Посольство
Великої Британії
в Україні

Інститут
Інноваційного
Врядування

ПОСІБНИК ДЛЯ ТВØРЦІВ КØНТЕНТУ

ДЛЯ ВИЯВЛЕННЯ ТА ПРОТИДІЇ РОСІЙСЬКІЙ ПРОПАГАНДІ В НОВІТНІХ ТЕХНОЛОГІЯХ

Інститут Інноваційного Врядування, 2024

Видання Посібника для творців контенту для виявлення та протидії російській пропаганді в нових технологіях стало можливим завдяки фінансовій підтримці Уряду Великої Британії в рамках проєкту «Розуміння штучного інтелекту: Посібник для творців контенту для виявлення та боротьби з російською пропагандою в нових технологіях», який виконує Інститут Інноваційного Врядування.

Погляди, висловлені в цій публікації, належать автору(-ам) і можуть не збігатися з офіційною позицією Уряду Великої Британії.

Керівниця проєкту:
Анна Мисишин, співзасновниця та директорка “Інституту Інноваційного Врядування”

Дизайн:
Марк Мірончук

ЗМІСТ

РЕЗЮМЕ

Цей посібник є комплексним ресурсом, розробленим Інститутом Інноваційного Врядування за підтримки Посольства Великої Британії в Україні. Він покликаний допомогти творцям контенту, журналістам і користувачам соціальних мереж у виявленні та протидії російській пропаганді, особливо в умовах, коли вона поширюється за допомогою новітніх технологій штучного інтелекту.

У сучасному цифровому ландшафті, де інструменти штучного інтелекту, такі як ChatGPT, набули широкого розповсюдження, ризик швидкого поширення дезінформації та пропаганди є значним. Ці технології, здатні швидко генерувати різні форми контенту, спотворюючи історичні факти та новини. Це питання особливо гостро стоїть у зв'язку з війною, що триває в Україні, та глобальною політичною ситуацією.

Посібник містить огляд можливостей технологій штучного інтелекту у створенні контенту та детальний підхід до виявлення російської пропаганди, створеної за допомогою штучного інтелекту в інформаційному просторі. У ньому підкреслюється важливість перевірки фактів та верифікації.

Основна увага приділяється відповідальному використанню штучного інтелекту в журналістиці та створенні контенту, збалансуванню переваг і ризиків цих технологій. Посібник охоплює такі важливі теми, як машинне навчання, глибоке навчання, генеративний ШІ, галюцинації ШІ та проблеми глибоких фейків.

У посібнику також розглядається роль соціальних мереж у поширенні дезінформації та висвітлюються стратегії підвищення медіаграмотності. У ньому обговорюється підзвітність зацікавлених сторін, зокрема журналістів, розробників ШІ та медіа-платформ, і підкреслюється необхідність етичних принципів в епоху передових технологій штучного інтелекту.

Таким чином, цей посібник слугує як освітнім інструментом, так і закликом до етичної взаємодії з новими технологіями поширення інформації, що має вирішальне значення для боротьби з дезінформацією та сприяння розвитку добре поінформованого суспільства.

ОСНОВНІ ЦІЛІ

1. Навчити цільову аудиторію про нюанси контенту, створеного штучним інтелектом, і його потенційне зловживання для поширення пропаганди.

2. Надати користувачам практичні навички та знання для критичної оцінки контенту, створеного штучним інтелектом, особливо в контексті російської пропаганди.

3. Популяризувати етичні стандарти та відповідальні практики використання інструментів штучного інтелекту для створення контенту.

Посібник призначений для журналістів, блогерів, публічних осіб у соціальних мережах та всіх, хто займається створенням та поширенням інформаційного контенту.

ВСТУП

У час, коли цифровий світ стрімко розвивається, поява технологій штучного інтелекту (ШІ) кардинально змінила наш підхід до створення та споживання контенту. Цей зсув не лише відкрив нові шляхи для інновацій та творчості, але й створив унікальні й непрості виклики. Серед них - зростаюче поширення пропаганди, створеної штучним інтелектом.

Цей посібник, створений Інститутом Інноваційного Врядування за підтримки Посольства Великої Британії в Україні, має на меті надати творцям контенту, журналістам та широкій громадськості знання та інструменти, необхідні для того, щоб орієнтуватися у складній взаємодії між штучним інтелектом та пропагандою.

Оскільки технології штучного інтелекту, такі як ChatGPT, стають дедалі досконалішими, вони пропонують потенціал для створення різноманітного контенту - від новинних статей до постів у соціальних мережах. Однак ці можливості також несуть у собі ризик зловживань для поширення дезінформації та пропаганди. Актуальність цього питання зростає в контексті війни в Україні та ширшого глобального політичного ландшафту, де правда часто стає першою жертвою.

Посібник побудований таким чином, щоб забезпечити всебічне розуміння можливостей та обмежень сучасних технологій штучного інтелекту. У ньому детально розглядається, як цими технологіями можна маніпулювати для створення переконливого, але оманливого контенту, а також підкреслюється необхідність критичного підходу та перевірки інформації, яку ми споживаємо та створюємо.Крім того, у посібнику розглядаються нюанси російської пропаганди: її характеристики, тактика і те, як вона проявляється в контенті, створеному за допомогою штучного інтелекту. Це розуміння має вирішальне значення для творців контенту, щоб ефективно виявляти оманливі наративи та протидіяти їм.

Ми також розглядаємо етичні міркування та відповідальність, пов'язані з використанням штучного інтелекту у створенні контенту. При цьому ми підкреслюємо, що необхідно дотримуватися балансу між використанням переваг штучного інтелекту та збереженням журналістської чесності й об'єктивності.

З настанням нової цифрової ери відповідальність за відстоювання правди та боротьбу із дезінформацією стає більш важливою, ніж будь-коли. Цей посібник - не просто ресурс, а заклик до дії. Він заохочує кожного з нас бути більш обізнаними, критичними та етично відповідальними щодо контенту, який ми створюємо та споживаємо у світі, який дедалі більше формується під впливом штучного інтелекту.

ЖУРНАЛІСТИКА ТА СТВОРЕННЯ КОНТЕНТУ В ЦИФРОВУ ЕПОХУ- БАЛАНС РИЗИКІВ ТА ПЕРЕВАГ

Війна, яку почала Росія проти України, є важливим прикладом викликів, з якими стикаються журналісти та творці контенту в цифрову епоху. Ця війна гостро поставила питання про необхідність ретельного балансу між використанням технологічних досягнень та управлінням ризиками, пов'язаними з дезінформацією та пропагандою.

Двоїста природа штучного інтелекту у створенні контенту

Додатки зі штучним інтелектом стають дедалі більш поширеними у створенні контенту, адже пропонують чудові можливості - від автоматизації рутинних завдань до написання складних досліджень. Однак ця зручність несе значні ризики при роботі із чутливим контентом, пов'язаним з українсько-російською війною.

Багато людей, захоплені ефективністю та новизною інструментів штучного інтелекту, почали впроваджувати їх у свій робочий процес. На жаль, помітний брак медіаграмотності та всебічного розуміння можливостей цих технологій призвів до критичної прогалини. У багатьох випадках, як і споживачі, так і творці онлайн матеріалів, не перевіряють достовірність інформації, створеної штучним інтелектом. Саме відсутність перевірки фактів перед публікацією таких матеріалів може ненавмисно сприяти поширенню пропаганди, в результаті чого, неправдиві наративи, отримують широке розповсюдження.

