

Міністерство освіти і науки України
Національний університет «Києво-Могилянська академія»
Факультет соціальних наук і соціальних технологій
Могилянська школа журналістики

Магістерська робота

Освітній ступінь – магістр

на тему: «**Технології штучного інтелекту в практиці виявлення
фейкових новин**»

Виконав: студентка 2-го року навчання,
спеціальності 061 Журналістика

Сюйва Євген Ігорович

Керівниця: Романюк В. С.
кандидатка наук з культурології, доцентка

Рецензент _____

Магістерська робота захищена
з оцінкою _____

Секретар ЕК _____
« ____ » _____ 2024р.

Київ 2024

Зміст

Вступ.....	3
Розділ 1: Інтеграція технологій штучного інтелекту в аналізі та виявленні фейкових новин.....	7
1.1 Машинне навчання.....	14
1.2 Навчання моделей та використання великих даних.....	16
1.3 Застосування моделей у реальному часі.....	17
1.4 Шляхи та перспективи використання штучного інтелекту в протидії фейковим новинам	19
1.5 Дезінформація та фейкові новини: аналіз та методи боротьби.....	23
1.6 Регулювання ШІ в світі.....	25
Розділ 2: Еволюція фейкових новин. Дослідження історії виникнення та розвитку фейкових новин у сучасному інформаційному просторі.....	28
2.1 Вплив технологій на інформаційне суспільство. Вплив технологій на інформаційне суспільство, аналіз ролі технологій у масовій комунікації та їх взаємодія з поширенням фейкової інформації.....	32

2.2	Штучний інтелект та виявлення фейкових новин.....	36
-----	---	----

Розділ 3: Аналіз наявних інструментів які є на сьогоднішній день на ринку та рекомендації для майбутніх розробок для фактчекерів в галузі штучного інтелекту.....	50
---	----

3.1 Рекомендації для майбутніх розробок.....	52
--	----

4. Висновки.....	58
------------------	----

Список використаної літератури.....	64
-------------------------------------	----

Додатки.....	68
--------------	----

Додаток 1.....	68
----------------	----

Додаток 2.....	74
----------------	----

Вступ

“...повертаємося до людського фактору. Кожна ситуація має бути оцінена окремо і оцінена з точки зору саме людського творчого креативного потенціалу. Тобто там, де буде збережений ваш людський вимір, а не через машинний, там ви собі зможете забезпечити безпеку від ШІ...”

МАРІАННА ПРИСЯЖНЮК

З розвитком інтернету та соціальних медіа фейкові новини стали не лише проблемою інформаційного простору, а й серйозною загрозою для суспільства та демократичних процесів. Їх вплив може бути деструктивним, спричиняючи масову дезінформацію, вплив на громадську думку, а навіть вплив на політичні рішення.

Щоб протистояти цій проблемі, потрібно використовувати нові підходи та інструменти, серед яких технології штучного інтелекту (ШІ) відіграють ключову роль. ШІ може бути використаний для виявлення фейкових новин шляхом аналізу великих обсягів даних та виявлення незвичайних патернів або спільних характеристик між ними (Human Compatible: Artificial Intelligence and the Problem of Control" by Stuart Russell). Алгоритми машинного навчання можуть навчитися розпізнавати типові ознаки фейкових новин, такі як заголовки з сенсаційними заявами, відсутність достовірних джерел, або низька якість контенту.

Крім того, ШІ може допомогти впроваджувати системи фактчекінгу, шляхом автоматичного перевірки фактів у статтях та новинах. Це дозволить швидко і ефективно виявляти недостовірну інформацію та реагувати на неї.

Протидія фейковим новинам потребує комплексного підходу, що поєднує в собі технології ШІ, спільні зусилля соціальних мереж, журналістів та громадськості. Тільки таким чином можна ефективно боротися з цією загрозою та зберегти довіру до інформації в цифровому віці.

Метою даної роботи є дослідження впливу технологій штучного інтелекту на процес виявлення фейкових новин. Подібне дослідження дозволить не лише розкрити потенційні можливості, а й зазначити обмеження та виклики, що виникають у контексті боротьби з дезінформацією. Штучний інтелект відкриває широкі можливості для автоматизації процесів аналізу великих обсягів інформації та виявлення паттернів, що вказують на можливі фейкові новини ("Fake News: Understanding Media and Misinformation in the Digital Age" by Brian L. Ott and Cara Louise Buckley). Зокрема, алгоритми машинного навчання можуть навчитися виявляти недостовірні новини на основі таких ознак, як відсутність підтверджуючих джерел, відмінності у стилі публікації в порівнянні з автентичними джерелами, або широке поширення інформації без перевірки її достовірності.

Проте, разом з потенційними перевагами технологій штучного інтелекту у боротьбі з фейковими новинами виникають і обмеження. Наприклад, існує ризик недостатньої точності алгоритмів виявлення фейкових новин, що може призвести до помилкового класифікування деяких автентичних новин як фейкових, або навпаки. Також існує проблема адаптації фейкових новин до алгоритмів виявлення, що може призвести до появи більш вдосконалених та важкодоступних форм дезінформації.

Зважаючи на складність та поширеність проблеми фейкових новин, проведення ретельного аналізу існуючих методів виявлення є критичним етапом дослідження. Це включає в себе огляд як традиційних підходів, так і новітніх стратегій, що базуються на аналізі великих обсягів даних та використанні алгоритмів машинного навчання. Порівняльний аналіз цих методів дозволить виявити їхні переваги та

недоліки, а також ідентифікувати найбільш ефективні підходи до виявлення фейкових новин.

Окрім того, огляд сучасних технологій штучного інтелекту в контексті їхнього застосування для виявлення фейкових новин дозволить з'ясувати, які конкретні методи та алгоритми можуть бути найбільш корисними. Цей аналіз включатиме огляд таких аспектів як обробка природної мови, класифікація текстів, виявлення відхилень у структурі, а також роботу з великими обсягами даних.

Після аналізу існуючих методів та технологій робота перейде до практичної частини це дозволить оцінити ефективність та точність вибраних методів та технологій у виявленні фейкових новин. Крім того, робота також буде вдосконалювати існуючі методи та розробляти нові рекомендації до виявлення фейкових новин, враховуючи отримані результати та використовуючи передові досягнення в області штучного інтелекту.

Об'єктом дослідження є процес виявлення фейкових новин, який стоїть на перехресті суспільних, технологічних та інформаційних викликів сучасності. З поширенням інтернету та соціальних медіа фейкові новини стали значущим інструментом маніпуляції масовою свідомістю. Їхнє виявлення та боротьба з ними стали нагальною проблемою, яка вимагає комплексного підходу та використання передових технологій.

Суспільні виклики

Фейкові новини впливають на суспільство на багатьох рівнях:

- Фейкові новини можуть бути використані для маніпуляції виборцями, впливу на вибори та політичні процеси. Вони здатні розпалювати ненависть, поляризувати суспільство та підривати довіру до демократичних інститутів.

- Неправдиві повідомлення можуть поширювати страх, паніку та недовіру. Це особливо помітно під час кризових ситуацій, таких як пандемії чи природні катастрофи.
- Економічні наслідки: Фейкові новини можуть впливати на фінансові ринки, ділову репутацію компаній та загальну економічну стабільність.

Технологічні виклики

Процес виявлення фейкових новин вимагає розробки та впровадження передових технологічних рішень:

- Соціальні медіа генерують величезну кількість даних щоденно. Виявлення фейкових новин потребує потужних інструментів для аналізу цих даних в режимі реального часу.
- Використання алгоритмів машинного навчання для автоматичної класифікації новин як правдивих або фейкових є ключовим елементом. Це включає аналіз тексту, зображень та навіть відео.
- Виявлення патернів поширення інформації та ідентифікація ключових вузлів у мережі, які сприяють поширенню фейкових новин.

Інформаційні виклики

Виявлення та боротьба з фейковими новинами вимагають також вирішення ряду інформаційних питань:

- Створення і підтримка баз даних, що містять перевірену інформацію, є необхідним для автоматизованих систем виявлення фейкових новин.
- Підвищення рівня медіаграмотності серед населення, щоб люди могли самостійно розпізнавати фейкові новини.
- Співпраця між соціальними медіа платформами для створення спільних стандартів і технологій боротьби з фейковими новинами.

Комплексний підхід

Ефективна боротьба з фейковими новинами вимагає комплексного підходу, який включає:

- Поєднання методів машинного навчання, аналізу великих даних і мережевого аналізу для створення потужних інструментів виявлення фейкових новин.
- Залучення фахівців з різних галузей — соціологів, психологів, фахівців з комп'ютерних наук і журналістів для розробки ефективних стратегій.
- Підтримка з боку уряду та громадських організацій створення політик та ініціатив, спрямованих на протидію фейковим новинам, підтримка незалежних організацій, що займаються перевіркою фактів.

Предмет дослідження, а саме вплив технологій штучного інтелекту на процес виявлення фейкових новин, зосереджується на розробці та застосуванні різноманітних алгоритмів, моделей та інструментів, спрямованих на автоматизований аналіз та класифікацію інформації. Це охоплює різноманітні аспекти, такі як обробка природної мови, машинне навчання, аналіз текстів, виявлення аномалій та інші. Розробка ефективних інструментів на базі штучного інтелекту може значно полегшити та покращити процес виявлення фейкових новин, забезпечуючи швидкий та точний аналіз інформації в реальному часі.

Таким чином, дослідження об'єкта та предмета нашої роботи дозволить не лише глибше зрозуміти проблему фейкових новин, а й розробити нові підходи навіть можливо інструменти для ефективного виявлення та протидії їм, використовуючи потенціал технологій штучного інтелекту.

Вплив штучного інтелекту (ШІ) на процес виявлення фейкових новин є багатограним і динамічним. З одного боку, ШІ пропонує потужні інструменти для

автоматизації аналізу даних, виявлення патернів та класифікації інформації, що може значно покращити ефективність боротьби з фейками. Алгоритми машинного навчання можуть навчатися на великих масивах даних, виявляти ознаки фейкових новин, такі як неправдиві заголовки, відсутність достовірних джерел або маніпулювання мовою.

Розділ 1: Інтеграція технологій штучного інтелекту в аналізі та виявленні фейкових новин

У світі інформаційної дійсності, де висвітлення подій та вираження думок переплітаються у безкрайній танцювальній грі, теоретичний ландшафт досліджень з виявлення фейкових новин виступає як віддзеркалення умілого танцю інтелекту та технології. На цьому етапі розуміння суті фейкових новин є ключовим фрагментом паззла, щоб увійти у витончений світ аналізу та виявлення. (Uscinski, Joseph P. Fake News and False Flags. Rowman & Littlefield, 2020.)

Фейкові новини, як свідомо неправдива або оманлива інформація, виходять за межі звичайної реальності, прагнучи здійснити танець введення в оману та маніпулювання колективною свідомістю.

Початки фейкових новин можна відслідкувати в історії засобів масової інформації. Ще в античні часи поширювалися неправдиві відомості для маніпуляції громадською думкою. Протягом століть такі новини набували різних форм і застосовувалися у політичних, економічних та соціальних контекстах. Із зростанням доступу до цифрових платформ, особливо соціальних мереж, поширення фейкових новин стало значно швидшим та ширшим. Соціальні мережі, завдяки своїй здатності до швидкого розповсюдження інформації, створили нові можливості для дезінформації. Алгоритми, які оптимізують контент для залучення користувачів, можуть ненавмисно сприяти поширенню фейкових новин, надаючи їм перевагу в потоці інформації.

Дослідження ранніх випадків фейкових новин дозволяє визначити ключові етапи їх формування та вдосконалення. Історичний аналіз показує, що такі новини еволюціонували разом із розвитком технологій масової комунікації. Наприклад, у XVIII-XIX століттях зростання популярності газет супроводжувалося появою сенсаційних та неправдивих повідомлень. У XX столітті з розвитком радіо та телебачення фейкові новини набули нових форм, але лише з появою інтернету та соціальних мереж вони отримали можливість миттєвого глобального поширення. Штучний інтелект пропонує нові можливості для боротьби з фейковими новинами. Технології на основі штучного інтелекту можуть автоматично виявляти та класифікувати фейкові новини, аналізувати джерела інформації та ідентифікувати патерни поширення дезінформації. Розробка таких технологічних рішень потребує глибокого розуміння еволюції фейкових новин та їх впливу на суспільство.

- Типологія фейкових новин розкриває широкий спектр їхніх танцювальних рухів. Від пристрасних акордів пропаганди, що линує як мелодія, вплетена в тканину масової свідомості, до енергійних сенсаційних заголовків клікбейту, які танцюють над змістом статей, немов веселі маріонетки. Введення в оману, як спотворення реальності, викликає в нас питання про справжність та намагається зіпсувати ритм нашого розуміння подій. Deepfakes, подібно майстерним танцюристам, перетворюють реальність у віртуозний виступ, де межі між правдою та вигадкою розтікаються.
- Розглядаючи існуючі методи виявлення фейкових новин, ми опиняємось на майданчику, де аналіз контенту виявляє неправдиві факти та маркери фейків, а

машинне навчання стає партнером у виконанні автоматизованої класифікації правдивих та фейкових текстів. Аналіз джерела, як перевірка репутації, та аналіз поведінки, що слідкують за танцем новин в соціальних мережах, додають ритму інтриг та наповнюють наше розуміння суті проблеми.

- Потреба в ручній перевірці, складність виявлення нових типів фейків та необхідність великих обсягів даних створюють ритмічні перешкоди для повного осмислення цього танцю інтелекту та технології. Віддзеркалюючи широкий пейзаж інтеграції технологій штучного інтелекту в аналіз та виявлення фейкових новин, ми крокуємо в арену, де розуміння справжньої природи та мети фейкових новин – це танець, який вимагає глибокого поринання у світ технологій та аналітичного мистецтва.

Визначаючи типологію фейкових новин, виявляються різні варіанти цього явища. Пропаганда, спрямована на маніпулювання громадською думкою та поширення певної ідеології, стає інструментом впливу. Клікбейт, з його сенсаційними заголовками, які не відображають справжній зміст статті, створює штучний інтерес. Введення в оману, через спотворення фактів та вигадкування історій, плутає реальність та хибну інформацію. Технологічний аспект представлений deepfakes - фальсифікацією відео або аудіозаписів для створення обманливих зображень або звукового контенту.

Поширення фейкових новин через соціальні мережі, онлайн-платформи та інші канали може мати значний вплив на суспільство, підриваючи довіру до достовірних джерел інформації, маніпулюючи громадською думкою та спричинюючи соціальні та політичні потрясіння. Виникнення фейкових новин не є новим явищем, але завдяки розвитку технологій та доступності Інтернету їх масштаби та вплив значно зросли. Це призвело до зростання потреби в ефективних методах виявлення та

протидії фейковим новинам. Зараз ми розберемо декілька досліджень які зробили вагомий внесок в дослідження штучного інтелекту у виявленні фейкових новин.

Дослідження з використання штучного інтелекту для виявлення фейкових новин зосереджується на різноманітних підходах, що охоплюють широкий спектр методів та технологій. Один із головних напрямків цих досліджень - аналіз тексту через природну обробку мови (Natural Language Processing, NLP). Застосування алгоритмів NLP дозволяє проводити структурний аналіз текстових матеріалів, виявляти аномалії в мові та враховувати контекст та взаємозв'язки між словами і фразами.

Ще одним важливим підходом є аналіз мережевих взаємодій, зокрема поширення новин через соціальні мережі. У такому випадку алгоритми спроможні виявляти патерни у розповсюдженні фейкових новин, а також враховувати характеристики користувачів та їхні мережі для встановлення ймовірності недостовірності інформації.

Більшість досліджень використовують методи машинного навчання для створення моделей розпізнавання фейкових новин. Ці моделі навчаються на основі різноманітних даних, включаючи як достовірні, так і фейкові новини. Вони використовують різноманітні ознаки, такі як структура речень, лексичний аналіз та внутрішні патерни текстів, для виявлення потенційно фейкових новин.

Окрім цього, деякі дослідження зосереджені на аналізі зображень та відео за допомогою комп'ютерного зору. Це дозволяє виявляти ознаки фотошопу, монтажу та перевіряти відповідність вмісту тексту та візуального матеріалу. Узагальнюючи, дослідження у цій області використовують різноманітні підходи та техніки, включаючи аналіз тексту, мережевий аналіз, машинне навчання та комп'ютерний зір, для ефективного виявлення фейкових новин. Ці підходи постійно розвиваються та удосконалюються для кращого забезпечення впевненості у надійності інформації в медіа.

Дослідження групи вчених з Массачусетського технологічного інституту (МІТ) було спрямоване на створення та впровадження системи, що базується на методах машинного навчання для виявлення фейкових новин. Це дослідження виявилось вкрай актуальним, оскільки розповсюдження фейкових новин через соціальні мережі, такі як Twitter та Facebook, стало серйозною проблемою. Для цього дослідження вчені з МІТ приділили особливу увагу аналізу широкої розповсюдженості фейкових новин в цих мережах. Завданням було створення моделі, яка могла б аналізувати текст новин, оцінювати їхню достовірність та враховувати рівень популярності їх поширення. Для досягнення цієї мети вчені використали навчання з учителем на основі набору даних, що містить як фейкові, так і реальні новини. Це дозволило системі навчитися розпізнавати характеристики фейкових новин та ефективно виявляти їх на соціальних мережах. Основне цього дослідження полягав у використанні методів машинного навчання для створення моделі, яка могла б автоматично аналізувати великі обсяги даних та виявляти підозрілі новини. Важливою перевагою було використання навчання з учителем, оскільки це дало можливість системі виявляти навіть ті новини, які могли бути важко розрізнити для людини.

Дослідження, проведене в рамках Microsoft Research, представляє собою важливий крок у використанні глибокого навчання для аналізу фейкових новин. У цьому дослідженні, вчені зосереджувалися на розробці та навчанні нейронних мереж, призначених для виявлення ознак фальсифікації в текстовому та візуальному контенті. Зокрема, аналіз мовлення та використання слів з певним емоційним забарвленням стали одними з ключових аспектів цього дослідження. Нейронні мережі були навчені розпізнавати відмінності у мовленні, які можуть вказувати на недостовірність інформації. Крім того, велика увага приділялася аналізу цілісності візуального матеріалу, такого як зображення та відео. Алгоритми виявлення фейкових новин були навчені розпізнавати ознаки фотошопу, монтажу та інших форм маніпуляцій з візуальним контентом.

Одним з ключових аспектів цього дослідження була обробка великого обсягу даних. Нейронні мережі потребували великої кількості прикладів фейкових та достовірних новин для ефективного навчання. Це дозволило створити моделі, які могли ефективно виявляти фейкові новини та надавати важливу підтримку у боротьбі з ними.

У рамках дослідження, проведеного університетом Стенфорда, вчені використовували природну обробку мови для виявлення фейкових новин. Їхня методика полягала в аналізі відповідності між заголовками новин та змістом відповідних статей. В результаті аналізу виявлено, що багато фейкових новин мають заголовки, що привертають увагу, але не відображають справжнього змісту матеріалу.

Основною метою цього дослідження було розроблення моделі, яка за допомогою методів машинного навчання здатна автоматично аналізувати текст новин та виявляти невідповідності між заголовком та змістом. Для цього використовувалися алгоритми, що навчалися на великому обсязі даних, які включали як достовірні, так і фейкові новини. У результаті експериментів вдалося створити модель, яка з високою точністю виявляла фейкові новини за їхніми заголовками та порівнювала їх зміст з відомими стандартами об'єктивності та достовірності. Це дослідження відкриває можливості застосування штучного інтелекту для боротьби з поширенням дезінформації та фейкових новин, допомагаючи забезпечувати користувачам більш об'єктивний та достовірний інформаційний простір.

Дослідження проведене університетом Карнегі-Меллона зосереджувалося на використанні методів мережевого аналізу для виявлення фейкових новин. Вчені аналізували різноманітні патерни поширення дезінформації через соціальні мережі, досліджуючи, як інформація поширюється між користувачами та як швидко вона набирає популярність. Ця робота передбачала не лише аналіз вірогідності

поширення, а й виявлення аномалій, що можуть вказувати на можливість фейкової інформації.

Одним із ключових аспектів дослідження було використання цих даних для створення моделі, яка автоматично розпізнавала аномалії у поширенні інформації та сприяла виявленню фейкових новин. Це включало аналіз активності користувачів, швидкості поширення новин та взаємодії між ними в мережі. В результаті, вони розробили ефективний інструмент, який допомагає вчасно виявляти та запобігати поширенню дезінформації в онлайн середовищі. Це дослідження відображає значний прорив у сфері боротьби з фейковими новинами, пропонуючи не лише технологічні рішення для виявлення дезінформації, але і високий рівень аналізу та розуміння механізмів поширення інформації в сучасному цифровому середовищі.

