ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА ЕКОНОМІЧНА БЕЗПЕКА

УДК 004.8:004.77:004.62 **DOI: 10.37332/2309-1533.2024.3.19**
JEL Classification: O310, O330

Дзюбановська Н.В.,
д-р екон. наук, професор,
професор кафедри прикладної математики,
Західноукраїнський національний університет, м.Тернопіль

ШТУЧНИЙ ІНТЕЛЕКТ ДЛЯ ВИЯВЛЕННЯ ФЕЙКОВОЇ ІНФОРМАЦІЇ: МІЖНАРОДНІ ТРЕНДИ І МОЖЛИВОСТІ ІМПЛЕМЕНТАЦІЇ

Dziubanovska N.V.,
dr.sc.(econ.), professor, professor at the
department of applied mathematics
West Ukrainian National University, Ternopil

ARTIFICIAL INTELLIGENCE FOR DETECTING FAKE INFORMATION: INTERNATIONAL TRENDS AND IMPLEMENTATION OPPORTUNITIES

Постановка проблеми. У сучасному інформаційному просторі, де обсяг доступної інформації безперервно зростає, проблема появи фейкових новин стала однією з найбільш актуальних. Фейкова інформація, яка може мати форму дезінформації, маніпулятивних повідомлень чи неправдивих новин, здатна впливати на громадську думку, формувати політичні настрої і навіть призводити до соціальних конфліктів. Ця проблема загострюється в умовах глобальних викликів, таких як вибори, пандемії, військові конфлікти, коли інформація може стати потужним інструментом впливу.

Традиційні методи виявлення фейків, такі як ручна перевірка фактів, вже не відповідають сучасним викликам через великий обсяг даних, що потребують обробки. У зв'язку з цим виникла необхідність використання новітніх технологій, зокрема штучного інтелекту (ШІ), для автоматизації процесу виявлення такого типу інформації.

Незважаючи на існування різних алгоритмів і систем, що базуються на ШІ, проблема залишається актуальною через: недостатню точність алгоритмів; потребу в постійному навчанні моделей для адаптації до нових форматів дезінформації; загрози, пов'язані з упередженістю в алгоритмах, що призводять до неправильного маркування контенту; низьку медіаграмотність населення, що ускладнює сприйняття інформації і підвищує вразливість до фейків тощо.

З огляду на це, аналіз міжнародного досвіду застосування ШІ для виявлення фейкової інформації є таким же важливим, як і оцінювання того, яким чином ці технології можуть бути адаптовані і впроваджені.

Аналіз останніх досліджень і публікацій. Протягом останніх декількох років спостерігається зростання інтересу до проблеми появи фейкових новин та потенціалу ШІ у їх виявленні. Адже стрімкий розвиток цифрових технологій значно спростив поширення інформації, але й водночас ускладнив процес перевірки фактів. Для подолання цих викликів потрібні нові підходи, що поєднують технологічні рішення та соціальні ініціативи.

Зокрема, Тищенко В. у науковій статті [1] аналізує різні методи навчання нейронних мереж для виявлення фейкових новин, охоплюючи текстові, візуальні та змішані дані. Розглядаються рекурентні, згорткові, глибокі нейронні мережі та генеративні змагальні мережі, а також методи навчання з вчителем і без. Виявлено, що якість і обсяг даних, а також складність фейкових матеріалів впливають на ефективність інструментів. Незважаючи на успіхи в застосуванні машинного та глибинного навчання, є потреба в подальших дослідженнях для покращення точності цих технологій у виявленні складних типів фейків.

Праздніков В. та Сугоняк І. у [2] досліджують проблему фейкового контенту та пропонують методи його виявлення за допомогою машинного навчання та аналізу текстових і мультимедійних даних. Автори акцентують увагу на доцільності застосування репрезентативного набору даних для навчання моделей, що здатні виявляти фейки, а також розглядають глибокі нейронні мережі, які автоматично можуть аналізувати лінгвістичні та візуальні ознаки для розпізнавання фейкового контенту. Також дослідники стверджують, що у наш час постала нагальна потреба для співпраці науковців, розвитку відкритих джерел даних та постійному вдосконаленні методів виявлення фейків з метою захисту інформаційного простору.