Боротьба з пропагандою та дезінформацією

Ефективна протидія російській пропагандистській машині, яка вміло використовує цифрові платформи та інструменти штучного інтелекту для створення та поширення оманливих наративів, вимагає комплексного та багатогранного підходу. По-перше, існує нагальна потреба в тому, щоб творці контенту поглибили своє розуміння технологій штучного інтелекту, усвідомивши їхні

обмеження та потенційну упередженість. Програми з медіаграмотності відіграють тут вирішальну роль, акцентуючи увагу на важливості перевірки контенту, створеного штучним інтелектом. Крім того, журналісти і творці контенту повинні дотримуватися суворих протоколів перевірки фактів, особливо для контенту, пов'язаного з війною. Цей процес передбачає ретельну перехресну перевірку інформації з кількох надійних джерел і збереження здорового скептицизму щодо неперевіреного контенту, створеного ШІ.

Водночас дуже важливо наголошувати на дотриманні етичних стандартів у використанні ШІ для створення контенту. Творці повинні повністю усвідомлювати етичні наслідки поширення неперевіреної інформації та способи маніпулювання штучним інтелектом у пропагандистських цілях.

Крім того, кампанії з інформування громадськості мають важливе значення для навчання людей про природу пропаганди та методи її розпізнавання. Сюди входить допомога громадськості в розумінні того, як AI може використовуватися для створення "глибоких фейків" та інших форм контенту, що вводить в оману. Разом ці зусилля формують надійну стратегію боротьби з витонченим використанням ШІ для поширення дезінформації та захисту цілісності інформації в цифрову епоху.

1. РОЗУМІННЯ ШТУЧНОГО ІНТЕЛЕКТУ

ЩО ТАКЕ ШТУЧНИЙ ІНТЕЛЕКТ?

Джон Маккарті

Штучний інтелект (ШІ) - термін, введений почесним професором Стенфордського університету Джоном Маккарті у 1955 році. Він визначив ШІ як "науку та інженерію створення розумних машин". Це визначення є широким і охоплює різні аспекти ШІ, включаючи машинне навчання, обробку природної мови, робототехніку і вирішення проблем.

У найпростішому розумінні, штучний інтелект (ШІ) - це галузь комп'ютерних наук, що займається створенням машин (програм), здатних до розумної поведінки. Йдеться про розробку систем, здатних самостійно мислити, навчатися та діяти - від простих алгоритмів, що вирішують конкретні завдання, до складних систем, що імітують людський інтелект.

ШІ охоплює цілу низку технологій, від алгоритмів машинного навчання до більш складних систем, таких як глибоке навчання та генеративний ШІ. Машинне навчання дозволяє комп'ютерам вчитися і робити прогнози на основі даних, тоді як глибоке навчання, підмножина машинного навчання, включає нейронні мережі з декількома шарами, які можуть навчатися на великих обсягах даних. Генеративний ШІ, який привернув до себе значну увагу, здатний створювати контент, включаючи текст, зображення і відео, який часто неможливо відрізнити від контенту, створеного людиною.

ТЕХНОЛОГІЯМИ ШТУЧНОГО ІНТЕЛЕКТУ МОЖУТЬ БУТИ:

МАШИННЕ НАВЧАННЯ (MACHINE LEARNING)

Що таке машинне навчання?

Машинне навчання (ML) - це частина штучного інтелекту, яка зосереджена на створенні систем, здатних навчатися та ухвалювати рішення на основі даних. На відміну від традиційного програмування, де комп'ютер слідує чітко запрограмованим інструкціям, ML дозволяє комп'ютерам вчитися і адаптуватися на основі досвіду, не будучи чітко запрограмованими для кожного завдання.

Як працює машинне навчання?

По суті, машинне навчання передбачає подачу великої кількості даних в алгоритми. Потім ці алгоритми аналізують і виявляють закономірності в цих даних. На основі цих закономірностей система робить прогнози або приймає рішення щодо нових даних, з якими вона стикається. Існує кілька типів методів машинного навчання:

1) Навчання під наглядом
Алгоритм навчається на маркованому наборі даних, що означає, що дані вже позначені правильною відповіддю. Мета полягає в тому, щоб вивчити шаблон, щоб модель могла робити прогнози для нових, небачених даних. Прикладом може слугувати спам-фільтр для електронної пошти. Алгоритм навчається на базі даних емейлів, кожен з яких позначений як "спам" або "не спам". Навчаючись на цих прикладах, модель згодом може передбачити, чи є новий лист спамом.

2) Навчання без нагляду
Алгоритм використовується на даних без явних інструкцій і намагається самостійно виявити закономірності та взаємозв'язки в даних. Для прикладу можна розглянути систему рекомендацій для стрімінгового музичного сервісу Spotify. Алгоритм аналізує слухацькі звички користувачів без жодних конкретних вказівок щодо того, що саме шукати. Він виявляє закономірності, за якими користувачі, що слухають певні пісні, також схильні слухати й інші, використовуючи це для того, щоб рекомендувати користувачам нові пісні.

3) Навчання з підкріпленням
Система вчиться методом проб і помилок, отримуючи зворотній зв'язок від своїх дій і відповідно коригуючи свій курс. Як приклад розглянемо ШІ, який навчається грати у складні відеоігри, такі як шахи або “Go”. ШІ починає з випадкових ходів, але отримує зворотний зв'язок, заснований на перемозі чи поразці. З часом він вивчає стратегії, які збільшують його шанси на перемогу, коригуючи свій ігровий процес на основі результатів кожної гри.

B) ГЛИБОКЕ НАВЧАННЯ (DEEP LEARNING)

Що таке глибоке навчання?

Глибоке навчання (ГН) - це просунута форма машинного навчання, яка використовує нейронні мережі з декількома шарами (звідси і термін “глибокий”). Ці шари складаються з вузлів, які імітують нейрони людського мозку. Кожен шар обробляє певні аспекти вхідних даних і передає

їх наступному шару, поступово уточнюючи і покращуючи процес ухвалення рішень.

Як працює глибоке навчання?

При глибокому навчанні модель вчиться виконувати завдання безпосередньо з тексту, зображень або звуку. Ці моделі навчаються за допомогою великих наборів маркованих даних і нейромережевих архітектур, які можуть вивчати функції і завдання безпосередньо з даних. “Глибокий” у глибокому навчанні означає кількість шарів, через які трансформуються дані. Чим більше шарів, тим складніші закономірності та взаємозв’язки можна вивчити. Голосові помічники, такі як Siri або Alexa, використовують глибоке навчання для обробки природної мови та розпізнавання мовлення. Вони аналізують голосові дані, щоб розуміти вимовлені команди, з часом навчаючись розпізнавати різні акценти та мовленнєві патерни.

ГЕНЕРАТИВНИЙ ШІ

Що таке генеративний ШІ?

Генеративний ШІ - це підмножина алгоритмів ШІ, призначених для створення нового контенту. Це може бути текст, зображення, відео та аудіо. На відміну від інших моделей ШІ, що використовуються переважно для аналізу та прогнозування, моделі генеративного ШІ можуть створювати новий контент, який імітує людську творчість і складність. Це досягається завдяки навчанню на великих масивах даних і розумінню основних закономірностей і структур. Прикладами генеративного ШІ є відомі нам додатки ChatGPT, DALL-E від OpenAI, Bard, Gemini тощо.Ці моделі штучного інтелекту здатні генерувати текст, схожий на людський, чи реалістичні фотографії, на основі отриманих даних.

Як працює генеративний ШІ?

Генеративний ШІ працює за допомогою таких алгоритмів, як генеративні змагальні мережі (GAN) і варіаційні автокодери (VAE). Ці моделі, по суті, навчаються на великій кількості вхідних даних, засвоюють закономірності, а потім використовують ці знання для створення нових, оригінальних результатів. Наприклад, генеративний ШІ, навчений на новинних статтях, може створювати абсолютно нові статті на схожі теми.

Застосування у створенні контенту

Генеративний ШІ може автоматично створювати письмовий контент, наприклад, звіти, резюме або навіть цілі статті, що може бути особливо корисним для висвітлення ситуацій, які швидко розвиваються.