Переходячи до існуючих методів виявлення фейкових новин, робиться акцент на різнобічному підході до проблеми. Аналіз контенту включає в себе виявлення неправдивих фактів, логічних помилок та емоційної мови, використовуючи машинне навчання для автоматизованої класифікації текстів на правдиві та фейкові.

Аналіз джерела інформації стає не менш важливим етапом виявлення фейків.

Перевірка репутації сайту та автора, а також виявлення ботів та фейкових акаунтів, визначають достовірність інформації. Аналіз поведінки в соціальних мережах, включаючи відстеження поширення новин та виявлення аномальних патернів, допомагає визначити потенційно обманливу інформацію.

Однак існуючі методи мають свої обмеження. Ручна перевірка виявляється трудомісткою, необхідність великих наборів даних для машинного навчання може

ускладнити процес, а складність виявлення нових типів фейків свідчить про постійний розвиток цього явища.

Сучасні методи виявлення фейкових новин характеризуються комплексним підходом до аналізу контенту. Цей підхід включає:

- Виявлення неправдивих фактів. За допомогою перехресних посилань на авторитетні джерела та використання фактологічних баз даних можна ідентифікувати неточну або вигадану інформацію.
- Аналіз логічних помилок. Логічні невідповідності, маніпуляції даними та некоректні узагальнення свідчать про фейковий характер тексту.
- Виявлення емоційної мови. Наявність сенсаційних заголовків, емоційно забарвлених слів та маніпулятивних прийомів може вказувати на фейк.

1.1 Машинне навчання

Алгоритми машинного навчання, навчені на великих масивах даних, здатні з високою точністю ідентифікувати фейкові новини на основі вищезазначених характеристик контенту.

Основні підходи в машинному навчанні для виявлення фейкових новин

1. Класифікація на основі ознак (Feature-based Classification) Машинне навчання використовує різні ознаки тексту для навчання моделей. Ці ознаки можуть включати лінгвістичні, семантичні та контекстуальні аспекти тексту.

- Моделі можуть враховувати частоту вживання слів, граматичні конструкції, довжину речень, та інші статистичні характеристики тексту.
- Використання методів обробки природної мови, таких як Word2Vec або GloVe, дозволяє моделі розуміти значення слів та їх взаємозв'язки у тексті.

- Аналіз контексту, в якому з'являється текст, може включати інформацію про автора, час публікації та платформу, на якій було опубліковано новину.

2. Наївні байєсівські класифікатори (Naive Bayes Classifiers) Це один з простих, але ефективних підходів до класифікації текстів. Наївний байєсівський класифікатор використовує теорему Байєса для обчислення ймовірності того, що текст належить до певної категорії (фейкові або правдиві новини).

3. Логістична регресія (Logistic Regression) Логістична регресія застосовується для бінарної класифікації і дозволяє оцінити ймовірність належності тексту до класу фейкових новин. Цей метод є популярним через свою простоту та ефективність.

4. Рішення дерев та випадкові ліси (Decision Trees and Random Forests).

- Алгоритм створює дерево рішень, де кожний вузол представляє вибір, заснований на одній з ознак тексту. Рішення дерева працюють швидко та інтуїтивно зрозумілі.
- Це ансамблевий метод, що використовує безліч рішень дерев для прийняття рішення. Випадкові ліси зменшують ризик переобучення та підвищують точність класифікації.

5. Підтримувальні векторні машини (Support Vector Machines, SVM) SVM використовують для побудови гіперплощини, що максимально відокремлює класи текстів. Цей метод особливо ефективний при роботі з великими обсягами текстових даних і високою розмірністю ознак.

6. Глибоке навчання (Deep Learning) Глибокі нейронні мережі, зокрема, рекурентні нейронні мережі (RNN) та конволюційні нейронні мережі (CNN), широко застосовуються для аналізу тексту.

- Рекурентні нейронні мережі (RNN): RNN, особливо LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit), добре справляються з послідовними даними, такими як текст, завдяки своїй здатності зберігати контекст попередніх слів у послідовності.
- Конволюційні нейронні мережі (CNN): CNN, хоча традиційно використовуються для обробки зображень, також можуть застосовуватися для текстової класифікації, виявляючи локальні патерни в тексті.

1.2 Навчання моделей та використання великих даних

Процес навчання моделей машинного навчання для виявлення фейкових новин включає кілька етапів:

- Збір великих обсягів даних, включаючи як фейкові, так і правдиві новини. Ці дані можуть бути зібрані з різних джерел, таких як соціальні медіа, новинні сайти та бази даних перевірки фактів.
- Очищення даних, включаючи видалення зайвих символів, нормалізацію тексту та видалення стоп-слів. Цей етап також може включати токенізацію тексту та його перетворення у векторні представлення.
- Розподіл даних на навчальний, валідаційний та тестовий набори для забезпечення надійної оцінки продуктивності моделі.
- Використання різних алгоритмів машинного навчання для навчання моделей на основі зібраних та оброблених даних.
- Оцінка продуктивності моделей за допомогою метрик, таких як точність, повнота, F1-міра. Оптимізація моделей шляхом налаштування гіперпараметрів та використання методів ансамблю.

1.3 Застосування моделей у реальному часі

Після навчання моделі можуть бути інтегровані у різні системи для виявлення фейкових новин у реальному часі. Це може включати:

- Автоматизовані системи, що сканують новинні сайти та соціальні медіа на наявність фейкових новин.
- Програмне забезпечення, що допомагає журналістам швидко перевіряти достовірність інформації.
- Інструменти, що надають аналітику та звіти про поширення фейкових новин.

Аналіз джерела інформації. Окрім аналізу контенту, важливим етапом виявлення фейків є ретельне вивчення джерела інформації. Це включає:

- Авторитетність та надійність видання, на якому опубліковано новину, можуть свідчити про її достовірність.
- Досвід, кваліфікація та попередні роботи автора дають уявлення про його надійність.
- Анонімність, штучність поведінки та масове поширення фейкових новин з фейкових акаунтів є ознаками фейку.

Аналіз поведінки в соціальних мережах. Соціальні мережі дають можливість відстежувати поширення новин та виявляти аномальні патерни, які можуть свідчити про фейк:

- Раптовий сплеск активності та масове поширення новин з неперевірених джерел може вказувати на фейк.
- Непропорційна кількість лайків, коментарів та репостів: Штучна активність та співвідношення лайків, коментарів та репостів, що відрізняється від норми, може бути ознакою фейку.

- Поширення ненависті, дезінформації та ворожнечі в коментарях до новин може свідчити про фейк.

Незважаючи на значні досягнення, існуючі методи виявлення фейкових новин мають ряд обмежень:

- Перевірка новин вручну трудомістка, що робить її неефективною для обробки великих обсягів інформації.
- Алгоритми машинного навчання потребують великих наборів даних для тренування, що може бути проблемою для нових типів фейків.
- Зловмисники постійно вдосконалюють методи створення фейків, що ускладнює їх виявлення.

Використання технологій штучного інтелекту, зокрема машинного навчання, має величезний потенціал у боротьбі з фейковими новинами. Завдяки здатності аналізувати великі масиви даних і виявляти складні патерни, алгоритми машинного навчання забезпечують автоматизований і точний аналіз інформації. Основні підходи, такі як наївні байесівські класифікатори, логістична регресія, рішення дерев, випадкові ліси, підтримувальні векторні машини та глибокі нейронні мережі, дозволяють створювати ефективні моделі для класифікації текстів. Моделі, навчені на великих даних, здатні ідентифікувати фейкові новини з високою точністю, використовуючи лінгвістичні, семантичні та контекстуальні ознаки тексту. Цей процес включає кілька етапів, від збору та передобробки даних до навчання та оптимізації моделей. Впровадження цих моделей у реальні системи моніторингу новин, інструменти для журналістів та аналітичні платформи забезпечує своєчасне виявлення та запобігання поширенню неправдивої інформації.

Таким чином, розвиток та застосування машинного навчання у сфері виявлення фейкових новин сприяє підвищенню достовірності інформаційного простору, зміцненню громадської довіри до медіа та зниженню негативного впливу фейкових новин на суспільство. Ефективні інструменти на базі ШІ допомагають забезпечити швидкий та точний аналіз інформації, що є критично важливим у сучасному світі з його великими обсягами даних та швидким обміном інформацією.

1.4 Шляхи та перспективи використання штучного інтелекту в протидії фейковим новинам

В умовах сучасного інформаційного оточення проблема фейкових новин та дезінформації набуває привертаючого увагу статусу, викликаючи значні загрози для суспільства та демократичних процесів. Одним із сучасних підходів до вирішення цієї проблеми є використання штучного інтелекту (ШІ), що стає інноваційною стратегією в протидії фейковим новинам. ШІ може ефективно застосовуватися для автоматизованого аналізу об'ємних наборів інформації з метою виявлення аномалій та неправдивих даних (Jones et al., 2020). Алгоритми машинного навчання можуть навчитися розпізнавати патерни фейкових новин, враховуючи лінгвістичні особливості та структуру інформації (Li et al., 2019). Цей підхід сприяє ефективному фільтруванню неправдивої інформації на рівні джерела та контенту, сприяючи зниженню її поширення (Wang & Zhang, 2022). Системи ШІ також можуть виявити застосування в аналізі соціальних мереж та медіа-платформ для виявлення та обмеження поширення фейкових новин (Gupta et al., 2021).

Автоматизовані агенти можуть взаємодіяти з користувачами, надаючи їм достовірну інформацію та підвищуючи рівень обізнаності стосовно феномену дезінформації.

Глибоке навчання та аналіз мови в ШІ можуть бути використані для розробки систем, спроможних автоматично виявляти маніпулятивні прийоми в фейкових новинах (Zhang & Liu, 2018). Визначення емоційно навантаженого мовлення та аналіз тональності тексту дозволяє виявляти маніпуляційно орієнтовані стилі написання (Kim et al., 2017).

Зрозуміло, що впровадження ШІ у стратегії протидії фейковим новинам вимагає розвитку технологій, співпраці з медіа-організаціями, а також урахування етичних аспектів та забезпечення прозорості в роботі систем (Bai et al., 2021; Wu et al., 2023). Надійність та успішність стратегії також залежать від співпраці з фактчекерами та активного розвитку освітніх ініціатив (Li & Chen, 2019). Використання ШІ в створенні інтерактивних педагогічних ігор та онлайн-курсів може сприяти підвищенню рівня критичного мислення та обізнаності громадян (Wang et al., 2020). Завершальним кроком у розвитку стратегії може бути створення міжнародних стандартів та спільних ініціатив для об'єднання зусиль різних країн у боротьбі з глобальною проблемою фейкової інформації (International AI Consortium, 2024). Це дозволить уніфікувати підходи та сприяти глобальному вирішенню проблеми дезінформації.

Все більшого значення набуває роль штучного інтелекту (ШІ) у сучасних стратегіях протидії фейковим новинам та дезінформації. Зокрема, ШІ може застосовуватися для автоматизованого аналізу великих обсягів інформації з метою виявлення аномалій та неправдивих даних. Його алгоритми машинного навчання навчаються розпізнавати патерни фейкових новин, аналізуючи лінгвістичні особливості та структуру інформації. Підхід, який передбачає використання ШІ, дозволяє ефективно фільтрувати та обмежувати поширення неправдивої інформації на рівні джерела та контенту. Особливу перспективність представляє аналіз соціальних мереж та медіа-платформ, де системи ШІ можуть виявляти та пригнічувати поширення фейкових новин.

Важливим внеском ШІ є його здатність аналізувати не лише текстову, а й емоційну складову мовлення. Системи глибокого навчання та аналізу мови можуть ефективно виявляти маніпулятивні прийоми, що використовуються в контексті фейкових новин. Розпізнавання емоційно навантаженого мовлення та аналіз тональності тексту стають інструментами для виявлення маніпуляційно орієнтованих стилів написання. При впровадженні ШІ в стратегії протидії фейковим новинам необхідно враховувати етичні аспекти та забезпечити прозорість у роботі систем. Співпраця з медіа-організаціями та розробка стандартів для інтеграції ШІ є важливим етапом у забезпеченні успішної реалізації цього підходу. Важливою частиною стратегії є також співпраця з фактчекерами та активний розвиток освітніх ініціатив для підвищення рівня критичного мислення серед громадян. Використання ШІ у створенні ефективних освітніх інструментів, таких як інтерактивні педагогічні ігри та онлайн-курси, сприяє формуванню обізнаності та усвідомленості стосовно проблем дезінформації.

Однією з ключових аспектів впровадження штучного інтелекту (ШІ) в стратегії протидії фейковим новинам є розробка та вдосконалення технологій, які лежать в основі алгоритмів машинного навчання для ефективного виявлення патернів фейкових новин. Однак для забезпечення успішного функціонування таких систем важливо також враховувати етичні аспекти використання ШІ, забезпечуючи прозорість та адаптабельність до змін у медіа-просторі. Інтеграція ШІ в аналіз соціальних мереж та медіа-платформ вимагає розробки інтелектуальних агентів, які взаємодіють з користувачами та надають вірогідну інформацію. Ці агенти можуть відігравати ключову роль у зменшенні поширення фейкових новин та сприяти інформаційній грамотності в суспільстві.

Додатково, розвиток систем на основі глибокого навчання та аналізу мови може виявити важливим у розпізнаванні маніпулятивних прийомів, що використовуються в фейкових новинах. Розпізнавання емоційно навантаженого мовлення та аналіз тональності тексту є складовими, які сприяють ефективному виявленню стилів

написання, спрямованих на маніпуляції. Поряд із технічними аспектами важливо враховувати роль фактчекерів та експертів у валідації інформації.

У словах Гарвардського професора Майкла Сміта, "Системи ШІ повинні бути узгоджені з людським експертним аналізом для забезпечення точності та достовірності виявлення фейкових новин". Забезпечення сталої та ефективної стратегії протидії фейковим новинам також включає розробку освітніх ініціатив та педагогічних ігор, використовуючи ШІ для підвищення критичного мислення громадян. Як відзначив Чжень Лі, "Посилення освітніх зусиль дозволить суспільству бути більш інформованим та відповідальним у сприйнятті інформації". Завершальним етапом стратегії може бути створення міжнародних стандартів та координація зусиль на міжнародному рівні. Як підкреслює Міжнародний Консорціум ШІ, "Лише через спільні зусилля та взаємодію на міжнародному рівні можна досягти значущих результатів у боротьбі з дезінформацією" (International AI Consortium, 2024).

Отже, використання ШІ у протидії фейковим новинам визнається як перспективна стратегія, яка вимагає не тільки розвитку технологій, але й уваги до етичних, освітніх та міжнародних аспектів. Засоби штучного інтелекту можуть стати потужним інструментарієм для ефективного вирішення проблеми дезінформації в інформаційному просторі.

1.5 Дезінформація та фейкові новини: аналіз та методи боротьби

Відділеною, але не менш важливою, частиною дослідження є аналіз сутності дезінформації та розробка ефективних методів для її боротьби. Описуючи феномен дезінформації, слід звертати увагу на різноманітні механізми та стратегії, використовувані для створення та поширення фейкових новин.

Дослідження включає аналіз лінгвістичних, культурних та соціальних аспектів дезінформації, зокрема вивчення психологічних механізмів, які зумовлюють прийняття неправдивої інформації. Важливо визначити ключові фактори, що сприяють ефективній розповсюдженості фейкових новин, зокрема вплив соціальних мереж та алгоритмів рекомендацій.

Методи боротьби повинні включати розробку навчальних програм, спрямованих на підвищення медіаграмотності, а також створення ефективних механізмів виявлення та блокування поширювачів фейкових новин. Важливо розглядати також правові та етичні аспекти використання алгоритмів та ШІ в цілях боротьби з дезінформацією для забезпечення справедливості та дотримання основних прав людини. Цей аспект дослідження допомагає інтегрувати підходи до використання ШІ та розробки стратегій боротьби з фейковими новинами в аспекті дезінформації. Щодо автоматизованої фільтрації та контролю інформації, ефективність сучасних алгоритмів машинного навчання, виявлення патернів та їхній потенцій уникнення пристосувань соціальних інженерних атак можуть бути досліджені згідно з результатами Conroy та його співавторів (2015).

Зокрема, важливо зазначити, що розробка адаптивних алгоритмів та систем, які постійно апдейтяться під впливом нових стратегій дезінформації (Horne & Adali, 2017), є важливою частиною дослідження в цьому напрямку. Результати такого аналізу можуть сприяти побудові більш динамічних та реактивних систем, спроможних адаптуватися до зростаючої різноманітності стратегій фейкових новин. Крім того, розгляд аспектів психології сприйняття інформації (Lewandowsky et al., 2017) може стати основою для розробки інструментів, спрямованих на покращення критичного мислення та аналізу інформації серед кінцевих користувачів. Врахування психологічних аспектів може покращити ефективність освітніх програм та розвивати підвищену свідомість серед громадськості щодо методів маніпуляції та розповсюдження дезінформації. Висвітлюючи розмаїття аспектів штучного інтелекту

та аналізу фейкових новин, а також враховуючи рекомендації щодо методів боротьби з дезінформацією, розробленими у зазначених дослідженнях, можна сформулювати висновки, що сприятимуть подальшій розробці та впровадженню інноваційних рішень в області інформаційної безпеки та боротьби з фейковими новинами.

Проведений аналіз літератури дозволив визначити ключові аспекти використання ШІ у протидії фейковим новинам та дезінформації. Зокрема, висвітлено ефективність алгоритмів машинного навчання для фільтрації неправдивої інформації та виявлення патернів фейкових новин. Розглянуті методи аналізу соціальних мереж та глибокого навчання дозволяють визначити маніпулятивні прийоми та створювати системи, що адаптуються до нових стратегій дезінформації. Дослідження також виявило, що розвиток методів боротьби з дезінформацією потребує інтеграції технічних, соціальних, етичних та правових аспектів. Аналіз психологічних механізмів сприйняття інформації підкреслює важливість розвитку інтегрованих стратегій, які поєднують технічні та соціокультурні підходи. Одним із ключових результатів досліджень є визначення перспектив використання ШІ для ефективного виявлення та боротьби з фейковими новинами.

На основі аналізу літератури було встановлено, що алгоритми машинного навчання виявляють високу ефективність у фільтрації неправдивої інформації. Роботи, такі як та Zannettou et al. (2019), акцентують увагу на використанні технологій машинного навчання для аналізу патернів та характеристик фейкових новин. Це дозволяє розробити системи, які навчаються розпізнавати та класифікувати інформацію на основі попередньої експозиції до дезінформаційних впливів.

Використання алгоритмів машинного навчання дозволяє створити системи, які можуть динамічно адаптуватися до нових стратегій фейкових новин та забезпечити більшу точність виявлення неправдивої інформації. (Russell, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, 2019.) Це робить

їх ефективним інструментом для автоматизованого виявлення та фільтрації дезінформації у великому обсязі інтернет-простору.

Результати досліджень також підкреслюють важливість аналізу соціальних мереж та використання глибокого навчання у виявленні та аналізі фейкових новин. Алгоритми глибокого навчання, зокрема в контексті аналізу зображень та текстів, можуть ефективно розпізнавати характеристики маніпуляційного контенту (Horne & Adali, 2017). Аналіз соціальних мереж, як вказано Guess et al. (2019), надає можливість виявлення та вивчення патернів поширення фейкових новин серед користувачів. Це важливо для розробки стратегій блокування поширювачів дезінформації та зменшення її впливу на громадську думку.

Розробка ефективних методів боротьби з дезінформацією вимагає інтеграції технічних рішень із соціокультурними стратегіями. Наприклад, розробка навчальних програм, які спрямовані на підвищення медіаграмотності (Conroy et al., 2015), може впливати на здатність громадян визначати та відрізняти дезінформацію. Засвоєння принципів критичного мислення та аналізу інформації сприяє формуванню високоінформованих та критично налаштованих громадян. Такі програми можуть стати ефективним доповненням технічних рішень, розвиваючи обізнаність та вміння визначати неправдиву інформацію.

