КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2024.25.449467
УДК 004.04

Данилик Віталій Миколайович
аспірант кафедри інформаційних систем та мереж
Національний університет «Львівська політехніка», Львів, Україна
ORCID ID: 0000-0001-5928-7235
vitalii.m.danylyk@lpnu.ua

Висоцька Вікторія Анатоліївна
доктор технічних наук, доцент, доцент кафедри інформаційних систем та мереж
Національний університет «Львівська політехніка», Львів, Україна
ORCID ID: 0000-0001-6417-3689
victoria.a.vysotska@lpnu.ua

Назаркевич Марія Андріївна
доктор технічних наук, професор, професор кафедри інформаційних систем та мереж
Національний університет «Львівська політехніка», Львів, Україна
ORCID ID: 0000-0002-6528-9867
mariia.a.nazarkevych@lpnu.ua

МЕТОДИ ІДЕНТИФІКАЦІЇ ДЕЗІНФОРМАЦІЇ, ФЕЙКІВ ТА ПРОПАГАГДИ В ЗАСОБАХ МАСОВОЇ ІНФОРМАЦІЇ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Анотація. У цій статті проводиться комплексне дослідження сучасних підходів, які використовуються для ідентифікації дезінформації/фейків/пропаганди. Дослідження починається з розгляду поширеності та впливу дезінформації, наголошуючи на необхідності передових методів ідентифікації. Простежуючи еволюцію методів, стаття заглиблюється в історичні підходи та їх адаптацію до динамічного медіаландшафту. Центральними для обговорення є передові методи, такі як обробка природної мови (NLP), яка аналізує лінгвістичні шаблони для виявлення невідповідностей у текстовому вмісті. У статті досліджуються переваги NLP, включаючи масштабованість і аналіз у реальному часі, разом з обмеженнями, пов'язаними з контекстною неоднозначністю та розвитком мовних нюансів. Мультимодальний аналіз займає центр уваги, пропонуючи цілісну перспективу завдяки інтеграції текстових і візуальних елементів. У статті з'ясовуються переваги мультимодального аналізу, підкреслюється його потенціал для всебічного розуміння та підвищеної точності, визнаючи при цьому проблеми, пов'язані зі складністю аналізу та контекстними нюансами. Машинне навчання постає як динамічний інструмент для розпізнавання образів і адаптації, що сприяє аналізу в реальному часі. У статті висвітлюються переваги машинного навчання, такі як масштабованість, і розглядаються його обмеження, включно з упередженнями в навчальних даних і вразливістю до агресивних атак. Додатково у статті поводиться аналіз пропаганди на основі емоційного забарвлення який виявляє відмінностей між пропагандою та не-пропагандою. Рекомендації щодо майбутніх досліджень підкреслюють важливість прозорості моделей і постійних зусиль для усунення упереджень. Оскільки цифровий ландшафт продовжує розвиватися, ці досягнення відкривають шлях до стійких стратегій у виявленні та протидії фейкам і пропаганді.

Ключові слова: дезінформація; фейк; пропаганда; ідентифікація дезінформації; NLP; кібербезпека; засоби масової інформації; опрацювання природної мови; мультимодальний аналіз; машинне навчання.

ВСТУП

У просторі цифрових комунікацій, що постійно розширюється, поширення інформації досягло безпрецедентних висот, уможливлюючи швидке поширення як точних, так і оманливих наративів [1]. Одним із найбільш тривожних побічних продуктів

ції цифрової ери є поширення фейкових новин і пропаганди, двох взаємопов'язаних явищ, які становлять значну загрозу цілісності інформації та основам демократичних суспільств [2]. Поширеність фейкових новин і пропаганди стала підступною силою, яка проникла на онлайн-платформи, новинні видання та канали соціальних мереж [3] – [5]. В епоху інформаційного перевантаження відрізнити факт від вигадки стає дедалі складнішим завданням для споживачів медіа. Наслідки дезінформації виходять за рамки простої плутанини, оскільки вони можуть формувати громадську думку, впливати на політичний ландшафт і навіть розпалювати соціальні розбрати. Вплив фейкових новин і пропаганди відчувається в різних сферах, від політичного дискурсу та охорони здоров'я до суспільної згуртованості. Фальшиві наративи можуть вплинути на вибори, підірвати довіру до інституцій і сприяти поляризації спільнот. Наслідки не тільки когнітивні, але й можуть проявлятися в реальних діях, про що свідчить випадки, коли дезінформація розпалювала громадські заворушення або перешкоджала ефективним реакціям на кризи.

Враховуючи серйозність цієї ситуації, існує нагальна потреба в передових методах виявлення та боротьби з дезінформацією. Традиційним процесам перевірки фактів важко встигати за швидким розповсюдженням оманливого контенту, якому сприяють цифрові платформи. З розвитком технологій змінюються і тактики розповсюджувачів дезінформації, що вимагає інноваційних та адаптивних підходів для захисту цілісності інформаційної екосистеми.

Постановка проблеми. Розроблення адаптивного підходу до проектування інформаційних систем ідентифікації дезінформації, фейків та пропаганди в засобах масової інформації на основі машинного навчання.

Аналіз останніх досліджень і публікацій. Нещодавні дослідження та публікації зробили значний внесок у поточний дискурс навколо ідентифікації дезінформації, запровадивши нові методології та вдосконаливши існуючі підходи [6] – [12]. У цьому розділі розглядаються ключові відкриття та досягнення, проливаючи світло на поточний стан галузі. Останні дослідження підкреслили важливість інтеграції мультимодального аналізу, поєднання NLP з техніками обробки зображень і відео. Цей цілісний підхід визнає мультимедійну природу сучасної дезінформації, вирішуючи проблеми, пов'язані з глибокими фейками та маніпуляціями на основі зображень. Одночасно перевіряючи текстовий і візуальний вміст, дослідники прагнуть підвищити загальну точність систем виявлення. Удосконалення зрозумілого штучного інтелекту привернули увагу в останніх публікаціях. Дослідники досліджують методи, щоб зробити процеси прийняття рішень NLP і алгоритмів машинного навчання більш прозорими та доступними для інтерпретації. Це не тільки підвищує надійність систем виявлення, але й дає змогу зрозуміти складні способи виявлення дезінформації, сприяючи ширшому розумінню цієї галузі. Остання література підкреслює зростаюче визнання необхідності міждисциплінарної співпраці. Хоча технологічні рішення залишаються ключовими, дослідники все більше виступають за партнерство між технологами, соціологами та політиками. Цей спільний підхід спрямований на боротьбу з дезінформацією не лише з технологічної точки зору, але й шляхом усунення її суспільних і політичних коренів. Етичні міркування в автоматизованих методах виявлення стали центральною темою останніх досліджень. Вчені досліджують потенційні упередження, вбудовані в алгоритми, і етичні наслідки автоматизованої модерації вмісту. Встановлення балансу між ефективним виявленням і збереженням індивідуальних свобод є центральною темою, яка відображає ширші суспільні наслідки застосування передових технологій для виявлення дезінформації [1].

Останні публікації висвітлюють проблеми в оцінці ефективності систем виявлення дезінформації. Розробка стандартизованих показників і контрольних показників, які враховують динамічний характер дезінформації, залишається актуальною проблемою. Дослідники активно досліджують способи створення комплексних систем оцінювання, які враховують різні параметри, від хибнопозитивних показників до здатності адаптуватися до тактики, що розвивається. Останні дослідження і публікації показують динамічний ландшафт, який характеризується інноваціями в мультимодальному аналізі, зосередженістю на зрозумілому ШІ та зростаючим акцентом на міждисциплінарній співпраці. Однак етичні міркування та труднощі в оцінці продуктивності системи підкреслюють складність вирішення проблеми дезінформації в цифровому середовищі, що постійно розвивається. Ці висновки закладають основу для майбутніх розробок і підкреслюють нагальність продовження досліджень у цій критичній галузі.

