УДК 004.8+316.472.4]:323.266 DOI: 10.31673/2412-4338.2024.044748

Легомінова Світлана Володимирівна
Державний університет інформаційно-комунікаційних технологій, Київ
ORCID ID: 0000-00002-4433-5123

Тищенко Віталій Сергійович
Державний університет інформаційно-комунікаційних технологій, Київ
ORCID ID: 0000-0003-3849-6243

Недодай Михайло Геннадійович
Державний університет інформаційно-комунікаційних технологій, Київ
ORCID ID: 0009-0000-0876-9971

Дьячук Олександр Станіславович
Державний університет інформаційно-комунікаційних технологій, Київ
ORCID ID: 0009-0006-5585-6393

Капелюшна Тетяна Вікторівна
Державний університет інформаційно-комунікаційних технологій, Київ
ORCID 0000-0001-7490-6751

ШТУЧНИЙ ІНТЕЛЕКТ ТА СОЦІАЛЬНІ МЕРЕЖІ: ПІДХОДИ ДО ВИЯВЛЕННЯ ФЕЙКОВОЇ ІНФОРМАЦІЇ

Анотація. *У статті розглядаються сучасні підходи до виявлення фейкової інформації в соціальних мережах із використанням штучного інтелекту (ШІ) та машинного навчання (МН). Зростання популярності соціальних мереж супроводжується глобальною проблемою дезінформації, яка має серйозні наслідки для суспільства, економіки, політики та навіть здоров'я населення, що стало особливо помітним під час пандемії COVID-19. Основна увага приділяється методам виявлення фейкових новин, які включають використання нейронних мереж, психологічних характеристик користувачів та аналіз тексту. Окремо досліджуються підходи до аналізу візуального контенту, зокрема зображень та відео, з метою визначення їх достовірності.*

Стаття також аналізує роль соціальних ботів у поширенні дезінформації, зокрема їхню здатність впливати на громадську думку через маніпуляції, а також інструменти, що допомагають їх ідентифікувати, наприклад графові моделі та евристичні методи. У роботі надається характеристика процесів збору та обробки даних, таких як API соціальних платформ, веб-скрапінг, моніторинг активності користувачів та їх взаємодій. Окрему увагу приділено попередній обробці даних, включно з етапами очищення, нормалізації, токенізації, лематизації та анотації, які суттєво впливають на якість результатів роботи алгоритмів виявлення.

Розглянуто також проблеми масштабованості та продуктивності таких систем в умовах великого обсягу даних, а також виклики, пов'язані із забезпеченням приватності користувачів. Крім того, підкреслюється необхідність адаптації алгоритмів до нових патернів дезінформації, що еволюціонують у відповідь на технологічний прогрес, і важливість міждисциплінарного підходу, що поєднує досягнення ШІ, когнітивних наук, лінгвістики та соціології.

Ключові слова: *штучний інтелект, соціальні мережі, фейкова інформація, машинне навчання, дезінформація, нейронні мережі, виявлення новин, обробка даних.*

Svitlana Lehominova
State University of information communication technologies, Kyiv
ORCID ID: 0000-0002-4433-5123

© Легомінова С. В., Тищенко В. С., Недодай М. Г., Дьячук О. С., Капелюшна Т.В. 2024

Vitalii Tyshchenko
State University of information and communication technologies, Kyiv
ORCID ID: 0000-0003-3849-6243

Mykhailo Nedodai
State University of information and communication technologies, Kyiv
ORCID ID: 0009-0000-0876-9971

Oleksandr Diachuk
State University of information and communication technologies, Kyiv
ORCID ID: 0009-0006-5585-6393

Tetiana Kapeliushna
State University of information and communication technologies, Kyiv
ORCID 0000-0001-7490-6751

ARTIFICIAL INTELLIGENCE AND SOCIAL NETWORKS: APPROACHES TO DETECTING FAKE INFORMATION

Abstract. *The article examines modern approaches to detecting fake information on social networks using artificial intelligence (AI) and machine learning (ML). The growing popularity of social networks is accompanied by a global disinformation problem, which has severe implications for society, the economy, politics, and even public health, as was especially evident during the COVID-19 pandemic. Particular attention is given to methods of detecting fake news, including the use of neural networks, user psychological profiling, and text analysis. Additionally, the study explores approaches to analyzing visual content, such as images and videos, to assess their authenticity.*

The article also analyzes the role of social bots in disseminating disinformation, including their ability to influence public opinion through manipulation, as well as tools for identifying them, such as graph models and heuristic methods. The study characterizes data collection and processing techniques, including social platform APIs, web scraping, user activity monitoring, and interaction analysis. Special attention is devoted to data preprocessing steps, including cleaning, normalization, tokenization, lemmatization, and annotation, which significantly impact the quality of algorithm performance in detecting fake information.