1.6 Регулювання ШІ в світі

Регулювання штучного інтелекту (ШІ) в світі стає все більш актуальним, особливо з урахуванням його потенціалу та впливу на суспільство і економіку. Оновлений GDPR та інші закони про дані від 13 березня 2024 року відводять особливе місце генеративному штучному інтелекту (наприклад, аудіо) та дідфейкам, виділяючи їх у відповідну категорію. Нові вимоги стосуються:

- Розкриття та маркування контенту, створеного штучним інтелектом. Цей аспект виникає з потреби прозорості та визначення джерела інформації для користувачів. Коли контент створюється за допомогою штучного інтелекту, особливо у випадках, коли він симулює людську мову або поведінку, користувачам слід бути повідомлено про це. Маркування такого контенту дозволяє людям розрізняти між автоматично згенерованим вмістом та тим, що створене реальними людьми. Це допомагає підвищити довіру до інформації та уникнути потенційних випадків маніпуляції.
- Розробка моделей для запобігання поширенню незаконного контенту. Однією з ключових викликів у сфері контенту в Інтернеті є поширення незаконного вмісту, такого як насильницький або дитяче порно, заклики до насильства, расизму тощо. Розробка технологій для виявлення та блокування цього контенту має велике значення для безпеки користувачів та забезпечення дотримання законодавства. Це може включати в себе використання алгоритмів машинного навчання для автоматичного виявлення незаконного вмісту, а також співпрацю з правоохоронними органами для вжиття необхідних заходів.
- Захист авторських прав. Зі зростанням використання штучного інтелекту для створення контенту стає важливим забезпечити захист авторських прав. ШІ може використовувати великі обсяги даних для створення нового контенту, який може бути захищений авторським правом. Законодавство повинно забезпечити, що автори отримують відповідне визнання та компенсацію за свою працю, навіть якщо цей контент створений за допомогою штучного інтелекту. Це може включати в себе розробку нових правових механізмів для визнання авторства та розподілу прибутку від використання такого контенту.

- Закон "ЄС оновлене GDPR та закон про дані" з 13 березня 2024 року визначає регулювання в галузі штучного інтелекту, зокрема генеративного штучного інтелекту (AI) та дідфейків. Цей закон призначений для захисту прав та інтересів користувачів, а також забезпечення етичного використання технологій штучного інтелекту, що можуть створювати аудіо контент.
- Розкриття та маркування контенту створеного штучним інтелектом. Закон вимагає, щоб будь-який контент, створений з використанням генеративного штучного інтелекту, був чітко відмічений як такий. Це означає, що будь-яка аудіо-продукція, яка була згенерована AI, повинна мати спеціальну маркування, яка інформує користувача про його походження.
- Розробка моделі для запобігання розвитку незаконного контенту. Закон вимагає від розробників штучного інтелекту розробки ефективних алгоритмів та моделей, спрямованих на виявлення та запобігання розвитку незаконного контенту, такого як шкідливі дідфейки, що можуть використовуватися для маніпулювання аудиторією або поширення дезінформації.
- Захист авторських прав. Закон також містить положення про захист авторських прав в контексті штучного інтелекту. Це означає, що розробники AI та дідфейків повинні дотримуватися авторських прав на вміст, який вони створюють, та не порушувати права інших осіб на їхні твори.

Регулювання штучного інтелекту (ШІ) в світі набуває все більшого значення у зв'язку з його потенціалом і впливом на суспільство та економіку. Оновлене законодавство, таке як GDPR та інші закони про дані, що були прийняті 13 березня 2024 року, визначає генеративний штучний інтелект (зокрема аудіо) та дідфейки як

окрему категорію, на яку поширюються специфічні вимоги. Серед них розкриття та маркування контенту, створеного ШІ, розробка моделей для запобігання поширенню незаконного контенту та захист авторських прав. Ці заходи спрямовані на забезпечення етичного та відповідального використання технологій ШІ, а також на захист прав і інтересів користувачів і авторів контенту.

Розділ 2: Еволюція фейкових новин. Дослідження історії виникнення та розвитку фейкових новин у сучасному інформаційному просторі

Дослідження еволюції фейкових новин є ключовим етапом у формуванні наукового підґрунтя для розуміння та протидії цьому явищу в сучасному інформаційному оточенні. Історія виникнення та розвитку фейкових новин відображає динаміку та трансформацію цього явища, надаючи ключові висновки для розробки технологічних рішень на базі штучного інтелекту.

Початки фейкових новин можна відслідкувати в історії засобів масової інформації, але із зростанням доступу до цифрових платформ, особливо соціальних мереж, їх поширення стало значно швидшим та ширшим. Дослідження ранніх випадків фейкових новин дозволить визначити ключові етапи їх формування та вдосконалення. Аналіз впливу соціальних мереж на поширення фейкових новин є необхідним для розуміння механізмів їх поширення та взаємодії з аудиторією. Розвиток інтернет-платформ, таких як Facebook, Twitter та інші, вплинув на структуру та інтенсивність поширення дезінформації. Це дослідження дозволить ідентифікувати основні тенденції та патерни, які супроводжують поширення фейкових новин через соціальні мережі. Використання технологій штучного інтелекту в дослідженні історії фейкових новин є ключовим компонентом методології. Алгоритми машинного навчання можуть аналізувати великі обсяги історичних даних, виявляти патерни та еволюцію методів поширення дезінформації.

Розробка моделей прогнозування та ідентифікації етапів розвитку фейкових новин дозволить забезпечити більш глибоке розуміння цього явища.

Дослідження еволюції фейкових новин є ключовим етапом у формуванні наукового підґрунтя для розуміння та протидії цьому явищу в сучасному інформаційному оточенні. Історія виникнення та розвитку фейкових новин відображає динаміку та трансформацію цього явища, надаючи ключові висновки для розробки технологічних рішень на базі штучного інтелекту.

Початки фейкових новин можна відслідкувати в історії засобів масової інформації, але із зростанням доступу до цифрових платформ, особливо соціальних мереж, їх поширення стало значно швидшим та ширшим. Дослідження ранніх випадків фейкових новин дозволить визначити ключові етапи їх формування та вдосконалення.

За словами Джозефа Тьюріта, визначення історії фейкових новин є необхідним для розуміння їх впливу на суспільство: "Розуміння корінь і етапи розвитку фейкових новин є критичним для ефективної боротьби із цим явищем. Історія виявлення та реакції на фейкові новини може стати основою для розробки стратегій протидії цьому інформаційному виклику." [Тьюріт, Д., "Реакція на фейкові новини", 2018]. Аналіз впливу соціальних мереж на поширення фейкових новин є необхідним для розуміння механізмів їх поширення та взаємодії з аудиторією. Розвиток інтернет-платформ, таких як Facebook, Twitter та інші, вплинув на структуру та інтенсивність поширення дезінформації.

Про це свідчить Джеймс Орсіні, який визначає роль соціальних мереж у поширенні фейкових новин: "Соціальні мережі стали ідеальним майданчиком для поширення фейкових новин, оскільки вони створюють сприятливе середовище для швидкого та широкого розповсюдження дезінформації." [Орсіні, Д., "Вплив соціальних мереж на інформаційний ландшафт", 2019]. Використання технологій штучного інтелекту в

дослідженні історії фейкових новин є ключовим компонентом методології. Алгоритми машинного навчання можуть аналізувати великі обсяги історичних даних, виявляти патерни та еволюцію методів поширення дезінформації. В цьому контексті, Кетрін Віссінгер стверджує: "Застосування штучного інтелекту дозволяє автоматизувати процес аналізу історії фейкових новин, надаючи можливість ефективно виявляти та аналізувати їхні патерни розвитку." [Віссінгер, К., "Штучний інтелект у боротьбі з дезінформацією", 2020].

Аналіз текстових характеристик фейкових новин є важливою складовою розробки технологій ШІ для їх виявлення. Тут ключовою роллю відіграє розробка алгоритмів, які дозволяють визначати маніпулятивні та неправдиві елементи у тексті. За словами Дейва Ховарда, експерта у галузі обробки природної мови, "Аналіз текстових особливостей, таких як вживання емоційно забарвлених слів, надмірна сенсаційність, та некоректне використання фактів, дозволяє розробляти алгоритми для автоматичного виявлення фейкових новин." [Ховард, Д., "Технологічні виклики виявлення фейкових новин", 2019].

Аналіз мережі, в якій поширюються інформаційні повідомлення, є ще однією ефективною стратегією виявлення фейкових новин. Тут використовуються методи соціального графу та глибокого навчання для визначення патернів поширення недостовірної інформації. Про це свідчить Аліша Баджан, дослідника соціальних мереж: "Аналіз взаємодії користувачів та розподілу інформації дозволяє визначити класові структури та ідентифікувати джерела, що часто поширюють фейкові новини." [Баджан, А., "Аналіз мережі в контексті фейкових новин", 2021].

Глибоке навчання виявляється ключовим елементом в розробці систем, спроможних ефективно розпізнавати фейкові новини. Методи глибокого навчання, такі як рекурентні нейронні мережі та трансформатори, застосовуються для аналізу семантичних взаємозв'язків у текстах та виявлення неправдивої інформації. За словами Емми Кларк, дослідниці у галузі машинного навчання, "Глибокі нейронні

мережі можуть ефективно виявляти субтельні семантичні зв'язки, які можуть свідчити про недостовірність інформації." [Кларк, Е., "Роль глибокого навчання у фільтрації фейкових новин", 2020]. За всією потужністю технологій ШІ в боротьбі з фейковими новинами стоять етичні питання. Розробка та застосування алгоритмів повинні дотримуватися принципів справедливості та прозорості. Про це зазначає Кетрін Ліндерманн, експертка з етики штучного інтелекту: "Використання ШІ в боротьбі з фейковими новинами вимагає розробки чітких етичних стандартів, щоб уникнути можливих негативних наслідків для свободи слова та приватності." [Ліндерманн, К., "Етика ШІ в контексті фільтрації інформації", 2021].

Зазначаючи успіхи сучасних технологій ШІ в боротьбі з фейковими новинами, експерти висловлюють думку про потенційні перспективи. Алан Маккензі, вчений з обробки природної мови, стверджує: "Майбутнє боротьби з фейковими новинами полягає у подальшому розвитку технологій ШІ, адаптації до нових стратегій дезінформації та удосконаленні методів аналізу текстів та мережевих структур." [Маккензі, А., "Майбутнє боротьби з фейковими новинами", 2022].

Злиття технічних та соціокультурних підходів є важливим етапом в боротьбі з фейковими новинами. Розвиток технічних рішень повинен іти паралельно із створенням навчальних програм та кампаній медіаграмотності. Такий синтез дозволяє забезпечити повноцінний підхід до проблеми дезінформації. За словами Лінди Ферріс, експертки з соціокультурного аналізу: "Використання ШІ у поєднанні із навчальними ініціативами дозволяє зміцнити оборону суспільства від фейкових новин не лише технічно, але й культурно та соціально." [Ферріс, Л., "Синергія технічних та соціокультурних підходів", 2021].

Аналіз отриманих результатів дослідження фейкових новин та ролі штучного інтелекту в їх протидії підтверджує важливість інтеграції технічних та соціокультурних стратегій. Використання алгоритмів машинного навчання та аналіз соціальних мереж спроможне ефективно виявляти та фільтрувати дезінформацію,

проте це повинно поєднуватися з розвитком медіаграмотності та етичними стандартами. Успіх у боротьбі з фейковими новинами полягає у взаємодії між технологічними розробками, основаними на штучному інтелекті, та соціокультурними стратегіями, спрямованими на навчання та усвідомлення громадськості. Тільки цей комплексний підхід може забезпечити ефективний контроль над поширенням фейкових новин та збереження довіри до інформації в сучасному цифровому світі.

2.1 Вплив технологій на інформаційне суспільство. Вплив технологій на інформаційне суспільство, аналіз ролі технологій у масовій комунікації та їх взаємодія з поширенням фейкової інформації

Вплив технологій на інформаційне суспільство є фундаментальним аспектом сучасного світу, вимагаючим уважного аналізу та розуміння. Особливу увагу слід приділити ролі технологій у масовій комунікації та їх взаємодії з поширенням фейкової інформації. Це есе розгляне вплив технологій на формування інформаційного простору сучасного суспільства, а також висвітлить ключові аспекти поширення фейкової інформації через ці технологічні механізми. Масова комунікація стала невід'ємною частиною сучасного інформаційного суспільства, а технології, зокрема інтернет та соціальні медіа, є головними каналами її поширення. (Chollet, F. (2017). Deep Learning with Python. Manning Publications.) Ці технології дозволяють надзвичайно швидко та ефективно розповсюджувати інформацію до широкої аудиторії. Проте, разом з цим виникає проблема поширення фейкової інформації. Фейкова інформація, або "фейки", часто має за мету маніпулювати громадською думкою, формувати певні переконання або викликати певні реакції у суспільстві.

Одним з головних чинників поширення фейкової інформації є зростаюча доступність інформаційних технологій для всіх шарів суспільства. Інтернет та соціальні медіа дозволяють будь-кому створювати та поширювати контент без значного контролю чи перевірки. Це відкриває двері для недостовірної інформації, яка може легко прийматися за правду через відсутність механізмів перевірки та фільтрації.

Більше того, алгоритми, які лежать в основі багатьох соціальних медіа та пошукових систем, часто спрямовані на максимізацію взаємодії користувачів з контентом. Це може призводити до упредметнення та підсилення фейкової інформації, оскільки вона часто викликає більшу реакцію та обговорення, ніж достовірна інформація. Таким чином, алгоритми стають свого роду катализаторами поширення фейкових новин. Зрозуміло, що боротьба з поширенням фейкової інформації вимагає комплексного підходу. По-перше, необхідно зміцнювати медійну грамотність серед громадян, навчаючи їх критично аналізувати інформацію, яку вони споживають. По-друге, потрібно розробляти та впроваджувати технологічні рішення, спрямовані на виявлення та фільтрацію фейкової інформації. Також важливо змінювати алгоритми соціальних медіа та пошукових систем таким чином, щоб вони сприяли підвищенню достовірності та об'єктивності інформації.

Розглядаючи вплив технологій на формування інформаційного простору сучасного суспільства, важливо враховувати різноманітні аспекти цього впливу. По-перше, технології, зокрема Інтернет та соціальні медіа, забезпечують широкий доступ до інформації для всіх шарів суспільства. Це сприяє демократизації доступу до знань та може підвищити рівень освіченості серед населення. Однак зростаюча доступність інформаційних технологій також створює серйозні виклики. Недостатнє регулювання та контроль може призвести до того, що інформаційний простір буде перенасичений фейковою або недостовірною інформацією.

Феномен фейкових новин може збентежити громадську думку та порушити соціальну стабільність. У світлі цих викликів, розуміння ролі технологій у масовій

комунікації та їх взаємодії з поширенням фейкової інформації стає критично важливим. Алгоритми соціальних медіа та пошукових систем, які спрямовані на максимізацію взаємодії користувачів з контентом, можуть стати важливими інструментами у боротьбі з фейками. Шляхом покращення алгоритмів і розробки нових технологічних рішень, можна знизити поширення фейкової інформації та підвищити рівень достовірності інформації, яка доходить до громадськості.

Крім того, важливо акцентувати увагу на розвитку медійної грамотності серед громадян. Навчання критичному мисленню та аналізу інформації може допомогти індивідуумам розрізнити фейкову інформацію від достовірної та вчасно виявляти маніпуляції в інформаційному просторі. Комплексний підхід, який об'єднує в собі технологічні інновації, регулювання та медійну грамотність, може стати ефективним інструментом у боротьбі з поширенням фейкової інформації та забезпечити створення інформаційного простору, який би був не лише відкритим та доступним, але й достовірним та об'єктивним.

Помітним аспектом впливу технологій на формування інформаційного простору є зміна ролі традиційних медійних джерел. Раніше головними джерелами новин та інформації були телебачення, радіо та газети, які зазвичай проходили певну перевірку та редакційний процес перед публікацією. Однак інтернет та соціальні медіа змінили цю динаміку, дозволяючи будь-кому створювати контент та розповсюджувати його без проміжництва традиційних редакційних процесів. Ця децентралізація інформаційного простору відкриває нові можливості для вираження різноманітних точок зору та голосів, що раніше можливо були пригнічені чи ігноровані.

Проте вона також створює середовище, де фейкова інформація може легко поширюватися без достатньої перевірки та підтвердження фактів. Наприклад, соціальні медіа дають можливість анонімним користувачам швидко

розповсюджувати маловідомі, а часто і неперевірені, історії, що створює потенційно небезпечні умови для поширення маніпуляційної чи недостовірної інформації.

Одним із способів боротьби з цим явищем є розвиток і впровадження штучного інтелекту (ШІ) та автоматизованих систем виявлення фейкової інформації. Технології машинного навчання можуть бути використані для аналізу великих обсягів даних та виявлення аномалій у розповсюдженні інформації, що може вказувати на можливість фейкової або маніпулятивної інформації. Також важливо залучити до цього процесу експертів з різних галузей, щоб забезпечити об'єктивність та ефективність аналізу.

Крім того, освітні заходи та програми медійної грамотності є ключовим елементом боротьби з поширенням фейкової інформації. Навчання громадян критично аналізувати інформацію, перевіряти джерела та переконуватися у достовірності фактів допоможе зменшити вплив фейків на суспільство. Ці освітні заходи повинні включати навички розрізнення між об'єктивною та прибічною інформацією, а також уміння розпізнавати маніпулятивні та дезінформаційні техніки.

Зокрема, дуже важливо враховувати роль медійних платформ у боротьбі з поширенням фейкової інформації. Соціальні мережі та інші медійні платформи мають велику відповідальність у забезпеченні якості та достовірності інформації, що публікується на їхніх платформах. Особливу вагу має розгляд ролі медійних платформ у контексті протидії поширенню фейкової інформації. Соціальні мережі та інші медійні платформи понесли на себе значну відповідальність у забезпеченні якості та достовірності інформації, що публікується на них. Це передбачає можливість розробки та впровадження алгоритмів та механізмів фільтрації, спрямованих на виявлення та блокування фейкової або недостовірної інформації, а також позначення такого контенту як спростування або сумнівну інформацію.

Проте важливо дотримуватися балансу між цінністю свободи слова та необхідністю протидії шкідливому контенту. Регулювання медійних платформ має бути прозорим та обґрунтованим, з урахуванням прав та свобод користувачів. Заборона фейкової інформації може стати об'єктом дискусій про свободу слова та цензуру, тому важливо забезпечити, щоб будь-яке регулювання було обґрунтованим та заснованим на доказах.

Крім того, варто пам'ятати про те, що протидія поширенню фейкової інформації потребує комплексного підходу, який включає як технологічні рішення, так і освітні та культурні заходи. Покращення медійної грамотності та критичного мислення серед громадян є ключовим чинником в цьому процесі. Навчання людей розрізняти фейкову інформацію від достовірної та аналізувати інформацію з критичним мисленням допоможе знизити вплив фейків на суспільство.

Отже, вплив технологій на інформаційне суспільство та проблема поширення фейкової інформації вимагають системного та комплексного підходу. Залучення всіх зацікавлених сторін, від громадян до технологічних компаній та урядових органів, є важливим для ефективного вирішення цих проблем. Тільки через спільні зусилля можна забезпечити створення інформаційного простору, який би був доступний, достовірний та безпечний для всіх користувачів.

2.2 Штучний інтелект та виявлення фейкових новин

1. "Automated Fact-Checking: A Comprehensive Review"

Автор: Smith, J., & Johnson, A.

Ця стаття пропонує глибокий аналіз та огляд різних підходів до автоматизованої перевірки фактів, зосереджуючись на використанні технологій штучного інтелекту для виявлення ненадійної інформації в сучасних новинах. Автори Сміт та Джонсон систематично досліджують та обговорюють різні аспекти та виклики, пов'язані з розвитком та впровадженням автоматизованих систем факт-чекінгу.

Стаття розглядає поточний стан справ у сфері автоматизованої перевірки фактів, визначає ключові підходи та методи, що використовуються в цьому напрямку досліджень. Велика увага приділяється технологіям штучного інтелекту, таким як машинне навчання, обробка природної мови та інші, які допомагають ефективно виявляти та аналізувати фактичну інформацію в новинах.

Автори також фокусуються на визначенні викликів та обмежень, з якими стикаються дослідники в цій галузі, таких як точність алгоритмів, виявлення контексту та здатність адаптації до швидкозмінюючихся новинних сценаріїв.

2. "Deep Learning for Fake News Detection: A Comprehensive Review"

Автор: Wang, L., & Zhang, F.

У цій роботі розглядаються різноманітні аспекти використання глибокого навчання для виявлення фейкових новин. Дослідження охоплює аналіз різних моделей глибокого навчання та їхню ефективність у виявленні дезінформації. Автори акцентують увагу на техніках, що поліпшують точність та надійність систем виявлення фейкових новин, враховуючи виклики, що впливають зі сучасного інформаційного середовища. В ході свого аналізу Wang, L., & Zhang, F. розглядають різні аспекти, такі як обробка природної мови, використання нейромереж звідних розмірів та аналіз інформації з соціальних мереж. Вони також докладно розглядають різноманітні датасети фейкових новин, враховуючи їх різноманітність та складність. Стаття також звертає увагу на етичні аспекти та виклики, пов'язані із визначенням терміна "фейкові новини" в сучасному інформаційному контексті. Дослідження

Wang, L., & Zhang, F. спрямоване на покращення нашого розуміння та використання глибокого навчання для боротьби з поширенням фейкових новин, враховуючи різноманітні технічні та концептуальні аспекти, що пов'язані із цією важливою проблемою в інформаційному суспільстві.

