

УДК 004.8:316.774

МАНІПУЛЯТИВНІ ВПЛИВИ ШТУЧНОГО ІНТЕЛЕКТУ В МЕДІА ТА ШЛЯХИ ЇХ ВРЕГУЛЮВАННЯ

Яблонь В. В.

e-mail: vasylyna.iablon@nure.ua

Харківський національний університет радіоелектроніки, каф. філософії,
м.Харків, Україна

The rapid spread of mass use of artificial intelligence increases the need to study its impact on the modern media space. In particular, understanding the role of Artificial Intelligence technologies in the field of information influence is the basis for predicting further trends and trajectories of the development of propaganda as a problematic social phenomenon. The basis for regulating the development of AI and its impacts, in particular in the media space, should be generally accepted ethical norms, consistency of values, impartiality in the management and development of algorithms, and an appropriate legal framework is mandatory for application.

Стрімке поширення масового використання штучного інтелекту збільшує потреби в дослідженні його впливів на сучасний медіапростір. Зокрема, розуміння ролі технологій Artificial Intelligence в сфері інформаційного впливу є основою для передбачення подальших тенденцій і траєкторій розвитку пропаганди як проблемного суспільного явища.

Штучний інтелект надає інструменти для швидкої, простої, якісної і дешевої генерації будь-якого типу медіаповідомлень, що широко використовується з метою створення пропаганди чи дезінформації. Отриманий контент може публікуватися та поширюватися практично миттєво через автоматизовані інструменти чи використання складно керованого інфополя соцмереж: через алгоритмізацію розповсюдження (як у Тік-Ток) чи відсутність об'єктивної модерації дезінформаційних повідомлень (як у X). Діапазон впливу здатний охоплювати надвелику кількість людей і, попри якісну протидію, залишає відбиток у масовій свідомості тих чи інших тем.

Окремої уваги заслуговує «двійництво», коли авторитет певної особи, компанії, інформаційного ресурсу тощо використовується з метою маніпулятивного впливу на аудиторію, зокрема, дипфейки, що відтворюють звук, зображення чи відеоряд, які важко відрізнити від реальних медіаповідомлень без втручання в їхній контекст (до прикладу фейкове відео з псевдозаявою Президента України Володимира Зеленського про капітуляцію). Водночас з'являється все більше інтернет-ресурсів та інших, сфальсифікованих за допомогою штучного інтелекту, джерел. Компанія Doppelganger фабрикує сайти відомих новинних агенцій, використовуючи невеликі відмінності у URL-адресі, й під виглядом

істинного, публікує видозмінений пропагандистський вміст з проросійськими наративами [1].

Не менш важливою є маніпулятивна роль чат-ботів, що призначені для загального масового користування: ChatGPT, Gemini, Copilot та інші. Дослідження журналістської групи аналітиків надійності інформації News Guard, у ході якого було протестовано десять широко використовуваних чат-ботів [2], показує, що загальнопоширені моделі штучного інтелекту з високою ймовірністю схильні повторювати дезінформацію і пропаганду, якщо користувач надає прямий запит такого характеру. Переконавання користувачів у правдивості заданих тверджень призводить до посилення вже існуючої інформаційної бульбашки.

Логічним видається рішення про додавання у моделі штучного інтелекту незалежних інструментів перевірки правдивості інформації, однак такий підхід може йти в розріз з політикою компаній-власників. Зокрема, І. Маск, власник чат-бота Grok, особисто просуває вільне поширення неперевіреної інформації у соцмережі X [3]. Доступ політиків та інших зацікавлених осіб до інструментів впливу ШІ унеможлиблює справедливу і неупереджену роботу самої технології як явища, а також чат-ботів як її загальнодоступної частини. Сам факт використання політиками штучного інтелекту для генерації пропагандистських меседжів, таких як відео про майбутнє Сектора Гази, опубліковане на сторінці чинного президента США, є неприпустимим у межах морально-етичного розуміння ролі технологій у соціально-політичному дискурсі.

