УДК 004.8:004.946:004.056

В. Я. Усатюк[1]
Л.М. Куперштейн[1]
С. О. Прокопенко[2]

ТЕХНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЯВЛЕННЯ ДЕЗІНФОРМАЦІЇ У СОЦІАЛЬНИХ МЕРЕЖАХ

[1]Вінницький національний технічний університет
[2]ГО «ҐВАРА МЕДІА»

Анотація

У роботі розглянуто архітектуру програмного засобу для автоматичного виявлення дезінформації у відеоконтенті соціальної мережі TikTok. Використано комбінацію моделей OpenAI Whisper для транскрибації аудіо та LLM для семантичного аналізу тексту.

Ключові слова: штучний інтелект, дезінформація, соціальна мережа, велика мовна модель

Abstract

The paper examines the architecture of a software tool for the automatic detection of disinformation in social media video content. A combination of OpenAI Whisper models for audio transcription and LLMs for semantic text analysis is used.

Keywords: artificial intelligence, disinformation, social media, Large language model.

ВСТУП

Стрімкий розвиток соціальних мереж призвів до значного збільшення обсягів інформації, яка поширюється в інтернеті. Разом із корисним контентом активно поширюється і дезінформація, що становить серйозну загрозу для користувачів. Особливо небезпечним є мультимедійний, створений за допомогою технологій штучного інтелекту, який може вводити в оману або спонукати до небажаних дій.

У зв'язку з цим актуальним є створення автоматизованих систем, здатних швидко аналізувати інформацію та виявляти неправдивий або маніпулятивний контент. Одним із ефективних і перспективних підходів є використання інтелектуальних систем на основі штучного інтелекту [1, 2]. Дослідження показують, що використання великих мовних моделей (Large Language Model, LLM) для виявлення фейків та дезінформації є ефективним інструментом для розв'язання таких задач [3, 4].

Серед найпопулярніших соціальних платформ в Україні — YouTube, TikTok та Instagram, які мають мільйони користувачів [5]. Саме на цих платформах спостерігається значне поширення недостовірної інформації, тому їхній контент є доцільним об'єктом для аналізу.

Крім того, дослідження показують, що використання штучного інтелекту суттєво підсилює поширення дезінформації та пропаганди, особливо в умовах гібридної війни, що актуалізує потребу в інструментах автоматичної перевірки контенту [6].

Результати дослідження

У межах дослідження було розроблено програмну систему для автоматизованого виявлення дезінформації у соціальній мережі TikTok. Вибір саме цієї платформи зумовлений її високою популярністю та швидкістю поширення контенту, що створює сприятливі умови для розповсюдження маніпулятивної або неправдивої інформації. В умовах активної інформаційної війни в кіберпросторі, зокрема між Україною та росією, проблема оперативної перевірки контенту набуває особливої актуальності.

Розроблена система реалізована у вигляді AI-агента на основі великої мовної моделі GPT-5.2 та інтегрована з месенджером Telegram у форматі чат-бота. Такий підхід забезпечує зручну взаємодію з користувачем, оскільки дозволяє швидко надсилати відео з TikTok для перевірки безпосередньо через інтерфейс месенджера.

Процес аналізу контенту складається з кількох етапів. На першому етапі здійснюється транскрибація аудіодоріжки відео з метою отримання текстового представлення мовлення. Це виконано із використанням API нейронної мережі Whisper [7]. Далі система аналізує текстовий опис відео, коментарі користувачів та інші доступні метадані, що дозволяє виявити ключові слова та контекст поширюваної інформації.

Окремо виконується обробка візуальної складової відео, зокрема покадровий аналіз із розпізнаванням тексту, що може бути розміщений безпосередньо у відеоряді. Такий підхід дозволяє врахувати всі можливі джерела інформації, оскільки зміст аудіо, тексту та візуальних елементів може відрізнятися або доповнювати один одного.

Після збору даних система використовує інструмент веб-пошуку, який ініціюється мовною моделлю для перевірки отриманої інформації за зовнішніми джерелами в мережі Інтернет. Додатково реалізовано механізм пошуку у внутрішній базі даних, яка містить раніше оброблені випадки дезінформації та їх «цифрові відбитки». Це дозволяє значно прискорити повторне виявлення схожого або ідентичного контенту.

