УДК 007 : 304 : 070
Островська Н. В.
канд. наук соц. ком., доц. НУ «Запорізька політехніка»

МОЖЛИВОСТІ ТА ОБМЕЖЕННЯ АВТОМАТИЗОВАНОЇ ПЕРЕВІРКИ ФАКТІВ У ФАКТЧЕКІНГОВИХ ПРОЄКТАХ

Швидке поширення дезінформації, зокрема ШІ-контенту, спонукає фактчекінгові організації активніше використовувати автоматизовані системи для перевірки фактів. Адже вони потенційно можуть допомогти оптимізувати

професійну діяльність фактчекерів, зробити процес перевірки швидшим та ефективнішим.

Л. Дірікс, К.-Г. Лінден та А. Л. Опдаль, узагальнивши дослідження про автоматизовану перевірку фактів, зазначають, що її можна використовувати на всіх етапах фактчекінгу: для пошуку тверджень, які потрібно перевірити, для виявлення вже перевірених раніше фактів, пошуку доказів та безпосередньо верифікації [1]. Це дозволить автоматизувати рутинні завдання, швидше ідентифікувати контент, який потребує уваги, та знаходити підтверджувальну інформацію. Особливо помічним серед перелічених можливостей, на нашу думку, є автоматичний моніторинг медіапростору; аналіз не лише текстів, а й інших типів контенту: зображень, аудіо та відео; дослідження великих обсягів даних; здійснення миттєвого порівняння тверджень з офіційними базами даних, державними реєстрами, перевіреними новинами тощо. Важливою є також можливість масштабувати процес перевірки до тисяч тверджень одночасно, що забезпечує високу швидкість реагування на загрози.

Українські фактчекінгові проєкти також активно впроваджують автоматизацію на різних етапах верифікації інформації. Зокрема, головна редакторка фактчекінгового проєкту «НотаЄнота» Альона Романюк зазначає про використання моніторингових систем для відслідковування вкидів і їхнього поширення. Також про ефективність інструментів ШІ для виявлення дипфейків та інших типів контенту, згенерованого ШІ. Аналітикиня фактчекінгового порталу «StopFake» Олена Чуранова говорить про активне використання таких інструментів, як AIornot, GeoSpy, Pimeyes, Facecheck.id, Sensity.ai, Sentinel тощо.

Олекса Шарабура з «БезБрехні» наголошує, що сучасні виклики, як-от масове створення та поширення інформації із застосуванням ШІ, вимагають новітніх рішень для їх вирішення. Без автоматизації процесів з пошуку, систематизації та перевірки контенту наразі неможливо здійснювати якісний моніторинг інформаційного простору. На думку фактчекера, конкретні назви інструментів можуть змінюватися, але групи завдань, які виконуються за допомогою програмних продуктів з ШІ, є сталими: це краулери соцмереж і вебсторінок, інструменти для пошуку особи та інформації про особу, аналізатори контенту на наявність втручань ШІ, сервіси виявлення геолокації.

Водночас дослідники (Л. Дірікс, К.-Г. Лінден та А. Л. Опдаль) виокремлюють низку суттєвих обмежень автоматизованої перевірки фактів. До технічних обмежень відносять зосередженість більшості сучасних розробок у сфері автоматизованої перевірки фактів переважно на аналізі текстового контенту, тоді як засоби для перевірки достовірності зображень і відеоматеріалів становлять лише незначну частку досліджень – близько 5%.

До того ж системи, побудовані на основі глибоких нейронних мереж, часто характеризуються недостатньою прозорістю, через що складно пояснити логіку ухвалення конкретного рішення або вердикту алгоритмом. Ще одним технічним викликом є схильність моделей машинного навчання до перенавчання, коли вони надто точно адаптуються до особливостей навчальних даних і через це демонструють нижчу ефективність під час роботи з новими, реальними кейсами. Окрему проблему становить і недосконаліть використовуваних метрик оцінювання: показники точності не завжди дають повне уявлення про рівень невизначеності результатів і можуть бути недостатньо надійними для об'єктивної оцінки якості роботи системи [1].