Боровик Д. та Бармак О. у [3] розглядають основні підходи до боротьби з фейками, що включають алгоритми машинного та глибокого навчання, а також аналізують сентименти й емоції користувачів соціальних мереж. У статті обговорюється авторська ідея щодо вдосконалення методу виявлення фейкових новин за допомогою нейромереж, зокрема шляхом збільшення кількості нейронів у згортковому шарі та додавання шару випадкового відключення для підвищення ефективності моделі.

Дацик С. у своїй праці [4] розробив інструмент за допомогою алгоритмів ШІ для виявлення дезінформації в новинах, що поширюються через соціальну мережу Facebook. Він акцентує увагу на ролі ШІ у виявленні фейкових новин, аналізуючи алгоритми для вивчення лінгвістичних шаблонів та обґрунтовуючи потребу в комбінованій стратегії, що поєднує людські зусилля та технології.

Дослідження Рабчун Д., Тищенко В. та Голобородько С. [5] зосереджене на розробці ефективних методів автоматизованого виявлення дезінформації з використанням нейронних мереж та обробки природної мови. Особливу увагу приділено аналізу емоційного впливу, що супроводжує дезінформаційний контент, та виявленню технік маніпуляції емоціями аудиторії. За результатами роботи розроблена система розпізнавань та категоризації емоційного забарвлення контенту, що дозволяє ефективно фільтрувати шкідливі матеріали та підвищувати стійкість інформаційного середовища. Дослідження також охоплює етичні аспекти використання нейронних мереж і пропонує стандарти захисту приватності користувачів.

Одним із аспектів вивчення дезінформації є її роль в інформаційно-психологічних атаках, які можуть загрожувати національній безпеці країни. Так, у праці [6] Дерюгін І. та Батько І. досліджують вплив таких атак на суспільство та безпеку держави, підкреслюючи їх здатність маніпулювати настроями громадян і поширювати дезінформацію. Особливу увагу приділено необхідності адаптації законодавчої бази для ефективного захисту від цих загроз. За результатами дослідження науковці рекомендують розвивати критичне мислення та ініціювати співпрацю між державними і приватними структурами для протидії інформаційно-психологічним атакам.

Незважаючи на значну кількість публікацій, присвячених виявленню фейкових новин за допомогою ШІ, подальші дослідження в цій галузі мають потенціал не лише для підвищення ефективності наявних рішень, а й для відкриття нових шляхів у боротьбі з дезінформацією. Це може допомогти інтегрувати сучасні технології та підходи, що забезпечить більш комплексне і гнучке реагування на виклики, які постають у інформаційному просторі.

Постановка завдання. Метою статті є аналіз сучасних міжнародних тенденцій у використанні ШІ для виявлення фейкової інформації, дослідження основних методів і технологій, що застосовуються для цього, а також оцінювання можливостей імплементації таких рішень у системи інформаційної безпеки.

Виклад основного матеріалу дослідження. Фейкова інформація, яка може виникати через неперевірені факти або помилки, часто поширюється випадково. Однак, коли неправдива інформація поширюється навмисно з метою маніпулювати думками або ввести людей в оману, вона перетворюється на дезінформацію.

Дезінформація поширюється таким самим чином, як і звичайна інформація, тобто через телебачення, радіо, інтернет та друковані матеріали. Друковані матеріали, такі як політичні буклети та рекламні листівки, активно використовуються під час виборів і можуть впливати на певні групи населення, особливо через їх безкоштовне розповсюдження. Радіо також є ефективним каналом дезінформації, особливо в регіонах з обмеженим доступом до телебачення чи інтернету. Телебачення – досить потужний інструмент поширення неправдивих новин через його широке охоплення та довіру серед глядачів. Однак саме інтернет і соціальні мережі є основними майданчиками для швидкого та масового поширення дезінформації, оскільки на цих платформах відсутній редакційний контроль і будь-який користувач може робити дописи.