Мета статті. Розробити інформаційну технологію ідентифікації дезінформації, фейків та пропаганди в засобах масової інформації на основі машинного навчання.

Основна мета цієї статті — комплексно дослідити сучасні підходи до виявлення та протидії фейковим новинам і пропаганді в медіатекстах. Заглиблюючись у методології та технології, використані в останніх дослідженнях, мета полягає в тому, щоб запропонувати детальне розуміння еволюції ландшафту виявлення дезінформації.

Об’єктом дослідження є процеси формування/ідентифікації оманливих наративів — тканини фейків і пропаганди, які насичують не лише величезний ландшафт сучасних медіа, але й пронизують самі основи публічного дискурсу, формуючи сприйняття, впливаючи на думки та кидаючи під сумнів правдивість. інформації в цифрову епоху.

Предмет дослідження — методи та засоби, що використовуються для виявлення дезінформації, фейків і пропаганди, слугуючи практичним інструментарієм, який дослідники, технологи та політики використовують для систематичного виявлення та протидії оманливим наративам у складному ландшафті сучасних засобах масової інформації.

Наукова новизна полягає в комплексному огляді та критичному аналізі існуючих методологій. Він сприяє синтезу різноманітних точок зору, визначенню прогалин або областей, які потребують удосконалення в поточних методах виявлення фейків та пропаганди, а також розуміння ефективності та обмежень цих підходів. Крім того, наукова новизна проявляється в здатності дослідження контекстуалізувати поточний стан виявлення дезінформації в середовищі цифрових медіа, що швидко розвивається, надаючи цінні знання для майбутніх розробок у цій галузі.

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Сучасні підходи до ідентифікації фейків і пропаганди використовують поєднання технологічних інновацій, міждисциплінарної співпраці та передових аналітичних методів для навігації у складному ландшафті дезінформації. Ці підходи охоплюють різні стратегії, кожна з яких спрямована на вирішення багатогранної природи оманливих наративів у сучасних ЗМІ [2].

- *Обробка природної мови (NLP).* Використовуючи NLP, ці підходи передбачають аналіз мовних шаблонів у текстовому вмісті. Алгоритми NLP аналізують речення, фрази та загальні мовні структури, щоб виявити аномалії, невідповідності та шаблони, що вказують на дезінформацію.

Аналіз настроїв і семантичний аналіз є невід'ємними компонентами, які допомагають розпізнати емоційний тон і значення слів.

- *Мультимодальний аналіз.* Визнаючи мультимедійний характер сучасної дезінформації, сучасні підходи включають мультимодальний аналіз. Це передбачає одночасний аналіз тексту, зображень і відео для виявлення маніпуляцій або невідповідностей у різних медіаформатах. Такі методи, як криміналістика зображень і глибоке навчання, сприяють більш повному розумінню оманливих наративів.
- *Алгоритми машинного навчання.* Розширені алгоритми машинного навчання навчаються на великих наборах даних, щоб розпізнавати шаблони, пов'язані з дезінформацією. Ці алгоритми можуть адаптуватися до нових тактик, які використовують постачальники обману, забезпечуючи динамічний і масштабований підхід до ідентифікації. У цьому контексті зазвичай використовуються контрольоване навчання, неконтрольоване навчання та ансамблеві методи.

Інтегруючи ці різноманітні стратегії, сучасні підходи до виявлення фейків і пропаганди прагнуть випереджати нові виклики. Поєднання технологічних досягнень, міждисциплінарних ідей і відданості етичним практикам формує міцну основу для пом'якшення впливу оманливих наративів на публічний дискурс і суспільний добробут.

МЕТОДИКА ДОСЛІДЖЕННЯ

Опрацювання природної мови (NLP) відіграє вирішальну роль у сфері виявлення фейків і пропаганди в текстовому вмісті. NLP — це підполе штучного інтелекту (ШІ), яке зосереджується на тому, щоб комп'ютери могли розуміти, інтерпретувати та генерувати людську мову [3]. У контексті виявлення оманливих наративів NLP застосовується для аналізу мовних моделей і текстового вмісту, щоб виявити невідповідності, оманливу інформацію або пропагандистські елементи. Техніки NLP для виявлення фейків і пропаганди включають кілька ключових аспектів:

- *Аналіз тексту:* Алгоритми NLP розбивають текстову інформацію, досліджуючи окремі слова, фрази та загальну структуру речень і абзаців. Цей процес допомагає ідентифікувати лінгвістичні моделі, пов'язані з дезінформацією.
- *Аналіз настроїв:* NLP використовується для оцінки настроїв, виражених у фрагменті тексту. Дезінформація часто несе в собі чіткі емоційні відтінки чи упередження, а аналіз настроїв допомагає позначати вміст із поляризованими або оманливими емоціями.
- *Семантичний аналіз:* розуміння значення слів і фраз має вирішальне значення. NLP дозволяє системам осягнути семантичний контекст, дозволяючи ідентифікувати тонкі зрушення в значенні або навмисні спотворення.
- *Контекстуальне розуміння:* Алгоритми NLP прагнуть зрозуміти контекстуальні нюанси мови, визнаючи, що значення слів може змінюватися залежно від навколишнього тексту. Це особливо важливо для виявлення сарказму, іронії чи інших форм непрямого спілкування, які можуть використовуватися в оманливих розповідях.

- *Розпізнавання іменованих сутностей (NER):* Техніки NLP включають NER, що передбачає ідентифікацію таких сутностей, як люди, організації та місця в тексті. Виявлення невідповідностей у представленні об'єктів може бути цінною підказкою для виявлення підробленої інформації.

Переваги NLP у виявленні фейків і пропаганди включають його масштабованість для обробки великих обсягів тексту, можливості аналізу в реальному часі та здатність адаптуватися до еволюційних лінгвістичних моделей. Однак обмеження можуть виникати через контекстуальну неоднозначність, коли значення слів значною мірою залежить від навколишнього контексту, і проблеми, пов'язані з оновленням моделей NLP для вирішення нових мовних нюансів. NLP пропонує кілька переваг у контексті виявлення фейків і пропаганди, сприяючи ефективності та ефективності ідентифікації дезінформації:

- *Масштабованість.* NLP дозволяє автоматизовано аналізувати величезну кількість текстового вмісту в масштабі. Враховуючи величезний обсяг інформації, що поширюється в Інтернеті, масштабованість NLP дозволяє своєчасно обробляти великі масиви даних, полегшуючи ідентифікацію оманливих наративів на різних платформах.
- *Аналіз у реальному часі.* Системи NLP працюють у режимі реального часу, надаючи можливість швидко аналізувати вхідну інформацію, щойно вона з'являється. Цей аналіз у режимі реального часу має вирішальне значення в динамічному ландшафті цифрових медіа, де дезінформація може швидко поширюватися. Швидке виявлення оманливого вмісту дає змогу вчасно втручатися та вживати заходів щодо пом'якшення.
- *Адаптивність.* Моделі NLP можна навчити та адаптувати до лінгвістичних моделей, що розвиваються, і до змін у способах поширення дезінформації. Ця адаптивність має важливе значення для вирішення динамічної природи мови та тактики, яку використовують пропагандисти, гарантуючи, що методи виявлення залишатимуться ефективними з часом.
- *Багатогранний аналіз.* NLP дозволяє багатогранний аналіз текстового вмісту, включаючи аналіз настроїв, семантичний аналіз і розуміння контексту. Враховуючи різні лінгвістичні виміри, NLP покращує здатність ідентифікувати не лише відверту брехню, але й витончені форми маніпуляції та упереджену мову, часто пов'язану з пропагандою.
- *Ефективне розпізнавання шаблонів.* NLP чудово розпізнає шаблони та аномалії в мовних структурах. Він може виявити мовні неузгодженості, відхилення від типових мовних шаблонів або зміни в настрої тексту, надаючи цінні показники потенційної дезінформації.
- *Покращене розпізнавання сутностей.* Розпізнавання іменованих сутностей (NER) є компонентом NLP, який може ідентифікувати такі сутності, як люди, організації та місця в тексті. Це особливо корисно для виявлення невідповідностей або викривлень, пов'язаних із ключовими об'єктами, додаючи додатковий рівень точності у виявленні обману.
- *Автоматизація та ефективність.* NLP автоматизує процес аналізу мови, зменшуючи потребу в ручному аналізі. Ця автоматизація підвищує ефективність і дозволяє аналітикам зосередитися на більш складних аспектах виявлення дезінформації, підвищуючи загальну точність.