Утім, застосування алгоритмів машинного навчання здатне виявляти зворотню дію при направленому зверненні розробників до технічних задач протидії маніпулятивним впливам. Прикладом використання ШІ з метою виявлення пропаганди та швидкого реагування на маніпулятивні та дезінформаційні впливи є український стартап Mantis Analytics, робота якого є відповіддю на гібридну війну, яку веде Росія проти України.

Щоб утримати баланс впливів, важливо чітко визначити ряд правил, яким має підкорятися ШІ. Свого часу А. Азімов сформував три базових правила робототехніки. Однак, такий підхід працюватиме лише за умови, що механізм, у нашому випадку – штучний інтелект, оперуватиме усіма необхідними знаннями і виключатиме можливість конфлікту інтересів на користь загальноприйнятих моральних норм. Згодом, Британська академія спільно з Королівським товариством внесли публічну пропозицію скоротити ці три закони до одного – «поліпшувати добробут людини» [4]. Подальша деталізація міститься у «Asilomar AI Principles», розробленого на конференції Beneficial AI 2017, у відкритому листі [5], де наголошуються: безпека, прозорість, відповідальність, дослідницька користь та інші вагомні людські цінності, які є обов'язковими умовами для сталого розвитку і гармонійного технологічного зростання. Слід відзначити суспільні ініціативи та проєкти, зокрема, листи, як «Pause Giant

AI Experiments: An Open Letter» [6], що закликає компаній-розробників призупинити активну фазу розробок ШІ задля якісного планування з урахуванням етичних потреб людства. Нарешті, 1 серпня 2024 року набув чинності запропонований Європейською Комісією та схвалений Європейським парламентом і Радою ЄС перший Закон про штучний інтелект (Artificial Intelligence Act) [7].

Отже, основу врегулювання розвитку ШІ та його впливів, зокрема у медіапросторі, мають складати загальноприйняті етичні норми, узгодженість цінностей, неупередженість в керуванні й розробці алгоритмів та відповідна правова база, обов'язкова для застосування. Додатковим чинником повинна стати медіаграмотність серед населення і широка обізнаність щодо мети, можливостей, завдань і ризиків пов'язаних з розвитком та використанням штучного інтелекту.

Список використаних джерел

1. Темний бік штучного інтелекту: як ШІ підсилює російську пропаганду. Інтернет Свобода. URL: <https://netfreedom.org.ua/article/temnij-bik-shtuchnogo-intelektu-yak-shi-pidsilyuye-rosijsku-propagandu> (дата звернення: 04.03.2025).

2. Top 10 Generative AI Models Mimic Russian Disinformation Claims A Third of the Time, Citing Moscow-Created Fake Local News Sites as Authoritative Sources. NewsGuard. URL:

<https://www.newsguardtech.com/special-reports/generative-ai-models-mimic-russian-disinformation-cite-fake-news/> (date of access: 04.03.2025).

3. Поліковська Ю. У соцмережі X прибрали функцію, яка дозволяла поскаржитися на дезінформацію про вибори. ms.detector.media. URL: <https://ms.detector.media/internet/post/33056/2023-09-27-u-sotsmerezhi-kh-prybraly-funktsiyu-yaka-dozvolyala-poskarzhytysya-na-dezinformatsiyu-pro-vyboru/> (дата звернення: 04.03.2025).

4. Data management and use: Governance in the 21st century. A joint report by the British Academy and the Royal Society. The British Academy. URL: https://www.thebritishacademy.ac.uk/documents/105/Data_management_and_use_-_Governance_in_the_21st_century.pdf (date of access: 04.03.2025).

5. Asilomar AI Principles. Future of Life Institute. URL: <https://futureoflife.org/open-letter/ai-principles/> (date of access: 04.03.2025).

6. Pause Giant AI Experiments: An Open Letter. Future of Life Institute. URL: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (date of access: 04.03.2025).

7. High-level summary of the AI Act | EU Artificial Intelligence Act. EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act. URL: <https://artificialintelligenceact.eu/high-level-summary/> (date of access: 04.03.2025).