3. "Natural Language Processing Techniques for Fake News Detection"

Автор: Chen, Y., et al.

Ця стаття вносить значний внесок у сферу виявлення фейкових новин, зосереджуючись на застосуванні технік обробки природної мови (NLP). Автори детально розглядають різні методи та підходи для аналізу текстової інформації з метою виявлення можливих ознак маніпуляції текстом у фейкових новинах. Ключовим аспектом статті є розгляд структур речень та використання мови у фейкових новинах. Автори досліджують синтаксичні та семантичні особливості тексту, такі як використання емоційно заряджених слів, несподівані зміни стилю письма, аномальні конструкції речень тощо. Це допомагає виявляти відмінності між аутентичними та фейковими новинами. У висновках статті наводяться потенційні переваги та обмеження використання різних методів для виявлення фейкових новин, а також пропонуються напрямки подальших досліджень у цьому напрямку. Ця стаття є цінним ресурсом для дослідників, які працюють у сфері виявлення та протидії поширенню фейкових новин.

4. "The Role of Machine Learning in Combating Disinformation"

Автор: Kim, H., & Lee, S.

Дослідження авторів Kim і Lee зосереджене на ролі методів машинного навчання в контексті боротьби з дезінформацією та виявлення штучно створених новин. Сучасний інформаційний ландшафт вимагає ефективних засобів виявлення та запобігання поширенню фальшивої інформації, яка може впливати на громадську думку, політичні процеси та соціальну стабільність. Стаття детально розглядає використання алгоритмів машинного навчання для аналізу текстових та мультимедійних джерел інформації з метою ідентифікації патернів, характерних для дезінформаційних кампаній. Автори обговорюють переваги та обмеження використання машинного навчання у цьому контексті та вказують на необхідність поєднання цих методів з експертною експертизою для досягнення максимальної точності виявлення дезінформації. У підсумку, дослідження Kim і Lee вносить важливий вклад у розуміння та розвиток стратегій протидії дезінформації за допомогою передових технологій машинного навчання.

5. "Fake News Detection Using Social Media Analytics and Machine Learning"

Автор: Gupta, R., & Singh, A.

Основною метою цього дослідження є розгляд ролі, яку відіграє машинне навчання у протидії дезінформації. Зокрема, воно зосереджується на використанні різноманітних алгоритмів машинного навчання для аналізу великих обсягів інформації та виявлення неправдивої чи маніпульованої інформації.

Машинне навчання виявляється дуже ефективним інструментом у боротьбі з дезінформацією завдяки своїй здатності автоматизовано аналізувати та класифікувати великі обсяги даних. Алгоритми можуть виявляти патерни та аномалії у поширенні інформації, допомагаючи вченим відокремлювати достовірні дані від штучно створених новин. Необхідно також враховувати етичні аспекти використання машинного навчання у сфері боротьби з дезінформацією. Сприятливі умови для розвитку та вдосконалення таких технологій можуть допомогти у забезпеченні

гармонійного та ефективного використання машинного навчання у важливій сфері інформаційної безпеки.

6. "Ethical Considerations in AI-Based Fake News Detection"

Автор: Patel, M., et al.

У статті розглядаються етичні аспекти цього процесу та можливі виклики, пов'язані з його впровадженням. Штучний інтелект (ШІ) стає все більш важливим інструментом у сфері медіа та інформаційних технологій. Однак, його використання для виявлення фейкових новин супроводжується рядом етичних питань, які варто ретельно розглянути.

Одним із ключових аспектів є прозорість алгоритмів, використовуваних у процесі виявлення фейків. Недостатня прозорість може призвести до виникнення проблем в галузі конфіденційності та захисту особистих даних. Додатково, важливо визначити, які джерела вважаються авторитетними та в основу яких алгоритми покладаються під час виявлення фейків. Іншим етичним викликом є можливість систематичної помилки, особливо в контексті культурних різниць та варіантів висловлення. У статті також розглядаються питання відповідальності за можливі помилки системи виявлення та вплив на інформаційний ландшафт. Визначення меж втручання штучного інтелекту у редакційні процеси та відносини зі споживачами інформації стає важливим завданням.

7. "A Survey of Deep Learning Approaches for Fake News Detection"

Автор: Liu, Y., et al.

Дана стаття розглядає різні підходи до розв'язання цієї проблеми, зосереджуючись зокрема на застосуванні глибокого навчання. Однією з ключових тем дослідження є аналіз різних архітектур нейронних мереж, які використовуються для виявлення фейкових новин. Автори розглядають та порівнюють їх переваги та недоліки з метою зрозуміння оптимальних стратегій для досягнення найкращих результатів. Вивчаючи ці архітектури, автори аналізують їхні можливості у виявленні паттернів та особливостей, які характеризують фейкові новини.

Загальною метою даного дослідження є надання огляду та критичного аналізу різних підходів до використання глибокого навчання у виявленні фейкових новин, щоб сприяти розвитку більш ефективних та точних методів боротьби з дезінформацією у сучасному інформаційному просторі.

8. "Semantic Analysis for Fake News Detection: A Literature Review"

Автор: Wang, Z., et al.

Автори звернули увагу на ключові аспекти, такі як глибина аналізу, врахування контексту та використання сучасних технік обробки природної мови. Результати дослідження вказують на перспективи семантичного аналізу у виявленні фейкових новин, зазначаючи високу точність при правильній налаштуванні параметрів та врахуванні специфіки мовлення. В порівнянні з іншими методами, семантичний аналіз проявляється як ефективний інструмент, здатний аналізувати не лише поверхневий зміст новин, але і їхній семантичний контекст. Також, автори відзначають важливість подальших досліджень у цьому напрямі, спрямованих на оптимізацію алгоритмів семантичного аналізу та розширення його застосування в реальних умовах.

9. "Robustness of AI Models in Fake News Detection: Challenges and Solutions"

Автор: Chen, Q., & Wang, Y.

Автори, Чен та Ванг, досліджують проблеми, які можуть виникнути при використанні інтелектуальних моделей для боротьби з розповсюдженням неправдивої інформації та запропонують ефективні шляхи їх вирішення. У статті докладно аналізується ступінь стійкості моделей штучного інтелекту до різноманітних видів фейкових новин, таких як маніпулювання зображеннями, текстова маніпуляція та інші хитрощі, які використовуються для обману алгоритмів. Автори також звертають увагу на проблеми з використанням неперевірених та неперевірених даних у процесі тренування моделей, що може призвести до недостатньої стійкості та надійності систем виявлення фейків. Помітний акцент робиться на необхідності вдосконалення алгоритмів усунення фейкової інформації та вдосконалення методів перевірки достовірності даних. Автори також обговорюють можливі рішення для підвищення стійкості та надійності моделей штучного інтелекту у виявленні фейкових новин, враховуючи високотехнологічні методи та підходи.

10. "Human-AI Collaboration in Fake News Detection"

Автор: Brown, E., et al.

Стаття висвітлює науковий підхід до проблеми фейкових новин і визначає ключові аспекти співпраці між людьми та штучним інтелектом в цьому контексті. Зокрема, досліджуються можливості використання людського досвіду та інтуїції в поєднанні з аналітичними можливостями систем штучного інтелекту.

Автори розглядають використання різних алгоритмів машинного навчання та нейронних мереж для автоматизованого виявлення фейкових новин, а також висвітлюють необхідність людського втручання для підтвердження та покращення точності таких систем. Дослідження базується на аналізі реальних прикладів фейкових новин та їхніх наслідків для суспільства. Ця стаття вносить важливий

внесок у розвиток стратегій виявлення та протидії фейковим новинам, підкреслюючи значущість співпраці між людьми та штучним інтелектом. Автори закликають до подальших досліджень та розвитку інноваційних методів для забезпечення вірогідного та ефективного виявлення дезінформації в епоху зростаючого обсягу інформації в мережі.

11. "Explainability and Interpretability in AI Models for Fake News Detection"

Автор: Wang, H., et al.

Це дослідження ставить за мету розкрити суть та роботу різних аспектів алгоритмів, які використовуються для виявлення штучно створених новин. Стаття починається з аналізу необхідності вирішення проблеми фейкових новин і ролі штучного інтелекту в цьому контексті. Вона висвітлює важливість створення моделей, які не лише ефективно виявляють фейки, але й надають пояснення своїх рішень. Автори аргументують, що лише такий підхід може забезпечити високий рівень довіри до систем виявлення фейкових новин. Далі в статті розглядається концепція прозорості в контексті штучного інтелекту. Автори аналізують, як різні моделі впроваджують механізми, які роблять їх рішення зрозумілими для кінцевого користувача. Це може включати в себе візуалізацію процесу виявлення фейкових новин, пояснення вагомості різних факторів чи демонстрацію логіки рішення.

По завершенню аналізу прозорості, стаття переходить до обговорення інтерпретованості. Автори розглядають методи та стратегії, що дозволяють робити моделі зрозумілими та прийнятними для інтерпретації. Це може включати в себе надання конкретних прикладів, які вказують, які конкретні аргументи чи фактори вплинули на рішення моделі. Загальний підхід статті полягає в тому, щоб вивести на передній план не тільки результати моделей для виявлення фейкових новин, але й їх

роботу і принципи, які керують їхніми рішеннями. Це сприяє не лише підвищенню ефективності, але і розумінню користувачами того, як саме працюють інструменти для боротьби з поширенням дезінформації.

12. "Deep Learning in Multimedia Forensics: Fake Image and Video Detection"

Автор: Liu, C., & Zhang, H.

В основі статті лежить докладне дослідження застосування глибокого навчання в області форензики, спрямоване на виявлення фейкових зображень та відео, які можуть бути використані для обману чи розповсюдження дезінформації. Автори висвітлюють ключові аспекти та технічні особливості глибокого навчання, застосованого до аналізу мультимедійного матеріалу. У статті надається детальний огляд методів виявлення фейкових зображень та відео, зокрема використання нейронних мереж та глибоких архітектур. Розглядаються існуючі виклики та труднощі, з якими стикаються дослідники у цій галузі, такі як підготовка невеликих об'ємів даних для навчання, адаптація до різних типів фейків та розпізнавання їх у реальному часі.

Автори статті роблять акцент на важливості подальших досліджень у цьому напрямі та вдосконалення методів виявлення фейків у мультимедійному контенті. Заключні висновки статті містять рекомендації щодо подальших можливостей розвитку глибокого навчання в мультимедійній форензиці та виявленні фейкового контенту.

13. "Blockchain Technology for Trustworthy News Sources"

Автор: Zhang, L., & Wang, Q.

В статті розглядається можливість використання технології блокчейн для створення довіреного середовища для новин та запобігання поширенню фейкових інформацій. Блокчейн відомий своєю безпекою та децентралізацією, що робить його ідеальним інструментом для боротьби з поширенням фейкових новин. Однією з основних переваг блокчейн-технології є можливість забезпечення прозорості та невідмінності інформації. Кожен блок даних, доданий до ланцюжка блоків, отримує унікальний ідентифікатор та пов'язується з попереднім блоком, що ускладнює можливість втручання чи видалення інформації.

Використання блокчейну може допомогти створити систему, в якій новини можна відстежувати від джерела до кінцевого користувача, гарантуючи їхню автентичність та невідмінність. Кожен користувач може перевірити джерело новин та шлях їхнього поширення, щоб переконатися в їхній достовірності.

14. "Cross-lingual Fake News Detection Using Machine Translation"

Автор: Kim, M., et al.

Стаття присвячена аналізу методів машинного перекладу для ефективного виявлення фейкових новин у міжмовному контексті. Один з головних викликів у сфері боротьби з поширенням фейкових новин є їхнє поширення в різних країнах та серед різних культур. Машинне навчання та переклад можуть стати ключовими інструментами у цьому процесі. Автори статті досліджують різні методи машинного перекладу та їхню ефективність у виявленні фейкових новин, особливо коли текст перекладається між різними мовами та культурами. Серед аналізованих методів виявлення фейкових новин зокрема звертається увага на техніку використання мовних особливостей та контексту при перекладі. Крім того, автори досліджують можливість використання нейромереж для поліпшення точності виявлення фейкових новин у крос-мовному середовищі.

15. "User Behavior Analysis and Fake News Propagation on Social Media"

Автор: Chen, X., & Liu, B.

Дана стаття присвячена вивченню впливу поведінки користувачів на поширення фейкових новин у соціальних мережах. Дослідження авторів, X. Chen та B. Liu, має на меті розкрити механізми поширення дезінформації в онлайн-середовищі та визначити можливості виявлення таких неправдивих інформаційних матеріалів.

Автори звертають увагу на важливість вивчення поведінкових аспектів користувачів у соціальних мережах, оскільки саме вони можуть визначати шляхи розповсюдження фейкових новин. Інтеграція аналізу поведінки користувачів дозволяє краще розуміти, які фактори сприяють виробленню та поширенню дезінформації в цьому онлайн-середовищі. Дослідження включає аналіз типових стратегій поширення фейкових новин, ідентифікацію ключових моментів, коли такі інформаційні матеріали найчастіше набувають популярності, а також виявлення потенційних "вразливих точок" в соціальних мережах, де може виникнути інтенсивне поширення дезінформації. Автори також пропонують методи виявлення фейкових новин на основі аналізу поведінки користувачів, що спрямовані на забезпечення більш ефективної фільтрації та виявлення неправдивої інформації в соціальних мережах.

16. "Legal and Regulatory Challenges in Combatting Fake News with AI"

Автор: Li, Y., & Wang, J.

У відділі цієї статті проводиться глибокий аналіз юридичних та регуляторних викликів, пов'язаних із використанням технологій штучного інтелекту для виявлення та протидії поширенню фейкових новин. Висвітлюються питання, пов'язані із

забезпеченням особистої конфіденційності та захистом свободи слова в контексті боротьби з дезінформацією. Стаття зосереджується на важливості ефективного використання штучного інтелекту для виявлення та аналізу фейкових новин, однак одночасно виявляє важливі правові та регуляторні виклики, які виникають у зв'язку з цією проблемою. Одним із основних аспектів є збалансованість між захистом особистої конфіденційності громадян і необхідністю виявлення та боротьби з поширенням дезінформації. Автори досліджують, які правові рамки і політики можуть забезпечити адекватний рівень захисту особистих даних під час використання штучного інтелекту в цілях виявлення фейкових новин.

Розглянемо основні прогалини в існуючих дослідженнях, присвячених використанню технологій штучного інтелекту (ШІ) для виявлення фейкових новин, та сформулюємо дослідницькі питання для подальшого удосконалення цієї області. По-перше, багато з досліджень зосереджені на технічних аспектах, ігноруючи контекстуальні особливості, культурні аспекти та політичні умови, що також мають вагомий роль у виявленні фейкових новин. Недостатня увага до цих аспектів ускладнює адаптацію моделей ШІ до різних соціокультурних середовищ. По-друге, проблемою є недостатня зрозумілість і пояснення результатів моделей ШІ. Багато алгоритмів виявлення фейкових новин базуються на складних архітектурах, що ускладнює їх інтерпретацію та пояснення результатів.

Третя прогалина полягає у відсутності універсального набору функцій для виявлення фейкових новин. Неоднорідність даних та методологій може призводити до варіацій у результатах між різними дослідженнями. Окрему увагу вимагають етичні аспекти використання ШІ для виявлення фейкових новин. Питання приватності, цензури та можливих викликів щодо свободи слова потребують більш ретельного розгляду. Крім того, розвиток ефективних методів виявлення та запобігання розповсюдженню контенту DeepFake залишається актуальним завданням у цій області.

Зважаючи на вищенаведені прогалини у дослідженнях та використанні технологій штучного інтелекту для виявлення фейкових новин, можна сформулювати додаткові дослідницькі питання, які варто розглянути для подальшого удосконалення цієї сфери:

1. Як враховувати взаємозв'язок між різними типами даних (текст, зображення, відео тощо) для більш точного виявлення фейкових новин?
2. Як можна застосувати ансамбль методів штучного інтелекту для підвищення точності та надійності виявлення фейкових новин?
3. Які стратегії можуть бути використані для виявлення і відстеження еволюції методів створення та поширення фейкових новин?
4. Як можна поєднати підходи машинного навчання та правові/політичні механізми для більш ефективного контролю над поширенням фейкових новин?
6. Які виклики і можливості існують у використанні технологій блокчейну для підтримки автентифікації новин та боротьби з фейками?
7. Як можна розвивати технології виявлення фейкових новин таким чином, щоб вони були здатні пристосовуватися до швидкозмінюючихся та нових форм фейків?
8. Яким чином можна покращити системи виявлення фейкових новин, щоб вони могли працювати в режимі реального часу на платформах соціальних медіа та новинних сайтах?
9. Які соціально-психологічні аспекти можуть впливати на ефективність та прийняття систем виявлення фейкових новин і як це можна врахувати у подальших дослідженнях?
10. Як виявлення фейкових новин може бути оптимізоване для адаптації до різноманітних мов, культур та контекстів в інтернеті?

Дослідницькі питання, висунуті для подальшого розвитку області виявлення фейкових новин за допомогою технологій штучного інтелекту, є важливими напрямками для майбутніх наукових досліджень. Виявлення фейкових новин є

складною проблемою у сучасному інформаційному суспільстві, і ефективні методи їх виявлення є критично важливими для забезпечення надійності та довіри до інформації. Для подальшого розвитку цієї наукової області необхідно проведення активних досліджень та усунення виявлених прогалин. Наприклад, важливою є розробка комплексних підходів, що враховують різноманітність типів даних (текстові, візуальні тощо) та культурних особливостей для покращення точності алгоритмів виявлення фейкових новин. Крім того, потрібен подальший розвиток етичних стандартів у використанні технологій штучного інтелекту в цілях боротьби з фейками та поглиблення досліджень соціально-психологічних аспектів впливу фейкових новин на суспільство.

Необхідність інтердисциплінарного співробітництва між вченими з інформаційних технологій та соціальних наук виходить на передній план, оскільки тільки такий підхід дозволить досягти суттєвих результатів у боротьбі з фейковими новинами. Взаємодія між різними галузями дозволить розробити більш ефективні та виправдані стратегії та методи для виявлення та протидії поширенню недостовірної інформації. Такий колективний підхід залучає різноманітність підходів та досліджень, що створює умови для більш глибокого розуміння феномену фейкових новин і розробки більш ефективних інструментів для їх виявлення та протидії. Тільки через цей спільний підхід можна досягти важливих досягнень у цій сфері і забезпечити більшу довіру до інформації, що сприятиме стійкому розвитку суспільства.

Розділ 3: Аналіз наявних інструментів які є на сьогоднішній день на ринку та рекомендації для майбутніх розробок для фактчекерів в галузі штучного інтелекту

На ринку існує безліч інструментів, які можуть допомогти фактчекерам в їх роботі.

Деякі з найпопулярніших інструментів включають:

- **Factmata** - це платформа на основі штучного інтелекту, яка допомагає фактчекерам автоматизувати деякі аспекти їх роботи, такі як виявлення потенційно неправдивої інформації та верифікація фактів.

Переваги:

1. Допомагає зменшити час, витрачений на перевірку фактів.
2. Постійно вдосконалює свої алгоритми на основі нових даних.

Обмеження:

1. Хоча алгоритми постійно вдосконалюються, існує ризик помилкових визначень.
2. Більшість моделей налаштовані на англійську мову, що може бути проблемою для інших мов.

- **Google Reverse Image Search** - це інструмент, який дозволяє користувачам шукати в Інтернеті за допомогою зображень. Це може бути корисно для фактчекерів, які намагаються перевірити походження фотографій або інших візуальних матеріалів.

Переваги:

1. Проста у використанні, доступна для всіх користувачів.
2. Доступ до великої кількості джерел та сайтів.

Обмеження:

1. Може видавати неточні або нерелевантні результати.
2. Результати обмежені тим, що є в базі Google.

- **InVid** - це інструмент з відкритим кодом, який дозволяє користувачам перевіряти автентичність відео. Він може використовуватися для виявлення глибоких фейків та інших маніпуляцій з відео.

Переваги:

1. Пропонує різноманітні функції, такі як фрагментація відео, пошук метаданих та зворотний пошук зображень.
2. Дозволяє користувачам вносити зміни та покращувати інструмент.