16 червня 2022 року Європейський Союз оновив Кодекс практики щодо протидії дезінформації [7]. Підписанти нового Кодексу зобов'язалися діяти у таких напрямках: демонетизація дезінформації, прозорість політичної реклами, забезпечення цілісності послуг, розширення прав користувачів, посилення ролі фактчекерів, розширення прав дослідників тощо. Окрім того, було створено Центр прозорості та Цільову групу для моніторингу виконання кодексу. Серед основних нововведень особливу увагу приділено активній участі приватних компаній у забезпеченні дотримання Кодексу. До переліку компаній станом на червень 2022 року, що повністю або частково прийняли умови

Посиленого кодексу та зобов'язалися вжити конкретних заходів проти дезінформації на своїх платформах, увійшли: Adobe, Avaaz, Clubhouse, Crisp, Demagog, DOT Europe, European Association of Communication Agencies (EACA), Faktograf, Globsec, Google, Interactive Advertising Bureau (IAB) Europe, Kinzen, Kreativitet & Kommunikation, Logically, Maldita.es, MediaMath, Meta, Microsoft, Neeva, Newsback, NewsGuard, PagellaPoltica, Reporters without Borders (RSF), Seznam, The Bright App, The GARM Initiative, TikTok, Twitch, Twitter, Vimeo, VOST Europe, WhoTargetsMe, World Federation of Advertisers (WFA).

Розробники безперервно працюють над вдосконаленням нових методів виявлення фейкової інформації та боротьби з дезінформацією в соціальних мережах. Серед найефективніших підходів виділяють [8]:

1) *аналіз текстового контенту*, що передбачає використання алгоритмів для перевірки дописів, які можуть виявляти стиль письма і лексичну схожість із відомими джерелами, що дозволяє знаходити фальсифікації. Також перевіряється контекст, таким чином, якщо фото або повідомлення не відповідають типовій поведінці користувача, вони помічаються як підозрілі;

2) *машинне навчання*, що дозволяє розробляти моделі, які тренуються на великих обсягах даних і це допомагає визначити патерни, які свідчать про дезінформацію, а офіційно визнані фейки використовуються як навчальні приклади для штучного інтелекту;

3) *відстеження аномалій*, націлене на пошук нетипових подій і різких змін в активності користувачів, тобто, якщо певний допис або зображення отримує значну увагу або надмірно високу кількість реакцій користувачів за короткий проміжок часу, це вказує на те, що інформація є фейковою або маніпулятивною;

4) *аналіз мультимедійного контенту* використовує алгоритми для визначення рівня редагування фото і відео, іншими словами, оцінюються технічні дані, такі як показники акселерометра смартфона, що можуть вказувати на те, чи зображення було створено програмою, чи ні. Для виявлення фейків ще застосовують аналізатори голосу та тексту, які сигналізують про підозрілу інформацію в аудіо- та відеозаписах;

5) *виявлення ботів*, що ґрунтується на використанні численних критеріїв і параметрів, таких як частота публікацій на день та співвідношення активності акаунтів у вихідні і будні, детальний аналіз тем, слів та інших аспектів їхньої поведінки в мережі. Це дозволяє визначити, чи у створенні та розповсюдженні контенту беруть участь спеціально розроблені програми.

Розробка і вдосконалення наведених методів спрямовані на підвищення ефективності боротьби з фейковими новинами і дезінформацією в соціальних мережах. Проте, їх інтеграція в єдину систему дозволяє створити комплексний підхід для протидії дезінформації на різних рівнях і в різних форматах.

Таким чином, постає необхідність чіткого розуміння технологічного змісту поняття «система виявлення фейкової інформації на основі ШІ». Ця система поєднує кілька рівнів аналізу, які дозволяють автоматизовано ідентифікувати потенційні загрози, аналізувати контент і поведінку користувачів, забезпечуючи всебічне оцінювання даних для виявлення дезінформації.

Враховуючи, сьогоднішні тенденції збільшення обсягу дезінформації, важливість розглянутих методів продовжує зростати, забезпечуючи більш ефективний захист інформаційного середовища. На глобальному рівні спостерігається зростаюча увага до викликів, пов'язаних із дезінформацією, що стимулює міжнародну співпрацю для вироблення стандартів і практик, які сприятимуть прозорості та надійності інформації.

