УДК 004.89:004.934

ОГЛЯД ЗАДАЧІ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТУ ДЛЯ РОЗПІЗНАВАННЯ ДЕЗІНФОРМАЦІЇ

Харіна О.С., Головянко М.В.
e-mail: oleksandra.kharina@nure.ua, mariia.golovianko@nure.ua
Харківський національний університет радіоелектроніки, каф. ШІ
м. Харків, Україна

This work discusses the challenges posed by the abundance of information available today, which facilitates the rapid spread of misinformation. The study focuses on developing machine learning algorithms for automated detection and classification of misinformation, addressing the importance of a combined approach that includes technological, social and psychological aspects. An objective and a plan have been developed to investigate machine learning tools for misinformation detection, taking into account the limitations of existing datasets.

На сьогоднішній день ми маємо доступ до необмеженої кількості інформації, більш того, для її отримання нам не потрібно прикладати жодних зусиль, оскільки численні масмедіа та соціальні мережі автоматично підберуть велику порцію нової інформації: останні гучні новини, поради та навчальний матеріал. Через прагнення до економії власних ресурсів в цьому нескінченному потоці знань люди схильні не перевіряти актуальність і достовірність даних чи то надійність їхніх джерел. З іншого боку, створення та поширення власної інформації теж стало необмеженим і легким: кожна людина з доступом до інтернету має можливість сказати чи показати будь-що на широку авдиторію, і оскільки цей процес не є контрольованим – вся відповідальність за вірогідність цих даних покладається на автора та його совість, які не є надійними гарантами якості інформації. Тож в результаті ми маємо купу неперевірених даних і величезну кількість людей готових вірити всьому, чи майже всьому, що вони бачать в інтернеті. Така ситуація на пряму призводить до збільшення поширення дезінформації (фейків), соціально деструктивних ідей та полегшення впливу на громадськість шляхом маніпуляції фактами.

Існує декілька способів протидії дезінформації:

– медіаграмотність (навички та знання, що дозволяють ефективно і безпечно користуватися медіасервісами);

– фактчекінг (перевірка достовірності інформації);

– контроль поширення дезінформації на законодавчому рівні або через саморегуляцію медіа.

Серед перелічених інструментів варто виділити саме фактчекінг, оскільки цей засіб є частиною медіаграмотності, а також є основою для контролювання поширення фейків різними установами. Процес фактчекінгу часто може потребувати значну кількість часу, а також специфічні знання та вміння. Тож коли мова йде про великий потік

інформації, перевірка її достовірності вимагає значних ресурсних витрат. Так ми підходимо до необхідності автоматизації цього процесу за допомогою інструментів машинного навчання. Такі розробки зараз набирають обертів і вже демонструють непогані результати.

Метою цього дослідження є визначення ефективних методів інтелектуального аналізу текстового медіа-контенту для автоматизованого виявлення та класифікації різних видів дезінформації, зокрема і маніпуляцій.

План дій для досягнення поставленої цілі виглядає наступним чином:

1. Провести аналіз існуючих методів інтелектуального аналізу та обробки природної мови, з точки зору їх застосування для виявлення дезінформації.
2. Виявити та класифікувати та дослідити ключові техніки маніпуляцій, що найчастіше використовуються в медіа-контенті (наприклад, емоційна маніпуляція, фейкові новини, пропаганда, підміна понять, проста брехня тощо).
3. Змоделювати та провести ряд експериментів, націлених на визначення впливу контексту та різних комбінацій тексту і метаданих на ефективність виявлення дезінформації.
4. Розробити алгоритми та моделі машинного навчання для автоматизованого виявлення та класифікації визначених маніпуляцій.
5. Оцінити існуючі датасети. Розібрати їхні окремі приклади з логічної та фактичної точки зору. Дослідити шляхи видобутку нових даних для навчання. Розробити модель фактичної бази знань.
6. Дослідити можливість використання методів пояснювального ШІ для інтерпретації результатів виявлення маніпуляцій.
7. Створити прототип системи, що реалізує розроблені методи, та провести його експериментальну оцінку на реальних даних медіа-контенту.
8. Оцінити ефективність розроблених методів з точки зору точності виявлення маніпуляцій чи навпаки правдивої інформації, швидкості обробки даних та стійкості до різних типів вхідних даних.
9. Порівняти ефективність розпізнавання фейкових новин машинним способом та людиною.

Варто зазначити що в рамках пошуку оптимальних рішень потрібно спиратися на міждисциплінарний підхід, що поєднує технологічні, соціальні та психологічні аспекти. Якщо говорити про деякі виклики в розробці подібних алгоритмів, то слід згадати також про складність відрізнити дезінформацію від сатири, гумору або особистої думки, а це є надзвичайно важливим, оскільки хороше рішення має не тільки знаходити брехню але й не помилятися щодо правдивої інформації.

На сьогодні вже існує достатня кількість великих публічних датасетів для розпізнавання фейкових новин, більшість з них містить багато параметрів контексту вхідної інформації, що значно покращує

ефективність готового рішення. Чудовим прикладом є датасет Liar, який вперше був опублікований в 2017 році. Він містить 12800 вручну позначених коротких тверджень з різних контекстів, зібраних з PolitiFact.com за десятиліття. Разом з ним було виставлено результати тренування моделі, це була гібридна згорткова мережа (Hybrid Convolutional Neural Network, CNN). Всередині навчальних даних кожна одиниця тексту супроводжується деякою додаткової інформацію, такою як контекст, автор слів, посада автора тощо. Таким чином один зразок даних може виглядати ось так.

Текст: У 1970-х роках спалахнув свинячий грип... за часів іншого демократа, президента Джиммі Картера.

Лейбл: брехня.

Контекст: охорона здоров’я.

Автор: Мішель Бахман.

Посада автора: Конгресвумен.

Партія автора: Республіканська.

Штат: Міннесота.

З переваг цього датасету крім наявності допоміжної інформації є градаційна система лейблів: тобто інформація може бути повною брехнею, частково брехнею з певним нахилом в одну чи іншу сторону, або ж повною правдою. Проте варто зазначити, що набір даних Liar націлений більш специфічно на американський контекст, тобто не є універсальним. У 2024 році був випущений LIAR2 – вдосконалений варіант з 23 тисячами записів, розширеною структурою та покращеним розподілом даних (8:1:1 для тренування/валідації/тестування). Серед існуючих сьогодні датасетів для виявлення фейкових новин, таких як LIAR2, Official-NV, Weibo Dataset, та FakeNewsNet, існує потреба в більш конкретній класифікації дезінформації та правдивих висловів. Хоча ці датасети забезпечують певний рівень класифікації, вони можуть бути обмежені в своєму охопленні та деталізації. Тож одним з потенційних рішень є поєднання, наприклад, датасету з фейковими новинами та датасету для розпізнавання жартів, сарказму.

Список використаних джерел:

1. Савченко О. В. Мас-медіа. Енциклопедія Сучасної України. URL: https://esu.com.ua/article-64254 (дата звернення: 05.03.2025).

2. Wang W. Y. "Liar, liar pants on fire": a new benchmark dataset for fake news detection. Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: short papers), м. Vancouver, Canada. Stroudsburg, PA, USA, 2017. URL: https://doi.org/10.18653/v1/p17-2067 (дата звернення: 05.03.2025).

3. Gifu D. An intelligent system for detecting fake news. Procedia computer science. 2023. Т. 221. С. 1058–1065. URL: https://doi.org/10.1016/j.procs.2023.08.088 (дата звернення: 05.03.2025).