Поєднання цих переваг робить NLP цінним інструментом в арсеналі проти фейків і пропаганди. Його здатність обробляти великі обсяги інформації, аналізувати вміст у режимі реального часу, адаптуватися до мінливих мовних нюансів і надавати детальну інформацію значною мірою сприяє поточним зусиллям з виявлення та боротьби з дезінформацією. Хоча NLP є потужним інструментом для виявлення фейків і пропаганди, вона не позбавлена своїх обмежень. Визнання цих обмежень має важливе значення для тонкого розуміння проблем, пов'язаних із використанням NLP для виявлення дезінформації:

- *Контекстуальна неоднозначність.* моделі NLP можуть мати проблеми з контекстною неоднозначністю, коли значення слів або фраз значною мірою залежить від навколишнього контексту. У оманливих наративах часто використовуються тонкі лінгвістичні нюанси, які можуть бути складними для точної інтерпретації алгоритмів, що може призвести до потенційних неправильних суджень.
- *Нюанси мови, що розвиваються.* Мова є динамічною та розвивається з часом. Моделі NLP, навчені на історичних даних, можуть не ефективно вловлювати нові мовні нюанси та зміни у виразі. Оскільки тактика дезінформації розвивається, нездатність швидко адаптуватися до мінливих лінгвістичних тенденцій може обмежити ефективність систем виявлення на основі NLP.
- *Сарказм та іронія.* Виявлення сарказму, іронії чи інших форм непрямого спілкування становить значну проблему для моделей NLP. Суб'єкти дезінформації можуть використовувати ці лінгвістичні засоби для передачі оманливих наративів, а автоматизованим системам може бути важко точно розпізнати передбачуване значення без контекстних підказок.
- *Відсутність розуміння контексту.* NLP, хоч і вправно аналізує текстовий вміст, може не мати глибокого розуміння ширших контекстуальних факторів. Розуміння наміру, що стоїть за частиною інформації, часто вимагає розуміння зовнішніх подій, культурних посилань або конкретної контекстної інформації, яка не може бути належним чином охоплена моделями NLP.
- *Упередженість у навчальних даних.* Моделі NLP настільки ефективні, наскільки ефективні дані, на яких вони навчаються. Якщо навчальні дані містять упередження або відображають певні точки зору, модель може ненавмисно закріпити та посилити ці упередження в своїх оцінках інформації. Це створює ризик упереджених результатів виявлення дезінформації.
- *Багатомовні виклики.* Моделі NLP, навчені одній мові, можуть не однаково добре розпізнавати дезінформацію кількома мовами. Адаптація моделей до різноманітних мовних ландшафтів потребує великих навчальних наборів даних і все ще може зіткнутися з труднощами в охопленні нюансів різних мов.
- *Семантичні зрушення.* моделям NLP може бути важко визначити тонкі зрушення в семантиці мови, особливо коли учасники дезінформації навмисно маніпулюють мовою, щоб передати оманливу інформацію. Здатність точно виявляти семантичні зрушення має вирішальне значення для того, щоб випереджати нові тактики, що використовуються в оманливих оповіданнях.

- *Змагальні атаки.* Суб'єкти дезінформації можуть використовувати змагальні атаки, щоб навмисно маніпулювати результатами моделей NLP. Це може включати тонкі зміни в тексті, призначені для того, щоб ввести в оману систему виявлення, підкреслюючи вразливість моделей NLP до складних маніпуляцій.

Розуміння цих обмежень має важливе значення для вдосконалення та розширення можливостей NLP у виявленні дезінформації. У міру розвитку цієї галузі вирішення цих проблем буде сприяти розробці більш надійних та адаптивних підходів до виявлення оманливих наративів у еволюції цифрового ландшафту.

Мультимодальний аналіз у контексті виявлення фейків і пропаганди передбачає одночасний аналіз різних форм медіа, таких як текст, зображення та відео. Цей комплексний підхід визнає, що в оманливих наративах часто використовується поєднання текстових і візуальних елементів для передачі оманливої інформації. Мультимодальний аналіз об'єднує методи обробки природної мови (NLP) для текстового аналізу, аналізу зображень для дослідження зображень і аналізу відео для ретельного вивчення відеовмісту. Оцінюючи узгодженість або невідповідність між різними способами, цей підхід підвищує точність виявлення оманливого вмісту. Мультимодальний аналіз також розглядає метадані, пов'язані з мультимедійним вмістом, надаючи додатковий контекст для оцінки автентичності. Загалом, він пропонує цілісний погляд на дезінформацію, визнаючи багатовимірну природу оманливих наративів у сучасних ЗМІ. Мультимодальний аналіз пропонує кілька переваг у контексті виявлення фейків і пропаганди завдяки застосуванню цілісного підходу, який враховує як текстові, так і візуальні елементи. Ось ключові переваги мультимодального аналізу:

- *Текстова та візуальна синергія.* Мультимодальний аналіз передбачає одночасне дослідження текстових і візуальних елементів у частині вмісту. Поєднуючи лінгвістичний аналіз (обробка природної мови або NLP) з техніками обробки зображень і відео, дослідники прагнуть вловити синергію між текстовою інформацією та супровідними візуальними медіа, визнаючи, що оманливі наративи часто включають скоординовану взаємодію слів і зображень.
- *Deepfake та Image Forensics.* Враховуючи розвиток технології deepfake та інструментів обробки зображень, мультимодальний аналіз включає методи ідентифікації маніпульованого або синтетичного вмісту. Алгоритми криміналістики зображень використовуються для виявлення змін, таких як глибокі фальшиві обличчя або маніпульовані зображення, підвищуючи загальну точність виявлення обману.
- *Крос-модальна узгодженість.* Мультимодальний аналіз прагне оцінити узгодженість або невідповідність між різними модальностями. В автентичному вмісті текстові та візуальні компоненти зазвичай узгоджені. В той же час, оманливі наративи можуть виявляти розбіжності, коли текстова інформація суперечить візуальним елементам або навпаки.
- *Перевірка метаданих.* Мультимодальний аналіз виходить за межі самого вмісту й включає перевірку метаданих, пов'язаних із зображеннями та відео. Аналіз метаданих, таких як мітки часу, геолокація та історія редагування, забезпечує додатковий контекст для оцінки автентичності та надійності мультимедійного вмісту.
- *Розпізнавання шаблонів різних модальностей.* Підхід передбачає розробку алгоритмів, здатних розпізнавати шаблони обману, які охоплюють кілька

модальностей. Це включає шаблони в тому, як певні зображення або візуальні елементи поєднуються з конкретними текстовими наративами, що дозволяє більш повно зрозуміти, як пропагандисти створюють оманливі повідомлення.