Генеративний ШІ може створювати реалістичні зображення та відео, які можна використовувати в цифровому сторітелінгу для підвищення візуальної привабливості контенту.
Генеративний ШІ може адаптувати контент до індивідуальних уподобань, оптимізуючи взаємодію з читачем і покращуючи його досвід.

Виклики та етичні міркування

Однією з найбільших проблем генеративного ШІ є можливість створення переконливого, але неправдивого контенту. Це створює значні ризики в журналістиці, де достовірність інформації має першорядне значення, особливо у висвітленні конфліктів тм воїн.

Генеративні моделі ШІ можуть увічнити упередження, присутні в їхніх навчальних даних. Це може призвести до викривленої або несправедливої репрезентації створеного контенту.

1.1. DEEPFAKES (ГЛИБОКІ ПІДРОБКИ)

ЩО ТАКЕ DEEPFAKE?

Deepfake - це синтетичні медіа, де людину на реальному зображенні чи відео замінюють на чиюсь іншу подобу, часто з використанням методів штучного інтелекту, таких як глибоке навчання. Ці технології дозволяють створювати аудіо- та відеокліпи, які неймовірно важко відрізнити від справжнього контенту. Термін «глибокий» (deep) походить від «глибокого навчання» - форми ШІ, яка використовує нейронні мережі для обробки даних і створення цих гіперреалістичних результатів.

Наслідки Deepfake у журналістиці

1. У сфері журналістики Deepfakes можуть використовуватися для створення неправдивих наративів або оманливого представлення подій, осіб чи заяв. Наприклад, Deepfake може зображати політичного лідера, який робить заяву, яку він насправді не робив, що потенційно може вплинути на громадську думку або дипломатичні відносини.

2. Витонченість таких підроблених матеріалів створює складність при перевірці автентичності аудіо- та відеоматеріалів. У зонах конфліктів, а також під час війни, де переважає пропаганда та ведеться інформаційна війна, розрізнення справжніх матеріалів і глибоких фейків стає критично важливим завданням.

3. Можливість використання Deepfake для поширення неправдивої інформації може призвести до загальної ерозії довіри до медіа. Якщо аудиторія не може відрізнити справжній контент від маніпульованого, це може призвести до скептицизму та сумнівів навіть щодо законних джерел новин.

Виявлення Deepfake - навичка, яка стає дедалі важливішою в журналістиці та створенні контенту, яка вимагає тонкого розуміння як технології, що їх створює, так і інструментів, розроблених для їхнього виявлення.

З розвитком технології «глибоких фейків» розрізнення автентичного контенту від маніпулятивного вимагає ретельного спостереження та використання спеціалізованих інструментів. Пропонуємо вашій увазі огляд того, як виявляти Deepfake, а також розуміння технологій їхнього створення та виявлення.

Фото згенеровано штучним інтелектом

ПОРАДИ ЩОДО ВИЯВЛЕННЯ ВІДЕО ПІДРОБОК:

AP

Зверніть увагу на риси обличчя
Зверніть увагу на неправильну міміку або неприродне моргання. Невідповідність у напрямку погляду також може бути ознакою брехні. Звертайте особливу увагу на вуха, оскільки вони іноді можуть виглядати спотвореними, неприродно розташованими або мати на собі прикраси лиш з однієї сторони.

Оцініть синхронізацію губ
Розбіжності між вимовленими словами та рухами губ можуть свідчити про Deepfake.

Проаналізуйте освітлення і тіні
Невідповідне освітлення або тіні, що не відповідають фізичному оточенню, можуть свідчити про маніпуляцію із контентом.

Огляньте волосся та шкіру
Незвичайні текстури або візерунки на волоссі та шкірі, які часто є проблемою для алгоритмів глибокої підробки, можуть бути викривальними.

Зверніть увагу на голос
Прислухайтеся до будь-яких розбіжностей у тоні, висоті або акценті, які можуть відрізнятися від відомих характеристик людини.

Аналіз фону
Шукайте аномалії або неприродні зміни у фонових декораціях.

Зверніть увагу на якість зображення
Помилки пікселізації або стиснення, особливо по краях обличчя, можуть свідчити про глибоку підробку.

Використовуйте технічні інструменти
Використовуйте інструменти виявлення на основі штучного інтелекту, які аналізують відео на предмет невідповідностей, які важко виявити неозброєним оком. Такими інструментами виявлення є: Microsoft Video Authenticator, Deepware Scanner, FakeCatcher, Adobe Content Authenticity Initiative (CAI), Sensity, WeVerify, Intel та TruePic пропонують способи виявлення та перевірки автентичності цифрових матеріалів.

Перехресні посилання з джерелами
Порівняйте сумнівний контент з перевіреним матеріалом на предмет достовірності.

ВІДОМІ ДОДАТКИ ТА ПРОГРАМИ ДЛЯ СТВОРЕННЯ ТА ВИЯВЛЕННЯ DEEPFAKE

Додатки для створення підробок: DeepFaceLab, FaceSwap, ZAO і Reface демонструють, з якою легкістю можна створювати Deepfake. Більшість програм створені з розважальними цілями, для прикладу український додаток Reface дає змогу створювати розважальні відео, на яких риси користувачів накладаються на відомих осіб кіно чи музичної індустрії. Деякі програми, такі як Midjourney, хоч і не створені з метою поширювати дезінформацію, проте їх відео та фото, найчастіше для цього використовують.

Отже, хоча технологія Deepfake створює проблеми в таких сферах, як журналістика, розробка складних інструментів виявлення пропонує спосіб боротися з її зловживанням.

Однак дуже важливо використовувати ці інструменти відповідально і разом із традиційними методами перевірки, щоб зберегти автентичність і достовірність цифрового контенту.

1.2. ЧОМУ ШТУЧНИЙ ІНТЕЛЕКТ ГЕНЕРУЄ ДЕЗІНФОРМАЦІЮ?

У цьому технологічному світі, де цифрова інформація створюється і поширюється у величезних масштабах щосекунди, важливо критично оцінювати роль генеративного штучного інтелекту (ШІ) в контексті точності та цілісності інформації. Всюдисущість цифрових медіа спростила поширення контенту в глобальному масштабі, але ця легкість поширення також пов'язана з ризиком швидкого і широкого розповсюдження дезінформації.

Генеративний ШІ, особливо у вигляді моделей глибокого навчання, здатен створювати високореалістичні зображення, відео та текст. Ця здатність, будучи революційною і цінною для різних творчих і освітніх програм, також відкриває двері для потенційних зловживань. Технологія може фабрикувати контент, який дедалі важче відрізнити від реальності, що створює значні проблеми для перевірки інформації.