Провідні міжнародні компанії, такі як Google, Facebook (нині Meta), Twitter (нині X), та інші великі технологічні організації активно застосовують ШІ для боротьби з фейковими новинами. Це пов'язано з тим, що дезінформація стала серйозною проблемою в цифровій епосі, коли швидкість поширення неправдивої інформації значно перевищує можливості її перевірки вручну.

Facebook (Meta) використовує подібні методи, зокрема технології машинного навчання для автоматичного виявлення фейкових новин і модерації контенту [9]. Компанія розробила цілу систему для виявлення та позначення фейків, а також зменшує видимість неправдивого контенту у стрічках новин користувачів. Для цього платформа співпрацює з перевіреними фактчекерами та організаціями, що аналізують новини.

У звіті 2021 року Гай Розен (Guy Rosen), голова відділу безпеки інформації (Chief Information Security Officer, CISO) корпорації «Meta» [9] акцентує увагу на жорсткій позиції компанії щодо боротьби з фейковими акаунтами та дезінформацією в соціальних мережах. У рамках ініціативи з очищення платформи було заблоковано понад 1,3 мільярда підроблених облікових записів, а також виявлено і ліквідовано більше 100 мереж скоординованої неавтентичної поведінки. З метою зменшення поширення неправдивої інформації, компанія вживає заходів, зокрема, зменшує видимість контенту, позначеного як неправдиве, і використовує попереджувальні ярлики. Протягом пандемії COVID-19 було видалено понад 12 мільйонів матеріалів, пов'язаних з дезінформацією щодо вірусу та вакцин. Проте, незважаючи на докладені зусилля, Гай Розен визнає, що повністю усунути дезінформацію неможливо, але керівництво «Meta» продовжує інвестувати в нові технології та команди для

покращення своєї стратегії боротьби з цим явищем. Зміни в алгоритмах та принципах функціонування платформи підтверджують намір запобігти поширенню фейків і забезпечити надійність контенту, надаючи користувачам достовірну інформацію.

Інша соціальна мережа, Twitter (нині X), впровадила інструмент під назвою Birdwatch, що дозволяє користувачам додавати примітки до потенційно фейкових твітів, надаючи додатковий контекст і допомагаючи іншим користувачам перевіряти інформацію. Алгоритми ШІ також аналізують вміст у реальному часі для автоматичного виявлення фейків і ненадійного контенту. Кіт Коулман (Keith Coleman), віцепрезидент Twitter, активно займається питанням дезінформації та способами її подолання на платформах соціальних медіа, у своєму дописі про інструмент Birdwatch у Twitter [10] зазначив, що Birdwatch отримала загальну підтримку користувачів за результатами понад 100 інтерв'ю з людьми з усього політичного спектру.

З 2021 року функціонує Центр безпеки контенту Google (Google Safety Engineering Center), який фокусується на боротьбі з незаконним і шкідливим вмістом [11] та розробляє правила, що забороняють шкідливий контент в своїх продуктах, таких як YouTube і Google Ads. Так, за перші 18 місяців пандемії YouTube видалив понад мільйон відео, що містили дезінформацію про COVID-19. При зборі достовірних даних ключовою стала співпраця з авторитетними партнерами, такими як Всесвітня організація охорони здоров'я (ВООЗ) і CDC (Centers for Disease Control and Prevention) – агентство Міністерства охорони здоров'я і соціальних служб США.

Окрім того, у 2021 році Google пожертвував 25 млн євро для запуску Європейського медіа та інформаційного фонду, що підтримує медіаграмотність та перевірку фактів. З березня 2020 року по березень 2021 року в Google відбулося понад 50 тисяч нових перевірок фактів. У листопаді 2022 року Google і YouTube надали грант розміром 13,2 млн доларів Міжнародній мережі перевірки фактів для підтримки 135 організацій.

Боротьба з дезінформацією є постійним викликом. Підрозділ Google Jigsaw досліджує нові підходи, такі як попереднє розвінчування (prebunking), що допомагає людям виявляти оманливі наративи. Цей підхід використовує попередження про маніпуляції та неправдиві заяви, що підвищує психологічну стійкість користувачів.