- *Інтеграція соціальних мереж.* Мультимодальний аналіз особливо актуальний у контексті соціальних мереж, де дезінформація часто поширюється через швидкий обмін зображеннями, відео та текстовим вмістом. Інтеграція мультимодальних методів в аналіз соціальних медіа покращує здатність ідентифікувати та протидіяти оманливим наративам на цих платформах.
- *Покращені індикатори обману.* Комбінуючи сигнали з різних модальностей, мультимодальний аналіз забезпечує розширені індикатори обману. Наприклад, текст, який стверджує, що зображує певну подію, може мати перехресні посилання з візуальним вмістом, щоб визначити, чи відповідає розповідь фактичним зображенням, додаючи рівні перевірки в процесі виявлення.

Мультимодальний аналіз представляє цілісний підхід до виявлення фейків і пропаганди, враховуючи взаємодію текстових і візуальних елементів. Інтегруючи методи NLP, криміналістики зображень і крос-модальних перевірок узгодженості, дослідники прагнуть створити більш надійні та комплексні системи для виявлення та пом'якшення оманливих наративів у різноманітному медіа-ландшафті. Хоча мультимодальний аналіз є цінним підходом для виявлення фейків і пропаганди, він має певні обмеження, які впливають на його ефективність. Ці обмеження включають:

- *Складність аналізу.* Інтеграція текстової, візуальної та іноді звукової інформації збільшує складність аналізу. Координація різноманітних аналітичних методів створює проблеми для розробки алгоритмів, які ефективно обробляють та інтерпретують інформацію в різних модальностях.
- *Ресурсомісткість.* Аналіз кількох модальностей вимагає значних обчислювальних ресурсів. Мультимодальний аналіз, особливо при роботі з великими наборами даних, може бути ресурсомістким і може вимагати передової обчислювальної інфраструктури, що обмежує її доступність у середовищах з обмеженими ресурсами.
- *Нові тактики обману.* У міру того, як тактика омани розвивається, мультимодальному аналізу може бути важко встигати за новими стратегіями. Адаптація алгоритмів для виявлення нових форм маніпуляції як у текстовому, так і у візуальному вмісті потребує постійних досліджень і розробок.
- *Контекстуальні виклики.* Розуміння контексту, в якому представлений мультимодальний контент, може бути складним завданням. Розрізнення між навмисним обманом і законним творчим вираженням чи сатирою вимагає детального розуміння контексту, який може бути важко повністю зрозуміти автоматизованим системам.
- *Складність крос-модальної інтеграції.* Інтеграція сигналів від різних модальностей і створення значущих зв'язків між текстовими та візуальними елементами представляє проблеми.

Забезпечення точного відображення аналізом передбачуваного повідомлення та наміру вмісту вимагає складної міжмодальної інтеграції.

- *Медіа-ландшафт, що швидко розвивається.* Динаміка медіа-ландшафту, включно зі швидким створенням і розповсюдженням контенту на платформах соціальних медіа, створює проблеми для багатомодального

аналізу. Виявлення обману в режимі реального часу стає дедалі складнішим, оскільки обсяг і швидкість обміну інформацією зростають.

- *Обмежене узагальнення.* моделі мультимодального аналізу, навчені на конкретних наборах даних, можуть мати проблеми з узагальненням у різних контекстах або різноманітних мовних і культурних нюансах. Створення надійних моделей у різних сценаріях залишається актуальною проблемою.
- *Змагальні атаки.* Пропагандисти можуть використовувати тактику змагальності, щоб навмисно маніпулювати виходом мультимодальних моделей аналізу. Це створює ризик незначних змін у вмісті, спрямованих на уникнення виявлення, підкреслюючи вразливість цих систем до складних атак противника.

Хоча мультимодальний аналіз значно розширює можливості виявлення фейків і пропаганди, усунення цих обмежень має вирішальне значення для підвищення надійності, адаптивності та етичних міркувань, пов'язаних із цим підходом. Поточні дослідження та розробки спрямовані на пом'якшення цих проблем і підвищення загальної ефективності мультимодального аналізу в мінливому ландшафті дезінформації.

Машинне навчання — це обчислювальний підхід, який використовується для аналізу закономірностей і прогнозування на основі даних. У сфері виявлення фейків і пропаганди алгоритми машинного навчання навчені розпізнавати особливості та характеристики, пов'язані з оманливими наративами. Ці алгоритми вивчають великі набори даних, що містять приклади як автентичного, так і оманливого вмісту, що дозволяє їм визначати шаблони, які можуть вказувати на дезінформацію [4].

Існує два основних типи підходів до машинного навчання:

- *Контрольоване навчання.* У цьому підході алгоритми навчаються на позначених наборах даних, де ідентифікуються випадки оманливого вмісту. Алгоритм вчиться узагальнювати на основі позначених прикладів, роблячи прогнози на основі нових, невідомих даних.
- *Неконтрольоване навчання.* Неконтрольоване навчання включає в себе алгоритми, які визначають закономірності та аномалії в даних без явного позначення. Цей підхід корисний для виявлення оманливого вмісту без попередньо визначених прикладів.

Моделі машинного навчання витягують із даних такі функції, як лінгвістичні структури, візуальні підказки або шаблони поширення. Ці ознаки служать основою для прогнозів або класифікацій щодо автентичності вмісту. Моделі машинного навчання є адаптивними, здатними адаптуватися до нових тактик, які використовують ті, хто поширює дезінформацію. Аналіз у реальному часі є ключовою перевагою машинного навчання в цьому контексті, що дозволяє швидко оцінювати вхідні дані для виявлення та протидії оманливим наративам. Крім того, масштабованість дозволяє моделям машинного навчання обробляти зростаючий обсяг інформації, що циркулює в Інтернеті.

Таким чином, машинне навчання виявлення фейків і пропаганди передбачає навчання алгоритмів для розпізнавання шаблонів, що вказують на оманливі наративи. Ці моделі адаптуються до нових тактик, працюють у режимі реального часу та масштабуються для аналізу великих наборів даних, сприяючи зусиллям із виявлення та пом'якшення дезінформації в цифрових медіа.

Машинне навчання пропонує кілька переваг у контексті виявлення фейків і пропаганди, сприяючи ефективності та ефективності виявлення оманливих наративів:

- *Розпізнавання шаблонів.* Машинне навчання чудово розпізнає шаблони в даних. У сфері фейків і пропаганди алгоритми можуть вивчати шаблони,

пов'язані з оманливими наративами, дозволяючи автоматично ідентифікувати подібні шаблони в нових, невидимих даних.

- *Адаптивність.* Моделі машинного навчання адаптуються та можуть розвиватися, щоб розпізнавати нові тактики, які використовують розповсюджувачі дезінформації. Оскільки оманливі стратегії змінюються, системи машинного навчання можна оновлювати, щоб випереджати шаблони, що розвиваються.
- *Аналіз у реальному часі.* Алгоритми машинного навчання можуть працювати в режимі реального часу, надаючи можливість аналізувати вхідні дані, коли вони з'являються. Це має вирішальне значення в динамічному середовищі цифрових медіа, де своєчасне виявлення оманливого вмісту має важливе значення для ефективного пом'якшення.
- *Масштабованість*. моделі машинного навчання можуть масштабуватися для обробки великих обсягів даних. Оскільки в Інтернеті циркулює величезна кількість інформації, масштабованість підходів машинного навчання дозволяє ефективно аналізувати великі набори даних.
- *Ефективність.* Автоматизовані системи машинного навчання можуть швидко обробляти й аналізувати дані, значно підвищуючи ефективність виявлення фейків і пропаганди. Ця ефективність необхідна для того, щоб не відставати від швидкого поширення дезінформації.
- *Ансамблеві методи.* Машинне навчання часто використовує ансамблеві методи, поєднуючи прогнози з кількох моделей для підвищення точності. Цей підхід використовує сильні сторони різних алгоритмів і пом'якшує слабкі сторони окремих моделей, створюючи більш надійні системи виявлення.
- *Безперервне навчання.* Системи машинного навчання можуть брати участь у безперервному навчанні. Коли вони стикаються з новими даними, вони можуть адаптувати та оновлювати свої моделі, забезпечуючи ефективність можливостей виявлення з часом.
- *Мультимодальний аналіз.* Машинне навчання полегшує мультимодальний аналіз шляхом обробки та аналізу різних типів даних, включаючи текст, зображення та відео. Цей комплексний підхід підвищує точність виявлення оманливих наративів, які можуть використовувати різні медіаформати.
- *Метрики оцінки.* Моделі машинного навчання можна оцінити за допомогою таких показників, як точність, запам'ятовування та оцінка F1, що забезпечує кількісну оцінку їх ефективності. Це дозволяє постійно вдосконалювати та оптимізувати можливості виявлення [5].

