УДК 004.8:316.774
DOI https://doi.org/10.32689/maup.it.2025.2.8

Володимир КАРАБЧУК
аспірант, Приватний вищий навчальний заклад «Європейський університет»,
volodymyr.karabchuk@e-u.edu.ua
***ORCID:** 0009-0002-4320-3169*

МЕТОДИ ВИЯВЛЕННЯ ТА БОРОТЬБИ З ДЕЗІНФОРМАЦІЄЮ ЗА ДОПОМОГОЮ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ

Анотація. *У статті розглядаються сучасні методи виявлення та боротьби з дезінформацією за допомогою технологій штучного інтелекту. Основна увага приділяється автоматизації аналізу текстової, візуальної та відеоінформації, використовуючи алгоритми машинного навчання, обробки природної мови та комп'ютерного зору. Розглянуто інструменти для моніторингу соціальних мереж, перевірки фактів і виявлення маніпуляцій у медіаконтенті. Окремо проаналізовано виклики, пов'язані з моральними аспектами, дефіцитом даних і розвитком технологій для створення фейків. Запропоновані підходи спрямовані на підвищення ефективності боротьби з дезінформацією, забезпечення інформаційної безпеки та стійкості суспільства до маніпуляцій.*

Мета роботи. *Дослідження сучасних методів виявлення та боротьби з дезінформацією за допомогою технологій штучного інтелекту, зокрема машинного навчання, обробки природної мови та комп'ютерного зору, для підвищення інформаційної безпеки суспільства.*

Методологія. *У роботі використано алгоритми машинного навчання, такі як згорткові нейронні мережі (CNN) та трансформерні моделі (наприклад, BERT), для аналізу текстового контенту. Для ідентифікації маніпуляцій у візуальному контенті застосовано технології комп'ютерного зору, а для аналізу поширення дезінформації в соціальних мережах – графові нейронні мережі (GNN).*

Наукова новизна. *Запропоновано інтегрований підхід, що включає аналіз текстового, візуального та відеоконтенту одночасно. Досліджено вплив емоційного забарвлення контенту на сприйняття фейкових новин і вперше вивчено роль мультимодальних систем у боротьбі з дезінформацією.*

Висновки. *Технології штучного інтелекту забезпечують ефективне виявлення дезінформації шляхом аналізу великих обсягів даних у реальному часі. Інтеграція мультимодальних підходів*

сприяє підвищенню стійкості інформаційного простору, хоча залишається необхідність розв'язання проблеми дефіциту якісних навчальних даних і врахування етичних аспектів.

Ключові слова: *дезінформація, штучний інтелект, обробка природної мови, машинне навчання, перевірка фактів, маніпулятивний контент, deepfake, кібербезпека.*

Volodymyr KARABCHUK. METHODS FOR DETECTING AND COMBATING DISINFORMATION USING ARTIFICIAL INTELLIGENCE TECHNOLOGIES

Abstract. *The article examines modern methods for detecting and combating disinformation using artificial intelligence technologies. The primary focus is on automating the analysis of textual, visual, and video information through the use of machine learning algorithms, natural language processing, and computer vision. Tools for monitoring social networks, fact-checking, and identifying manipulations in media content are discussed. Challenges related to ethical aspects, data scarcity, and the development of technologies for creating fake content are also analyzed. The proposed approaches aim to enhance the effectiveness of combating disinformation, ensure information security, and increase societal resilience to manipulation.*

Purpose. *To study modern methods for detecting and combating disinformation using artificial intelligence technologies, particularly machine learning, natural language processing, and computer vision, to enhance societal information security.*

Methodology. *The study employs machine learning algorithms, such as convolutional neural networks (CNN) and transformer models (e.g., BERT), for textual content analysis. Computer vision technologies are applied to detect manipulations in visual content, while graph neural networks (GNN) are used to*

analyze the dissemination of disinformation in social networks.