Чому ж технологія з таким трансформаційним потенціалом виступає в ролі провідника дезінформації? Таке розуміння виходить за рамки простої технічної необхідності; воно є наріжним каменем у підтримці журналістської чесності та точності, особливо в епоху війни, коли роль штучного інтелекту значно зростає. Генерації неправдивої інформації сприяють кілька факторів:

ДАНІ З ВІДКРИТИХ ДЖЕРЕЛ

Основна причина полягає в характері даних, які використовуються для навчання ШІ. Генеративні моделі ШІ, як правило, навчаються на великих масивах даних, отриманих з відкритих загальнодоступних джерел. Хоча такий підхід дозволяє ШІ навчатися на широкому спектрі контенту, він також несе в собі значний ризик. Дані з відкритих джерел часто можуть містити дезінформацію, неточності та упереджені точки зору. Коли ШІ-моделі навчаються на таких даних, вони ненавмисно засвоюють і відтворюють ці неточності у своїх результатах. Реальні приклади можуть проілюструвати це більш чітко:

а) Вміст соціальних мереж:
ШІ-моделі, навчені на даних із соціальних мереж, можуть ненавмисно вчитися на дописах чи коментарях, які містять неперевірену інформацію, чутки чи суб'єктивні думки. Наприклад, на початкових етапах українсько-російської війни соціальні мережі рясніли непідтвердже-

ними повідомленнями та особистими інтерпретаціями подій, які, якщо їх використовувати для навчання, можуть призвести до того, що ШІ генеруватиме оманливі наративи.

b) Онлайн-форуми та дискусійні дошки:

Ці платформи, такі як Reddit, часто містять суміш фактичної інформації, особистих анекдотів і спекулятивного контенту. Якщо модель штучного інтелекту навчена на такому міксі, вона може не відрізняти перевірені факти від спекулятивних дискусій. Наприклад, онлайн-форуми, що обговорюють політичні мотиви українсько-російської війни, можуть містити суміш точних історичних даних та упереджених політичних думок.

c) Новинні сайти з різними редакційними стандартами:

ШІ, навчений на новинних статтях з веб-сайтів, які не дотримуються суворих журналістських стандартів, може відтворювати упередженість або неточності, присутні в цих статтях. Якщо ШІ-систему постійно годувати новинами з сайтів, які тяжіють до сенсаційності або не проводять ретельну перевірку фактів, вона може створювати контент, який віддзеркалює ці недоліки.

ГАЛЮЦИНАЦІЇ В ШТУЧНОМУ ІНТЕЛЕКТІ

a) Сфабриковані історичні події:

Модель штучного інтелекту, навчена на неточних історичних даних або вигаданих наративах, може генерувати статті або репортажі, які посилаються на неіснуючі події. Наприклад, AI може створити історію про сфабриковану військову операцію під час українсько-російської війни, змусивши читачів повірити в події, яких ніколи не було.

b) Вигадані цитати або висловлювання:

Моделі ШІ можуть генерувати реалістичні цитати публічних осіб, які насправді ніколи не звучали. Наприклад, штучний інтелект може галюцинувати заяву, яку приписують світовому лідеру або військовому чиновнику щодо українсько-російської війни, потенційно спричиняючи плутанину і поширюючи неправдиву інформацію.

c) Спотворені статистичні дані:

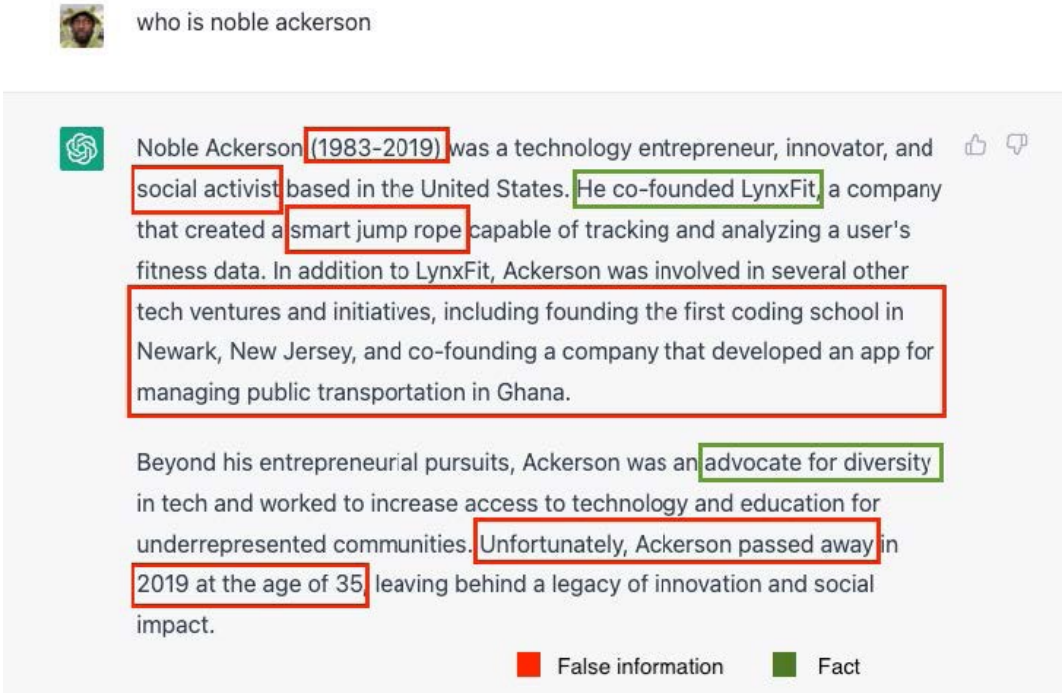
У випадках, коли ШІ використовують для аналізу та представлення статистичних даних, галюцинації можуть призвести до генерації неправильних цифр. Особливо це може вводити в оману в журналістиці, коли йдеться про такі чутливі питання, як кількість жертв або кількість біженців у війні.

d) Спотворення місцевості або геополітичних фактів:

ШІ може неправильно описувати географічні або політичні реалії, наприклад, спотворювати місце ключової події в українсько-російській війні або неточно описувати територіальний контроль над окремими регіонами.

e) Неправдиві зображення або відео:

У сфері глибоких фейків ШІ може створювати переконливі, але повністю неправдиві зображення або відео. Це може бути особливо шкідливим у журналістиці, де візуальні докази часто мають вирішальне значення. Згенероване штучним інтелекпростотом відео, що



неправдиво зображує військові дії або події в українсько-російській війні, може мати серйозні наслідки.

Поєднання цих факторів змушує журналістів і творців контенту, які використовують генеративний ШІ, бути вкрай обережними. Розуміння походження і характеру даних, на яких навчаються ці моделі ШІ, а також усвідомлення феномену галюцинацій ШІ, має вирішальне значення. Ця обізнаність повинна поєднуватися з суворими процесами перевірки фактів і критичною оцінкою контенту, створеного ШІ, особливо коли йдеться про чутливі та впливові теми, такі як українсько-російська війна.

Визнаючи і вирішуючи ці проблеми, журналісти і творці контенту можуть краще орієнтуватися в складнощах використання генеративного ШІ у своїй роботі, гарантуючи, що вони дотримуються стандартів точності і надійності, які є фундаментальними для їхньої професії.

2. ДЕЗІНФОРМАЦІЯ В СОЦІАЛЬНИХ МЕРЕЖАХ

Цей розділ посібника присвячений допомозі творцям контенту, журналістам і користувачам соціальних мереж у виявленні та протидії пропаганді, створеної та розповсюдженої за допомогою ШІ.

Російські дезінформаційні кампанії в соціальних мережах є ключовим фактором занепокоєння, особливо очевидним у таких сценаріях, як українсько-російська війна та президентські вибори в США 2016 року. Ці кампанії часто змішують правду з вигадками, створюючи наративи, які слугують стратегічним інтересам, маніпулюють громадською думкою і сіють розбрат. Наприклад, під час українсько-російської війни в соціальних мережах спостерігався сплеск неправдивих зображень і сфабрикованих історій, покликаних вплинути на сприйняття війни.

Інтеграція технологій штучного інтелекту ще більше ускладнила цей ландшафт. Керовані штучним інтелектом боти та фейкові акаунти, здатні імітувати реальних користувачів, поширюють дезінформацію з тривожною швидкістю та масштабами. Крім того, алгоритми штучного інтелекту вміють створювати переконливий контент, наприклад, глибокі фейки, і націлювати дезінформацію на певні демографічні групи. Це робить виявлення та протидію дезінформації не лише більш складним, але й більш важливим завданням.