Переваги машинного навчання у виявленні фейків і пропаганди включають його здатність розпізнавати закономірності, адаптуватися до нових тактик, працювати в режимі реального часу, ефективно масштабувати та полегшувати використання методів ансамблю. Ці можливості сприяють розробці надійних систем для виявлення та пом'якшення оманливих наративів у цифрових медіа.

Хоча машинне навчання є потужним інструментом для виявлення фейків і пропаганди, воно має певні обмеження, які впливають на його ефективність у певних контекстах:

- *Зміщення в навчальних даних*. Моделі машинного навчання значною мірою залежать від якості та репрезентативності навчальних даних. Якщо навчальні дані містять упередження або відображають певні перспективи,

модель може ненавмисно закріпити та посилити ці упередження, що призведе до спотворення результатів у виявленні фейків і пропаганди.

- *Відсутність контекстуального розуміння.* Моделі машинного навчання, особливо ті, що базуються на вузьких навчальних даних, можуть мати проблеми з розумінням ширшого контексту, в якому подається інформація. Розуміння сарказму, іронії чи культурних нюансів, які часто використовуються в оманливих оповіданнях, може бути складним для цих моделей.
- *Обмежене узагальнення.* моделі, навчені на конкретних наборах даних, можуть мати труднощі з узагальненням для нових і різноманітних сценаріїв. Ефективність моделі машинного навчання може бути обмеженою, якщо зіткнутися з варіаціями в мові, культурному контексті або еволюції тактики, що використовується для поширення дезінформації.
- *Вразливість до агресивних атак.* Моделі машинного навчання можуть бути вразливими до агресивних атак, коли люди навмисно маніпулюють вхідними даними, щоб обдурити систему. Змагальні атаки в контексті виявлення фейків і пропаганди можуть включати тонкі зміни вмісту, щоб уникнути виявлення.
- *Брак пояснення.* Багатьом моделям машинного навчання, особливо складним, таким як глибокі нейронні мережі, бракує прозорості та пояснення. Природа «чорної скриньки» цих моделей може ускладнити розуміння того, як вони приходять до конкретних рішень, зменшуючи довіру та інтерпретацію в контексті виявлення фейків і пропаганди.
- *Перенавчання та недонавчання.* Перенавчання відбувається, коли модель навчена надто близько до навчальних даних, захоплюючи шум, а не базові моделі. З іншого боку, недостатнє навчання трапляється, коли модель занадто проста, щоб захопити відповідні візерунки. Баланс між надмірним і недостатнім навчанням є вирішальним для узагальнення моделі на нові дані.
- *Динамічний характер дезінформації.* Тактики дезінформації швидко розвиваються, і моделям машинного навчання може бути важко встигати за новими стратегіями. Модель, навчена на історичних даних, може не вловлювати нові тактики та шаблони, які використовуються для поширення фейків і пропаганди.
- *Високі вимоги до ресурсів.* Деякі моделі машинного навчання, особливо складні, потребують значних обчислювальних ресурсів для навчання та висновків. Це може обмежити їх доступність у середовищах з обмеженими ресурсами.

Розуміння цих обмежень має важливе значення для розробки більш надійних і відповідальних моделей машинного навчання для виявлення фейків і пропаганди. Вирішення цих проблем передбачає безперервне дослідження, етичні міркування та міждисциплінарний підхід для підвищення надійності систем виявлення.

Аналіз пропаганди на основі емоційного забарвлення. Сентимент-аналіз є ключовим інструментом в аналізі тексту, спрямованим на визначення емоційного забарвлення текстових даних. Цей метод полягає у визначенні тональності тексту як позитивної, негативної або нейтральної. Використовуючи сентимент-аналіз, можна автоматично визначити, наскільки певний текст емоційно забарвлений, що є корисним для багатьох завдань аналізу тексту, включаючи виявлення пропагандистських матеріалів. Сентимент-аналіз використовується для ідентифікації емоційно забарвлених

ISSN 2663 - 4023

текстів, таких як пропагандистські матеріали, які зазвичай мають виражений емоційний характер. Пропагандистські тексти, спрямовані на маніпуляцію аудиторією та виклик певних реакцій, часто містять відзначені позитивні або негативні настрої, які можуть бути виявлені за допомогою сентимент-аналізу. Під час аналізу пропаганди важливо враховувати не лише зміст тексту, а й його емоційне навантаження. Сентимент-аналіз дозволяє об'єктивно визначити емоційне забарвлення текстів та порівняти його між пропагандистськими та не-пропагандистськими матеріалами.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

У нашому дослідженні ми використовуємо сентимент-аналіз для аналізу емоційного забарвлення пропагандистських матеріалів порівняно з не-пропагандистськими текстами. Ми обчислюємо середнє значення сентименту для кожної категорії текстів та порівнюємо їхні емоційні характеристики. Використання сентимент-аналізу є важливим етапом у виявленні пропагандистських матеріалів та розумінні їхнього емоційного впливу на аудиторію. Цей метод допомагає здійснити об'єктивну оцінку текстів та виявити їхній потенційний вплив на сприйняття інформації.

main.py

```
def main():
    (propaganda, non_propaganda) = get_datasets()

    (propaganda_emotions, propaganda_positive, propaganda_negative, propaganda_subjective, propaganda_mixes) = analyze_emotion(propaganda)
    (non_propaganda_emotions, non_propaganda_positive, non_propaganda_negative, non_propaganda_subjective, non_propaganda_mixes) =
analyze_emotion(non_propaganda)

    propaganda_result_emotions = sum(propaganda_emotions) / len(propaganda_emotions)
    non_propaganda_result_emotions = sum(non_propaganda_emotions) / len(non_propaganda_emotions)

    propaganda_result_positive = sum(propaganda_positive) / len(propaganda_positive)
    non_propaganda_result_positive = sum(non_propaganda_positive) / len(non_propaganda_positive)

    propaganda_result_negative = sum(propaganda_negative) / len(propaganda_negative)
    non_propaganda_result_negative = sum(non_propaganda_negative) / len(non_propaganda_negative)

    propaganda_result_subjective = sum(propaganda_subjective) / len(propaganda_subjective)
    non_propaganda_result_subjective = sum(non_propaganda_subjective) / len(non_propaganda_subjective)

    propaganda_result_mix = sum(propaganda_mixes) / len(propaganda_mixes)
    non_propaganda_result_mix = sum(non_propaganda_mixes) / len(non_propaganda_mixes)
```

Рис. 1. Програма для аналізу емоційного забарвлення пропагандистських та не-пропагандистських матеріалів

Розроблено програму для аналізу текстів з двох різних датасетів (рис. 1): пропагандистських новин і не-пропагандистських новин. Ця програма використовує бібліотеку TextBlob для аналізу емоційного забарвлення кожного тексту та отримання об'єктивної оцінки емоційної тонікації. Програма аналізує кожен текст для визначення його емоційного забарвлення, позитивного та негативного емоційного забарвлення, ступеня суб'єктивності та комплексного емоційного забарвлення. Для цього вона використовує методи бібліотеки TextBlob, яка простежує емоційний тон тексту та визначає його характеристики.