Інші міжнародні технологічні компанії також розробляють подібні рішення, прагнучи зменшити вплив дезінформації. Зокрема, вони використовують моделі глибокого навчання, обробку природної мови (NLP) і технології кластеризації для виявлення патернів у фейкових новинах і спроб маніпуляцій. Наприклад, ці компанії аналізують великі обсяги даних з соціальних мереж та новинних платформ, щоб виявити і зупинити поширення неправдивої інформації. Таким чином, технологічні інновації та партнерства стають важливими інструментами у боротьбі з дезінформацією у сучасному світі.

Висновки з проведеного дослідження. Дослідження дезінформації та фейкових новин вказує на складність і багатогранність проблеми, що вимагає комплексного підходу до її вирішення. Важливо зазначити, що дезінформація є серйозним викликом сучасному суспільству, адже вона може маніпулювати громадською думкою і впливати на соціально-політичні процеси. Використання нових технологій, зокрема ШІ, стає ключовим фактором у боротьбі з дезінформацією, оскільки дозволяє швидше і ефективніше виявляти неправдиву інформацію, аналізувати контент і поведінку користувачів, а також автоматизувати процеси модерації.

Незважаючи на значні зусилля технологічних компаній у виявленні фейкових новин, проблема залишається актуальною, і необхідно вдосконалювати методи моніторингу і виявлення дезінформації. Це включає розвиток алгоритмів машинного навчання, які здатні розпізнавати патерни фальсифікацій та аномальну активність користувачів, а також аналіз мультимедійного контенту, щоб відстежувати зміни у формі та якості інформації.

Крім того, важливим аспектом є співпраця між різними технологічними компаніями, фактчекерами, міжнародними організаціями та урядами держав. Така взаємодія сприяє розробці стандартів і практик, що покращують прозорість і надійність інформації в цифровій епосі. Необхідно також звернути увагу на розвиток критичного мислення у суспільстві, що допоможе користувачам самостійно відрізняти надійні джерела інформації від недостовірних.

Загалом, ефективна боротьба з дезінформацією вимагає поєднання технологічних інновацій, етичних стандартів та соціальної відповідальності. Впровадження комплексних підходів до виявлення та моніторингу фейкових новин може суттєво знизити їхній вплив на суспільство, забезпечуючи більш якісне інформаційне середовище.

Імплементація заходів щодо боротьби з дезінформацією та фейковими новинами розширить можливості у сфері інформаційної безпеки, громадянської освіти, технологічного розвитку та суттєво підвищить ефективність цих зусиль. Варто продовжувати впроваджувати технології ШІ та машинного навчання, які здатні автоматизувати процеси виявлення та оцінювання інформації. Це дозволить значно зменшити час реагування на фейкові новини і підвищити точність їх виявлення. Компанії можуть інтегрувати алгоритми, які аналізують текст, зображення та відео, виявляючи аномалії або патерни, що свідчать про маніпуляції з інформацією. Окрім того, важливо забезпечити активну співпрацю між технологічними компаніями, державними органами та громадськими організаціями.

Створення міжвідомчих платформ для обміну даними та кращими практиками може допомогти у формуванні єдиного фронту в боротьбі з дезінформацією. Такі платформи можуть включати механізми для спільної перевірки фактів, розвитку відкритих даних та ресурсів для навчання користувачів.

Ще однією, вартою уваги, ініціативою є імплементація освітніх програм на всіх рівнях, від шкіл до університетів, що сприятиме вихованню критичного мислення. Навчання громадян розпізнавати фейкові новини, розуміти механізми дезінформації і використовувати надійні джерела є важливим елементом для створення стійкого суспільства.

І, звичайно, необхідне функціональне правове регулювання для ефективної роботи платформ обміну інформацією, зокрема, встановлення чіткої відповідальності за розповсюдження фейків. Важливим кроком у цьому процесі є забезпечення прозорості алгоритмів, що використовуються для ранжування контенту, адже, це підвищить довіру користувачів до інформаційних майданчиків та зменшить ризик маніпуляцій інформацією.