Scientific novelty. *An integrated approach is proposed, combining the analysis of textual, visual, and video content simultaneously. The study investigates the impact of emotional coloring on the perception of fake news and explores the role of multimodal systems in combating disinformation for the first time.*

Conclusions. *Artificial intelligence technologies provide effective disinformation detection by analyzing large volumes of data in real-time. The integration of multimodal approaches enhances the resilience of the information space, though challenges such as the lack of high-quality training data and ethical considerations remain.*

Key words: *disinformation, artificial intelligence, natural language processing, machine learning, fact-checking, manipulative content, deepfake, cybersecurity.*

Вступ. Дезінформація залишається однією з найважливіших загроз інформаційній безпеці, впливаючи на політичні вибори, соціальну стабільність і економічний розвиток. Наприклад, під час пандемії COVID-19 значна кількість фейкових новин про вакцини викликала масову дезорієнтацію громадськості, що, за даними ВООЗ, призвело до зниження рівня вакцинації в окремих регіонах. Інформаційні війни, як-от ті, що розгортаються під час міжнародних конфліктів, підкреслюють, наскільки важливою є швидка та точна перевірка достовірності даних.

Дослідження в цій галузі вказують, що штучний інтелект здатний ефективно протидіяти дезінформації. Наприклад, алгоритми обробки природної мови (NLP) вже застосовуються для автоматизованого аналізу текстів у реальному часі. Дослідження, [2] опубліковане в журналі *Information Sciences* (2022), демонструє, як моделі трансформерів, такі як BERT та GPT, можуть аналізувати мільйони текстів для виявлення маніпуляцій і упереджених тверджень. У візуальній сфері технології комп'ютерного зору, такі як методи аналізу пікселів і метаданих, використовуються для ідентифікації deepfake-зображень, що підтверджується [4, 10] результатами проєкту DeepFake Detection Challenge.

Водночас існує низка викликів, які ускладнюють ефективне застосування штучного інтелекту для боротьби з дезінформацією. Зокрема, спостерігається недостатність якісних навчальних даних, які необхідні для створення універсальних моделей, здатних працювати в багатомовних і культурно різноманітних середовищах. Крім того, швидке вдосконалення технологій створення фейкового контенту, таких як deepfake, значно ускладнює їх своєчасне виявлення. Окрему проблему становлять етичні аспекти використання ШІ, зокрема ризик хибного маркування правдивої інформації як дезінформації через контекстні помилки. Наприклад, дослідження [9] показують, що автоматизовані системи можуть неправильно інтерпретувати сарказм або сатиру, що призводить до помилкових результатів (Mazurczyk et al., 2023).

У цій роботі акцент зроблено на створенні інтегрованого підходу до виявлення дезінформації, що включає аналіз текстового контенту, моніторинг соціальних мереж та ідентифікацію маніпуляцій у відеоматеріалах. Використання сучасних моделей машинного навчання та NLP сприятиме підвищенню ефективності аналізу, забезпечуючи стійкість інформаційного простору до загроз дезінформації.

Постановка проблеми. У сучасному інформаційному просторі, де кожна секунда має вирішальне значення, дезінформація стала серйозною загрозою для соціальної стабільності, політичних процесів і довіри до медіа. Поширення фейкових новин, маніпулятивного контенту та візуальних фейків (наприклад, deepfake) створює значні ризики для громадського порядку та безпеки.

Використання технологій штучного інтелекту, зокрема машинного навчання та обробки природної мови, стало ефективним підходом до виявлення дезінформації. Однак виникає низка проблем, які ускладнюють реалізацію цих рішень. Серед основних викликів:

- Нестача якісних навчальних даних для розпізнавання маніпуляцій;
- Висока швидкість створення нових технологій для дезінформації, таких як deepfake;
- Необхідність аналізу великих обсягів даних у реальному часі.

Особливу увагу привертає емоційний вплив дезінформації. Дослідження [12] показують, що фейкові новини часто створюються з метою викликати страх, ненависть або інші сильні емоції, які роблять аудиторію більш вразливою до маніпуляцій. Існуючі підходи, такі як використання згорткових нейронних мереж (CNN), аналіз настроїв та ідентифікація емоційних патернів, є ефективними, але їхня реалізація вимагає подальшого вдосконалення.