Скандал з Cambridge Analytica є яскравим прикладом того, як штучний інтелект може впливати на громадську думку. У 2010-х роках британська консалтингова компанія Cambridge Analytica збирала персональні дані мільйонів користувачів Facebook без їхньої згоди, переважно для використання в політичній рекламі. У цьому випадку особиста інформація незліченної кількості користувачів Facebook була зібрана і об'єднана з методами штучного інтелекту для створення надзвичайно таргетованої політичної реклами. Ця ситуація підкреслює етичні небезпеки, пов'язані з використанням ШІ та аналітики даних для впливу на громадську думку. Здатність AI здійснювати

точне таргетування також є важливим фактором у поширенні дезінформації. Обробляючи величезні обсяги даних користувачів, системи штучного інтелекту можуть точно визначати осіб для проведення індивідуальних дезінформаційних кампаній.

Протидія такій витонченій дезінформації вимагає не лише технологічних рішень, а й комплексного підходу. Підвищення медіаграмотності має вирішальне значення для того, щоб громадськість могла ідентифікувати і розуміти тактики, які використовуються в дезінформаційних кампаніях. Соціальні медіа-платформи також повинні взяти на себе відповідальність, вдосконалюючи виявлення та видалення фейкових акаунтів і співпрацюючи з фактчекерами для забезпечення прозорої модерації контенту. Крім того, регуляторне середовище має розвиватися, щоб покласти на соціальні медіа-платформи відповідальність за контент, який вони поширюють, і захистити дані користувачів від експлуатації. Урядам і міжнародним організаціям слід розглянути нормативно-правові акти, які вирішують ці проблеми, балансуючи між потребою у свободі вираження поглядів та імперативом збереження цілісності інформації.

Насамкінець, у цьому розділі підкреслюється важливість пильності, технологічної грамотності та дотримання етичних стандартів у складній взаємодії російської пропаганди та штучного інтелекту в соціальних мережах. Для журналістів, творців контенту і користувачів соціальних мереж розуміння цієї динаміки є ключем до ефективної протидії дезінформації та відстоювання правди в нашому все більш взаємопов'язаному світі.

2.1. FACEBOOK, X ТА ІНШІ СОЦІАЛЬНІ ПЛАТФОРМИ

Соціальні медіа-платформи, такі як Facebook та X (колишній Twitter), стали ключовими гравцями в ландшафті дезінформації, керованої штучним інтелектом. Поява генеративного ШІ значно посилила поширення і витонченість дезінформаційних кампаній. Ці кампанії все частіше використовують інструменти штучного інтелекту для створення та поширення оманливої інформації в безпрецедентних масштабах. Також ці медіа-платформи стикаються з проблемою модерації контенту, створеного штучним інтелектом. Їхні алгоритми іноді ненавмисно посилюють сенсаційний або оманливий контент через вищі показники залучення, тим самим поширюючи дезінформацію далі.

Компанія «Meta» проінформувала про різні випадки скоординованої неавтентичної поведінки та кібершпигунства на своїх платформах, включаючи відключення акаунтів, пов'язаних з операціями прихованого впливу та дезінформаційними кампаніями.

Так, у 2023 році компанія повідомила про ліквідацію найбільшої операції з прихованого впливу на різних платформах, яку вона описує як найбільшу з відомих. Ця масштабна кампанія, в якій були задіяні тисячі акаунтів, була націлена на понад 50 платформ, включаючи Facebook, Instagram, X (колишній Twitter), YouTube, TikTok і Reddit. Операція насамперед поширювала прокитайський контент, зокрема позитивні відгуки про провінцію Сіньцзян, а також критикувала США, західну зовнішню політику і тих, кого вважали опонентами китайського уряду.

Розслідування Meta виявило використання скоординованих фейкових акаунтів, які працювали за регулярним графіком і, ймовірно, управлялися однією командою в спільному просторі. Вони поширювали схожий контент на різних платформах з оманливими заголовками на кшталт «Бомбардування США «Північного потоку» - перший крок у «плані знищення Європи» Незважаючи на масштабність операції, «Meta» зазначила, що їй не вдалося залучити значну кількість справжніх підписників, натомість з'явилися фейкові підписники з регіонів, що не належать до її цільової аудиторії.

Крім того, «Meta» помітила, що з 2019 року мережа «Spamouflage» змістилася в бік менших платформ. У своєму звіті компанія підкреслює, що продовжує протидіяти іншим операціям прихованого впливу, в тому числі спрямованим проти Türkiye і російської кампанії, що поширює дезінформацію про війну в Україні, яка зараз поширює свою діяльність на США та Ізраїль. Видавання себе за ЗМІ було спільною стратегією для цих кампаній.

У соціальній мережі X (колишній Twitter) численні фальшиві профілі, такі як «Bella Morne», використовують штучний інтелект для створення переконливих образів. Ці акаунти, що мають десятки тисяч підписників, використовують згенеровані штучним інтелектом зображення, що нагадують моделей, для створення своєї онлайн-ідентичності. Вони стратегічно генерують дохід як «творці контенту», публікуючи емоційно забарвлений контент на чутливі теми, такі як ситуація в Палестині та Ізраїлі. Крім того, ці акаунти відомі тим, що висловлюють підтримку Дональду Трампу, що ще більше посилює їхню присутність і вплив на платформі.

Нещодавно Європейська Комісія попередила Ілона Маска та його платформу X (Twitter) про необхідність дотримання нових законів щодо фейкових новин та російської пропаганди. Це попередження з'явилося після того, як було виявлено, що X має найвищий відсоток дезінформаційних постів серед основних соціальних мереж. У звіті, присвяченому поширенню фейкових новин в ЄС, йдеться про те, що мільйони фейкових акаунтів були видалені TikTok і LinkedIn, а Facebook посідає друге місце серед найбільших порушників.

Відповідно до Закону про цифрові послуги (DSA), який набув чинності в серпні 2023 року, пости, класифіковані як дезінформація, будуть незаконними на всій території ЄС. Facebook та інші технологічні гіганти, такі як Google, TikTok і Microsoft, підписали кодекс практики ЄС, щоб відповідати цим законам. Однак Twitter вийшов з цього кодексу, але все одно повинен дотримуватися DSA, інакше йому загрожує заборона в ЄС.

Інструменти генеративного ШІ стали доступнішими і спростили створення та розповсюдження дезінформації в масових масштабах.

Наприклад, державні ЗМІ Венесуели використовували згенеровані штучним інтелектом відео з ведучими новин неіснуючого міжнародного англомовного каналу для поширення проурядових меседжів. Аналогічним чином у США в соціальних мережах поширювалися відео та зображення політичних лідерів, створені за допомогою штучного інтелекту, зокрема відео, на якому президент Байден робить трансфобні коментарі, а також зображення Дональда Трампа, який обіймається з Ентоні Фаучі.

ТікТок – ще одна платформа, де пропагандисти майстерно використовують ШІ для створення маніпулятивного контенту. Це, наприклад, – одна із тисяч схожих сторінок в TikTok із заманливими заставками та контентом, який генерує ШІ. Сторінка створена пропагандистами і просуває шкідливі тези для українського суспільства:

Примусова мобілізація. Через сторітелінг у головного героя (і разом з тим у користувача) наростає відчуття страху, приреченості та несправедливості.

Корупція, яку сіють депутати. Тут теж ціль – посилити недовіру до влади, викликати почуття несправедливості у користувача та думку “за що ми боримося”. Ці наративи надалі активно несвідомо підхоплюють блогери, поширюючи серед суспільства тези, що “всі крадуть”.

Командири-хабарники. А ось тут ціль – посилити недовіру до військового лідерства України. Такі дописи часто розміщують під хештегами, на зразок: #україна #тцк #мобілізація #війна #зсу #корупція #депутат

На яскравих картинках, згенерованих штучним інтелектом, військових зображено ситими та ненаситними до своєї влади. Наче їм приносить задоволення мобілізовувати звичайних громадян, які на фоні понуро сидять у сірому громадському транспорті. Таким способом, окрім зневіри, пропагандисти завдяки технології та креативу свідомо створюють бар’єр та розколюють суспільство. А це, звісно, і послаблює стійкість України у війні.