Обмеження:

1. Може вимагати певних технічних знань для ефективного використання.
2. Можливо, відсутність постійної підтримки та оновлень.

- **Polygraph** - це платформа, яка використовує штучний інтелект для аналізу тексту та виявлення потенційно неправдивої інформації.

Переваги:

1. Може аналізувати великі обсяги тексту швидко і ефективно.
2. Виявляє підозрілі твердження, які потребують додаткової перевірки.

Обмеження:

1. Можливі обмеження у роботі з текстами на різних мовах.
2. Як і всі алгоритми ШІ, може давати помилкові результати.

- **Snoopes** - це веб-сайт, який публікує фактичні перевірки новин та інформації. Він має велику базу даних статей, які можна використовувати для перевірки фактів.

Переваги:

1. Відомий своєю ретельною роботою і надійністю.
2. Легко доступний для широкого кола користувачів.

Обмеження:

1. Залежність від людських ресурсів для перевірки та написання статей.
2. Менше автоматизованих функцій порівняно з іншими інструментами.

3.1 Рекомендації для майбутніх розробок

Існує багато можливостей для вдосконалення інструментів, які використовуються фактчекерами. Деякі з найважливіших напрямків досліджень та розробок включають:

- Розробка більш точних алгоритмів для виявлення неправдивої інформації. Алгоритми штучного інтелекту, які використовуються для виявлення неправдивої інформації, постійно вдосконалюються, але вони все ще можуть робити помилки. Необхідно продовжувати дослідження та розробку нових алгоритмів, які є більш точними та надійними.
- Розробка інструментів для перевірки фактів у форматах соціальних мереж. Соціальні мережі є одним із основних джерел неправдивої інформації. Необхідно розробити інструменти, які допоможуть фактчекерам перевіряти факти в публікаціях у соціальних мережах.
- Розробка інструментів для перекладу фактичних перевірок на різні мови. Неправдива інформація може поширюватися мовами. Необхідно розробити інструменти, які допоможуть фактчекерам перекладати свої перевірки на різні мови, щоб вони могли охопити ширшу аудиторію.
- Розробка інструментів для освіти людей про те, як виявляти неправдиву інформацію. Важливо навчити людей, як виявляти неправдиву інформацію та

оцінювати достовірність інформації, яку вони бачать в Інтернеті. Необхідно розробити інструменти, які допоможуть людям зробити це.

- Інтеграція розширеної реальності (AR) та віртуальної реальності (VR). Розширена реальність та віртуальна реальність можуть відкрити нові можливості для фактчекерів. Наприклад, за допомогою AR фактчекери можуть отримувати інтерактивну допомогу під час перевірки фактів на місці події, шляхом проєкції додаткової інформації на зображення або відео. VR може бути використана для створення іммерсивних навчальних досліджень для користувачів, які навчаються виявляти неправдиву інформацію.
- Використання блокчейн технології. Блокчейн може забезпечити безпеку та недоторканість даних, що є критично важливим для фактчекінгу. З використанням технології блокчейн можна створити розподілену базу даних, яка дозволить фактчекерам перевіряти джерела інформації та забезпечити їх невідредагованість.
- Створення спільнот та платформ співпраці. Розвиток спільнот та платформ співпраці для фактчекерів може значно полегшити обмін інформацією, ресурсами та кращими практиками. Це може включати форуми, чати, відеоконференції та спільні проєкти, які допоможуть фактчекерам працювати разом для боротьби з дезінформацією.
- Розвиток алгоритмів для виявлення контексту та мотивацій. Покращення алгоритмів для аналізу контексту та мотивацій за поширенням неправдивої інформації може допомогти фактчекерам краще розуміти причини та способи боротьби з дезінформацією. Це може включати розпізнавання маніпулятивних

технік, політичних агенд та інших чинників, що впливають на поширення неправдивої інформації.

Усе враховуючи, розвиток інструментів фактчекінгу є критично важливим у світі, де поширення дезінформації може мати серйозні наслідки для суспільства. Наявність широкого спектру інструментів, включаючи ті, що базуються на штучному інтелекті, технології блокчейн, інтерактивність розширеної та віртуальної реальності, дозволяє фактчекерам ефективніше виявляти та спростовувати неправдиву інформацію. Однак, для подальшого вдосконалення систем фактчекінгу потрібні постійні дослідження та розвиток. Це включає в себе покращення алгоритмів, розширення можливостей перевірки інформації у соціальних мережах, мультимедійність та навчання користувачів навичкам критичного мислення.

Зростання співпраці між фактчекерами, створення спільнот та платформ для обміну досвідом також може сприяти більш ефективній боротьбі з дезінформацією. В цілому, поєднання технологічних інновацій, спільної роботи та освіти може сприяти створенню більш безпечного та інформованого цифрового простору.

Завдяки проведеним інтерв'ю з експертами, ми можемо вивести наступні висновки.

- Штучний інтелект (ШІ) може бути корисним інструментом для журналістів та фактчекерів, але його не слід використовувати як заміну людській експертизі. ШІ може допомогти дослідникам швидше знаходити інформацію, виявляти закономірності та генерувати нові ідеї, але він не може замінити людську здатність розуміти контекст, оцінювати достовірність інформації та приймати етичні рішення.
- Журналістам та фактчекерам важливо критично оцінювати інформацію, яку вони отримують від ШІ, та перевіряти її з інших джерел. ШІ може генерувати упереджену або неточну інформацію, тому важливо використовувати його лише як допоміжний інструмент.

- Гібридні впливи — це серйозна загроза для демократії та свободи слова. Ці впливи можуть включати пропаганду, дезінформацію, кібератаки та інші методи, спрямовані на підрив довіри до інституцій та роз'єднання суспільства.
- Журналістам та фактчекерам важливо бути обізнаними про гібридні впливи та знати, як їм протистояти. Вони можуть робити це, перевіряючи факти, розкриваючи упередження та пропаганду, а також надаючи аудиторії точну та неупереджену інформацію.
- Майбутнє інформаційної війни буде залежати від того, як ШІ буде використовуватися як журналістами, так і урядами. Журналісти можуть використовувати ШІ для створення більш переконливого контенту та більш ефективної боротьби з дезінформацією, а уряди можуть використовувати ШІ для маніпулювання громадською думкою та підриву довіри до своїх супротивників.

Рекомендації:

- Журналістам та фактчекерам слід пройти навчання з використання ШІ в своїй роботі. Це навчання повинно включати в себе те, як використовувати ШІ для знаходження інформації, перевірки фактів, виявлення упередженості та генерування нових ідей.
- Урядам слід розробити політику щодо використання ШІ в інформаційній війні. Ця політика повинна бути спрямована на запобігання використанню ШІ для маніпулювання громадською думкою та підриву довіри до демократичних інституцій.
- Громадськість повинна бути обізнана про ризики та переваги використання ШІ в журналістиці та інформаційній війні. Це допоможе їм критично оцінювати

інформацію, яку вони отримують, та приймати обґрунтовані рішення щодо того, кому вірити.

Результати фокус-групи з молоддю (20-29 років) свідчать про те, що фейкові новини стали значною проблемою в їхньому онлайн-середовищі.

- Частота. Усі учасники (100%) стикаються з фейковими новинами дуже часто, особливо на популярних платформах, таких як Facebook, Instagram та TikTok.
- Вплив. Ця дезінформація викликає занепокоєння щодо поширення неправдивих знань, недовіри до авторитетних джерел та негативного впливу на прийняття рішень.

1. ШІ потенційний інструмент для боротьби з дезінформацією

Більшість учасників (80%) позитивно ставляться до використання ШІ для виявлення фейкових новин.

Переваги. ШІ може швидко та ефективно аналізувати великі обсяги інформації, виявляти складні закономірності та звільняти час для глибшого аналізу користувачам. Проте. Учасники висловлюють стурбованість можливими ризиками, пов'язаними з використанням ШІ, такими як:

- Цензура. ШІ може обмежувати доступ до інформації та призводити до втрати свободи слова.
- Упередженість. Алгоритми ШІ можуть відтворювати існуючі упередження в суспільстві.
- Маніпулювання. ШІ може використовуватися для поширення дезінформації з певними цілями.

2. Необхідність відповідального та етичного підходу до ШІ

Використання ШІ для виявлення фейкових новин потребує відповідального та етичного підходу:

- Прозорість. Розробники ШІ-систем повинні забезпечувати прозорість роботи алгоритмів та відкритий доступ до їх коду.
- Відкритість. Важливо заохочувати відкриту дискусію про етичні аспекти використання ШІ.
- Баланс. Потрібно знайти баланс між ефективністю виявлення фейкових новин та збереженням свободи слова та доступу до інформації.

3. Наступні кроки: дослідження, освіта та співпраця

- Дослідження. Потрібні додаткові дослідження для вдосконалення ШІ-алгоритмів та оцінки їх ефективності.
- Освіта. Важливо навчати користувачів критичного мислення та навичок перевірки інформації.
- Співпраця. Потрібна співпраця між розробниками ШІ, журналістами, дослідниками та громадськістю для розробки ефективних стратегій боротьби з фейковими новинами в епоху ШІ.

4. Майбутнє: ШІ як допоміжний інструмент

Учасники фокус-групи (80%) готові використовувати ШІ-інструменти для перевірки новин за умови їх прозорості та неупередженості.

- ШІ може стати допоміжним інструментом для користувачів, журналістів та медіа у боротьбі з фейками, звільняючи час для більш глибокого аналізу та перевірки інформації.
- Однак, остаточне рішення щодо достовірності інформації має залишатися за людиною.

Використання ШІ для виявлення фейкових новин володіє значним потенціалом, але водночас несе в собі певні ризики. З одного боку, ШІ може значно прискорити та масштабувати процес аналізу інформації, виявляючи закономірності та потенційні ознаки фейків в текстах, зображеннях та відео. Алгоритми ШІ, вільні від емоцій та

упереджень, можуть об'єктивно оцінювати інформацію, адаптуючи методи виявлення фейків до індивідуальних потреб користувачів. Це може звільнити час для більш глибокого аналізу та перевірки інформації людьми. З іншого боку, існує ризик упередженості алгоритмів, що може призвести до необ'єктивного виявлення фейків. Неконтрольоване використання ШІ може обмежити свободу слова та призвести до цензури. Зловмисники можуть використовувати ШІ для створення більш складних фейків, важких для виявлення. Непрозорість роботи ШІ-систем може підірвати довіру користувачів до їх результатів.

Рекомендації щодо відповідального використання ШІ:

- Прозорість та неупередженість. Розробка алгоритмів з відкритим кодом та заходи щодо усунення упередженості.
- Залучення користувачів та експертів. Активна участь користувачів та експертів у тестуванні та вдосконаленні ШІ-систем.
- Співпраця. Об'єднання зусиль розробників ШІ, журналістів, дослідників, медіаплатформ та громадськості.
- Освіта та навички. Навчання користувачів критичного мислення, навичок перевірки інформації та розуміння роботи ШІ.
- Регулювання та етика. Розробка етичних норм та правил використання ШІ для виявлення фейків.

Поєднання ШІ з критичним мисленням людей, перевіркою фактів журналістами та медіа, а також співпрацею між усіма учасниками інформаційного простору може допомогти створити більш об'єктивне та надійне онлайн-середовище. Важливо пам'ятати, що ШІ - це інструмент, який може використовуватися як для добра, так і для зла. Відповідальне та етичне використання ШІ - це ключ до його ефективності в боротьбі з фейковими новинами.

4. Висновки

У ході дослідження використання алгоритмів машинного навчання для виявлення фейкових новин було виявлено як значні переваги, так і певні обмеження даної технології. Ключовим аспектом є інтеграція різноманітних наборів даних та розробка методів для виявлення і корекції упередженостей, що забезпечить об'єктивність і точність алгоритмів. Використання широкого спектру джерел даних дозволить охопити різні аспекти інформації та зменшити ризик упередженості, яка може виникати при обробці обмежених або однобоких даних. Це не лише підвищить надійність результатів, але й сприятиме більшій довірі користувачів до систем виявлення фейкових новин. З огляду на це, можна сформулювати наступні висновки.

Зменшення упередженості

Використання різноманітних наборів даних. Для того, щоб зменшити упередженість алгоритмів машинного навчання, необхідно використовувати різноманітні набори даних, які представляють різні точки зору, культурні контексти та соціальні групи. Наприклад, включення даних з різних регіонів світу, різних мов та різних соціальних верств допоможе забезпечити, що алгоритми отримують більш повне уявлення про різноманітність інформаційного простору. Це також допоможе зменшити ризик упередженого трактування новин, які можуть бути специфічними для певних культур або груп населення.

Розробка методів виявлення та видалення упередженостей. Необхідно розробляти і впроваджувати методи для виявлення та видалення упередженостей у даних. Це може включати алгоритми, які автоматично аналізують дані на предмет упередженості, та інструменти для корекції цих упередженостей. Наприклад, можна використовувати методи перевірки і балансування наборів даних, щоб забезпечити рівномірне представлення різних груп.

Створення алгоритмів, стійких до упередженостей. Слід зосередитися на розробці алгоритмів, які мають вбудовані механізми для мінімізації упередженостей. Це можуть бути адаптивні алгоритми, які можуть змінювати свої параметри в

залежності від виявлених упередженостей, або алгоритми, які використовують додаткові шари аналізу для виявлення та усунення упередженостей у реальному часі.

Підвищення стійкості

Розробка методів виявлення та запобігання атакам. Важливо розробляти технології та методи, які здатні ефективно виявляти та запобігати різним типам атак на алгоритми машинного навчання. Наприклад, методи аналізу аномалій можуть допомогти виявити незвичну активність, яка може свідчити про атаку. Крім того, важливо створювати алгоритми, які можуть протистояти атакам, зокрема отруєнню даних (коли зловмисники вводять шкідливі дані в набір для навчання) та атакам на інферування (коли зловмисники намагаються маніпулювати виходами моделі).

Створення алгоритмів, стійких до маніпуляцій. Необхідно розробляти алгоритми, які мають вбудовану стійкість до маніпуляцій. Це можуть бути алгоритми, які використовують різноманітні перевірки і верифікації на всіх етапах обробки даних. Наприклад, використання криптографічних методів для захисту даних та результатів аналізу може забезпечити додатковий рівень безпеки.

Розробка прозорих алгоритмів. Слід зосередитися на створенні алгоритмів, які дозволяють користувачам зрозуміти, як і чому приймаються ті чи інші рішення. Це може включати розробку моделей з пояснювальними властивостями, які надають зрозумілі описи процесів прийняття рішень. Наприклад, використання методів візуалізації даних може допомогти зробити процес прийняття рішень більш зрозумілим для користувачів.

Пояснення рішень алгоритмів. Важливо, щоб алгоритми машинного навчання могли надавати зрозумілі пояснення своїх рішень. Це може бути реалізовано через інтерфейси користувача, які надають детальну інформацію про те, які фактори були враховані при прийнятті рішення і як вони вплинули на кінцевий результат. Наприклад, створення пояснювальних моделей, які можуть надавати зворотний зв'язок у вигляді текстових пояснень або графіків.

Етичне використання алгоритмів. Необхідно розробляти та впроваджувати етичні стандарти для використання алгоритмів машинного навчання. Це включає дотримання принципів прозорості, відповідальності та справедливості при розробці та застосуванні алгоритмів. Наприклад, створення етичних кодексів для розробників та користувачів алгоритмів може допомогти забезпечити відповідальне використання технологій.

Підвищення довіри до алгоритмів. Для підвищення довіри до алгоритмів машинного навчання важливо забезпечити їх прозорість та пояснюваність, а також гарантувати, що вони діють відповідно до етичних норм. Це може включати створення незалежних комітетів з етики, які б оцінювали алгоритми та їх використання, а також залучення громадськості до процесу розробки та впровадження алгоритмів.

Таким чином, для покращення ефективності та надійності алгоритмів машинного навчання у виявленні фейкових новин необхідно зосередитися на зменшенні упередженості, підвищенні стійкості до атак, забезпеченні прозорості та пояснюваності, а також дотриманні етичних стандартів. Це дозволить створити більш надійні та ефективні інструменти для боротьби з дезінформацією у сучасному інформаційному просторі.

Загалом, результати цього дослідження наголошують на важливості всебічного підходу до створення алгоритмів машинного навчання для виявлення фейкових новин, враховуючи технічні, етичні та соціальні аспекти. Такий підхід сприятиме розробці більш ефективних і надійних інструментів для боротьби з дезінформацією, що є вкрай важливим у сучасному інформаційному середовищі. Ключовим аспектом є інтеграція різноманітних наборів даних та розробка методів для виявлення і корекції упередженостей, що забезпечить об'єктивність і точність алгоритмів. Це не лише підвищить надійність результатів, але й сприятиме більшій довірі користувачів.

Стійкість до маніпуляцій та атак також є критично важливою. Розробка надійних алгоритмів, здатних протистояти зовнішнім загрозам, гарантує їх безпеку і надійність. Це охоплює не тільки захист від технічних атак, але й запобігання спробам використання алгоритмів для поширення фейкових новин.

Прозорість та пояснюваність алгоритмів є важливими для етичного використання технологій машинного навчання. Створення моделей, які можуть надавати зрозумілі пояснення своїх рішень, допоможе знизити ризики, пов'язані з їх використанням, і підвищить довіру до цих технологій. Це також підвищить загальний рівень медіаграмотності серед користувачів, які зможуть краще розуміти процеси, що лежать в основі виявлення фейкових новин.

Важливо також враховувати міждисциплінарний підхід до розробки алгоритмів, що включає співпрацю з експертами в галузі соціології, психології, журналістики та права. Така взаємодія дозволить створити комплексніші рішення, які враховують різноманітні аспекти проблеми дезінформації. Соціологічні та психологічні дослідження допоможуть краще зрозуміти поведінку користувачів та мотиви поширення фейкових новин, що є важливим для розробки більш цілеспрямованих та ефективних алгоритмів.

Залучення журналістів до процесу розробки алгоритмів може забезпечити точність і достовірність інформації, оскільки вони мають глибоке розуміння принципів журналістської етики та стандартів. Правові аспекти також не можна ігнорувати: розробка алгоритмів повинна здійснюватися у відповідності до існуючих законодавчих норм і стандартів, що регулюють питання конфіденційності, авторських прав та захисту даних. Для подальшого підвищення ефективності алгоритмів виявлення фейкових новин важливо зосередитися на безперервному навчанні моделей та адаптації до нових форм дезінформації. Це включає регулярне оновлення наборів даних та алгоритмів для врахування нових трендів і методів поширення фейкових новин. Постійний моніторинг і оцінка ефективності алгоритмів забезпечать їх актуальність і здатність оперативно реагувати на нові виклики.

Крім того, розвиток міжнародного співробітництва та обмін знаннями між дослідниками і розробниками з різних країн сприятиме створенню більш універсальних і ефективних рішень для боротьби з дезінформацією на глобальному рівні. Це також допоможе врахувати культурні та мовні особливості, що можуть впливати на процес поширення і сприйняття фейкових новин.

Дотримання етичних стандартів у розробці та використанні алгоритмів є не менш важливим. Впровадження етичних кодексів та залучення незалежних експертів і громадськості до цього процесу допоможе уникнути потенційних зловживань і забезпечить відповідальне використання технологій.

Отже, підхід, що поєднує технічну досконалість з етичними та соціальними зобов'язаннями, дозволить створити ефективні інструменти для протидії дезінформації. Це забезпечить надійність інформаційного середовища, підвищить довіру до медіа та сприятиме формуванню інформованого та критично мислячого суспільства. Таким чином, розробка та впровадження алгоритмів машинного навчання для виявлення фейкових новин є важливим кроком на шляху до створення більш справедливого та прозорого інформаційного простору.

Список використаної літератури

1. "FactCheck.org".
2. "Snopes.com".
3. "PolitiFact.com".
4. Negnevitsky, Michael. Artificial Intelligence: A Guide to Intelligent Systems. Pearson Education, 2005.
5. Luger, George F. Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Addison-Wesley, 2009.
6. Bostrom, Nick. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2014.
7. Murphy, Kevin P. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
8. Goodfellow, Ian, et al. Deep Learning. MIT Press, 2016.
9. Russell, Stuart. Human Compatible: Artificial Intelligence and the Problem of Control. Viking, 2019.
10. Tegmark, Max. Life 3.0: Being Human in the Age of Artificial Intelligence. Vintage, 2018.
11. Poole, David L., and Alan K. Mackworth. Artificial Intelligence: Foundations of Computational Agents. Cambridge University Press, 2017.
12. Nilsson, Nils J. Artificial Intelligence: A New Synthesis. Morgan Kaufmann Publishers, 1998.
13. Craft, Stephanie, and Seth C. Lewis. "The Anatomy of Fake News." Journalism Practice, vol. 12, no. 4, 2018, pp. 365-381.

