

УДК 004.8:004.912:316.774

О. Г. ЧЕРЕП

доктор економічних наук, професор,
Запорізький національний університет,
м. Запоріжжя, Україна

Т. А. ЧЕПУРНА

Запорізький національний університет,
м. Запоріжжя, Україна

ТЕХНОЛОГІЇ ТА ПІДХОДИ ШТУЧНОГО ІНТЕЛЕКТУ У БОРОТБІ З ДЕЗІНФОРМАЦІЄЮ

Для боротьби з дезінформацією використовуються різні технології та підходи ІІІ:

1. Обробка природної мови (Natural Language Processing, NLP). Це напрям ІІІ, що дозволяє комп'ютерам розуміти, інтерпретувати та генерувати людську мову. В контексті протидії дезінформації NLP використовується для:

- аналізу тональності (Sentiment Analysis): визначення емоційного забарвлення тексту (позитивне, негативне, нейтральне), що допомагає виявляти маніпулятивні та пропагандистські матеріали;

- виявлення мови ворожнечі (Hate Speech Detection): ідентифікація образливих, дискримінаційних висловлювань;

- тематичне моделювання (Topic Modeling): визначення основних тем, що обговорюються у великих масивах тексту, для відстеження поширення конкретних наративів;

- аналіз стилю письма: виявлення ознак, характерних для пропагандистських текстів або згенерованих ботами повідомлень (надмірна емоційність, повторювані фрази, специфічні мовні конструкції);

- перевірка фактів (Fact-Checking Support): автоматичне зіставлення тверджень у тексті з базами даних відомих фактів або надійними джерелами [1].

2. Машинне навчання (Machine Learning, ML). Алгоритми ML дозволяють системам навчатися на основі даних без явного програмування. У боротьбі з фейками ML використовується для:

- класифікації контенту: навчені моделі можуть автоматично визначати, чи є новина, стаття або допис у соцмережі правдивими, неправдивими чи маніпулятивними, базуючись на різних ознаках (джерело, стиль, наявність посилань, реакції користувачів);

- виявлення аномалій: ідентифікація нетипової активності, наприклад, скоординованих інформаційних атак, масового поширення однакових повідомлень бот-мережами [2];

- кластеризації: групування схожих новин або повідомлень для виявлення масштабних дезінформаційних кампаній.

3. Комп'ютерний зір (Computer Vision). Ця галузь ІІІ дозволяє комп'ютерам «бачити» та інтерпретувати візуальну інформацію (зображення, відео). Застосування:

- виявлення маніпуляцій із зображеннями: ідентифікація фотомонтажу, редагування зображень;
- виявлення дипфейків (Deepfake Detection): аналіз відео на наявність артефактів, невідповідностей у міміці, освітленні, що можуть вказувати на підробку [3];
- зворотний пошук зображень: перевірка, чи використовувалося зображення раніше в іншому контексті.

4. Мережевий аналіз (Network Analysis). Аналіз зв'язків між користувачами, групами та джерелами інформації в соціальних мережах для виявлення скоординованих мереж поширення дезінформації (бот-ферм).

Однією з ключових переваг ІІІ є можливість здійснювати безперервний та масштабний моніторинг інформаційного простору – вебсайтів, новинних порталів, соціальних мереж, форумів, месенджерів. Алгоритми ІІІ можуть:

- відстежувати появу нових наративів: швидко ідентифікувати нові дезінформаційні повідомлення та теми, що починають поширюватися;
- аналізувати швидкість та шляхи поширення: визначати, як швидко поширюється фейк, через які канали (соцмережі, ЗМІ, месенджери) та які групи користувачів є основними розповсюджувачами;
- ідентифікувати ключові джерела: виявляти вебсайти, акаунти або групи, які є першоджерелами або активними розповсюджувачами дезінформації;
- оцінювати охоплення та вплив: аналізувати кількість переглядів, репостів, коментарів, щоб оцінити потенційний вплив дезінформаційної кампанії;
- прогнозувати розвиток подій: на основі аналізу поточних трендів моделі ІІІ можуть намагатися прогнозувати появу нових фейків або напрямки майбутніх інформаційних атак.

Література

1. Generative AI is already helping fact-checkers. But it's proving less useful in small languages and outside the West. URL: <https://reutersinstitute.politics.ox.ac.uk/news/generative-ai-already-helping-fact-checkers-its-proving-less-useful-small-languages-and> (дата звернення: 06.04.2025).
2. How Meta Is Planning for Elections in 2024. URL: <https://about.fb.com/news/2023/11/how-meta-is-planning-for-elections-in-2024/> (дата звернення: 06.04.2025).
3. Діпфейки: як розпізнати та захиститися? URL: <https://netfreedom.org.ua/article/dipfejki-yak-rozpiznati-ta-zahistitisya> (дата звернення: 06.04.2025).