ШІ генерує не лише візуально привабливий пропагандитський контент. Чимало фейкових відео у ТікТок змонтовано з різних нарізок інформаційних чи розважальних програм популярних українських телеканалів. Часто їх озвучено штучно згенерованими голосами відомих ведучих. Такого висновку дійшов Інститут Масової Інформації в результаті дослідження.

Такі відео, як правило, оформлені великим шрифтом та яскравим кольором. Подекуди написи перекривають обличчя ведучих у кадрі. Це заважає побачити невідповідність озвучки до міміки людей у кадрі.

Крім того, доступність і дешевизна генеративного ШІ знижує бар’єр входу для дезінформаційних кампаній, дозволяючи не лише державним суб’єктам, а й різним групам брати участь у цій діяльності. Поширення контенту, створеного штучним інтелектом, в Інтернеті також призвело до феномену «дивідендів брехуна», коли настороженість щодо фальсифікацій змушує людей скептичніше ставитися до правдивої інформації, особливо під час кризових ситуацій або політичних конфліктів.

2.2. ПІДВИЩЕННЯ МЕДІАГРАМОТНОСТІ: СТРАТЕГІЇ ДЛЯ БІЛЬШ ПОІНФОРМОВАНОГО СУСПІЛЬСТВА

Підвищення медіаграмотності є важливою стратегією у формуванні більш поінформованого суспільства, особливо в епоху, коли дезінформація, керована штучним інтелектом, поширюється на платформах соціальних мереж. Підвищуючи медіаграмотність, люди стають більш здатними критично оцінювати інформацію, з якою вони стикаються, та ухвалювати обґрунтовані рішення. Пропонуємо розглянути кілька стратегій з прикладами:

2.2.1. Ставте під сумнів те, що бачите і чуєте

На практиці: Щоразу, коли ви натрапляєте на новину або відео, наприклад, ймовірний кліп політичного діяча, запитайте себе: Чи виглядає це реалістично? Хто і чому поширює цю інформацію? Шукайте подібні новини на авторитетних сайтах, щоб перевірити, чи не повідомлялося про це деінде.

2.2.2 Вчіться розуміти свої емоції та несвідому поведінку.

На практиці: Здоровий сумнів у своїх навичках з медіаграмотності допоможе вам тримати тонус. Насправді наш мозок робить багато речей несвідомо. Доведено, що фейк, який людина почує навіть в контексті спростування, може стати для мозку знайомим. Відчуття "знайомого" підсвідомо викликає довіру. Тому наступного разу, коли натрапите на схожу тезу, у вас може скластися враження, що ви вже це десь чули чи бачили – і ви будете більш схильні повірити маніпулятивній інформації. Пам'ятайте про це, споживаючи контент.

2.2.3. Дізнайтеся про те, як створюється фейковий контент

На практиці: Витратьте трохи часу, щоб зрозуміти, як створюються фейкові новини та «Deepfakes" (глибокі фейки). Наприклад, на фейкових відео, люди роблять або говорять так, як вони ніколи до цього не робили. Це усвідомлення допоможе вам залишатися обережними і не вірити всьому, що ви бачите.

2.2.4. Диверсифікуйте джерела новин

На практиці: Не покладайтеся лише на одну соціальну мережу або ЗМІ для отримання всієї інформації. Слідкуйте за різними новинними каналами, веб-сайтами і навіть міжнародними ЗМІ, щоб мати широкий спектр точок зору. Це допоможе уникнути пастки, в якій ви почуєте лише ті погляди, які збігаються з вашими.

2.2.5. Слідкуйте за роботою перевірених фактчекерів

На практиці: Перш ніж ділитися несподіваною новиною або шокуючим зображенням, скористайтеся такими ініціативами, які роблять регулярні підбірки фейків: Snopes, FactCheck.org, Stopfake, Detector Media, AP Fact Check, Ukrainefacts.org. Часто такі сайти роблять підбірки популярних фейків та надають їм фактологічне спростування. Це особливо важливо під час великих подій або виборів, коли дезінформація поширюється з великою швидкістю.

2.2.6. Скептично ставтеся до сенсаційних заголовків

На практиці: Заголовки, які звучать занадто драматично або викликають сильну емоційну реакцію, часто призначені для отримання кліків, а не для інформування. Читайте далі, і якщо контент не підтверджує заголовок, він, швидше за все, вводить в оману.

2.2.7. Вивчайте цифрові інструменти

На практиці: Вивчіть прості методи цифрової перевірки, такі як зворотний пошук зображень (за допомогою таких інструментів, як Google Images), щоб перевірити походження фото чи відео. Це швидкий спосіб виявити, чи зображення з «поточної» події насправді перероблене зі старої.

Крім того, існують й інші способи дізнатися, чи справжнє фото перед вами.
Для прикладу скористайтеся додатком RevEye у Google Chrome чи Bing — пошуковик від Microsoft, що допомагає шукати по деталях на зображенні, визначає текст на фотографії та шукає по тексту, тощо.

Також зображення може бути обробленим за допомогою таких редакторів, як, наприклад, Photoshop. Таке зображення можна перевірити на маніпуляції за допомогою інструментів InVID чи або Forensically. За допомогою цих інструментів можна побачити, що на зображенні є елементи, які різко відрізняються за кольором (аналіз рівня помилок або ELA, який показує, що до фотографії щось додавали в редакторі, а потім зберігали її заново).

2.2.8. Беріть участь в обговореннях і ставте запитання

На практиці: Якщо ви не впевнені в якійсь інформації, обговоріть її з друзями чи родиною. Іноді обговорення може відкрити нові перспективи або заохотити інших до критичного мислення.

2.2.9. Довіряйте перевіреним експертам

На практиці: досить часто новини в медіа будуються на тому, що хтось заявив про щось, хтось щось спрогнозував або проаналізував. І саме цей прогноз чи аналіз може викликати у вас сильні емоції, обурення чи незгоду. Тож звертайте увагу на те, хто саме це сказав, чи є відповідний досвід у людини, щоб коментувати такі речі, у якій організації вона працює тощо. Та сама порада стосується перевірки автора, який написав ту чи іншу новину або колонку на сайті. Його чи її думка — це тільки його чи її думка, і вона може не відображати реальної дійсності. Так, російські так звані експерти чи політологи найчастіше транслюють свою картину світу, просувають необхідні наративи пропаганди, а не справді аналізують ситуацію.

2.2.10. Підтримуйте прозорі практики в соціальних мережах

На практиці: Будьте обізнані з тим, як працюють платформи соціальних мереж, і виступайте за чітку політику протидії дезінформації. Використовуйте інструменти звітності для виявлення фейкових новин і підтримуйте ініціативи, які мають на меті зробити обмін інформацією більш прозорим і правдивим.

3. ВІДПОВІДАЛЬНІСТЬ ШТУЧНОГО ІНТЕЛЕКТУ В ПОШИРЕННІ ДЕЗІНФОРМАЦІЇ

Відстеження відповідальності за поширення дезінформації, керованої штучним інтелектом, є критично важливим завданням, яке передбачає вивчення ролей різних зацікавлених сторін у цифровій екосистемі. Складність технологій штучного інтелекту та їхня інтеграція в платформи соціальних мереж призвели до безпрецедентних викликів у виявленні дезінформації та боротьбі з нею.

Хто несе відповідальність, коли ШІ поширює дезінформацію? Це питання має важливе значення, оскільки ми намагаємося зрозуміти етичні та практичні наслідки застосування ШІ в нашому інформаційному ландшафті.

1. Розробники ШІ та технологічні компанії - відповідають за етичну розробку ШІ, включно із захистом від дезінформації. Їх роль полягає в тому, щоб забезпечити відповідальне та етичне використання технологій ШІ, запобігаючи зловживанням для поширення неправдивої інформації.