14. Ott, Brian L., and Cara Louise Buckley. *Fake News: Understanding Media and Misinformation in the Digital Age*. Routledge, 2019.
15. Wardle, Claire, and Hossein Derakhshan. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making." Council of Europe Report, 2017.
16. Benkler, Yochai, et al. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press, 2018.
17. Uscinski, Joseph P. *Fake News and False Flags*. Rowman & Littlefield, 2020.
18. O'Connor, Cailin, and James Owen Weatherall. *The Misinformation Age: How False Beliefs Spread*. Yale University Press, 2019.
19. Papacharissi, Zizi. "The Hoaxer Project: Understanding and Countering Fake News." *Communication Theory*, vol. 28, no. 4, 2018, pp. 411-426.
20. Kavanagh, Jennifer, and Michael D. Rich (eds.). *Disinformation and Democracy: The Global Challenge of Fake News*. Oxford University Press, 2018.
21. Keyes, Ralph. *The Post-Truth Era: Dishonesty and Deception in Contemporary Life*. St. Martin's Press, 2004.
22. Barclay, Donald A. *Fake News, Propaganda, and Plain Old Lies: How to Find Trustworthy Information in the Digital Age*. Rowman & Littlefield, 2018.
23. "Artificial Intelligence: A Guide to Intelligent Systems" by Michael Negnevitsky
24. "Artificial Intelligence: Structures and Strategies for Complex Problem Solving" by George F. Luger
25. "Superintelligence: Paths, Dangers, Strategies" by Nick Bostrom
26. "Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy
27. "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
28. "Human Compatible: Artificial Intelligence and the Problem of Control" by Stuart Russell
29. "Life 3.0: Being Human in the Age of Artificial Intelligence" by Max Tegmark
30. "Artificial Intelligence: Foundations of Computational Agents" by David L. Poole and Alan K. Mackworth
31. "Artificial Intelligence: A New Synthesis" by Nils J. Nilsson

a.

32. "The Anatomy of Fake News" by Stephanie Craft and Seth C. Lewis
33. "Fake News: Understanding Media and Misinformation in the Digital Age" by Brian L. Ott and Cara Louise Buckley
34. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making" by Claire Wardle and Hossein Derakhshan
35. "Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics" by Yochai Benkler, Robert Faris, and Hal Roberts
36. "Fake News and False Flags" by Joseph P. Uscinski
37. "The Misinformation Age: How False Beliefs Spread" by Cailin O'Connor and James Owen Weatherall
38. "The Hoaxer Project: Understanding and Countering Fake News" by Zizi Papacharissi
39. "Disinformation and Democracy: The Global Challenge of Fake News" edited by Jennifer Kavanagh and Michael D. Rich
40. "The Post-Truth Era: Dishonesty and Deception in Contemporary Life" by Ralph Keyes
41. "Fake News, Propaganda, and Plain Old Lies: How to Find Trustworthy Information in the Digital Age" by Donald A. Barclay
42. Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd ed.). Pearson.
43. Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
44. Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf.
45. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
46. Silver, D., Schrittwieser, J., Simonyan, K., et al. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354-359.
47. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

48. Marcus, G. (2018). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon.
49. Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.
- a.
50. "Legal and Regulatory Challenges in Combatting Fake News with AI" Li, Y., & Wang, J.
51. "User Behavior Analysis and Fake News Propagation on Social Media" Chen, X., & Liu, B.
52. "Cross-lingual Fake News Detection Using Machine Translation" Kim, M., et al.
53. "Blockchain Technology for Trustworthy News Sources" Zhang, L., & Wang, Q.
54. "Deep Learning in Multimedia Forensics: Fake Image and Video Detection" Liu, C., & Zhang, H.
55. "Explainability and Interpretability in AI Models for Fake News Detection" Wang, H., et al.
56. "Human-AI Collaboration in Fake News Detection" Brown, E., et al.
57. "Robustness of AI Models in Fake News Detection: Challenges and Solutions" Chen, Q., & Wang, Y.
58. "Semantic Analysis for Fake News Detection: A Literature Review" Wang, Z., et al.
59. "A Survey of Deep Learning Approaches for Fake News Detection" Liu, Y., et al.
60. "Ethical Considerations in AI-Based Fake News Detection" Patel, M., et al.
61. "Fake News Detection Using Social Media Analytics and Machine Learning" Gupta, R., & Singh, A.
62. "The Role of Machine Learning in Combating Disinformation" Kim, H., & Lee, S.
63. "Natural Language Processing Techniques for Fake News Detection" Chen, Y., et al.
64. "Deep Learning for Fake News Detection: A Comprehensive Review" Wang, L., & Zhang, F.
65. "Automated Fact-Checking: A Comprehensive Review" Smith, J., & Johnson, A.
66. Turet, D., "Reaction to fake news", 2018.
67. Orsini, D., "The impact of social media on the information landscape", 2019

68. Wissinger, K., “Artificial intelligence in the fight against disinformation”, 2020
69. Howard, D., “Technological Challenges of Detecting Fake News”, 2019
70. Bajan, A., “Network analysis in the context of fake news”, 2021
71. Clark, E., “The role of deep learning in filtering fake news”, 2020
72. Lindermann, K., “Ethics of AI in the context of information filtering”, 2021
73. Mackenzie, A., “The future of fighting fake news”, 2022
74. Ferris, L., “Synergy of technical and socio-cultural approaches”, 2021

Додатки

Додаток 1

Мета: Фокус-група: Використання ШІ для виявлення фейкових новин

Вік: 20-29 років

Кількість учасників: 5

Тривалість: 60 хвилин

Мета: Дослідити ставлення та думки молоді щодо використання технологій штучного інтелекту (ШІ) для виявлення фейкових новин.

Аня (22 роки, студентка, активний користувач соцмереж):

Фейкові новини:

- З фейковими новинами стикаюся досить часто, особливо в соцмережах.

- Зазвичай я перевіряю джерело інформації, шукаю підтвердження в інших ЗМІ та звертаю увагу на емоційний тон заголовка.
- Мене найбільше турбують фейкові новини про політику та здоров'я.
- Я бачила, як фейкові новини призводили до розпалювання ворожнечі та дезінформації.

ШІ та фейкові новини:

- Про використання ШІ для виявлення фейкових новин я чула, ідея здається мені цікавою.
- Я позитивно ставлюся до цієї мети, адже ШІ може обробляти великі обсяги даних значно швидше, ніж люди.
- Перевагою ШІ я вважаю його неупередженість та можливість виявляти складні закономірності.
- Мене турбує можливість цензури та маніпулювання інформацією з боку ШІ.
- ШІ може допомогти автоматизувати процес перевірки новин, звільняючи час для більш глибокого аналізу.

Практичний досвід:

- Я б перевірила джерело новини, пошукала б інші статті на цю тему та запустила б інструмент на основі ШІ для перевірки фактів.
- Мені б хотілося мати доступ до розширення браузера, яке позначає фейкові новини та пропонує альтернативні джерела інформації.
- ШІ може допомогти мені економити час на перевірці новин та бути більш впевненою у їх достовірності.

Завершення:

- Моя думка про ШІ не змінилася, я вважаю, що це ефективний інструмент для боротьби з фейковими новинами.
- Мені б хотілося дізнатися більше про алгоритми роботи ШІ та про те, як він навчається.
- Так, я готова використовувати ШІ-інструменти для перевірки новин.

Максим (25 років, програміст, користувач Reddit):

Фейкові новини:

- На жаль, фейкові новини стали невід'ємною частиною онлайн-середовища, і я стикаюся з ними досить часто, особливо на Reddit.
- Щоб перевірити достовірність інформації, я зазвичай шукаю підтвердження в авторитетних джерелах, таких як наукові статті або перевірені ЗМІ.
- Мене найбільше турбують фейкові новини, пов'язані з наукою та технологіями, адже вони можуть мати значний вплив на суспільство.
- Я бачив, як фейкові новини призводили до поширення конспірологічних теорій та поляризації суспільства.

ШІ та фейкові новини:

- Про використання ШІ для виявлення фейкових новин я знаю і вважаю, що це перспективний напрямок розвитку.
- Я позитивно ставлюся до цієї мети, адже ШІ може аналізувати великі обсяги даних і виявляти закономірності, недоступні людському оку.
- Перевагою ШІ я вважаю його швидкість та об'єктивність при оцінці інформації.
- Мене турбує можливість упередженості алгоритмів ШІ та їхня залежність від даних, на яких вони навчаються.
- ШІ може допомогти автоматизувати процес виявлення фейкових новин, роблячи перевірку інформації більш ефективною.

Практичний досвід:

- Для початку я б перевіряв джерело новини та проаналізував коментарі інших користувачів. Потім я б запустив інструмент на основі ШІ для перевірки фактів та шукав би додаткову інформацію на авторитетних сайтах.
- Мені б хотілося мати доступ до розширення браузера, яке візуалізує ймовірність фейковості новини та пропонує альтернативні джерела інформації.
- ШІ може допомогти мені економити час на перевірці новин та приймати більш обґрунтовані рішення щодо споживання інформації.

Завершення:

- Моя думка про використання ШІ для виявлення фейкових новин зміцнилася. Я вважаю, що це необхідний інструмент для боротьби з дезінформацією в Інтернеті.
- Мені б хотілося дізнатися більше про те, як розробляються алгоритми ШІ для виявлення фейкових новин та про їхню ефективність.
- Так, я готовий використовувати ШІ-інструменти для перевірки новин, але за умови їх прозорості та неупередженості.

Ірина (27 років, маркетолог, користувач Twitter):

Фейкові новини:

- На жаль, фейкові новини стали поширеною проблемою, особливо в Twitter, де інформація поширюється дуже швидко.
- Щоб перевірити достовірність інформації, я зазвичай шукаю підтвердження в інших користувачів, перевірених ЗМІ та профільних експертів.
- Мене найбільше турбують фейкові новини, пов'язані з політикою та економікою, адже вони можуть мати значний вплив на прийняття рішень людьми.
- Я бачила, як фейкові новини призводили до поширення паніки, розпалювання ворожнечі та маніпулювання громадською думкою.

ШІ та фейкові новини:

- Про використання ШІ для виявлення фейкових новин я чула і вважаю, що це може бути ефективним інструментом.
- Я позитивно ставлюся до цієї мети, адже ШІ може аналізувати великі обсяги даних та виявляти патерни, які складно побачити людині.
- Перевагою ШІ я вважаю його масштабованість та можливість працювати в режимі реального часу.
- Мене турбує можливість цензури та обмеження свободи слова при використанні ШІ.

- ШІ може допомогти автоматизувати процес перевірки новин, звільняючи час для аналітики та глибокого розуміння інформації.

Практичний досвід:

- Для початку я б перевірила профіль автора новини, проаналізувала б його активність та інші твіти. Потім я б запустила інструмент на основі ШІ для перевірки фактів та шукала б додаткову інформацію від авторитетних джерел.
- Мені б хотілося мати доступ до розширення браузера, яке візуалізує ймовірність фейковості новини та пропонує альтернативні джерела інформації, а також можливість отримувати повідомлення про фейкові новини, які я вже поширила.
- ШІ може допомогти мені зберегти репутацію в онлайн-середовищі та уникнути поширення недостовірної інформації.

Завершення:

- Моя думка про використання ШІ для виявлення фейкових новин зміцнилася. Я вважаю, що це важливий крок до створення більш інформативного та безпечного онлайн-простору.
- Мені б хотілося дізнатися більше про етичні аспекти використання ШІ для перевірки новин та про заходи щодо запобігання упередженості алгоритмів.
- Так, я готова використовувати ШІ-інструменти для перевірки новин, але за умови прозорості їх роботи та незалежного аудиту.

Тарас (29 років, дизайнер, користувач Facebook):

Фейкові новини:

- На жаль, фейкові новини трапляються й на моїх улюблених платформах, особливо у Facebook, де багато контенту генерується користувачами.
- Щоб перевірити достовірність інформації, я зазвичай шукаю підтвердження в перевірених ЗМІ, аналізую коментарі та звертаю увагу на дату публікації.
- Мене найбільше турбують фейкові новини про здоров'я та особисте життя, адже вони можуть мати негативний вплив на людей.

- Я бачив, як фейкові новини призводили до поширення страхів, паніки та недовіри до авторитетних джерел інформації.

ШІ та фейкові новини:

- Про використання ШІ для виявлення фейкових новин я чув, але не до кінця розумію, як це працює.
- Я з обережністю ставлюся до цієї мети, адже штучний інтелект може помилятися.
- Перевагою ШІ я вважаю його швидкість та можливість аналізувати великі обсяги даних.
- Мене турбує можливість маніпулювання інформацією з боку ШІ та його вплив на свободу слова.
- ШІ може допомогти полегшити процес перевірки новин, але остаточне рішення все одно має залишатися за людиною.

Практичний досвід:

- Для початку я б перевіряв джерело новини, проаналізувала її стиль та мову написання. Потім я б запустила інструмент на основі ШІ для перевірки фактів, але ставилася б до його результатів з критикою.
- Мені б хотілося мати доступ до розширення браузера, яке позначає фейкові новини та пропонує альтернативні джерела інформації з обґрунтуванням їх достовірності.
- ШІ може допомогти мені економити час на перевірку новин та отримувати більш збалансовану картину подій.

Завершення:

- Моя думка про використання ШІ для виявлення фейкових новин не остаточна. Я вважаю, що потрібно більше інформації про ефективність та безпеку таких інструментів.
- Мені б хотілося дізнатися більше про те, як люди навчають алгоритми ШІ виявляти фейкові новини, а також про можливість контролювати їх роботу.
- Я готовий використовувати ШІ-інструменти для перевірки новин, але за умови їх прозорості та відкритого коду.

Оля (20 років, студентка, користувач TikTok):

Фейкові новини:

- На жаль, фейкові новини стали невід'ємною частиною онлайн-середовища, і я стикаюся з ними досить часто, особливо на TikTok та Instagram.
- Щоб перевірити достовірність інформації, я зазвичай шукаю підтвердження в авторитетних джерелах, таких як Wikipedia або YouTube-канали відомих блогерів.
- Мене найбільше турбують фейкові новини про політику та історію, адже вони можуть спотворити моє уявлення про світ.
- Я бачила, як фейкові новини призводили до конфліктів, поляризації суспільства та недовіри до історичних фактів.

ШІ та фейкові новини:

- Про використання ШІ для виявлення фейкових новин я чула, ідея здається мені цікавою, але я не певна, як це буде працювати.
- Я позитивно ставлюся до цієї мети, адже ШІ може аналізувати великі обсяги інформації швидко та ефективно.
- Перевагою ШІ я вважаю його неупередженість та можливість виявляти приховані патерни.
- Мене турбує можливість цензури інформації з боку ШІ та його вплив на свободу слова.
- ШІ може допомогти автоматизувати процес перевірки новин, звільняючи час для глибшого вивчення теми.

Практичний досвід:

- Для початку я б перевірила джерело новини, проаналізувала її стиль та мову написання. Потім я б запустила інструмент на основі ШІ для перевірки фактів, але ставилася б до його результатів з обережністю.

- Мені б хотілося мати доступ до розширення браузера, яке позначає фейкові новини та пропонує альтернативні джерела інформації у форматі коротких відео або інфографіки.
- ШІ може допомогти мені економити час на перевірку новин та бути більш впевненою у своїх знаннях.

Завершення:

- Моя думка про використання ШІ для виявлення фейкових новин зміцнилася. Я вважаю, що це важливий крок до створення більш об'єктивного інформаційного простору.
- Мені б хотілося дізнатися більше про те, як боротися з упередженістю алгоритмів ШІ та зробити їх більш універсальними.
- Так, я готова використовувати ШІ-інструменти для перевірки новин, але за умови їх прозорості та відкритого доступу до коду.

Додаток 2

Інтерв'ю 1

- **МАРІАННА ПРИСЯЖНЮК**
- Національна академія внутрішніх справ за спеціальністю «Психологія» та Національний університет політичних та адміністративних студій (SNSPA, București) за спеціальністю «Комунікації та Європейський союз». Наразі досліджує зовнішню політику Російської Федерації в аспірантурі Бухарестського національного університету. З липня 2023 року очолює Групу з аналізу гібридних впливів Українського кризового медіацентру (HWAG\UCMC) й продовжує співпрацювати з фактчекінговим проектом StopFake.
-
- І чи ви даєте дозвіл на те, щоб я міг писати ваше ім'я та прізвище батькові в своїй дипломній роботі? Звісно. Дякую. Окей, добре.

-

Тоді виявлення фейкових новин?

- Я користуюся доступними інструментами, доступними методами для загальної перевірки інформації в контексті фактчекінгу і в контексті протидії інформаційним загрозам. як остаточна інстанція. Він може бути допоміжним

- наразі,

-

- 0:52

- але я б не визначала його як основним в роботі і в центрі фактчекінгової роботи. Тому в загальних інформаційних моментах його можна використовувати. Щось специфічне в контексті гібридної журналістики, якою є фактчекінг, в принципі чати, штучний інтелект можуть допомогти у формулюванні структурних моментів,

-

- 1:27

- у формулюванні якогось наративної послідовності. Тобто це допомагає з точки зору формальної логіки побудови тексту, визначення якихось семантичних особливостей, тобто, підвести теоретичну базу, яка вже точно є включеною в інтелектуальну базу, яка стоїть за штучним інтернетом, з якої штучний інтернет генерує свій контент. Коли ми пишемо статтю чи новину чи якийсь дебанк, ми генеруємо якийсь новий смисл в будь-якому разі покладатися. В цьому, звісно, на штучний інтелект не варто, тому що нам треба захистити і свою інтелектуальну унікальність цього контенту, який ми продукуємо. Це по-перше, тому що коли ми перевіряємо штучним інтелектом свій контент, ми його скормлюємо фактично в цю інтелектуальну базу, яка стоїть за цим інструментом. Ну, по-друге, штишний генетик не формулює сенси.

-

- 2:36

- Він витягує вже старі зв'язки із цієї бази, яку йому вже скормили до нас. Тому щось нове навряд чи. Окей.

-

Які, на вашу думку, найбільші переваги та виклики штучного інтелекту?

-

- Переваги – це швидкий доступ до величезного масиву теоретичних знань, швидкий доступ, підбір відповідної теорії когнітивного впливу, тобто ви можете спитати у штучного інтелекту про ефект,

- 3:19

- яку він думку, наприклад, має, яка б теорія доречніше пояснила той чи інший прояв, як певний інформаційний вплив діє на суспільство. Ви можете спитати у штучного інтелекту, яка б методологія була доречною в тому чи іншому випадку, але все одно вам самому її треба розробляти. Тобто швидший інтелект може запропонувати певні структурні зв'язки, використовуючи масив знань доступних йому. Але сприяти створенню якоїсь нової ідеї з цього, я думаю, що без відскоку креативності неможливо бігти. Ну і, власне, факт-чекінг. Якщо ми покладаємось на

-

- 4:09

- таку тонку аргументацію в певних моментах, яка би сприяла тому, щоб розбити певний шкідливий народ, певний шкідливий. Дуже важливий елемент людської творчості. І це не є характеристиками штучного інтелекту. Штучний інтелект допомагають вже людині, оперуюючи цим, сформуванню новий кут, нове бачення, нову ідею.

-

- 4:56

- Тобто, в ідеалі це має працювати разом. Як раніше, мені в школі забороняли користуватися калькулятором, коли там п'ятий клас, я пам'ятаю, що стосовно ставлення людей до цих інструментів, це щось подібне. Тобто нам треба зберегти

навичку вміти і бути здатними напоруватися самим погнітивно і генерувати якісь ідеї. Але при цьому певні процеси перекласти на швидший інтелект, що нам би допомогло черпати якусь інформаційну базу. Тобто, ну, так само, як маючи інтернет, ми маємо швидкий доступ до інформації, до соціальних мереж,

- 5:44
- і це вплинуло дуже на комунікацію, на комунікаційні процеси взагалі, і там інтерперсональні, інтергрупові, і на дуже багато таких речей. Раніше дисертації писались, люди поділили в бібліотеку, сиділи, переписували абзаци з книжок,
- 6:00
- які були доступні тільки, ну, якщо ви прийдете і фізично замовите цю книгу. Тобто ми маємо черговий етап комунікації, комунікаційну революцію. Якщо розглядати з точки зору журналістики, з точки зору інформаційних наук, звісно, з точки зору національної безпеки і інші виклики. Тобто з нашої вузької перспективи це черговий етап розвитку комунікації як феномену.
- 6:34
- Дякую за відповідь. Перейдемо в технічні питання.
-

Які типи алгоритм, який побудований за принципами кібербезпеки і за принципами кінестичних, тобто традиційних воєн.

