.05.2024).

51) What is Suno? The AI music generator everyone is talking about. *daily.dev | Where developers grow together*. URL: https://daily.dev/blog/what-is-suno-the-ai-music-generator-everyone-is-talking-about (дата звернення: 05.05.2024).

52) The algorithmic composition for music copyright protection under deep learning and blockchain / N. Wang та ін. *Applied Soft Computing*. 2021. Т. 112. С. 107763. URL: https://doi.org/10.1016/j.asoc.2021.107763 (дата звернення: 05.05.2024).

53) Відеофейк: Головнокомандувач ЗСУ Залужний готує військовий переворот. *StopFake*. URL: https://www.stopfake.org/uk/videofejk-golovnokomanduvach-zsu-zaluzhnij-gotuye-vijskovij-perevorot/ (дата звернення: 05.05.2024).

54) Для чого російська пропаганда використовує діпфейк В.Залужного | Центр протидії дезінформації. *Центр протидії дезінформації | Головна сторінка*. URL: https://cpd.gov.ua/main/dlya-chogo-rosijska-propaganda-vykorystovuye-dipfejk-v-zaluzhnogo/ (дата звернення: 05.05.2024).

55Gamage D. The Emergence of Deepfakes and its Societal Implications: A Systematic Review. *Academia.edu - Share research*. URL: https://www.academia.edu/74762342/The_Emergence_of_Deepfakes_and_its_Societal_Implications_A_Systematic_Review (дата звернення: 05.05.2024).

56) Baudrillard J. Simulacra and Simulations. https://web.stanford.edu/class/history34q/readings/Baudrillard/Baudrillard_Simulacra.html?source=post_page . URL: https://web.stanford.edu/class/history34q/readings/Baudrillard/Baudrillard_Simulacra.html?source=post_page (дата звернення: 05.05.2024).

57) A Guide to Jean Baudrillard's Simulacra and Simulation. Media Studies. URL: https://media-studies.com/baudrillard/ (дата звернення: 05.05.2024)

58) Kalberg S. Max Weber's Types of Rationality: Cornerstones for the Analysis of Rationalization Processes in History. Chicago Journals. 2010. The American Journal of Sociology, Vol. 85, No. 5 (Mar., 1980), pp. 1145-1179. С. 8. URL: https://www.bu.edu/sociology/files/2010/03/Weberstypes.pdf (дата звернення: 05.05.2024).

59) LANG K. Mass Society, Mass Culture, and Mass Communication: The Meaning of Mass. International Journal of Communication 3 (2009), 998-1024. 2024. 5 трав.

С. 5. URL: https://ijoc.org/index.php/ijoc/article/viewFile/597/407 (дата звернення: 05.05.2024).

60) Patarlapati N. Unmasking Reality: Exploring the Sociological Impacts of Deepfake Technology. *International Journal & Research Paper Publisher | IJRASET*. URL: https://www.ijraset.com/best-journal/unmasking-reality-exploring-the-sociological-impacts-of-deepfake-technology (дата звернення: 05.05.2024).

61) Hess A. How a Virtual Assistant Taught Me to Appreciate Busywork. The New York Times. URL: https://www.nytimes.com/2024/04/24/arts/artificial-intelligence-assistants-parents.html (дата звернення: 13.05.2024).

62) РБК-Украина. Крутіше Софії: в Китаї презентували реалістичного робота-телеведучого. РБК-Украина. URL: https://www.rbc.ua/ukr/lite/it_i_tehnologii/diktorov-kitae-mogut-zamenit-roboty-1541797774.html (дата звернення: 13.05.2024).

АНОТАЦІЯ

Дипломна робота на тему «Deepfakes як інструмент маніпулятивних медіакомунікацій» присвячена аналізу впливу технологій штучного інтелекту, зокрема генеративно-змагальних мереж (GANs), на створення маніпулятивного контенту в медіа. В роботі розглядаються наслідки використання deepfakes у контексті політичних подій та військових конфліктів, з акцентом на випадках використання цих технологій для ведення військової пропаганди та дискредитації політичних діячів. У ході цієї роботи я досліджую потенційні загрози від створення і використання deepfakes, включаючи можливості для маніпуляцій громадською думкою та впливу на політичні процеси. Також розглядаються сучасні стратегії та технологічні рішення для ідентифікації та маркування синтетичного контенту. Важливим аспектом є аналіз законодавчих ініціатив щодо регулювання використання deepfakes, враховуючи розвиток нормативно-правової бази в різних країнах. Об’єктом дослідження виступають медіакомунікації, що використовують технології штучного інтелекту для створення маніпулятивних матеріалів. Предметом – deepfakes як засіб маніпуляцій у медійному просторі. Метою дипломної роботи є аналіз технологічних і правових підходів до протидії впливу deepfakes на суспільство та політику. За результатами дослідження автор пропонує комплексні стратегії для виявлення та обмеження розповсюдження маніпулятивних deepfakes, акцентуючи на необхідності посилення міжнародного співробітництва у цій галузі. Подальші наукові дослідження можуть розширити розуміння впливу синтетичних медіа на глобальні політичні та соціальні процеси.

Ключові слова: deepfakes, штучний інтелект, медіакомунікації, політична пропаганда, маніпуляції, регулювання синтетичного медіа контенту.

ABSTRACT

The thesis titled "DEEPFAKES AS A TOOL FOR MANIPULATIVE MEDIA COMMUNICATIONS" is dedicated to analyzing the impact of artificial intelligence technologies, specifically Generative Adversarial Networks (GANs), on creating manipulative content in media. This work examines the consequences of using deepfakes in the context of political events and military conflicts, focusing on instances where these technologies are used for military propaganda and discrediting political figures.

The author investigates the potential threats posed by the creation and use of deepfakes, including their ability to manipulate public opinion and influence political processes. The study also explores current strategies and technological solutions for identifying and labeling synthetic content. An important aspect is the analysis of legislative initiatives aimed at regulating the use of deepfakes, considering the development of legal frameworks in various countries.

The object of the study is media communications that employ artificial intelligence technologies to create manipulative materials. The subject is deepfakes as a means of manipulation in the media space. The aim of the thesis is to analyze technological and legal approaches to counteracting the impact of deepfakes on society and politics.

Based on the research findings, the author proposes comprehensive strategies for detecting and limiting the spread of manipulative deepfakes, emphasizing the need for enhanced international cooperation in this field. Further scientific research could expand understanding of the influence of synthetic media on global political and social processes.

Keywords: deepfakes, artificial intelligence, media communications, political propaganda, manipulation, regulation of synthetic media content.