Проблема полягає у розробці інноваційних методів виявлення дезінформації, які б забезпечували високу точність та ефективність навіть за умов обмеженості ресурсів і швидких змін технологій.

Аналіз останніх досліджень і публікацій. Застосування технологій штучного інтелекту для виявлення та протидії дезінформації активно вивчається в останні роки. Зокрема, дослідники приділяють увагу використанню алгоритмів обробки природної мови (NLP), машинного навчання та комп'ютерного зору. Одним із ключових напрямків є аналіз емоційного впливу дезінформаційного контенту, оскільки фейкові новини часто спрямовані на викликання сильних емоцій, таких як страх, ненависть або гнів. Це підтверджено дослідженнями [5], які вказують, що такі емоції підвищують довіру до неправдивого контенту.

Наукові публікації активно досліджують [6] автоматизацію виявлення дезінформації, зокрема застосування згорткових нейронних мереж (CNN) для класифікації текстів та аналізу візуального контенту. Дослідження демонструють [15], що CNN ефективно виявляють фейкові новини через аналіз структурних особливостей тексту та візуальних маніпуляцій. Інтеграція подвійних емоційних ознак із традиційним аналізом настроїв значно підвищує точність виявлення маніпулятивних текстів. Крім того, графові нейронні мережі (GNN) показують перспективи у виявленні координованих дезінформаційних кампаній у соціальних мережах, покращуючи аналіз структурованих взаємодій у цифровому просторі.

Особливий інтерес викликають підходи до автоматичного аналізу візуального контенту. Технології комп'ютерного зору, наприклад, системи для виявлення deepfake-зображень і відео, використовуються для розпізнавання маніпуляцій у пікселях і метаданих. Дослідження [14] підтверджує ефективність цих технологій у реальних умовах.

Попри досягнення, дослідники зазначають проблеми, зокрема нестачу якісних навчальних даних та труднощі аналізу багатомовного контенту. В роботах [11, 13] запропоновано створення універсальних корпусів даних для навчання моделей, що дозволяють адаптувати рішення до специфіки певних регіонів або мов.

Мета. Метою статті є дослідження сучасних методів виявлення та боротьби з дезінформацією за допомогою технологій штучного інтелекту, зокрема машинного навчання, обробки природної мови та комп'ютерного зору. Для досягнення цієї мети необхідно вирішити наступні завдання:

1. Провести аналіз останніх досліджень та публікацій щодо застосування штучного інтелекту для виявлення дезінформації.
2. Дослідити ефективність алгоритмів машинного навчання, таких як згорткові нейронні мережі (CNN), та їх роль у розпізнаванні фейкових новин.
3. Вивчити підходи до аналізу емоційного впливу текстів, зображень і відеоконтенту як одного з ключових чинників розпізнавання маніпуляцій.
4. Оцінити вплив якості та обсягу навчальних даних на точність методів виявлення дезінформації.
5. Запропонувати перспективні напрями вдосконалення технологій штучного інтелекту для підвищення ефективності боротьби з дезінформацією.

Результати дослідження сприятимуть кращому розумінню ролі штучного інтелекту у забезпеченні інформаційної безпеки та підвищенні стійкості суспільства до дезінформаційних загроз.

Виклад основного матеріалу. Дезінформація, або свідоме поширення неправдивої інформації, є однією з головних загроз сучасного цифрового простору. Завдяки доступності технологій та стрімкому розвитку соціальних мереж, фейкові новини стали невід'ємною частиною інформаційного середовища. Їхня здатність впливати на думки мільйонів людей призводить до серйозних соціальних, політичних і економічних наслідків.