Усі зібрані дані — транскрибований текст, коментарі, метадані, результати розпізнавання та пошуку — передаються до LLM для комплексного аналізу. На основі цього формується підсумковий висновок щодо достовірності контенту, який повертається користувачу.

Результати аналізу зберігаються у базі даних, що забезпечує накопичення знань системи та підвищення ефективності її роботи в майбутньому.

Висновки

У роботі запропоновано підхід до виявлення дезінформації в соціальній мережі TikTok із використанням технологій штучного інтелекту. Розроблено Telegram-бот, який автоматизує процес аналізу відеоконтенту та перевірки достовірності поширюваної інформації.

Показано, що застосування мультимодального аналізу, який поєднує оброблення відео, аудіо та текстових даних, забезпечує вищу ефективність виявлення фейкового контенту порівняно з підходами, що базуються лише на одному типі вхідних даних. Такий підхід дає змогу підвищити повноту аналізу інформаційних матеріалів і зменшити ймовірність пропуску маніпулятивних або сфальсифікованих повідомлень.

Отримані результати підтверджують доцільність використання інтелектуальних мультимодальних систем у сфері інформаційної безпеки, зокрема для автоматизованого моніторингу інформаційного простору, виявлення дезінформації та підтримки процесів протидії інформаційно-психологічним впливам.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Kupershtein L., Zalepa O., Sorokolit V., Prokopenko S. AI-agent-based system for fact-checking support using large language models // *Proceedings of the 7th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2024)*. CEUR Workshop Proceedings. 2025. Vol. 3917. P. 321–331.
2. Rahman, S.S., Islam, M.A., Alam, M.M. et al. Hallucination to truth: a review of fact-checking and factuality evaluation in large language models. Artif Intell Rev 59, 70 (2026). https://doi.org/10.1007/s10462-025-11454-w.
3. Карабчук В. Методи виявлення та боротьби з дезінформацією за допомогою технологій штучного інтелекту // *Інформаційні технології та суспільство*. Київ: Міжрегіональна академія управління персоналом, 2025. Вип. 2 (17). DOI: https://doi.org/10.32689/maup.it.2025.2.8.

4. Куперштейн Л.М., Сороколіт В.О., Прокопенко С.О. Аналіз можливостей великих мовних моделей для автоматизації фактчекінгу // Матеріали міжнародної науково-практичної інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи (МН-2024)», 11-20 травня 2024 р. Електрон. текст. https://conferences.vntu.edu.ua/index.php/mn/mn2024/paper/viewFile/20855/17997 (дата звернення 20.03.2026).
5. Рейтинг найпопулярніших серед українців соцмереж. URL: https://skilky-skilky.info/reytynh-naypopuliarnishykh-sered-ukraintsiv-sotsmerezh-1-shi-mistsia-zaymaiut-youtube-ta-tiktok (дата звернення 20.03.2026).
6. Драбюк С. С. Штучний інтелект і пропаганда та дезінформація: основні виклики // *Науковий вісник Ужгородського національного університету. Серія: Право*. 2025. Т. 5, № 90. DOI: https://doi.org/10.24144/2307-3322.2025.90.5.45
7. Пропонуємо до вашої уваги Whisper. URL: https://openai.com/uk-UA/index/whisper.

Усатюк Віталій Ярославович – студент групи 1БКС – 22Б, факультет інформаційних технологій комп'ютерної інженерії, Вінницькйи національний технічний університет, Вінниця, e-mail: vitalijusatuk582@gmail.com

Куперштейн Леонід Михайлович - доцент кафедри захисту інформації, Вінницький національний технічний університет, Вінниця, e-mail: kupershtein@vntu.edu.ua.

Прокопенко Сергій Олександрович – керівник ГО «ҐВАРА МЕДІА», email: serhii@gwaramedia.com.

Vitaly Yaroslavovych Usatyuk – student of group 1BKS - 22B, Faculty of Information Technologies of Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: vitalijusatuk582@gmail.com

Kupershtein Leonid Mikhailovich — Associate Professor of the Department of Information Protection, Vinnytsia National Technical University, Vinnytsia, e-mail: kupershtein@vntu.edu.ua.

Prokopenko Serhii - Head of the NGO «GWARA MEDIA», email: serhii@gwaramedia.com.