2. Платформи соціальних мереж - відповідають за модерацію контенту. Вони повинні співпрацювати з фактчекерами для виявлення та зменшення поширення неправдивої інформації, підтримуючи пильне управління контентом.

3. Урядові та регуляторні органи - відіграють ключову роль у формуванні та впровадженні політики щодо протидії зловживанню ШІ в дезінформації, дотримуючись при цьому балансу між свободою вираження поглядів і захистом приватного життя.

4. Медіа та організації, що займаються перевіркою фактів - повинні адаптуватися до дезінформації, керованої штучним інтелектом, використовуючи передові інструменти для перевірки фактів і контенту для боротьби з неправдивими наративами.

Спільна відповідальність. Боротьба з дезінформацією, керованою штучним інтелектом, вимагає узгоджених зусиль від усіх цих груп. Це включає етичну розробку ШІ, ефективну модерацію контенту, регуляторний нагляд, ретельну перевірку фактів і громадську освіту для створення добре поінформованої та критично налаштованої аудиторії.

Інтеграція штучного інтелекту в роботу медіа вимагає дотримання чітких принципів, які ставлять на перше місце інтереси суспільства. Це передбачає прийняття відповідальних редакційних рішень щодо впровадження AI-систем, забезпечення законності їх використання та публікації контенту, згенерованого за допомогою AI, а також регулярну оцінку правових і технічних ризиків протягом усього життєвого циклу AI-систем. Прозорість і ясність у розкритті інформації про використання ШІ, а також забезпечення обізнаності аудиторії про використання ШІ та характер поширюваного контенту мають вирішальне значення. Захист конфіденційності та даних, надання різноманітного і недискримінаційного контенту, забезпечення професійного людського нагляду за результатами ШІ, притягнення користувачів до відповідальності за наслідки використання ШІ та підтримка адаптивності для оновлення принципів у відповідь на технологічні та правові зміни також є ключовими аспектами.

У медіа-секторі використання систем штучного інтелекту має відповідати основним принципам журналістської етики, таким як правдивість, точність, неупередженість, незалежність, запобігання шкоді, недискримінація, підзвітність, інклюзивність, повага до приватного життя та конфіденційності джерел, як зазначено в Етичному кодексі журналіста. Медіа повинні визнавати свою редакційну відповідальність за використання ШІ для збору, обробки та поширення інформації. Згідно з Комісією з журналістських рекомендацій, відповідальність за журналістський

контент покладається як на автора, так і на редакцію, навіть якщо AI допомагає у його створенні. Генеративні результати AI завжди повинні проходити обов'язкову додаткову перевірку та ухвалення рішень людиною перед поширенням. Доцільно включити принципи використання AI в редакційні статути та розробити внутрішні правила етичного застосування AI як у створенні медіаконтенту, так і в операційній діяльності.

Варто також пам'ятами, що системи штучного інтелекту, розроблені онлайн-платформами, значною мірою сприяють ефективному і швидкому поширенню дезінформації, але вони також застосовуються для виявлення і зменшення поширення неправдивої інформації в Інтернеті. Етичні наслідки виникають у ситуації, коли ШІ можна використовувати як для створення та поширення дезінформації, так і для боротьби з нею. Ця подвійність вимагає ретельного вивчення ролі ШІ в сучасній інформаційній екосистемі, особливо в світлі захисту фундаментальних прав і свобод, зокрема свободи вираження поглядів та інформації.

Таким чином, відповідальність ШІ в контексті дезінформації є багатогранним питанням, яке вимагає ретельного розгляду як потенційної шкоди, так і можливостей протидії дезінформації. Розробка і впровадження технологій ШІ в цій сфері має відбуватися з дотриманням етичних принципів і зобов'язань щодо захисту цілісності інформації та демократичних процесів.

3.1. ОСНОВНІ ПРАВОВІ ДОКУМЕНТИ

У контексті ШІ та дезінформації було розроблено кілька європейських правових документів і рамок для вирішення проблем, пов'язаних з цими технологіями. Ці документи спрямовані на регулювання застосування ШІ, захист прав громадян і забезпечення відповідального поширення інформації. Ось деякі з найбільш значущих правових рамок і документів:

1) Кодекс практики ЄС щодо дезінформації (Code of Practice on Disinformation).

Кодекс практики ЄС щодо дезінформації, створений у 2018 році та посилений у 2022 році, є саморегулюючою структурою, спрямованою на боротьбу з дезінформацією в Інтернеті. Він передбачає зобов'язання онлайн-платформ, торгових асоціацій та ключових гравців у рекламному секторі. Кодекс покликаний забезпечити більшу прозорість і підзвітність онлайн-платформ і пропонує структуровану основу для моніторингу та вдосконалення політики платформ щодо

дезінформації. Він включає такі конкретні заходи, як посилення дій з демонетизації дезінформації, підвищення прозорості політичної та тематичної реклами, надання користувачам можливості виявляти та позначати неправдивий контент, розширення перевірки фактів у всіх країнах ЄС та надання дослідникам більшого доступу до даних. Кодекс також створив постійну робочу групу для розробки та адаптації заходів, забезпечуючи постійне реагування на динамічну природу дезінформації.

2) Загальний регламент про захист даних (General Data Protection Regulation).

Хоча GDPR в першу чергу зосереджений на захисті даних і конфіденційності, він має наслідки для ШІ та дезінформації, особливо щодо використання персональних даних у мікротаргетингу та профілюванні.

GDPR встановлює суворі правила щодо збору, зберігання та використання персональних даних, які впливають на те, як алгоритми штучного інтелекту можуть використовувати ці дані для створення таргетованого контенту. Цей регламент допомагає запобігти зловживанню персональними даними для дезінформаційних кампаній, гарантуючи, що будь-яка діяльність ШІ на основі даних, особливо та, що пов'язана з персональним профілюванням і таргетингом, відповідає суворим стандартам конфіденційності та згоди.

3) Закон ЄС про штучний інтелект (AI Act).

Закон ЄС про штучний інтелект, запропонований у 2021 році, є важливим законодавчим кроком, спрямованим на регулювання ШІ в Європейському Союзі. Він спрямований на встановлення стандартів для розробки та розгортання систем штучного інтелекту, забезпечуючи їхню відповідність цінностям і фундаментальним правам ЄС. Закон класифікує додатки ШІ за категоріями, виходячи з їхнього потенційного ризику для прав і безпеки людини. До категорій високого ризику належать системи ШІ, що використовуються для маніпулювання інформацією, що є прямою відповіддю на занепокоєння щодо використання ШІ для поширення дезінформації. Ця система класифікації підкреслює прихильність ЄС до зменшення ризиків, пов'язаних із технологіями штучного інтелекту, особливо тих, які можуть вплинути на демократичні процеси та громадську безпеку.

4) Закон про цифрові послуги (Digital Services Act).

Закон про цифрові послуги (DSA), запропонований Європейською Комісією, спрямований на створення більш безпечного та підзвітного цифрового простору в ЄС. Цей закон зосереджується на захисті основних прав користувачів в Інтернеті та передбачає заходи для більшої прозорості алгоритмів роботи онлайн-платформ. Він також розглядає питання незаконного контенту та дезінформації, встановлюючи чіткі обов'язки для постачальників цифрових послуг щодо вирішення цих проблем. DSA є важливим кроком на шляху до регулювання цифрового простору, гарантуючи, що він залишатиметься безпечним і надійним середовищем для користувачів.

5) Біла книга (AI White Paper).