Такі платформи, як Facebook, Twitter та TikTok, дозволяють користувачам швидко ділитися інформацією, часто без її попередньої перевірки. Це створює ідеальне середовище для поширення маніпулятивного контенту, зокрема таких типів дезінформації:

- Фейкові новини: Статті або повідомлення, що свідомо викривляють факти.
- Візуальні фейки: Зображення або відео, змінені за допомогою технологій, таких як deepfake.
- Маніпулятивні повідомлення: Контент, створений для викликання сильних емоцій, таких як страх, ненависть чи гнів.

Ці види дезінформації часто поширюються через бот-мережі або цілеспрямовані кампанії, спрямовані на підрив довіри до офіційних джерел інформації.

Сучасні виклики у боротьбі з дезінформацією вимагають застосування інноваційних технологій, таких як штучний інтелект. Завдяки розвитку алгоритмів машинного навчання, обробки природної мови (NLP) та комп'ютерного зору, з'являються нові можливості для автоматичного виявлення дезінформації. Наприклад, аналіз емоційного забарвлення текстів дозволяє ідентифікувати маніпулятивні елементи, тоді як технології комп'ютерного зору успішно застосовуються для розпізнавання deepfake-зображень та відео.

Таким чином, боротьба з дезінформацією є ключовою складовою забезпечення інформаційної безпеки. Її ефективне вирішення вимагає комплексного підходу, що включає використання сучасних технологій, медіаосвіти та міжнародної співпраці.

Інструменти та алгоритми боротьби з дезінформацією. Одним із ключових інструментів у боротьбі з дезінформацією є автоматизовані системи перевірки фактів. Наприклад, ClaimBuster [7] використовує алгоритми обробки природної мови (NLP) для ідентифікації тверджень у текстах, які потребують перевірки. Ця система може працювати в реальному часі, аналізуючи публікації політиків чи новинних ресурсів.

Ще одним важливим напрямком є аналіз візуального контенту. Технології комп'ютерного зору дозволяють виявляти маніпуляції у зображеннях та відео. Наприклад, **DeepTrace** [3] застосовує алгоритми згорткових нейронних мереж (CNN) для розпізнавання deepfake, аналізуючи такі характеристики, як текстура шкіри, синхронізація рухів губ із аудіо, а також метадані файлів. Схожим чином працюють інструменти на кшталт **Amped Authenticate** [1], які зосереджені на перевірці метаданих та ідентифікації підробок у візуальному контенті.

ISSN 2786-5460 (Print), ISSN 2786-5479 (Online)

Соціальні мережі, такі як TikTok, Instagram чи LinkedIn, є ключовими платформами для поширення інформації, включно з дезінформацією. Інструменти, як-от Hoaxy [8], аналізують маршрути поширення контенту, відображаючи мережі користувачів, які діляться маніпулятивною інформацією. Платформи на кшталт Hootsuite чи Brandwatch дозволяють відстежувати та аналізувати вплив контенту, включно з дезінформацією, використовуючи сучасні алгоритми обробки природної мови (NLP).

Комплексний підхід також включає мультимодальні системи, такі як DeepMind Gopher [16], які поєднують можливості великих мовних моделей та алгоритмів машинного навчання для виявлення, аналізу й протидії дезінформації. Це дозволяє ефективно працювати з текстовим, візуальним і відеоконтентом одночасно, використовуючи передові методи штучного інтелекту.

Основні виклики в боротьбі з дезінформацією

Незважаючи на значний прогрес у розробці технологій для виявлення та протидії дезінформації, існує низка викликів, які ускладнюють досягнення високої ефективності в цій сфері. Ці виклики охоплюють як технічні, так і етичні аспекти, що вимагають комплексного підходу до їх вирішення.

Недостатність якісних навчальних даних

Алгоритми машинного навчання та обробки природної мови значною мірою залежать від доступності якісних і репрезентативних навчальних даних. Проте створення таких корпусів є складним завданням через:

- Відсутність доступу до перевірених джерел, особливо в регіонах із обмеженою свободою слова.
- Проблему багатоаспектності та багатомовності дезінформації, що вимагає створення моделей, адаптованих до різних культурних контекстів.
- Нестачу структурованих наборів даних для навчання алгоритмів аналізу візуального контенту, таких як deepfake-зображення.