Загалом імплементація заходів для боротьби з фейковою інформацією має великий потенціал. Поєднання технологічних рішень, освітніх ініціатив та правових зусиль сприятиме істотним позитивним змінам у формуванні якісного інформаційного середовища та забезпечення інформованості громадян.

Література

1. Тищенко В. Аналіз методів навчання та інструментів нейромереж для виявлення фейків. *Кібербезпека: освіта, наука, техніка*. 2023. № 4(20). С. 20-34. DOI: https://doi.org/10.28925/2663-4023.2023.20.2034.

2. Праздніков В. О., Сугоняк І. І. Моделі та методи машинного навчання для розпізнавання фейкового контенту. *Технічна інженерія*. 2023. № 2(92). С. 131-136. DOI: https://doi.org/10.26642/ten-2023-2(92)-131-136.

3. Боровик Д., Бармак О. Удосконалений метод виявлення фейкових новин на основі використання CNN нейромережі. *Вісник Хмельницького національного університету*. 2023. Том 2. № 5. С. 19-24. DOI: 10.31891/2307-5732-2023-327-5-19-24.

4. Дацик С. В. Використання штучного інтелекту для виявлення дезінформації в новинах соціальної мережі Facebook : кваліфікаційна робота на здобуття освітнього ступеня магістр за спеціальністю 122 – комп'ютерні науки. Тернопіль : ТНТУ, 2024. 90 с.

5. Рабчун Д. І., Тищенко В. С., Голобородько С. О. Ефективне розпізнавання дезінформації за допомогою нейронних мереж: фокус на виявлення емоційного впливу. *Телекомунікаційні та інформаційні технології*. 2024. № 2(83). С. 37-48. DOI: 10.31673/2412-4338.2024.024860.

6. Дерюгін І. К., Батько І. І. Інформаційно-психологічні атаки як загроза державній безпеці. *Аналітично-порівняльне правознавство*. 2023. № 05. С. 315-319. DOI: https://doi.org/10.24144/2788-6018.2023.05.57.

7. Оновлення кодексу ЄС щодо боротьби з дезінформацією: основні положення. URL: https://jurliga.ligazakon.net/news/212519_onovlennya-kodeksu-s-shchodo-borotbi-z-deznformatsyu-osnovn-polozhennya (дата звернення: 02.08.2024).

8. Сучасні методи виявлення фейків у соціальних мережах. URL: https://optima.study/blog/suchasni-metodi-viyavlennya-feykiv-u-socialnih-merezhah. (дата звернення: 02.08.2024).

9. Guy Rosen. How We're Tackling Misinformation Across Our Apps. URL: https://about.fb.com/news/2021/03/how-were-tackling-misinformation-across-our-apps/ (дата звернення: 02.08.2024).

10. Keith Coleman. Introducing Birdwatch, a community-based approach to misinformation. URL: https://blog.x.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation (дата звернення: 02.08.2024).

11. Google's approach to fighting misinformation online. URL: https://safety.google/intl/en_uk/stories/fighting-misinformation-online/ (дата звернення: 02.08.2024).

References

1. Tyshchenko, V. (2023), "Analysis of Training Methods and Neural Network Tools for Fake Detection", *Kiberbezpeka: osvita, nauka, tekhnika*, no. 4(20), pp. 20-34, DOI: https://doi.org/10.28925/2663-4023.2023.20.2034.

2. Prazdnikov, V.O. and Suhoniak, I.I. (2023), "Models and Methods of Machine Learning for Recognizing Fake Content", *Tekhnichna inzheneriia*, no. 2(92), pp. 131-136, DOI: https://doi.org/10.26642/ten-2023-2(92)-131-136.

3. Borovyk, D. and Barmak, O. (2023), "An Improved Method for Detecting Fake News Based on the Use of CNN Neural Networks", *Visnyk Khmelnytskoho natsionalnoho universytetu*, Vol. 2, no. 5, pp. 19-24, DOI: 10.31891/2307-5732-2023-327-5-19-24.

4. Datsyk, S.V. (2024), "The Use of Artificial Intelligence for Detecting Disinformation in News on the Facebook Social Network" : Master's degree work in specialty 122 - computer science, TNTU,Ternopil, Ukraine, 90 p.