-
- запрограмувати інформаційний вплив і зобразити інформаційний вплив як кампанію. Фреймворк не ідеальний, але його перевага в тому, що він дозволяє побачити інформаційний вплив з точки зору загрози, не з точки зору, наприклад, маркетингу в якихось інструментів, не з точки зору виключно факт-чекінгу, а з точки зору війни. І я думаю, що це дуже допоможний інструмент. Якщо ви знайомі з концептом хімії,
- 7:47

- foreign information, як я звуком, як це шифрується, foreign information malign interference, якось так, то, в принципі, якщо не помиляється єдиний концепт, який включає офлайн-активності, активні заходи. І це дуже важливо. Це саме те, про що StopFake говорив і звертав на це увагу роками. Тому що інформаційні впливи, пропаганда, дезінформація, це і створення артефактів, так званих розтаєвання листів, тобто створювання інформаційних приладів, на основі яких йде подальша дестабілізаційна робота.

- 8:47

- Таких два моменти. Дякую.

-

- А як оцінюєте ефективність алгоритмів?

- Ну, я думаю, що це... Ефективність треба оцінювати комплексно, яку ви ставите перед собою мету. Ну, алгоритми, ну я не знаю, якщо засновникам чату GPT, вони ставили перед собою мету зробити крутий чат, який захопить світ. Вони свої мети досягли, вони ефективні. Ми з ПУКРІК ставимо перед собою мету створювати зміцнення резіліансу українського інформаційного поля, відбувати інформаційні наративи

-

- 9:30

- Російської Федерації, які Росія намагається просунути в українське інформаційне поле. Ну і це наша мета і наша робота чи досягаємо ми цієї мети, визначає, чи ми ефективні чи ні. Я думаю, що в якійсь мірі в нас виходить, в якійсь мірі ні. Звісно, ми при цьому використовуємо певні алгоритми, певні фреймворки і дуже багато інструментів штучного інтелектом, коли мені треба знайти певну теорію, яка би пояснювала той чи інший ефект. Тобто я можу собі дозволити звернутися до того ж чата, щоб він мені порадив методологію. Далі вже я її сама опрацьовую, перевіряю і приймаю рішення, чи мені ця структура підходить для моєї подальшої унікальної роботи, для мого унікального тексту.

-

- 10:29

- Тобто я можу позичити в алгоритму рамку, можу звернутися до нього як до бази знань і вже потім ці знання використовувати для свого матеріалу. Також колеги, я знаю, використовують алгоритми для перевірки зображень, для перевірки відео. Коли ми аналізуємо традиційними способами зображення, якими ілюструються певні матеріали, ми звертаємо увагу наскільки вони...

-

- 11:00

- чи вони згенеровані штучно, спеціально, з такою метою, щоб підсилити той чи інший наратив. Тобто, ну, тому що штучний інтелект, він не дуже сильний в плані точності генерації певних зображень. Може бути ще кілька пальців, можуть бути помилки в буквах, особливо в цифрах. Ну і звичайна людська уважність.

-

- 11:25

- Вона допомагає визначити, наскільки певні алгоритми були використані проти нас, чи ми використовуємо їх. Тобто це рух вдві сторони. Треба певно використовувати свою користь ці речі, ну і звісно звертати увагу, як ці речі використані проти нас. Ну, дідфейк, такий феномен, цього року, минулого року, це просто по вихідній річі ці речі відбуваються. І ми фіксуємо, що це застосовується найчастіше наразі проти перших осіб держави, проти топових політиків. Ну і як така технологія делегітимізації держави через дискредитацію перших осіб держави,

-

- Ви значите наскільки певні алгоритми були використані проти нас чи ми використовуємо їх. Тобто це рух в дві сторони. Треба повільно використовувати свою користь ці речі, ну і звісно звертати увагу як ці речі відбувається. І ми фіксуємо, що це застосовується найчастіше наразі проти перших осіб держави, проти топових політиків. Ну і як така технологія делегітимізації держави через дискредитацію перших осіб держави, ну, зараз використання дідфейку фіксується все частіше і частіше.

-

- 11:58
- І не тільки проти України. Але я від себе можу відмітити, що також ці відео наразі дуже низької якості, і це незбройним оком видно, що це згенерований контент. І за звуковими якимись індикаторами, за візуальними індикаторами, також за деталями, які неадекватно зображені деталі, неадекватно витворені емоції на обличчях. Тобто багато таких моментів, які людина ще здатна відрізнити, навіть не маючи якоїсь спеціального обладнання, чи спеціальної програми, чи спеціальної підготовки. Але ризик в тому, що звісно технології будуть здосконалюватися швидше, ніж людські когнітивні здібності.

-
- 12:40
- Тому, ну, нам, я думаю, потрібні технології, або потрібні нові підходи для використання технологій на свій захист від них же самих. Як певний якийсь фаєргол, я не знаю. Тому що очевидно, що розвиток людського мозку, він за часом рамками відставатиметься за революцією технології. Ну і до цього треба знову ж таки підходити креативно і покладатися не на рух людини вперед, а на рух людських ідей вперед, що ми можемо придумати унікальний нове для того,

-
- 14:17
- щоб воно нас захищало в подальшому від нашого ж виробу, від наших технологій. Дякую. Зараз ми вже перейдемо до етичних питань.

•

Ви вже трішечки затронули тему етики в штучному інтелекті, але які етичні проблеми виникають при використанні?

-
- Я думаю, що етичні проблеми, звісно, виникають, тому що це плагіат, в першу чергу, мені здається. І я думаю, якщо автори розуміли б

-
- 2:54

- рух інформації взагалі в цій нейромережі, цих етичних би викликів було менше. Тому що, наприклад, для журналіста унікальний текст, це його заробіток, це його ім'я, це його рух по професійному житті взагалі. І коли постають такі виклики перед вами особисто як спеціалістам, вам менше хочеться скормлювати свого контенту в штучному інтелекту, тому що ви розумієте, що в якийсь мірі штучний інтелект – це ваш конкурент. І навіть якщо ви адекватно оцінюєте, що цю битву навіть не варто вв'язуватися, хочеться свій контент берегти і генерувати самостійно певні ідеї.

-

- 3:43

- Але відмовлятися від можливостей, які надає штучний інтелект інформаційникам, інформаційним спеціалістам, журналістам, аналітикам, це також безглуздо, тому що, окрім викликів нових, це нові можливості, це оптимізація роботи – це доступ швидкий до світової бази знань. Я нагадаю, що коли ЧАТ, наприклад, ну це мій особисто улюблений інструмент, і коли ЧАТ став доступним в нашому регіоні, якщо я не помиляюсь,

-

- 4:21

- його база знань включає в себе і ситуацію стосовно російської агресії після 22-го року, 24-го лютого. Тобто швидкість наповнення цієї бази неймовірна і всі ці, так чи інакше, всі ці знання будуть включені. яким вам вдасться зберегти свій контент, ексклюзивність свого контенту, це буде, я думаю, впливати і на вашу успішність, на ваш заробіток, на вашу ексклюзивність, на вашу професійність, ну і взагалі на сенс вашої присутності в професії. Я думаю, що врешті ми зіштовхнемося з такою дилемою, що в професії, в інформаційному полі зможуть залишитися лише якісь спеціалісти,

-

- 5:20

- які певною мірою будуть здатні захистити свій контент від поглинання штучним інтелектом. І це цікаво, тому що інформаційні технології вскоріть, і штучний

інтелект – це не тільки чат, це і створення презентацій, це і визначення облич, це і системи безпеки, і дрони, і це все. Штучний директ проник в кожен сфери, інформаційно-комунікаційної сфери, мабуть, краще сказати, то це питання контенту, те, що ми генеруємо,

-
- 6:12
- те, що ми зміємо зберегти в людській формі, це буде мати сенс. Окей.
-

А як можна захистити приватність саме? Бо коли ти використовуєш штучний інтелект, ти вписуєш якісь дані особисті, ну і як це захистити?

-
- Дивіться, я приватність розглядаю, ну взагалі навіть не знаю як почати, в мене колись був дуже класний теренінг з безпеки для журналістів,
-

- 6:46
- і з тих пір я приватність як елемент безпеки розглядаю з двох перспектив. Перша перспектива, чи варто її захищати. І друга перспектива, що варто її захищати. У нас дуже багато таких ситуацій, коли захищаючись своєю приватністю, ми створюємо собі загрозу. І дуже багато водночас таких ситуацій, коли надмірно відкриваючись,
-

- 7:10
- ми створюємо знову ж таки собі загрозу. Тобто повертаємося до людського фактору. Кожна ситуація має бути оцінена окремо і оцінена з точки зору саме людського творчого креативного потенціалу. Тобто там, де буде збережений ваш людський вимір, оцінка ситуації через ваш людський вимір, а не через машинний, там ви собі зможете забезпечити безпеку. Я думаю, що ми такий міні-закон навіть з вами вивели. Тому що коли ми, наприклад, в соціальних мережах, ну я наведу

такий приклад, тому що мені ця дискусія навколо світлого інтелекту дуже нагадує дискусію навколо зростання

- соціальних мереж теж.

- 7:52
- Етичні дилемми, наскільки це впливає на комунікацію, на національну безпеку. простір навколо себе, там де зберігаємо людську природу. Там де ми віддаємо це на відкуп технології, ми це вже не можемо контролювати. Там немає етики, в машинах, в комп'ютерах немає етики. Там є математична доцільність. Ну а це не завжди відповідає інтересам справедливості, етики, естетики.

-
- 8:18
- Рухають суспільство по коридору світлої сторони історії. От власне безпека. Якщо ви впевнені в своїй роботі, впевнені в своїх контактах, якщо ваша робота не ставить під загрозу ваші джерела, якщо ви можете захистити свою позицію, тобто вам доречно бути, якомога відкритіше. Вам доречно робити селфі із людьми, з якими ви спілкуєтесь, вам доречно їх постити, викладати в мережу, вам доречно підкріплювати свої дії артефактами. під загрозу життя чи вплинути негативно на інших суб'єкти вашої цієї діяльності,

-
- 9:12
- якоїсь спільної, на інших людей, то звісно ви маєте відповідно оцінювати і свою приватність. Я, наприклад, в своїй роботі придержуюсь з першої такої перспективи. Я максимально відкрита. Звісно, я не публікую якихось таких матеріалів, які стосуються виключно мого якогось приватного, гігієнічного життя, але там хто мої друзі, хто мої контакти, куди я подорожую, де я відпочиваю, що я пишу, з ким я сьогодні говорила. Ну, тобто мене не можна дискредитувати, наприклад, засобами якогось таємничого стеження чи щось таке. Але в принципі

це з іншого боку відкриває перспективу для генерування якогось файкового контенту,

-
- 10:10
- базуючись на моїй публічності. З іншого боку це відкриває можливість для того, щоб вкрасити мою цифрову ідентичність, використати її на іншому боці планети, в якійсь рекламі якогось спрею, відмук. Ну, тобто, це також впливає на мою безпеку в загальній перспективі. Тобто, це ваш вибір, як ви оцінюєте ризики для себе.
-
- 10:38
- Це, по-перше. По-друге, який би вибір ви не зробили, є ще інші моменти, соціальні, інженерії, психологічні моменти, які можна використати, ну, тобто, любого ситуацію можна використати проти людини. Ну це гра вдовго і тут немає цілі і немає пункту Б. Дякую. Ну і вже поговоримо про майбутнє, чуть-чуть дуже.
-
- Це які нам виклики та можливості виникнуть у зв'язку з розвитком ШІ?

• 11:09

Можливості і виклики?

- Так. Ну, я думаю, що виклики, можливості... Я думаю, що основним викликом є задача перед людиною якось все-таки прилаштуватися, встигнути за неймовірним розвитком технологій і озброїтися певними навичками, які дозволяли б контролювати, а не конкурувати,
-
- 11:42
- тому що конкурувати з технологіями людина ніколи не зможе, але оскільки технології все одно створила людина, знання, навички по створенню, контролю,

управлінню таких процесів, це, я думаю, задача і ринку праці №1, і задача загальної освіти №1. І це мають бути навички на рівні таблиць множення в майбутньому. Ми маємо років 10 для того, щоб наступні покоління володіли вже на якомусь середньому рівні мовами програмування, так як наше покоління більш-менш вже виговорилося англійською мовою.

-
- 12:28
- Тому що, наприклад, мої батьки, для них іноземні мови – це щось таке недосяжне. Моє покоління, не всі, але скажу прямо так, мені 32 роки, в моєму колі спілкування 70% володіють англійською мовою нормально, на нормальному професійному рівні. Наші діти вже, це для них буде базова навичка в мультикультурному середовищі. Такою ж самою базовою навичкою повинні здати навички з користування інформаційними технологіями, з безпеки в інтернеті, з безпеки користування ші.

-
- 13:07
- Це щодо задач. Ми повинні остерігатися того, що наш просерв для людської творчості буде згужуватися. І наша задача зберегти якомога довше, якомога більше простору для маневрування своїми, як сказати, людським потенціалом. Дякую.

І останнє питання вже на сьогодні, це який Ваш прогноз щодо впливу штучного інтелекту на боротьбу з дезінформацією?

-
- Я думаю, що він буде застосовуватися і фахівцями, і агентами інформаційних шпифів в достатньо великій мірі. в новолювій сфері, тому що, наприклад, ті ж самі діпфейки, вони поки що не статні ігнорувати такий контент, який би був узгоджений за цими індикаторами.

- 14:09
- реакції, знання якоїсь невербальної мови, як це все зображається під час мовлення. Це допомогло фактчекерам професійніше, наприклад, працювати з діфейк-відео, діфейк-зображеннями. І це саме той момент, там де штучний інтелект, мені здається, не назужений людям, принаймні не так Вербальну, невербальну використання – це досить така сфера, яку людину і можна вкерувати одночасно.

ІНТЕРВ'Ю

2

ІРИНА СУБОТА

З 2023 року - викладач курсу «Журналістська етика» у Могилянській школі журналістики.

З 2022 року працює в Центрі стратегічних комунікацій та інформаційної безпеки при Міністерстві культури та інформаційної політики України.

З 2020 року спеціалізується на фактчекінгу, медіаграмотності, стратегічних комунікаціях та аналізі інформаційних загроз. Була фактчекером в українських фактчекінгових проєктах (StopFake та проєкті «За лаштунками новин» (Інститут розвитку регіональної преси).

З 2020 по 2022 рік працювала в ГО «Інтерньюз-Україна» редакторкою проєкту «Інтернет-свобода в Україні». Аналітик Центру Стратегічних комунікацій при МКІП

Добре, то почнемо з першого блоку, це загальні питання. Який Ваш досвід роботи з технологіями штучного інтелекту в контексті виявлення фейкових новин?

Якщо ми говоримо про виявлення фейкових новин, зокрема, зараз я працюю в Центрі стратегічних комунікацій Моніторингові кабінети, які створені на основі штучного інтелекту, вони можуть відстежувати і аналізувати, по суті, або наративи, швидше ми говоримо про використання самих моніторингів. Чуково.

Які, на вашу думку, найбільші переваги використання штІт для виявлення фейкових новин?

- 1:02
- От найбільші. це відстеження і аналіз, але я все одно не давала 100% гарантію, що це прямо те, чому ми можемо 100% довіряти. Є певні все одно сумніви і, наприклад, навіть коли я щось відстежую, то в мене дуже багато може виникати сумніви стосовно того, наскільки той штучний інтелект правильно інтерпретував ту інформацію, яку він там отримав. І, наприклад, коли ми подивимося, є певні програми, так, чи онлайн ресурси, які так само допомагають, коли ми говоримо про фейки, вони теж допомагають нам відстежити, наприклад, чи це дїпфейк чи ні,
- 1:55
- чи було певне втручання, використання штучного інтелекту у створення цього дїпфейку, то навіть в тих програмах, ми побачимо, там буде така примітка завжди, що вони не дають такої 100% гарантії по тому, чи дійсно повноцінно було створено за допомогою штучного інтелекту чи ні. Так, вони показують певні відсотки, що ось такий відсоток там дійсно відстежується, що було втручання штучного інтелекту, але в деяких програмах завжди буде примітка про те, що ці програми там не можуть бути 100% гарантією, чи дійсно там є створення з допомогою штучного інтелекту. А якщо говорити про відстеження і про аналіз, то у мене там будуть виникати більше сумнівів стосовно правильності інтерпретації. Так як я вже сказала, чи правильно зумів прочитати штучний інтелект,
- 2:54

- тому що іноді важко проникнути в ту кількість інформацій, яку він зумів проаналізувати. І відповідно я вже бачу саме результат, який надає мені цей моніторинговий кабінет. Ого, цікаво дуже.

-

Як ще може допомогти покращити якість інформації в інтернеті? І чи може він її покращити?

-

- Так, ну тут треба, що ви самі маєте на увазі. Ну в принципі, якщо розробляти певні тех програми штучного інтелекту для покращення?

-

- 3:39

- Цікаве запитання, я просто не думала в цьому плані. Можливо, це б стосувалося більше. Навіть цікаво була б ідея, щоб так, як зараз там відстежуються оці, все одно є певний аналіз, чи було створено там зображення чи відео за допомогою штучного інтелекту, а можливо, могли б бути певні програми, які автоматично відстежували, аналізували, і відповідно були якісь позначки стосовно того, що це діпфейк,

- 4:10

- чи це зображення створене за допомогою штучного інтелекту, так згенеровано штучним інтелектом. Тому, напевно, знаєте, є певна така думка, що сам штучний інтелект, я б схилилася до думки, що штучний інтелект більше є нейтральним. Я б не говорила, що він чи негативний, чи позитивний. Це залежить від того, як його використовувати. Якщо ми говоримо про те, що штучний інтелект використовує пропагандисти з метою щось обманути, посіяти якісь інформаційні вкиди, то це вже буде негативне використання.

- 4:50

- Коли ми говоримо про штучний інтелект у контексті того, щоб якось протидіяти, виявляти, або якось попереджувати, це вже буде позитивне його використання. Тому, швидше за все, я б говорила про штучний інтелект як такий нейтральний

інструмент, який вже з метою, з якою він використовується, залежатиме його відповідна конотація. Окей, дякую. З першою частинкою ми завершили, зараз буде друга, де технічні питання.

- 5:23
- От, то які типи алгоритмів всі використовують саме для виявлення? Як воно працює? Так, ну тут такі вже технічні. Швидше, те що я бачу, я не розробляю ці програми, тому я б не могла сказати точно про ті алгоритми, як саме виявляють ці ресурси для аналізу певного зображення, там будуть певні відсотки, за якими... Ну, там дуже вже технічні такі показники в плані того, що саме використовувалось, як було сконструйоване це зображення, як відбувалася генерація за допомогою штучного інтелекту. А якщо, наприклад, говорити про використання, як алгоритми, наприклад, аналізують наративи,
- 6:36
- коли ми говоримо про моніторингові кабінети і відстеження, то в них є певна база певних джерел інформації, з яких аналізується. і відповідно так само там, умовно я навожу тільки на одному прикладі, так, моніторингово кабінети якісь саме базуються на штучному інтелекті, тому що вони є різні, ось і в моїй практиці, ну, оце один із прикладів, бо інші це швидше в такому ручному форматі, де вони просто там, вводиш як в Google ключові слова і воно просто надає певні публікації, які вже треба самостійно аналізувати. Але є і такі приклади, коли теж за ключовими словами відбувається пошук інформації
- 7:27
- і далі штучний інтелект намагається прочитати, це такі моніторинги на кабінеті, які знаходяться, можна сказати, більше в тестовому форматі, тобто, що вони розраховані, що зробили певну роботу, тоді там ще хтось дає певний фідбек, як покращити, тобто, це те, що я казала, що 100% гарантію я особисто не можу дати, що штучний інтелект це правильно прочитав. Але розробники намагаються, тобто вони почуть, зараз дуже багато є, якщо ви подивитесь, дуже багато є організацій, ІТ-організацій, які намагаються саме працювати в сфері протидії дезінформації.

Вони максимально почнуть розробити ті продукти, які будуть просто автоматично там відстежувати, попереджати про певні інформаційні вкиди.

- 8:29
- Тому робота над цим ведеться. Дуже багато і українських, і є приклади американських ІТ-компаній, які безпосередньо розробляють свої моніторингові кабінети. Поки що вони поєднують ручну роботу і поєднують використання штучного інтелекту. Тому що це ще поки що все вимагає, по суті, такої перевірки людської, наскільки це все правильно було подано в результат.
- 9:02
- Дякую. Дякую за відповідь. А зараз підемо вже в етичні питання. Оце от прям дуже цікаво. І тут, за словами своїх, то тут дуже так, 50 на 50 думки розходяться.
-

Які етичні проблеми виникають у використанні щі?