The challenges of scalability and system efficiency in the context of large data volumes are discussed, along with issues related to ensuring user privacy. Furthermore, the article highlights the necessity of adapting algorithms to new patterns of disinformation, which evolve in response to technological advancements, and the importance of an interdisciplinary approach that combines achievements in AI, cognitive science, linguistics, and sociology.

Keywords: *artificial intelligence, social media, fake news, machine learning, disinformation, neural networks, news detection, data processing.*

1. Вступ.

Штучний інтелект (ШІ) та соціальні мережі стали невід'ємною частиною сучасного суспільства, значно впливаючи на спосіб спілкування та обміну інформацією. Однак, разом із зростанням популярності соціальних мереж, з'явилася проблема поширення фейкової інформації, яка може мати суттєві наслідки для суспільства, економіки та політики [1]. Фейкові новини, дезінформація та маніпуляції стали викликом для дослідників та фахівців з кібербезпеки, які намагаються розробити ефективні методи їх виявлення та протидії.

Одним із основних підходів щодо виявлення фейкової інформації є використання методів машинного навчання та глибокого навчання. Результати досліджень свідчать, що найбільш ефективними є комбінації класичних методів машинного навчання, координованих нейронними мережами. Крім того, важливу роль відіграють психологічні та лінгвістичні характеристики користувачів, які можуть бути використані для прогнозування схильності до поширення фейкової інформації.

Соціальні боти, які автоматично або напівавтоматично імітують поведінку людей також є значним джерелом фейкової інформації. Для боротьби з ними розробляються спеціальні інструменти, такі як Botometer, які допомагають користувачам ідентифікувати та блокувати боти. Однак, постійна еволюція технологій вимагає регулярного оновлення алгоритмів та навчальних даних для забезпечення їх ефективності [2].

Таким чином, дослідження в області виявлення фейкової інформації в соціальних мережах є багатостороннім процесом, що включає в себе різні підходи, методи та міждисциплінарну співпрацю, а також поєднує знання з соціальних наук, психології, когнітивних наук, нейронауки, ШІ та комп'ютерних наук для розробки дієвих комплексних рішень щодо управління інформаційною безпекою в організаціях [3].

2. Постановка проблеми.

Проблема виявлення фейкової інформації в соціальних мережах актуалізується через зростання обсягу дезінформації, яка може спричиняти соціальні, політичні та економічні наслідки. Штучний інтелект (ШІ) відіграє ключову роль у розробці інструментів і методів для автоматизованого аналізу контенту з метою виявлення неправдивих повідомлень. Застосування алгоритмів машинного навчання, обробки природної мови та аналізу великих даних дозволяє не лише ідентифікувати фейки, але й проводити моніторинг їх поширення, визначати вплив на користувачів. Попри розвиток технологій, ефективність таких систем залежить від здатності адаптування до нових способів створення дезінформації, що постійно вдосконалюються та стають дедалі складнішими для ідентифікації.

3. Аналіз останніх досліджень і публікацій.

Із розвитком інтернету та соціальних мереж проблема поширення фейкової інформації наростає, набуває серйозних наслідків. Соціальні мережі, такі як Facebook, X та інші, стали основними платформами для обміну новинами та інформацією, що значно полегшує поширення фейкових новин [4]. Фейкові новини можуть мати руйнівні наслідки для суспільства, включаючи дезінформацію громадськості, вплив на політичні процеси та поширення соціальної напруги.

Однією з основних проблем є те, що фейкові новини часто виглядають дуже схожими на правдиві, що ускладнює їх виявлення традиційними методами. Крім того, фейкові новини часто використовують емоційно забарвлену мову та сенсаційні заголовки, щоб привернути увагу користувачів, що, в свою чергу, призводить до швидкого викривлення інформації та негативно позначається на суспільно-економічних процесах..

Для вирішення цієї проблеми дослідники активно розробляють методи виявлення фейкових новин за допомогою штучного інтелекту (ШІ) та машинного навчання (МН) [5]. Сучасні підходи включають використання глибоких нейронних мереж, таких як: бінаправлені довгострокові короткочасні пам'яті (MHS-BiLSTM) [5], графові згорткові мережі (GCN) та ансамблеві моделі глибокого навчання [5]. Ці методи дозволяють автоматично аналізувати великі обсяги даних та виявляти фейкові новини з високою точністю.