5. Rabchun, D.I., Tyshchenko, V.S. and Holoborodko, S.O. (2024), "Effective Detection of Disinformation Using Neural Networks: Focus on Identifying Emotional Impact", *Telekomunikatsiini ta informatsiini tekhnolohii*, no. 2(83), pp. 37-48, DOI: 10.31673/2412-4338.2024.024860.

6. Deriuhin, I.K. and Batko, I.I. (2023), "Information-Psychological Attacks as a Threat to National Security", *Analitychno-porivnialne pravoznavstvo,* no. 05, pp. 315-319, DOI: https://doi.org/10.24144/2788-6018.2023.05.57.

7. "Update of the EU Code of Practice on Disinformation: Key Provisions", available at: https://jurliga.ligazakon.net/news/212519_onovlennya-kodeksu-s-shchodo-borotbi-z-deznformatsyu-osnovn-polozhennya (access date August 02, 2024).

8. "Modern Methods of Detecting Fake News in Social Media", available at: https://optima.study/blog/suchasni-metodi-viyavlennya-feykiv-u-socialnih-merezhah (access date August 02, 2024).

9. Rosen, Guy (2021), "How We're Tackling Misinformation Across Our Apps", available at: https://about.fb.com/news/2021/03/how-were-tackling-misinformation-across-our-apps/ (access date August 02, 2024).

10. Coleman, Keith (2021), "Introducing Birdwatch, a community-based approach to misinformation", available at: https://blog.x.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation (access date August 02, 2024).

11. Google's approach to fighting misinformation online, available at: https://safety.google/intl/en_uk/stories/fighting-misinformation-online/ (access date August 02, 2024).

Дзюбановська Н.В.

ШТУЧНИЙ ІНТЕЛЕКТ ДЛЯ ВИЯВЛЕННЯ ФЕЙКОВОЇ ІНФОРМАЦІЇ: МІЖНАРОДНІ ТРЕНДИ І МОЖЛИВОСТІ ІМПЛЕМЕНТАЦІЇ

Мета. Аналіз сучасних міжнародних тенденцій у використанні штучного інтелекту (ШІ) для виявлення фейкової інформації, дослідження основних методів і технологій, що застосовуються для цього, а також оцінювання можливостей імплементації таких рішень у системи інформаційної безпеки.

Методика дослідження. Теоретичною та методологічною основою дослідження є наукові праці вчених із питань застосування штучного інтелекту для виявлення фейкової інформації, а також дослідження в галузі машинного навчання та аналізу текстових і мультимедійних даних. У процесі дослідження використовувалися загальнонаукові та спеціальні методи. Зокрема, аналіз літератури та огляд досліджень дозволили вивчити сучасні міжнародні тренди у сфері застосування ШІ для виявлення фейкових новин; порівняльний аналіз існуючих рішень дав змогу оцінити ефективність таких методів, як аналіз текстів, машинне навчання, відстеження аномалій, аналіз мультимедійного контенту та виявлення ботів для боротьби з дезінформацією.

Результати дослідження. Визначено зміст поняття «системи виявлення фейкової інформації на основі штучного інтелекту», під якою розуміється інтегрована система, що використовує автоматизовані алгоритми для аналізу та верифікації інформації, спрямована на боротьбу з дезінформацією. Основним завданням такої системи є виявлення фейкових новин через використання методів машинного навчання, аналізу текстів, відстеження аномалій та аналізу мультимедійного контенту, що дозволяє підвищити точність і швидкість ідентифікації фейкових повідомлень. Ідентифіковано основні виклики, пов'язані з поширенням дезінформації, які вимагають нових підходів у застосуванні ШІ: зростання обсягу даних, поява нових форм фейків, розвиток дезінформаційних кампаній та необхідність швидкої реакції на них. Проведено аналіз міжнародних практик застосування штучного інтелекту для виявлення фейкової інформації, у результаті чого узагальнено основні принципи використання алгоритмів для автоматичного аналізу даних, навчання моделей, виявлення ботів та відстеження аномальних поведінкових патернів.