Розвиток технологій створення дезінформації

Зловмисники активно використовують новітні технології для створення складного маніпулятивного контенту. Наприклад, генеративні моделі, такі як GPT, дозволяють створювати правдоподібні тексти, які важко відрізнити від достовірних. Технології deepfake, з іншого боку, ускладнюють виявлення змінених зображень та відео через їхню високу якість. Цей постійний розвиток технологій створює «гонку озброєнь», де інструменти для боротьби з дезінформацією повинні постійно вдосконалюватися.

Етичні аспекти та помилкові маркування

Автоматизовані системи, такі як фактчекінгові платформи, ризикують помилково маркувати правдиву інформацію як дезінформацію. Це може мати серйозні наслідки, зокрема:

- Підрив довіри до систем боротьби з дезінформацією.
- Викликання соціальної напруги через спотворення достовірних даних. Крім того, виникають питання конфіденційності даних, оскільки алгоритми часто вимагають доступу до великої кількості особистої або чутливої інформації для навчання моделей.

Швидкість поширення дезінформації

Дезінформація поширюється значно швидше, ніж інструменти її виявлення встигають зреагувати. Соціальні мережі надають можливість охоплення широкої аудиторії у дуже короткі строки. Наприклад, фейкова новина може досягти мільйонів користувачів упродовж кількох годин. Це створює значний виклик для систем, які працюють у режимі реального часу.

Проблеми міжнародної координації

Дезінформація часто має глобальний характер, але боротьба з нею ускладнюється через різні підходи до регулювання та обмеження інформації в різних країнах. Брак координації між державами та міжнародними організаціями обмежує ефективність глобальних заходів протидії.

Запропонований інтегрований підхід

Для ефективної боротьби з дезінформацією пропонується інтегрований підхід, який охоплює аналіз текстового контенту, моніторинг соціальних мереж та ідентифікацію маніпуляцій у візуальному і відеоконтенті. Основні компоненти підходу включають:

Моніторинг соціальних мереж:

- Застосування NLP-алгоритмів для автоматичного виявлення ключових слів і фраз, які сигналізують про можливе поширення фейкових новин.
- Використання графових моделей для виявлення бот-мереж та координованих кампаній із поширення дезінформації.

Аналіз візуального та відеоконтенту:

- Технології комп'ютерного зору дозволяють виявляти deepfake-зображення та відео шляхом аналізу піксельних патернів, метаданих та синхронізації аудіо з відео.

● Алгоритми згорткових нейронних мереж (CNN) ефективно розпізнають візуальні маніпуляції, зокрема змінену текстуру зображень чи неприродні рухи в відео.

Інтеграція мультимодальних даних:

● Поєднання аналізу тексту, зображень та поведінкових патернів у єдину систему дозволяє не лише ідентифікувати дезінформацію, але й оцінити її потенційний вплив на аудиторію.

● Використання мультимодальних нейронних мереж для одночасного аналізу кількох типів контенту.

Постійний моніторинг та вдосконалення:

● Впровадження механізмів автоматичного збору зворотного зв'язку від користувачів для вдосконалення моделей.

● Постійне оновлення алгоритмів для адаптації до нових форм дезінформації.

Запропонований підхід дозволяє створити комплексну систему, яка ефективно протидіє поширенню дезінформації, підвищуючи стійкість інформаційного простору до маніпуляцій. Інтеграція різних технологій штучного інтелекту сприятиме зменшенню впливу фейкових новин і забезпеченню інформаційної безпеки.

Перспективи розвитку

Розвиток технологій штучного інтелекту відкриває нові горизонти у боротьбі з дезінформацією, дозволяючи створювати інноваційні підходи та системи для її виявлення й запобігання. Проте зростання складності маніпулятивного контенту потребує не лише вдосконалення існуючих методів, а й впровадження стратегій, які враховують майбутні виклики.