У дослідженні [6] автори розглядають можливість виявлення фейкових новин у соціальних мережах шляхом поєднання алгоритмів машинного навчання з методами інженерії. Під час аналізу дослідники-практики використовували набори даних із твітів, пов'язаних із президентськими виборами у США 2016 року, що дозволило їм з високою точністю ідентифікувати фейкові новини.

Однак, незважаючи на значні досягнення, існує ряд викликів, які потребують подальшого дослідження. Наприклад, більшість моделей виявлення фейкових новин залежать від конкретних наборів даних, що ускладнює їх застосування до нових подій. Крім того, існує потреба в розробці інтерпретованих моделей, які можуть пояснити чому певна новина була класифікована як фейкова.

Таким чином, проблема виявлення фейкової інформації в соціальних мережах є складною та багатогранною, що вимагає подальших досліджень та розробки нових методів для підвищення точності та інтерпретованості моделей виявлення фейкових новин.

4. Мета і задачі дослідження

Мета даного дослідження полягає в аналізі існуючих методів і технологій штучного інтелекту, які використовуються для виявлення фейкової інформації в соціальних мережах та розробці рекомендацій щодо їх вдосконалення й ефективного застосування. Це включає оцінку методів обробки природної мови, систем аналізу даних і стратегій виявлення дезінформації, з метою розробки практичних рекомендацій, що підвищують точність та адаптивність систем у боротьбі з поширенням неправдивих повідомлень в інтернет-середовищі.

5. Виклад основного матеріалу.

Методи збору та підготовки даних є фундаментальним етапом у процесі виявлення фейкової інформації в соціальних мережах за допомогою штучного інтелекту. Сучасні підходи базуються на комплексному використанні API соціальних платформ, спеціалізованих інструментів веб-скрапінгу та систем моніторингу соціальних медіа. Особливу увагу приділяють автоматизованим методам збору даних, які дозволяють обробляти великі обсяги інформації в режимі реального часу, враховуючи різноманітні формати контенту: текстові повідомлення, зображення, відео та аудіоматеріали.

Процес підготовки даних включає декілька критично важливих етапів:

- очищення даних від нерелевантної інформації;
- нормалізацію текстового контенту;
- токенізацію;
- створення структурованих наборів даних для подальшого аналізу.

Важливим аспектом є також анотація та розмітка даних, які можуть здійснюватися як експертами вручну, так і за допомогою напівавтоматичних методів із використанням попередньо навчених моделей машинного навчання. Якість розмітки даних безпосередньо впливає на ефективність роботи алгоритмів виявлення фейкової інформації. Для забезпечення високої якості аналізу застосовуються методи збагачення даних додатковими метриками, такими як: часові мітки, геолокаційні дані, інформація про взаємодію користувачів із контентом та мережеві зв'язки між джерелами інформації. Особлива увага приділяється валідації та верифікації зібраних даних, що включає перевірку їх достовірності, повноти та актуальності.

Основним джерелом даних для аналізу та виявлення фейкової інформації є соціальні мережі та платформи, які надають доступ до контенту через спеціалізовані програмні інтерфейси (API). Twitter API забезпечує доступ до публічних твітів, користувацьких профілів та метаданих взаємодій. Facebook Graph API дозволяє отримувати дані з публічних сторінок, груп та постів. Telegram Bot API надає можливість моніторити повідомлення у публічних каналах та групах. YouTube Data API забезпечує доступ до відеоконтенту, коментарів та метаданих каналів. Instagram Basic Display API дозволяє отримувати дані про публічні пости та профілі користувачів.

Попередня обробка даних є критично важливим етапом для підготовки зібраної інформації до подальшого аналізу, даний процес включає кілька ключових кроків, а саме:

- здійснюється очищення даних від нерелевантної інформації, включаючи видалення HTML-тегів, спеціальних символів та виправлення орфографічних помилок.
- відбувається нормалізація текстового контенту, яка передбачає приведення тексту до нижнього регістру, лематизацію слів та видалення стоп-слів.

– токенізація - розбиття тексту на окремі елементи (токени), що можуть бути словами, фразами або реченнями.