Наукова новизна результатів дослідження. Набув подальшого розвитку технологічний зміст поняття «система виявлення фейкової інформації на основі штучного інтелекту», що включає використання машинного навчання, аналізу текстів, відстеження аномалій та аналізу мультимедійного контенту для ефективного виявлення дезінформації. Обґрунтовано основні напрямки застосування штучного інтелекту для ідентифікації фейкових новин і визначено їхню ефективність у вирішенні проблем інформаційної безпеки та боротьби з дезінформаційними кампаніями на міжнародному рівні. Узагальнено практики виявлення фейків, які можуть бути адаптовані для різних контекстів, що робить систему більш гнучкою та адаптивною.

Практична значущість результатів дослідження. Результати дослідження мають важливе практичне значення і можуть бути використані для розробки нових інноваційних рішень у сфері виявлення фейкової інформації. Це дозволить медіаплатформам, державним та приватним організаціям підвищити точність і швидкість ідентифікації дезінформації. Зменшення поширення фейкових новин сприятиме забезпеченню інформаційної безпеки. Дослідження може стати важливим ресурсом для розробки державних стратегій у сфері протидії дезінформації, а також корисним для компаній, що займаються моніторингом інформаційних потоків і кібербезпекою, у їхній роботі щодо запобігання дезінформації та захисту суспільства від негативних інформаційних впливів.

Ключові слова: ШІ, фейкова інформація, машинне навчання, аналіз текстів, виявлення ботів, відстеження аномалій, інформаційна безпека, дезінформація, мультимедійний контент, міжнародні тренди.

Dziubanovska N.V.

ARTIFICIAL INTELLIGENCE FOR DETECTING FAKE INFORMATION: INTERNATIONAL TRENDS AND IMPLEMENTATION OPPORTUNITIES

Purpose. The aim of the article is to analyse the current international trends in the use of artificial intelligence (AI) to detect fake information, study the main methods and technologies used for this, as well as the assessment of the possibilities of implementing such solutions in the systems of information security.

Methodology of research. The theoretical and methodological foundation of the study is based on scientific works by researchers on the application of artificial intelligence for detecting fake information, as well as studies in machine learning and the analysis of textual and multimedia data. General scientific and special methods were used during the research process. In particular, literature analysis and research reviews enabled the study of current international trends in the application of AI for detecting fake news. Comparative analysis of existing solutions allowed for the assessment of the effectiveness of methods such as text analysis, machine learning, anomaly detection, multimedia content analysis, and bot detection in combating disinformation.

Findings. The content of the concept of “AI-based fake information detection systems” was defined, understood as an integrated system that employs automated algorithms to analyse and verify information aimed at combating disinformation. The main task of such a system is to identify fake news through the use of machine learning methods, text analysis, anomaly tracking, and multimedia content analysis, which enhances the accuracy and speed of identifying fake messages. The main challenges associated with the spread of disinformation were identified, requiring new approaches in the application of AI: the growing volume of data, the emergence of new forms of fakes, the development of disinformation campaigns, and the necessity for rapid response to them. An analysis of international practices in applying artificial intelligence for detecting fake information was conducted, resulting in the generalization of the main principles for using algorithms for automatic data analysis, model training, bot detection, and tracking anomalous behavioural patterns.

Originality. The technological content of the concept of “AI-based fake information detection systems” has been further developed, encompassing the use of machine learning, text analysis, anomaly tracking, and multimedia content analysis for effective disinformation detection. The main areas of application of artificial intelligence for identifying fake news were substantiated, and their effectiveness in addressing information security issues and combating disinformation campaigns at the international level was determined. Practices for detecting fake news were generalized, which can be adapted for different contexts, making the system more flexible and adaptive.

Practical value. The research results have significant practical implications and can be used to develop new innovative solutions in the field of fake information detection. This will allow media platforms, government, and private organizations to enhance the accuracy and speed of disinformation identification. Reducing the spread of fake news will contribute to ensuring information security. The research may serve as an important resource for developing government strategies in combating disinformation and could also be useful for companies involved in monitoring information flows and cybersecurity in their efforts to prevent disinformation and protect society from negative informational impacts.

Key words: AI, fake information, machine learning, text analysis, bot detection, anomaly tracking, information security, disinformation, multimedia content, international trends.