Досліджуючи методологічні проблеми, науковці окрему увагу приділяють аналізу наборів даних для навчання ШІ. Адже вони часто містять помилки, а використання неспеціалістів через краудсорсинг призводить до перенесення людських упереджень в алгоритми. Про «алгоритмічні непередбачуваності» таких систем застерігають також А. Д. Ернандес, Р. Оуен, Д. С. Нільсен, Р. Макконвілл. Дослідники наголошують, що дані для навчання моделей є соціально сконструйованими й часто ґрунтуються на висновках фактчекінгових організацій, які можуть бути упередженими. Надмірна орієнтація на показники точності створює ризики неправомірної цензури, закріплення хибних наративів і замовчування системних проблем платформ [2].

До методологічних обмежень, які ускладнюють автоматизовану перевірку фактів, науковці також відносять відсутність єдиного стандартизованого визначення основних понять, як-от «твердження» або «верифікація». Також у редакціях використовують різні системи вердиктів, що ускладнює уніфікацію підходів до автоматизованої перевірки. Крім того, проблемою є прив'язаність анотованих наборів даних до конкретної мови, тематичної сфери або контексту. Через це системи втрачають гнучкість і можуть демонструвати нижчу ефективність у разі появи нових інформаційних приводів, змін медійного середовища чи роботи з іншими мовними й тематичними контекстами.

Додаткові труднощі виникають і всередині редакційного середовища, де існують психологічні та культурні бар'єри для прийняття нових технологій. Зокрема, журналістам часто бракує алгоритмічної грамотності, що ускладнює розуміння принципів роботи таких систем і знижує рівень довіри до них. Окрім цього, інтеграція нових інструментів часто є складним завданням через необхідність їх адаптації до вже сформованих і багаторівневих робочих процесів редакцій [1].

Практика редакцій українських фактчекінгових ресурсів свідчить про розрив між технічними рішеннями та реальними потребами редакцій. Як

зауважує Олекса Шарабура, не всі інструменти OSINT можуть бути оприлюднені і тоді постає питання використання результатів у форматі фактчек, адже читач не зможе перевірити їх достовірність. Подвійної уваги та постійної перевірки результатів вимагає застосування для аналізу знайденої інформації LLM-моделей через відомі проблеми з втратою фокусу та «галюцинуванням». На переконання фактчекера, повністю автоматизувати фактчекінг за допомогою ШІ не можна, оскільки ШІ часто не бачить або не розуміє контекст.

За спостереженням Олекси Шарабури, сучасний інструментарій для виявлення згенерованого штучним інтелектом контенту характеризується високою динамічністю оновлень, проте забезпечує лише ймовірну, а не остаточну оцінку. А це часто призводить до технічних помилок і суперечливих результатів (наприклад, ігнорування візуальних аномалій). Через таку недосконаліть алгоритмів фактчекери змушені застосовувати комплексний підхід, використовуючи декілька детекторів одночасно.

Отже, автоматизація фактчекінгу є необхідною відповіддю на швидкість поширення дезінформації, проте вона не є нейтральною через технічні, методологічні та професійно-культурні бар'єри. Загалом використання автоматизованих систем дозволяє поєднувати обчислювальну потужність ШІ з людською інтуїцією, критичним мисленням та професійним досвідом. Водночас надійний захист медіапростору можливий лише через гібридні системи, де ШІ масштабує роботу, але не замінює критичне мислення та інтуїцію людини. Тому ключова роль у верифікації та фінальному аналізі даних незмінно залишається за людиною-фахівцем.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Dierickx L., Lindén C.-G., Opdahl A. L. Automated Fact-Checking to Support Professional Practices: Systematic Literature Review and Meta-Analysis. *International Journal of Communication*. 2023. №17. P. 5170– 5190. URL: https://ijoc.org/index.php/ijoc/article/view/21071/4287

2. Hernández A. D., Owen R., Nielsen D. S., McConville R. Addressing contingency in algorithmic (mis)information classification: Toward a responsible machine learning agenda. *arXiv.* 2022. URL: https://doi.org/10.48550/arXiv.2210.09014