- *Emotion (Емоційне забарвлення)* (рис. 2): Середнє значення емоційного забарвлення для пропагандистських новин становить 0.151, а для не-пропагандистських — 0.116.

ISSN 2663 - 4023

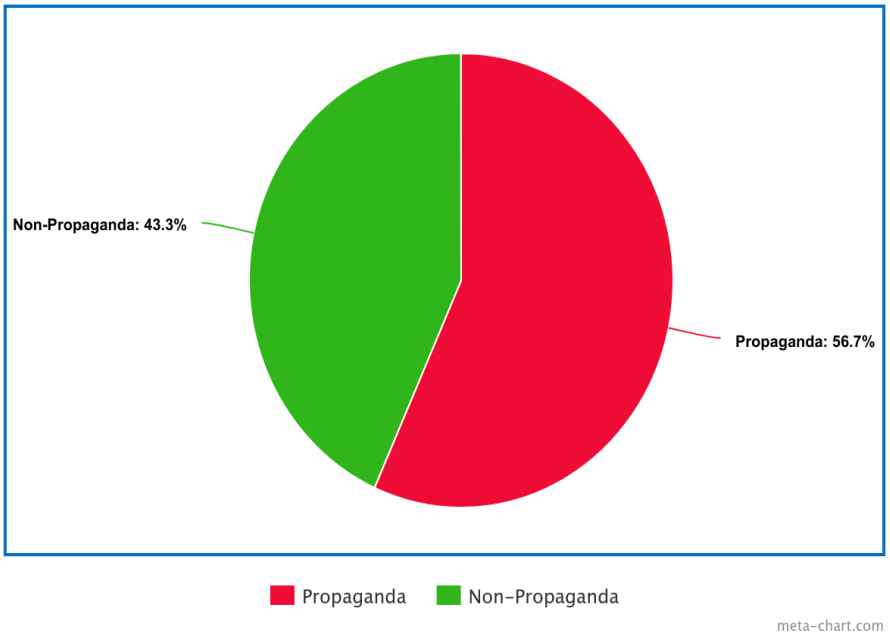

Рис. 2. Емоційність

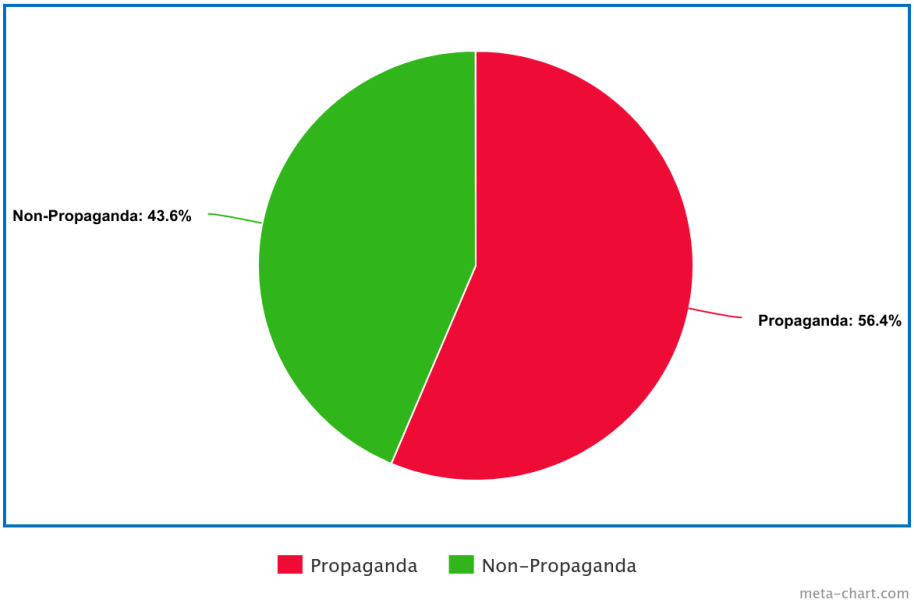

Рис. 3. Суб'єктивність

- *Subjectivity (Суб'єктивність)* (рис. 3): Середнє значення ступеня суб'єктивності для пропагандистських новин становить 0.365, а для не-пропагандистських — 0.283.
- *Positive emotion (Позитивне емоційне забарвлення)* (рис. 4): Середнє значення позитивного емоційного забарвлення для пропагандистських новин становить 0.087, а для не-пропагандистських — 0.082.