Таблиця 1

Рекомендації щодо обробки даних соціальних мереж для виявлення фейкової інформації з використанням ШІ

Етап	Метод	Зміст
1. Основні джерела даних із соціальних платформ	а) API соціальних мереж	Twitter API (тепер X API) Facebook Graph API Instagram Basic Display API Telegram Bot API YouTube Data API
	б) Спеціалізовані інструменти збору даних	Web-скрапери Системи моніторингу соцмедіа
2. Попередня обробка даних	а) Очищення даних	Видалення HTML-тегів Видалення спеціальних символів Видалення дублікатів
	б) Нормалізація	Приведення тексту до нижнього регістру Лематизація Видалення стоп-слів
	в) Токенізація	Розбиття тексту на слова Створення словника унікальних токенів
3. Анотація та розмітка даних	а) Типи розмітки	Бінарна (фейк/не фейк) Розмітка емоційного забарвлення
	б) Методи розмітки	Ручна експертна розмітка Використання існуючих розмічених датасетів
4. Машинне навчання	Навчання моделей	Використання розмічених даних для тренування алгоритмів
	Класифікація контенту	Автоматичне визначення категорій контенту та ймовірності його недостовірності
5. Верифікація результатів	Перехресна валідація	Оцінка якості моделей шляхом тестування на різних підмножинах даних
	Експертна оцінка	Перевірка результатів роботи системи фахівцями з фактчекінгу
6. Візуалізація та звітність	Відображення графів	Візуалізація зв'язків та часової динаміки
7. Моніторинг та оновлення	Адаптація моделей до нових типів фейків	Перенавчання алгоритмів з урахуванням нових патернів дезінформації

Для підвищення якості аналізу застосовуються методи стандартизації формату даних, що включає уніфікацію дат, чисел та одиниць виміру. Важливим аспектом є також обробка багатомовного контенту, яка може включати переклад або специфічну обробку для кожної мови. Після завершення попередньої обробки дані структуруються у форматі, придатному для застосування алгоритмів машинного навчання та інших методів аналізу. Якість попередньої обробки даних безпосередньо впливає на ефективність виявлення фейкової інформації та точність роботи аналітичних систем.

Наступний етап передбачає забезпечення високої ефективності систем виявлення фейкових новин методом машинного навчання. Спочатку проводиться навчання моделей із

використанням розмічених даних, що дозволяє алгоритмам вивчати патерни, властиві як достовірному, так і недостовірному контенту. Одним із основних завдань у цій сфері є класифікація контенту — автоматичне визначення його категорій та ймовірності недостовірності, що забезпечує можливість оперативно і надійно виявляти потенційно фейкову інформацію в масштабних масивах даних.

Верифікація результатів є важливою складовою процесу, що забезпечує надійність моделей і полягає у перехресній валідації, що передбачає оцінку якості алгоритмів шляхом тестування на різних підмножинах даних. Додатково проводиться експертна оцінка, в процесі якої фахівці з фактчекінгу перевіряють результати роботи системи, що дозволяє підтвердити або спростувати їх точність.

Ефективність роботи таких систем значно підвищується завдяки візуалізації та звітності. Відображення мережевих графів дозволяє візуалізувати зв'язки між користувачами, а також часову динаміку поширення новин, що допомагає аналітикам краще розуміти структуру та характеристики розповсюдження дезінформації.

Останнім етапом є моніторинг та оновлення системи. Алгоритми повинні адаптуватися до нових типів фейкових новин, що вимагає перенавчання моделей з урахуванням нових патернів дезінформації. Регулярне оновлення забезпечує збереження високої точності та актуальності системи, що є критичним фактором у боротьбі з поширенням фейкової інформації в динамічному середовищі соціальних мереж.

Для проведення ефективного аналізу даних у соціальних мережах та виявлення фейкової інформації важливо дотримуватись ключових етапів обробки. У таблиці 1 представлено основні етапи даного процесу та методи, що застосовуються для обробки даних соціальних мереж для виявлення фейкової інформації з використанням ШІ.

6. Висновки та перспективи подальших досліджень.

Проблема виявлення фейкової інформації в соціальних мережах залишається складним і багатогранним завданням, що вимагає інтегрованого підходу та постійного вдосконалення аналітичних методів. У цій статті детально розглянуто основні етапи збору, підготовки та обробки даних, зокрема застосування машинного навчання, верифікацію результатів і безперервний моніторинг систем. Наголошено, що ефективність моделей виявлення фейкових новин значною мірою залежить від якості розмітки даних, регулярного оновлення алгоритмів та адаптації до нових патернів дезінформації.

Незважаючи на досягнуті успіхи, існують численні виклики, які потребують подальшого дослідження. Серед ключових напрямів — розробка більш інтерпретованих моделей, здатних пояснювати свої висновки та узагальнення, удосконалення алгоритмів для роботи в багатомовному середовищі та підвищення їхньої гнучкості у відповідь на зміни форматів подання інформації. Особливу увагу слід приділити аналізу контекстуальних маніпуляцій і гібридних форм дезінформації, які поєднують правдиві та вигадані елементи.