1. Інтеграція мультимодальних систем

Майбутнє боротьби з дезінформацією полягає у створенні інтегрованих систем, які аналізують текстовий, візуальний та поведінковий контент одночасно. Поєднання можливостей обробки природної мови (NLP), комп'ютерного зору та аналізу поведінки користувачів дозволить створити більш комплексні рішення, які будуть адаптовані до нових форм маніпуляцій.

Приклад: Інтеграція даних із соціальних мереж, фактчекінгових платформ і візуального контенту в єдину систему моніторингу значно підвищить ефективність виявлення дезінформації на ранніх етапах її поширення.

2. Розвиток алгоритмів генерації контрнаративів

Замість блокування дезінформації ефективним підходом може стати створення контрнаративів, які пояснюють аудиторії неправдивий характер маніпуляцій. Це дозволить не лише запобігати поширенню фейків, а й підвищити рівень медіаграмотності суспільства.

Перспектива: Використання генеративних моделей, таких як GPT, для створення текстів, які пояснюють аудиторії небезпеку дезінформації та надають достовірну інформацію.

3. Використання блокчейн-технологій

Блокчейн може стати ефективним інструментом у забезпеченні прозорості джерел інформації. Технологія дозволяє зберігати історію змін у контенті, підтверджувати його автентичність і надавати користувачам можливість перевіряти достовірність інформації.

Приклад: Використання блокчейну для створення системи цифрового підтвердження автентичності новинних статей, де кожна зміна буде зафіксована у розподіленому реєстрі.

4. Залучення користувачів до процесу виявлення дезінформації

Системи, що дозволяють користувачам маркувати підозрілий контент, можуть стати важливим інструментом у боротьбі з фейками. Такі підходи не лише залучають аудиторію, а й створюють додаткове джерело даних для автоматизованих систем аналізу.

Ініціатива: Платформи, що стимулюють користувачів до маркування контенту, можуть бути інтегровані із системами штучного інтелекту для подальшого навчання моделей.

5. Підвищення рівня медіаграмотності

Освітні програми, спрямовані на розвиток критичного мислення та аналізу інформації, залишаються однією з найбільш перспективних стратегій. Впровадження курсів із цифрової грамотності у шкільні та університетські програми дозволить зменшити вплив дезінформації на суспільство.

Приклад: Програми для молоді, які вчать розпізнавати фейкові новини за допомогою аналізу джерел, риторики та емоційного забарвлення контенту.

6. Міжнародна співпраця

Дезінформація має глобальний характер, що потребує координації зусиль між країнами. Створення міжнародних платформ для обміну даними та технологіями дозволить швидше реагувати на нові виклики.

ISSN 2786-5460 (Print), ISSN 2786-5479 (Online)

Приклад: Ініціативи, подібні до EUvsDisinfo, спрямовані на об'єднання ресурсів і спільне розроблення технологій для виявлення дезінформації.

Висновки. Сучасні технології штучного інтелекту відкривають значні можливості для виявлення та протидії дезінформації. Алгоритми машинного навчання, обробки природної мови та комп'ютерного зору дозволяють автоматизувати аналіз текстового, візуального та поведінкового контенту. Інтеграція цих методів може значно підвищити ефективність боротьби з маніпулятивною інформацією.

Однак існують численні виклики, які ускладнюють впровадження таких систем. Серед них – нестача якісних навчальних даних, швидкість поширення дезінформації та розвиток технологій, які її створюють. Важливими залишаються й моральні аспекти, зокрема забезпечення прозорості та коректності роботи автоматизованих систем.

У перспективі ефективна боротьба з дезінформацією вимагає поєднання технічних рішень із освітніми ініціативами. Підвищення рівня медіаграмотності серед населення, впровадження міжнародної співпраці та розвиток інноваційних технологій можуть стати ключовими факторами у зменшенні впливу фейкових новин.