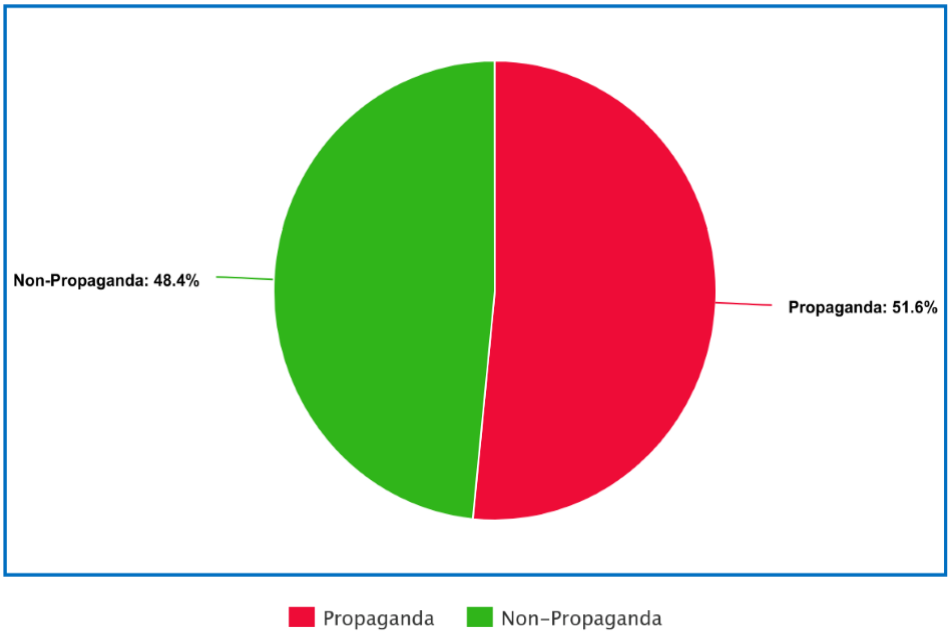

Рис. 4. Позитивні емоції

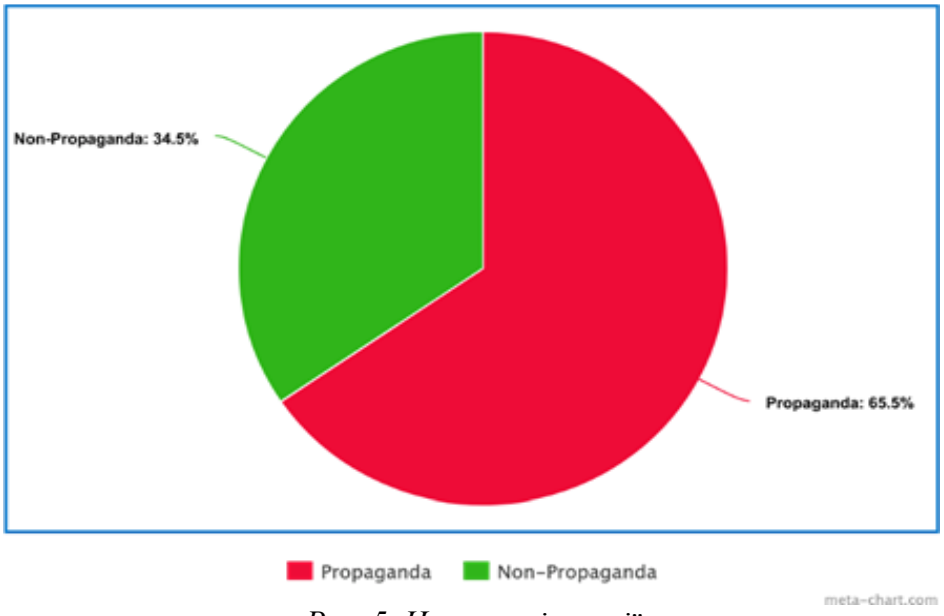

Рис. 5. Негативні емоції

- *Negative emotion (Негативне емоційне забарвлення)* (рис. 5): Середнє значення негативного емоційного забарвлення для пропагандистських новин становить 0.064, а для не-пропагандистських — 0.034.
- *Complex (Комплексна оцінка)* (рис. 6): Середнє значення комплексного емоційного забарвлення для пропагандистських новин становить 0.021, а для не-пропагандистських — 0.010.

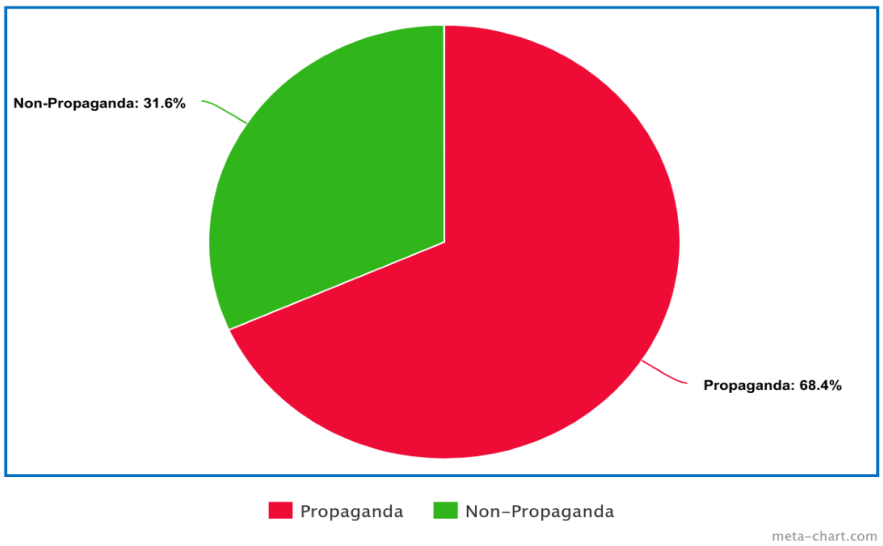

Рис. 6. Комплексна оцінка

Результати аналізу емоційного забарвлення текстів вказують на кілька ключових відмінностей між пропагандистськими та не-пропагандистськими матеріалами. Зокрема, у пропагандистських новинах спостерігається значно більше негативного емоційного забарвлення порівняно з не-пропагандистськими. Середнє значення негативного емоційного забарвлення для пропагандистських новин складає 0.064, тоді як для не-пропагандистських це значення лише 0.034. Це свідчить про те, що пропагандистські матеріали мають тенденцію до акцентуації негативних аспектів, що може бути спрямовано на формування певної негативної думки чи емоційного реакції у читачів.

Однак, щодо позитивного емоційного забарвлення, спостерігається, що його рівень у пропагандистських та не-пропагандистських новинах приблизно однаковий. Середнє значення позитивного емоційного забарвлення для пропагандистських новин складає 0.087, а для не-пропагандистських — 0.082. Це може свідчити про те, що позитивні емоції використовуються у пропагандистських матеріалах для збільшення привабливості та привертання уваги аудиторії.

Найбільша відмінність спостерігається в показнику комплексного емоційного забарвлення. У пропагандистських новинах цей показник виявляється в два рази більшим, ніж у не-пропагандистських матеріалів. Це може свідчити про те, що емоційне забарвлення пропагандистських текстів є більш складним і включає в себе більше різних аспектів, таких як негативність, емоційність та суб'єктивність. Ця відмінність може бути використана для розробки більш точних алгоритмів виявлення та аналізу пропагандистських матеріалів у текстах, що дозволить покращити ефективність виявлення таких матеріалів та розуміння їхнього впливу на аудиторію.

ПОДЯКА

Ця стаття підготовлена завдяки грантової підтримки Національного Фонду Досліджень України, реєстраційний номер проєкту 187/0012 від 1/08/2024 (2023.04/0012) «Розроблення інформаційної системи автоматичного виявлення джерел дезінформації та неавтентичної поведінки користувачів чатів» за конкурсом «Наука для зміцнення обороноздатності України».

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

В епоху цифрових технологій поширений вплив фейків і пропаганди вимагає складного та багатогранного підходу для виявлення та пом'якшення. У цій статті розглядаються різні методи, з особливим акцентом на обробці природної мови (NLP), мультимодальному аналізі та машинному навчанні в контексті виявлення оманливих наративів. Кожен метод надає унікальні сильні сторони досягненню головної мети — захисту цілісності розповсюдження інформації.

Дослідження обробки природної мови підкреслило її ефективність у аналізі мовних шаблонів, тоді як мультимодальний аналіз визнає мультимедійну природу сучасної дезінформації. Машинне навчання з його адаптивністю та можливостями в реальному часі надає динамічний інструмент для розпізнавання тактик, що розвиваються.

Міждисциплінарна співпраця має першочергове значення. Завдяки об'єднанню технологів, соціологів, політиків і медіа-професіоналів виникає більш повне розуміння соціально-політичних та етичних аспектів дезінформації.

Контекстуальні проблеми, пов'язані з мовою, культурними нюансами та швидко розвиваючим ландшафтом дезінформації, підкреслюють необхідність постійних досліджень для покращення контекстуального розуміння систем виявлення.

Враховуючи динамічний характер тактики дезінформації, необхідні постійні дослідження для розробки моделей, які можуть динамічно адаптуватися до нових стратегій. Це передбачає створення систем, які випереджають шаблони, що розвиваються, шляхом постійного навчання та оновлень.

Поліпшення інтеграції різних способів виявлення дезінформації має вирішальне значення. Дослідницькі зусилля повинні вивчати методи, які плавно поєднують текстову, візуальну та, можливо, звукову інформацію для створення більш цілісних систем виявлення.

Емоційне забарвлення в пропагандистських матеріалах виявляється більш вираженим у відношенні до не-пропагандистських. Пропагандистські новини мають схильність до більшого негативного емоційного висловлювання, що може бути використано для маніпуляції аудиторією. Також спостерігається більша складність емоційного забарвлення в пропагандистських матеріалах.

Підсумовуючи, боротьба з фейками та пропагандою вимагає комплексної стратегії, що розвивається. Синергія технологічних досягнень, етичних міркувань, міждисциплінарної співпраці та розширення можливостей користувачів є важливою. По мірі просування вперед важливо розглядати виявлення фейків і пропаганди не як окремий виклик, а як динамічну сферу, де постійні дослідження та розробки є ключовими для того, щоб випередити тих, хто прагне маніпулювати інформацією для своїх цілей. Лише шляхом узгоджених та адаптивних зусиль ми можемо сподіватися побудувати стійкий захист від повсюдного впливу оманливих наративів у нашу цифрову епоху.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Гайдайчук, В., & Смаль, М. (2021). Методичні підходи до виявлення фейків та пропаганди в мережі Інтернет. Н*ауковий вісник Ужгородського національного університету. Серія: Філологія, (2),* 27–31.
2. Zhang, Z., Chen, Y., & Chen, L. (2020). Methods for detecting fake news: A survey. *ACM Computing Surveys, 53(6),* 1–37.
3. Яковлева, О. (2021). Методи виявлення фейків та пропаганди в текстових даних. *Науковий вісник Дніпровського університету ім. Олеся Гончара. Серія: Комп'ютерні науки, (2),* 28–32.