Подальші дослідження мають бути спрямовані на розробку моделей, що враховують рекомендації, викладені в статті, зокрема, інтеграцію методів штучного інтелекту з підходами когнітивних наук для глибшого розуміння механізмів поширення дезінформації. Крім того, перспективними напрямами вирішення проблеми дезінформації є створення інструментів для аналізу мультимедійного контенту та розробка платформ для автоматизованої співпраці з аналітиками та фактчекерами. Впровадження запропонованих рішень сприятиме підвищенню ефективності боротьби з дезінформацією та відновленню довіри до інформаційного простору.

Список використаних джерел

1. Silva, F., Vieira, R., & Garcia, A. Can machines learn to detect fake news? A survey focused on social media. Hawaii international conference on system sciences - HICSS. 2019. P. 1–8. URL: https://doi.org/10.24251/HICSS.2019.332.

2. Arming the public with artificial intelligence to counter social bots / K. Yang et al. Human behavior and emerging technologies. 2019. Vol. 1, no. 1. P. 48–61. URL: https://doi.org/10.1002/hbe2.115 (date of access: 03.11.2024).

3. The mass, fake news, and cognition security / B. Guo et al. Frontiers of computer science. 2020. Vol. 15, no. 3. URL: https://doi.org/10.1007/s11704-020-9256-0 .

4. Nistor A., Zadobrischi E. The influence of fake news on social media: analysis and verification of web content during the COVID-19 pandemic by advanced machine learning methods and natural language processing. Sustainability. 2022. Vol. 14, no. 17. URL: https://doi.org/10.3390/su141710466

5. Bio-Inspired artificial intelligence with natural language processing based on deceptive content detection in social networking / A. A.Albraikan et al. Biomimetics. 2023. Vol. 8, no. 6. P. 449. URL: https://doi.org/10.3390/biomimetics8060449.

6. Fake news detection on social media / K. Shu et al. ACM SIGKDD explorations newsletter. 2017. Vol. 19, no.1. P. 22–36. URL: https://doi.org/10.1145/3137597.3137600 .

7. Shevchenko O., Bondarchuk A., Polonevych O., Zhurakovskyi B., Korshun N. (2021) Methods of the objects identification and recognition research in the networks with the IoT concept support. Workshop on Cybersecurity Providing in Information and Telecommunication Systems, (pp. 197–209)

8. Moshenchenko, M., Zhurakovskyi, B., Poltorak, V., Bondarchuk, A., Korshun, N. (2021) Optimization Algorithms of Smart City Wireless Sensor Network Control. Workshop on Cybersecurity Providing in Information and Telecommunication Systems, (pp. 32–42)

References

1. Silva, F., Vieira, R., & Garcia, A. Can machines learn to detect fake news? A survey focused on social media. Hawaii international conference on system sciences - HICSS. 2019. P. 1–8. URL: https://doi.org/10.24251/HICSS.2019.332.

2. Arming the public with artificial intelligence to counter social bots / K. Yang et al. Human behavior and emerging technologies. 2019. Vol. 1, no. 1. P. 48–61. URL: https://doi.org/10.1002/hbe2.115 (date of access: 03.11.2024).

3. The mass, fake news, and cognition security/ B. Guo et al. Frontiers of computer science. 2020. Vol. 15, no. 3. URL: https://doi.org/10.1007/s11704-020-9256-0.

4. Nistor A., Zadobrischi E. The influence of fake news on social media: analysis and verification of web content during the COVID-19 pandemic by advanced machine learning methods and natural language processing. Sustainability. 2022. Vol. 14, no. 17. URL: https://doi.org/10.3390/su141710466

5. Bio-Inspired artificial intelligence with natural language processing based on deceptive content detection in social networking/ A. A. Albraikan et al. Biomimetics. 2023. Vol. 8, no. 6. P. 449. URL: https://doi.org/10.3390/biomimetics8060449.

6. Fake news detection on social media/ K. Shu et al. ACM SIGKDD explorations newsletter. 2017. Vol. 19, no. 1. P. 22–36. URL: https://doi.org/10.1145/3137597.3137600 .

7. Shevchenko O., Bondarchuk A., Polonevych O., Zhurakovskyi B., Korshun N. (2021) Methods of the objects identification and recognition research in the networks with the IoT concept support. Workshop on Cybersecurity Providing in Information and Telecommunication Systems, (pp. 197–209)

8. Moshenchenko, M., Zhurakovskyi, B., Poltorak, V., Bondarchuk, A., Korshun, N. (2021) Optimization Algorithms of Smart City Wireless Sensor Network Control. Workshop on Cybersecurity Providing in Information and Telecommunication Systems, (pp. 32–42)