Подальший розвиток цієї сфери може включати створення мультимодальних систем, що поєднують аналіз текстів, зображень і поведінкових даних, а також використання технологій блокчейну для підтвердження автентичності інформації. Ці підходи здатні сприяти формуванню стійкого інформаційного простору, що відповідає сучасним викликам.

Список використаних джерел:

1. Фільтруйте недоброчесних блогерів, дипфейки, маніпуляції – користуйтесь інтернетом з розумом. *Хмарочос*. 2023. 4 грудня. URL: https://hmarochos.kiev.ua/2023/12/04/filtrujte-nedobrochesnyh-blogeriv-dipfejky-manipulyacziyi-korystujtes-internetom-z-rozumom/ (дата звернення: 12.06.2025).

2. Anggrainingsih R., Hassan G. M., Datta A. Evaluating BERT-based Pre-training Language Models for Detecting Misinformation. 2022. URL: https://arxiv.org/abs/2209.13594 (дата звернення: 12.06.2025).

3. Battiato S. DeepFake Detection by Analyzing Convolutional Traces. *Journal of Imaging*. 2020. Vol. 6. No. 7. P. 1–14.

4. Dolhansky B., Bitton J., Pflaum B., Lu J., Howes R., Wang M., Canton Ferrer C. The DeepFake Detection Challenge. *arXiv preprint*. 2020. arXiv:2006.07397.

5. Ge X., Zhang M., Wang X. A., Liu J., Wei B. Emotion-Driven Interpretable Fake News Detection. *International Journal of Data Warehousing and Mining*. 2022. Vol. 18. No. 3. P. 1–15.

6. Goldani M. H., Momtazi S., Safabakhsh R. Detecting Fake News with Capsule Neural Networks. *Computational Intelligence and Neuroscience*. 2020. Vol. 2020. Article ID 2165391.

7. Hassan N., Arslan F., Li C., Tremayne M. Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster. *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM 2017)*, Singapore, 6–10 Nov. 2017. P. 2179–2182.

8. Hoaxy – офіційний сайт додатку. URL: https://hoaxy.osome.iu.edu/ (дата звернення: 12.06.2025).

9. Mazurczyk W., Lee D., Vlachos A. Disinformation 2.0 in the Age of AI: A Cybersecurity Perspective. *IEEE Security & Privacy*. 2023. Vol. 21. No. 3. P. 40–49.

10. Meta AI. Deepfake Detection Challenge Dataset. 2020. URL: https://ai.facebook.com/blog/deepfake-detection-challenge/ (дата звернення: 12.06.2025).

11. Mohtaj Salar, Ata Nizamoglu, Premtim Sahitaj, Vera Schmitt, Charlott Jakob, Sebastian Möller (2024). NewsPolyML: Multi-lingual European News Fake Assessment Dataset

12. Rae J., Irving G., Weidinger L. Language Modelling at Scale: Gopher, Ethical Considerations, and Retrieval. *DeepMind Blog*. 2021. URL: https://deepmind.google/discover/blog/language-modelling-at-scale-gopher-ethical-considerations-and-retrieval/ (дата звернення: 12.06.2025).

13. Shahi G. K., Nandini D. FakeCovid – A Multilingual Cross-domain Fact Check News Dataset for COVID-19. *arXiv preprint*. 2020. arXiv:2011.11459.

14. Wang T., Liao X., Chow K. P., Lin X., Wang Y. Deepfake Detection: A Comprehensive Survey from the Reliability Perspective. *Journal of Cybersecurity*. 2024. Vol. 12. No. 1. P. 15–36.

16. Yang Y., Zheng L., Zhang J., Cui Q., Li Z., Yu P. S. TI-CNN: Convolutional Neural Networks for Fake News Detection. *arXiv preprint*. 2018. arXiv:1806.00749.

15. Zhang X., Cao J., Li X., Sheng Q., Zhong L., Shu K. Mining Dual Emotion for Fake News Detection. *Proceedings of the Web Conference*. 2021. P. 3465–3476.

Дата надходження статті: 12.06.2025
Дата прийняття статті: 17.06.2025
Опубліковано: 23.09.2025