-
- Якщо це в плані, може бути, виявлення фейсів.
- 9:26
- Так, якщо ми говоримо про виявлення фейсів, тут буде трошки складніше, бо все дуже просто, коли ми говоримо про чат GPT чи інші ресурси на основі штучного інтелекту, які там пишуть нам певні відповіді, створюють певний текст, там будуть свої етичні питання стосовно виявлення фейків і використання штучного інтелекту, тут більше етичного я б, напевно, поклалась лише на такий нюанс, що повністю я б не покладалась або завжди зазначала, що цей висновок був поданий за допомогою штучного інтелекту. Тобто, що це зображення, наприклад, є згенероване штучним інтелектом
- 10:31
- і завжди зазначало про те, що такий висновок був наданий певним онлайн ресурсом, через який перевірялися. Чи якщо ми говоримо про формування нарративів, ми говоримо про те, що цей висновок був виданий за допомогою аналізу

- Загалом про оці позначки, вони якраз є, ну зараз тут говориться про обов'язковість їхнього використання, навіть коли ми говоримо про журналістику і говоримо про там певний, наприклад, що якийсь шматок тексту чи ще щось було зроблено за допомогою штучної текти, чи зображення було згенеровано, умовно, хотіли просто візуально оформити якоесь статтю, то це обов'язково має зазначатися. В іншому випадку це знову ж таки буде обман аудиторії. Це вже є вимоги, не в одних правилах, про позначення використання штучного інтелекту для створення певної навіть частинки, якщо це відбувалося.
- 0:44
- І тому я думаю, що стосовно ось виявлення, про етичні, Я б більше говорила про цей взаємозв'язок саме людини і штучного інтелекту, що якщо людина все-таки оформлює вже певні висновки щодо того, є це фейком чи ні, то тоді потрібно в такому разі зазначати, що ці висновки, або частина цих висновків, була зроблена на основі використання штучного інтелекту. Це вже більш загальне питання, не конкретно фейкові новини. використання цього штучного інтелекту, що мають бути певні обмеження на ту інформацію, яку може аналізувати штучний інтелект.
- 1:36
- Бо коли, наприклад, взяти чат GPT, то ми побачимо, здається, при першому використанні, там буде навіть попередження, що чат GPT для подальшого свого навчання використовує ту інформацію, тобто такі питання, які ви йому задаєте, і там є попередження, що будьте уважні з якоюсь щитливою інформацією, яку ви вносите в чат GPT. Але це зі сторони людини, тому що людина сама має бути обережна з тим, як вона працює з інтелекту, то тут дійсно треба говорити про те, що мають бути введені певні норми, так само і етичні, про які ви запитували, про те, що може використовувати штучний інтелект в ході свого навчання, що не може використання статей певних видань.
- 2:59

- От коли, здається, був з New York Times історія, що використання штучного інтелекту. Ці всі спроби є, є вже дуже багато розроблено і по суті етичних певних таких собі кодексів, і зараз розробляється на рівні ЄС використання саме штучного інтелекту і якраз те, про що ви ключове запитуйте, про збереження оцих конфіденційних даних, тобто те, яким чином використовувати і самими організаціями, і самими людьми, як взаємодіяти в ширшому розумінні, і яким чином оці дані можуть тоді використовуватися. Дякую. І вже останній блок на майбутнє.

- 4:05

Які нові виклики та можливості виникнуть у зв'язку з розвитком ШІ?

-
- От що в нас зараз штучний інтелект дуже активно розвивається. І які нові виклики, напевно, я б тут повернулася до тої думки, що як саме будуть використовувати це щіт. Тобто, якщо ми будемо його використання. Але, звісно, ми будемо говорити і про виклики, що завжди будуть залишатися. А відповідно, якщо ми говоримо про штучний інтелект як те, що може самостійно навчатися,

- 5:25

- тобто постійно удосконалювати свою технологію, то відповідно ці фейки теж будуть ускладнюватися і це вже буде трошечки важче відстежувати, відповідно потрібні будуть нові ресурси, програми, які зможуть щитувати це втручання і генерацію певних текстів чи зображень за допомогою штучного інтелекту. Саме негативну частину, це дійсно про ускладнення з цього поширення, умовно будемо говорити в контексті дезінформації, то стосовно фейків може бути абсолютне ускладнення, потім щитування і розуміння, де правдиве, де реальне зображення чи текст,

- 6:10

- а де сконструйоване і згенеровано за допомогою штучного інтелекту. Добре.

-

Який Ваш прогноз щодо впливу щі?

- Це загальне питання. Так. Мені здається, що ШІ потребує досить такого удосконалення зі сторони їхнього використання, зі сторони політики, зі сторони взагалі, як працюють ці самі GPT, не знімає проблеми, що їх будуть використовувати з негативною метою, і це буде трошки і ускладнювати з іншого боку. Тому я б тут швидше говорила про те, що в будь-якому разі ми повинні брати цю технологію,

- 7:28

- Просто вона має бути унормована і дійсно, там, подальшому удосконалена. І це все-таки буде нам використано більше як така допоміжна. Тобто я б говорила про штучний інтелект як певний інструмент. Я б не говорила про певну заміну там всього, що він буде замість нас там посрати, думати і так далі. Я б швидше говорила про те, що його б варто було використовувати як певний інструмент для полегшення певної роботи, тобто яка більш заключається в певній структуризації роботи з певними даними чи цифрами. Те, що більш таке вимагає. навчити комп'ютер і зробити це набагато швидше, звести якісь дані,

- 8:27

- зробити певний результат, визначити і так далі. Я б говорила швидше про використання саме як інструменту штучного інтелекту для полегшення певної роботи, а не про загалом заміну його людському знанню, людській роботі, а швидше з розумом використання. Ось так би я сказала. Дякую. І останнє питання.

- 8:56

Чи хотіли би ви мати програму на телефоні, яка буде автоматично щитувати всю інформацію, чи це фейк, і буде показувати, чому це саме фейк?

-
- Так. Тут цікава ситуація. Я уявила, якщо встановити таку програму на телефон, то швидше за все вона буде поруч з тим, щоб перевіряти зображення чи текст, чи це фейк, чи ні. щоб ви бачили, що я заходжу на цей сайт, я читаю цю інформацію і так далі. Можливо, або така програма мала б тоді обов'язково мати певні обмеження і правила щодо конфіденційності,

- 9:46
- тобто як будуть використати, завантаживши перевіриши щось. Ну, зокрема, в онлайн-просторі. Тому тут теж питання. Або він просто існує не конкретно на моєму телефоні, а просто як додаток до Google Home, чи ще щось, де він просто сканує сам контент без того, що він там щитує, конкретно просто дивитися загалом, це вже як свої там... В інтернеті без вчитування моєї інформації, моїх дій в цифровому просторі. Дякую, дякую, що дали відповідь. Це дуже-дуже цінно мені. Дякую, що знайшли час.
-
- 11:09
- Це вже всі питання, які були на сьогодні. І маю запитати ще одну річ. Чи даєте ви погодження на те, що я можу вас цитувати завдяки нашому інтерв'ю в своїй дипломній і писати ваше ім'я? Так, звісно. Так, так. Дякую. Тоді дякую вам ще раз велике, дякую, що знайшли час, що приділили його і дякую, що відповіли на всі питання. Так, прошу вам успішного завершення написання.

ІНТЕРВ'Ю 3

Тетяна Андрійчук

Кандидат політичних наук, старший викладач (сумісник) Факультет правничих наук

- Окей. І хотів ще уточнити одне питання. Чи даєте ви згоду, щоб я міг, господи, вписувати вас в свій диплом, саме тому, що ви це казали? Тобто прямо взагалі, щоб Тетяна Андрійчук щось сказала? Так. Або цитування вас. Ну в принципі можна, так, але питання в тому, що... Я не певна, що я вам по цій темі так багато зможу... Ну, такого корисного можу сказати, що цим прямо можна буде цитувати.
- 0:42

- Але тут всі питання такі доволі загальні. то тут не буде ніякого, що от пані Тетяна таке сказала, а це означає, що це отак працює. Ні, такого не буде в жодному випадку. По питанням в нас буде три блоки, це загальні, етичні і максимальні. Вони дуже-дуже загальні. Окей, тож почнемо.
- 1:07
- Які, на вашу думку, найбільші переваги використання шіри, проявлення фейку? Ну дивіться, насправді, так сказати, що ми випробовували використання штучного інтелекту для виявлення дезінформації. Але якщо уявити, що це буде впроваджуватися, то я думаю, що найбільша перевага може бути економія часу. Ми знаємо, що саме штучний інтелект дозволяє дуже швидко робити те, що людина або декілька людей може моніторити ручками, і шукати, і аналізувати. Знаючи, як працюють ці програми, зараз є багато програм, які за хвилину можуть проаналізувати великий масив інформацій, то все-таки мені здається, що якщо це буде впроваджуватися, то найперше це буде економія часу, так і економія ресурсів,
- 2:37
- в тому числі людських, тобто набагато менше треба буде зусиль людей. Будуть звичайно якісь працівники, які будуть потім ще це перевіряти, якоюсь людським аналізом, але загалом це буде набагато швидше і потребувати нам менше людей. Дякую. І можливо ще додати, можливо, як я розумію, що штучний інтелект дуже швидко навчається, можливо він буде навіть давати можливість, там умовно кажучи, дуже швидко реагувати на появу якихось нових видів фейків,
- 3:16
- використання якихось таких технологій, можливо, ще якби вони тільки починають використовуватися, щоб люди їх ще не зовсім будуть відчувати, а штучний інтелект за рахунок якихось алгоритмів може, умовно кажучи, Щось є там схоже на дезінформацію. А які існують виклики при використанні штучного інтелекту? Виклики?
- 3:44

- Так. Знову звук пропадав. Виклики... Ну, напевно, найперший виклик це в тому, що все одно треба його перевіряти. Повністю покладатися в будь-якій сфері користання штучного інтелекту,
- 4:05
- а тим більше, якщо ми говоримо про такі речі, повністю покладатися на те, що він буде правий. Принаймні зараз це занадто самовпевно. можуть бути помилки і відповідно має бути якась система контролю за цими помилками і зовнішнього людського аналізу. Ну, напевно, це найбільший виклик. Найбільший виклик того, що всі можуть помилятися, люди можуть помилятися. Особливо на початку. Ми ж розуміємо, що ці програми вчась-вчась, вони самовдосконалюються.
- 4:57
- Але на початку може бути такий. І плюс те, що зараз ще постає питання, навіть, можливо, ви чули в контексті оцієї, публічного використання штучного інтелекту, та що МЗС запусив віртуальну жінку, для консульських роз'яснень, консульських питань. Один з найперших викликів, про які почали всі говорити експерти,
- 5:30
- це про те, що можуть бути якісь несанкціоновані втручання і відповідно це виклик. Наскільки можна повністю убезпечитися, щоб ті програми, які використовуються, не зазнавали якогось втручання. Добре, дякую. А які методи інтелекту Ви вважаєте найпокиднішими? Це можете зараз розглядати так само і в держслужбі. Ну, конкретних програм я не скажу, бо, на жаль, аж так близько не стикалися. Знаємо, ці програми, які допомагають визначити, чи відео було створено, справжнє відео, чи це відео тим же самим штучним інтелектом створено,
- 6:37
- або воно зазнавало якихось втрачень. Так само по фотографіям. Дезінформаційні, які будуть найближчими роками перед органами влади, це дещо більше контенту відео, фотографій. Коли можуть підніматися якісь фотографії фейкові, чи відео, якісь записи військових, які військові щось говорять про те,
- 7:21

- що влада нас здала, а це насправді буде згенерована дезінформація. Програми, мені здається, саме будуть найбільш корисними, ті, які будуть допомагати працювати саме з такими питаннями,
- 7:39
- відеофотофейків, так, і давати можливість дуже рано це виявляти і, ну, як це є, здійснювати спростування, так, оперативно. Тому що ми ж бачимо в соціальних мережах, як воно, десь там якісь відео, де щось казали військові, воно починає віруситися, воно починає кудись розходитися, і відповідно саме програми, які давали можливість дуже швидко відслідкувати і сказати, що це фейкова дезінформація, щоб можна було державі,
- 8:16
- державним органам дуже швидко це спростувати, здається, це зараз буде одне з найбільш важливих, що треба було знати. Дякую. Зараз підемо вже в етичні запитання. Перше, які етичні проблеми виникають при використанні штучного інтелекту? Ну, да, складно сказати. Ну напевно все таки що, на мою думку, здається що етика це взагалі більше те що
- 8:55
- властиво людям, тобто не комп'ютерним програмам, а саме людям і відповідно до питань комунікації там людина людина і чому ж навіть дуже багато ще й критики було з приводу цієї штучної представниці МЗС. Це про те, що люди потребують все-таки людської уваги. Як це конкретніше сказати, прямо термінами про етику, але якщо так описово, це про те, що ми більше довіряємо людям, що ми хотіли б комунікувати з людьми, бачити, що дуже сильно...
- і пояснювати, чому органи влади, наприклад, вирішили саме для таких конкретних питань використовувати штучний інтелект. Щоб не наражатися на критику, як зараз казали, от вони створили цю жінку, тому що вони самі бояться повідомляти інформацію. Треба дуже сильно пояснювати, чому це. Пояснювати, до речі, і те, що це безпечно. Питання там всяких і даних, і доступу до якоїсь

інформації, і питання безпеки даних, і персональних якихось даних. Якщо, можливо, в роботі програм штучного інтелекту буде якийсь доступ до цього, то це такий найбільший виклик, який можуть виникати, що буде критикуватися так само в такій частині. Дякую. От те, що штучний інтелекту?

- 1:18
- Ну це взагалі питання, які складні і навіть і не в штучному інтелекті до кінця у нас не вирішено. Це навіть є експерти, які критикують навіть теж саме дію, яка діє не штучний інтернет, але за стосунок де достатньо великий доступ до даних персональних і як кажуть, що все таки були взлами цієї дії і можливо там ставалися витіки цієї інформації. Хоча з іншого боку мені здається, що відповідні органи, вони старалися і все що могли робили. Вони там, я пам'ятаю, робили ці забули, як вони називаються, коли навіть спеціально запрошують таких, умовно кажучи, хакерів, яких пропонують зламати і, умовно кажучи, таким чином шукають якісь сладкі місця.
- 2:16
- Тобто, воно все робилося, але питання в тому, що все одно воно до кінця не гарантує. Якщо в нас є ворог такий як Росія, яка дуже сильно може фінансувати ці речі і цих фахівців і це можуть зламати. Тому тут таке питання. Мені здається, що наші органи, які там Держспектзв'язково або Мінцифри, то вони все що можуть від себе робить, вони роблять по захисту. Але мені здається, що до кінця... Я сама до кінця не впевнена, що це сто відсотково можна захистити, враховуючи саме такі спеціальні...
- 3:08
- Спеціальну увагу спецслужб, наприклад, Росії. Тут як і буде цілесформовано це робити. Ну, але з іншого боку, можливо, якщо це якась там може бути співпраця з якимись міжнародними організаціями, корпораціями, так, які, можливо, там більше вже досвіду мають і щодо штучного інтелекту, і щодо захисту даних, щоб бути вирішено через співпрацю з найкращими фахівцями міжнародними. Не лише українські органи влади будуть захищати, а й провідні технології.
- 3:49

- І вже зібраємо в майбутнє. Останні два запитання. Перший, які виклики та можливості виникнуть в найближчі роки саме в контексті розвитку штучного інтелекту? Що ви думаєте, що може статися? Мені здається, що найбільший виклик, який може бути, це якраз виклик складності переходу на використання тими ж саме державними органами, особливо якщо це буде відкрито, що громадській буде знати, що органи влади починають використовувати саме якісь механізми штучного інтелекту, а тим більше якщо там для комунікації буде,
- 4:53
- як з тим самим МЗФ, тому що воно ніби такий сміливий крок, а з іншого боку, суспільство може виявитися не до кінця готовим, і не до кінця розуміти, чому, наприклад, вчора я ще міг подзвонити з людиною, поговорити в кулцентрі, А тепер я заходжу на сайти і мені пропонують зі штучним інтелектом комунікувати. Я знаю не в контексті штучного інтелекту, але колись потрапляло дослідження за кордоном. Не штучне інтелекто стосувалося, а автоматизованих відповідей скул-центр, так як роботи, такі алгоритми теж допомагають.
- 5:35
- і досліджувала, що люди навіть на захоті, вони дуже так негативно ставилися, вони хотілися такого людського спілкування, але не розуміли, чому вони з роботом мають розмовляти. Тому мені здається, що такий один з викликів може бути те, що не до кінця буде продуманий перехід в якихось там сферах на застосування штучного інтелекту і буде наражатися на критику того, що люди будуть не розуміти, вони будуть вважати, що це погіршення для них певне.
- 6:12
- Але це до того, що треба, мовно кажучи, придумувати стратегію стратегію застосування таких технологій штучного інтелекту, щоб не... Знову ж таки, це МЗФ, я згадую. Для нас це одна з подій, тому що саме в органі державної влади. Це в Міністерстві сталося. Тобто вони вирішили, і вони так зробили.
- 6:38
- Як на мене, напевно, було би дуже результативно, якби була якась така, можливо, стратегія державна, застосування якихось програм штучного інтелекту і там,

умовно кажучи, якийсь там і перехідний період, так, і вже якісь провідні сфери обумовлені, так, що отут, отут, отут, в таких та таких сферах це може бути там ефективно, зоційно, томуто, томуто, так, і, умовно кажучи,

- 7:06
- це буде проаналізовано і це буде системний перехід, так, тобто не те, що там один захотів в цій сфері, так, штучний інтелект, і орган менший, і орган в іншій сфері захотів. От, тому, мені здається, що, ну, але оскільки це є, і це, ну, ми не можемо закритися, так, сказати, що в нас там всього не існує і ми цього не будемо робити, так, все одно органи, скоріш за все, будуть, будуть застосування.
- 7:39
- І ті сфери, які якраз будуть найкраще підпадати, з них можна випробувати. Можливо, це якраз найперші сфери будуть ті, які будуть мати не так багато доступу до даних, власне, ті ж саме персональні дані, про які ви згадували. Тобто спробувати спочатку в якихось таких сферах, потім подивитися і вже думати, як впроваджувати в інші сфери.
- 8:03
- Плюс звичайна комунікаційна кампанія, пояснення людям. Можливо перші сфери це ті будуть, які знову ж таки, там десь стосуватися тієї ж самої молоді. Яка молодь вона ж краще сприймає технології, вона з більшим розумінням. Молодь навряд чи буде ставити питання, а чому ви вирішили штучну жінку заробити. Можливо, перші сфери будуть більше для молоді,
- 8:32
- потім подивитися і поступово розвивати, впроваджувати ці технології. Ну і теж можна, напевно, сказати, що певний виклик може бути і серед державних службовців. Тобто, якщо, наприклад, починати впроваджуватися оці засоби штучного інтелекту, наприклад, для тієї ж боротьби з дезінформацією, то можливо, якщо раніше моніторинг робила 10 людей, а тепер це може так само робити штучний інтелект і достатньо дві людини, які будуть це переглядати і потім приймати якісь рішення, які певні перестороги у цих державних службовців,

- 9:15
- чому це їх вимагається замінити якимось штучним інтелектом. І з цим тоді теж треба думати, як працювати. Переводити їх на інший напрямок роботи, або аналітичний. Думати, як так зробити, бо якщо нерозуміння державних службовців, то це неправильно, тому що воно може теж там саботуватися. Дякую. І останнє ще. Який ваш прогноз, що в найближчі роки це буде ще більше використовуватися. Я думаю, що є таке припущення, що можливо навіть в Україні буде пробування нових механізмів.
- 10:19
- Можливо, знову ж таки, зараз під час війни вона найбільше зазнає впливу дезінформації і знаєте, як зброю у нас, звичайну зброю, ну зараз дуже часто оцю нову зброю, яку там виробляють за кордоном то одночасно її випробування саме цих інструментів штучного інтелекту. Тому що йому потрібні дані, йому потрібна навала цієї дезінформації, а Україна, я статистики не маю, але маю підозру, що Україна може бути країна, яка найбільше зараз зазнає дезінформаційного впливу. А в штучному інтелекту треба чим більше інформації, чим більше можливостей навчатися на ті же інформації, яка заходить.
- 11:26
- Може бути так, що у нас будуть випробувані найближчими роками це буде дуже активно. Добре, це все.