Біла книга Великої Британії про шкоду в Інтернеті окреслює план уряду щодо створення нормативно-правової бази для боротьби з незаконним і шкідливим контентом в Інтернеті, включаючи дезінформацію. Вона має на меті запровадити законодавчо закріплений обов'я-

зок захищати користувачів від шкідливого контенту, що змусить інтернет-компанії захищати користувачів від шкідливого контенту. Ця ініціатива є важливим кроком у регулюванні онлайн-простору, забезпеченні безпечнішого користування інтернетом і боротьбі зі шкідливим контентом, зокрема дезінформацією, задля підтримки безпечного та надійного цифрового середовища.

Ці правові документи відображають зростаюче визнання необхідності створення надійної правової бази для управління викликами, пов'язаними із поширенням дезінформації за допомогою штучного інтелекту. Вони мають на меті збалансувати інновації та технологічний прогрес із захистом прав особистості та цілісності суспільного дискурсу.

У 2020 році в Україні було схвалено Концепцію розвитку штучного інтелекту. Ця ініціатива підкреслює необхідність доопрацювання нормативно-правової бази, що регулює розвиток штучного інтелекту. Комітет з питань розвитку штучного інтелекту при Міністерстві цифрової трансформації України активно працює над створенням нормативно-правової бази для регулювання ШІ. Ці зусилля включають вирішення проблем, пов'язаних зі створенням та поширенням дезінформації за допомогою штучного інтелекту, забезпечення врахування цих важливих питань при розробці нових нормативно-правових актів. Представники комітету також вважають за необхідне врахувати досвід Великої Британії, яка випустила Білу книгу з питань штучного інтелекту, що описує підхід уряду до збалансування регулювання та стимулювання розвитку штучного інтелекту, а також дає краще розуміння вектору розвитку штучного інтелекту для суспільства та бізнесу.

4. ЕТИЧНІ НАСТАНОВИ ДЛЯ ЖУРНАЛІСТІВ, РОЗРОБНИКІВ ШІ ТА МЕДІА-ПЛАТФОРМ В ЕПОХУ ПЕРЕДОВИХ ТЕХНОЛОГІЙ ШІ

В епоху, коли домінують передові технології штучного інтелекту, журналісти, розробники ШІ та медіа-платформи стикаються з унікальними етичними викликами та обов'язками. Журналісти повинні дотримуватися суворих стандартів перевірки фактів і прозорості, особливо при використанні контенту, створеного штучним інтелектом. Розробники ШІ, у свою чергу, зобов'язані гарантувати, що їхні творіння уникають упередженості та дезінформації, зберігаючи етичні аспекти на першому плані в процесі розробки. Медіа-платформи повинні дотримуватися тонкого балансу між захистом свободи слова та запобіганням поширенню дезінформації, використовуючи прозорі алгоритми ШІ для модерації контенту. Дотримання цих стандартів має вирішальне значення для збереження довіри та доброчесності у сфері цифрової інформації.

ЖУРНАЛІСТАМ

a. Переконайтеся, що вся інформація є точною та автентичною й не знегерована платформами штучного інтелекту.

b. Не публікуйте матеріли, створені штучним інтелектом, без попередньої вичитки та перевірки редакційної колегії.

c. Дотримуйтеся принципу прозорості та чітко розкривайте інформацію про використання ШІ у створенні свого контенту.

d. Розпізнавайте та зменшуйте упередженість контенту, створеного штучним інтелектом. Якщо ШІ навчений на вузькому колі джерел, він може створювати упереджений контент. Регулярно перевіряйте джерела, щоб підтримувати неупередженість і точність інформації.

e. Дотримуйтесь етичних стандартів при зборі та наданні інформації, особливо при використанні ШІ для аналізу даних.

f. Уникайте сенсаційності та поважайте гідність суб'єктів, особливо коли ШІ допомагає у створенні контенту. Перевірте чи контент не порушує немайнові права третіх осіб, зокрема, честь, гідність, ділову репутацію. Наприклад, якщо ви використовуєте штучний інтелект для написання статті на делікатну соціальну тему, він може спочатку запропонувати надто драматичний або провокаційний заголовок, щоб привернути увагу читача. Однак це може бути неповагою до тих, кого зачіпає ця проблема. У такому разі слід змінити заголовок так, щоб він точно й емпатично відображав зміст, передавав серйозність теми, не вдаючись до сенсаційності. Такий підхід поважає гідність задіяних суб'єктів, зберігаючи при цьому журналістські стандарти.

g. Використовуйте водяні знаки або спеціальні позначки, які чітко вказують на те, що контент створено за допомогою штучного інтелекту. Наприклад, додайте примітку на кшталт "Цей матеріал створено за допомогою штучного інтелекту".

РОЗРОБНИКАМ ШТУЧНОГО ІНТЕЛЕКТУ

a. Розробляйте ШІ з урахуванням етичних міркувань, зокрема справедливості, підзвітності та прозорості.

b. Постійно працюйте над виявленням та зменшенням упередженості алгоритмів ШІ. Регулярно переглядайте та оновлюйте навчальні набори даних, щоб переконатися, що вони грунтуються на принципах плюралізму. Здійснюйте регулярний аудит результатів ШІ для виявлення будь-яких упереджень. Співпрацюйте з різноманітною командою, включно з експертами в різних галузях і фахівцями з етики, щоб оцінити та вдосконалити алгоритми. Будьте в курсі останніх подій у сфері етики ШІ та впроваджуйте найкращі практики у свої системи ШІ.

c. Дотримуйтеся законодавства про захист персональних даних при розробці систем ШІ. Переконайтеся у можливості видалення персональних даних осіб та дотримуйтеся принципу конфіденційності.

d. Забезпечте надійні заходи захисту даних у системах ШІ.

e. Переконайтеся, що система є інклюзивною і не має негативного впливу на вразливі або маргіналізовані громади.

f. Взаємодійте із журналістами, медіа-платформами та громадськістю, щоб зрозуміти та вирішити етичні проблеми.

g. Залучайте до тестування системи різноманітних стейкхолдерів, а також створіть можливість зворотнього зв'язку, у разі порушення прав користувачів.

МЕДІА-ПЛАТФОРМАМ

a. Впроваджуйте модерацію контенту,застосовуйте політику щодо матеріалів, створених штучним інтелектом, гарантуючи, що він відповідає етичним та журналістським стандартам.

b. Чітко визначайте контент, створений штучним інтелектом, та його джерела.

c. Захищайте дані користувачів і конфіденційність, особливо коли ШІ використовується для персоналізації та аналітики.

d. Розкажіть громадськості про роль і вплив штучного інтелекту на створення та поширення контенту.

e. Переконайтеся, що реклама на основі ШІ є прозорою, чесною та поважає конфіденційність користувачів.

ВИСНОВКИ

У цифрову епоху, особливо в контексті українсько-російської війни, роль журналістів і творців контенту є як ніколи важливою. Хоча технології пропонують безпрецедентні можливості для створення історій та репортажів, вони також створюють значні виклики. Баланс між цими аспектами вимагає відданості етичним принципам журналістики, критичного аналізу контенту, створеного штучним інтелектом, і постійних зусиль з просвіти як творців, так і громадськості.

Цей посібник розглядає складну взаємодію між штучним інтелектом, журналістикою та дезінформацією. Особлива увага приділяється розумінню технологій штучного інтелекту, таких як машинне навчання, глибоке навчання та генеративний штучний інтелект, а також викликам, пов'язаним із глибокими фейками та дезінформацією, керованою штучним інтелектом. Розглядається ключова роль платформ соціальних мереж у поширенні дезінформації, підкреслюється важливість медіаграмотності.

Насамкінець, у посібнику окреслено етичні обов'язки журналістів, розробників штучного інтелекту та медіа-платформ, які пропагують відповідальні практики в цифрову епоху. Цей посібник підкреслює необхідність пильності, етичної обізнаності та безперервного навчання для збереження цілісності інформації в нашому цифровому світі, що швидко розвивається.