4. Gupta, M., Mishra, A., & Kumar, V. (2020). A survey on fake news detection using machine learning techniques. *Journal of Big Data, 7(1),* 1–21.
5. Висоцька, В. А. (2024). Інформаційна технологія розпізнавання пропаганди, фейків та дезінформації у текстовому контенті на основі методів NLP та машинного навчання. *Радіоелектроніка, інформатика, управління, (2),* 126. https://doi.org/10.15588/1607-3274-2024-2-13
6. Vysotskyi, O., & Vysotska, O. (2020). Technologies of public diplomacy. *Епістемологічні дослідження в філософії, соціальних і політичних науках, 3(1*), 139–147.
7. Vysotska, V., Mazepa, S., Chyrun, L., Brodyak, O., Shakleina, I., & Schuchmann, V. (2022). NLP tool for extracting relevant information from criminal reports or fakes/propaganda content. *IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT),* 93–98. https://doi.org/10.15588/10.1109/CSIT56902.2022.10000563
8. Mykytiuk, A., Vysotska, V., Markiv, O., Chyrun, L., & Pelekh, Y. (2023). Technology of Fake News Recognition Based on Machine Learning Methods. *CEUR Workshop Proceedings, vol. 3387,* 311–330.
9. Vysotska, V., Chyrun, L., Chyrun, S., & Holets, I. (2024). Information technology for identifying disinformation sources and inauthentic chat users' behaviours based on machine learning. *CEUR Workshop Proceedings, vol. 3723,* 466–483.
10. Vysotskyi, O. Y., Holovko, I. K., & Vysotska, O. Y. (2023). Theory of geocultural technologies in international relations. Journal of Geology, *Geography and Geoecology, 32(2),* 426–435.
11. Vysotska, V. (2024). Modern State and Prospects of Information Technologies Development for Natural Language Content Processing. *CEUR Workshop Proceedings, vol. 3668,* 198–234.
12. Lynnyk, R., Vysotska, V., & Andrunyk, V. (2024). Study of the Problems of Determining Public Opinion of the Israeli-Palestinian War in Social Networks. *Qeios.* https://doi.org/10.32388/0BIA5E

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

Vitalii Danylyk
PhD student,
Information Systems And Networks Department,
Lviv Polytechnic National University, Lviv, Ukraine
ORCID ID: 0000-0001-5928-7235
vitalii.m.danylyk@lpnu.ua

Victoria Vysotska
Associate Professor,
Information Systems And Networks Department,
Lviv Polytechnic National University, Lviv, Ukraine
ORCID ID: 0000-0001-6417-3689
victoria.a.vysotska@lpnu.ua

Mariia Nazarkevych
Professor,
Information Systems And Networks Department,
Lviv Polytechnic National University, Lviv, Ukraine
ORCID ID: 0000-0002-6528-9867
mariia.a.nazarkevych@lpnu.ua

DISINFORMATION, FAKES AND PROPAGANDA IDENTIFICATION METHODS IN MASS MEDIA BASED ON MACHINE LEARNING

Abstract. This article provides a comprehensive research of modern approaches used to identify fakes and propaganda. The study begins by examining the prevalence and impact of misinformation, emphasising the need for advanced identification methods. Tracing the evolution of methods, the article delves into historical approaches and their adaptation to the dynamic media landscape. Central to the discussion are advanced techniques such as natural language processing, which analyses linguistic patterns to detect inconsistencies in textual content. The article explores the benefits of natural language processing, including scalability and real-time analysis, along with the limitations associated with contextual ambiguity and the development of linguistic nuances. Multimodal analysis takes centre stage, offering a holistic perspective through the textual and visual elements integration. The article explores the benefits of multimodal analysis, highlighting its potential for comprehensive understanding and increased accuracy, while acknowledging the challenges associated with analytical complexity and contextual nuance. Machine learning is emerging as a dynamic tool for pattern recognition and adaptation that facilitates real-time analysis. The paper highlights the advantages of machine learning, such as scalability, and discusses its limitations, including biases in the training data and vulnerability to brute-force attacks. In addition, the article provides a propaganda analysis based on emotional colouring, which reveals the differences between propaganda and non-propaganda. Recommendations for future research emphasise the importance of transparency of models and ongoing efforts to eliminate biases. As the digital landscape continues to evolve, these advances pave the way for sustainable strategies in detecting and countering fakes and propaganda.

Keywords: disinformation; fake; propaganda; identification of disinformation; NLP; cyber security; mass media; natural language processing; multimodal analysis; machine learning.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Haydaychuk, V., & Smal, M. (2021). Methodological approaches to detecting fakes and propaganda on the Internet. *Scientific Bulletin of Uzhhorod National University. Series: Philology, (2),* 27–31.
2. Zhang, Z., Chen, Y., & Chen, L. (2020). Methods for detecting fake news: A survey. *ACM Computing Surveys, 53(6),* 1–37.
3. Yakovleva, O. (2021). Methods for detecting fakes and propaganda in text data. Scientific Bulletin of Oles Honchar Dnipro University. *Series: Computer Science, (2),* 28–32.

4. Gupta, M., Mishra, A., & Kumar, V. (2020). A survey on fake news detection using machine learning techniques. *Journal of Big Data, 7(1),* 1–21.
5. Vysotska, V.A. (2024). Information technology for recognizing propaganda, fakes and disinformation in textual content based on NLP and machine learning methods. *Radioelectronics, Informatics, Management, (2),* 126. https://doi.org/10.15588/1607-3274-2024-2-13
6. Vysotskyi, O., & Vysotska, O. (2020). Technologies of public diplomacy. *Epistemological studies in philosophy, social and political sciences, 3(1),* 139–147.
7. Vysotska, V., Mazepa, S., Chyrun, L., Brodyak, O., Shakleina, I., & Schuchmann, V. (2022). NLP tool for extracting relevant information from criminal reports or fakes/propaganda content. *IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT),* 93–98. https://doi.org/10.15588/10.1109/CSIT56902.2022.10000563
8. Mykytiuk, A., Vysotska, V., Markiv, O., Chyrun, L., & Pelekh, Y. (2023). Technology of Fake News Recognition Based on Machine Learning Methods. *CEUR Workshop Proceedings, vol. 3387,* 311–330.
9. Vysotska, V., Chyrun, L., Chyrun, S., & Holets, I. (2024). Information technology for identifying disinformation sources and inauthentic chat users' behaviours based on machine learning. *CEUR Workshop Proceedings, vol. 3723,* 466–483.
10. Vysotskyi, O. Y., Holovko, I. K., & Vysotska, O. Y. (2023). Theory of geocultural technologies in international relations. Journal of Geology, *Geography and Geoecology, 32(2),* 426–435.
11. Vysotska, V. (2024). Modern State and Prospects of Information Technologies Development for Natural Language Content Processing. *CEUR Workshop Proceedings, vol. 3668,* 198–234.
12. Lynnyk, R., Vysotska, V., & Andrunyk, V. (2024). Study of the Problems of Determining Public Opinion of the Israeli-Palestinian War in Social Networks. *Qeios.* https://doi.org/10.32388/0BIA5E