УДК 316.6

А. В. Павловська

ШІ ЯК ЗАСІБ МАНІПУЛЯЦІЇ СВІДОМІСТЮ

Вінницький національний технічний університет

Анотація

У дослідженні проаналізовано феномен штучного інтелекту (ШІ) як потужного інструменту впливу на масову суспільну свідомість. Особливу увагу зосереджено на можливостях провідних генеративних моделей, таких як GPT-4, Gemini і Midjourney, у створенні правдоподібного текстового та візуального контенту. Розкрито механізми застосування цих технологій для поширення дезінформації та маніпулювання громадською думкою. Робота містить конкретні приклади використання діпфейків і автоматизованих систем у політичному та соціальному контекстах, що підкреслює актуальність і ризики сучасних ШІ-технологій у соціально-політичній сфері.

Ключові слова: штучний інтелект, маніпуляція свідомістю, генеративні моделі, діпфейк, дезінформація, інформаційна безпека.

Abstract

The study analyzes the phenomenon of artificial intelligence (AI) as a powerful tool for influencing mass public consciousness. Particular attention is paid to the capabilities of leading generative models, such as GPT-4, Gemini, and Midjourney, in creating believable text and visual content. The mechanisms of using these technologies to spread disinformation and manipulate public opinion are revealed. The work contains specific examples of the use of deepfakes and automated systems in political and social contexts, which emphasizes the relevance and risks of modern AI technologies in the socio-political sphere.

Keywords: artificial intelligence, consciousness manipulation, generative models, deepfake, disinformation, information security.

Вступ

Сучасний світ перебуває на порозі безпрецедентних змін, пов’язаних із переходом до нової епохи інформаційного суспільства, де визначальну роль відіграють технології штучного інтелекту. Якщо раніше алгоритми виконували переважно допоміжні завдання, зокрема автоматизацію рутинних обчислень чи оптимізацію пошукових процесів, то сьогодні ми стаємо свідками виникнення генеративного штучного інтелекту, здатного до творчої та майже автономної діяльності. Цей технологічний прорив кардинально змінив способи взаємодії людей з інформаційним середовищем.

Завдяки можливостям створювати переконливі тексти, надреалістичні візуалізації та імітувати голоси, штучний інтелект став могутнім інструментом впливу. Технології дозволяють не лише викривляти окремі факти, а й конструювати цілісні наративи, здатні маніпулювати політичними поглядами, провокувати соціальну напругу й підірвати довіру до державних структур.

Тривогу викликає зокрема доступність і масштабованість подібних технологій. Те, що раніше вимагало значних ресурсів і узгоджених зусиль численних пропагандистських структур, зараз може бути автоматизоване і навіть адаптоване до конкретного користувача.

Результати дослідження

ШІ — це така технологія, яка дає змогу комп'ютерам думати та робити рішення, схожі на людські. Замість того, щоб просто виконувати завдання, задані людьми, комп'ютери зі штучним інтелектом можуть навчатися, аналізувати інформацію та ухвалювати самостійні рішення. Вони використовують алгоритми та моделі для обробки величезних масивів даних, щоб витягувати з них корисну інформацію [1].

На сьогодні провідними моделями в сфері пошуку інформації та генерації зображень є:

1. GPT-4 від OpenAI: Мультимодальна модель, здатна генерувати тексти, що стилістично та змістовно майже не відрізняються від написаних людиною. Вона може аргументовано підтримувати будь-яку точку зору, що робить її ідеальним інструментом для створення пропагандистських наративів [2].
2. Gemini від Google: Конкурентна модель з величезними контекстними вікнами, здатна аналізувати та синтезувати великі обсяги даних, адаптуючи контент під конкретну аудиторію.

3. Midjourney, DALL-E 3, Sora: Інструменти для генерації гіперреалістичних зображень та відео. Вони дозволяють створювати візуальні «докази» подій, яких ніколи не існувало.

Саме ці можливості відкривають шлях до маніпуляції свідомістю через такі механізми:

– персоналізована пропаганда (Micro-targeting): ШІ може аналізувати цифровий слід користувача і генерувати повідомлення, що впливають на його специфічні страхи чи упередження;
– астротурфінг (Astroturfing): Створення ілюзії масової підтримки певної ідеї за допомогою тисяч ботів, керованих LLM, які ведуть «живі» дискусії у соцмережах;
– розмивання реальності (Deepfakes): Підробка голосу та зображення авторитетних осіб.

Ці загрози вже реалізуються в реальному життя, особливо шахраями. Вони як ніхто інший першими зрозуміли, що ШІ може генерувати відео чи голосові повідомлення, від лиця іншої людини. Прикладом цього є те як у 2024 році, шахраї зламували соціальні мережі людей, та генерували відео чи голос людини, та просили гроші в близьких друзів чи родичів, на цю провокацію міг би багато хто піддатися, тому що навряд чи ми б змогли відмовити в нагальній потребі близькій людині [3].

Ще одним яскравим прикладом втручання ШІ у наш світ, є випадок на передодні виборів у 2024 році в США. У Нью-Гемпишері виборці отримували телефоні дзвінки, ніби то від Джо Байдена, який закликав їх не приходити на вибори. Це створило яскраву маніпуляцію, коли через авторитет людини, мож-на маніпулювати великою кількість виборців. Тому що при розмові з такою поважною людиною, ти стаєш впевнений що так правильно буде вчени, навіть не думаючи про наслідки і не аналізуючи, чому тобі простій людині дзвонить політ такого рівня як Джо Байден. [4]

Також показовим є випадок із фейковим зображенням вибуху біля Пентагону у травні 2023 року, згенерованим ШІ. Поширення цього зображення у Twitter верифікованими акаунтами призвело до короткочасного падіння фондового ринку, що доводить здатність ШІ впливати не лише на думки, а й на економічну поведінку мас [5].

Протидія маніпулятивному впливу, здійснюваному за допомогою штучного інтелекту, вимагає інтегрованого та міждисциплінарного підходу, який поєднує новітні технологічні розробки, комплексне правове регулювання та підвищення рівня соціальної обізнаності. На технологічному рівні особливої ваги набуває впровадження автоматизованих засобів виявлення синтетичного контенту, зокрема використання цифрових водяних знаків і криптографічних механізмів для забезпечення автентичності й походження медіаматеріалів. Одночасно необхідна адаптація національних та міжнародних законодавчих норм для зобов'язання онлайн-платформ маркувати матеріали, створені штучним інтелектом, а також для запровадження юридичної відповідальності за розповсюдження деструктивних діпфейків.

Важливим вектором захисту є впровадження новітніх освітніх методик, зокрема концепції пребанкінгу. Ефективним інструментом тут виступають гейміфіковані симуляції, такі як гра «Get Bad News», розроблена Кембриджським університетом, де користувачі вчаться розпізнавати маніпуляції, створюючи їх у віртуальному середовищі [6]. Окрім цього, необхідно навчати населення користуватися технічними засобами верифікації, наприклад, сервісом Deepware Scanner для виявлення діпфейків, та застосовувати метод SIFT (Stop, Investigate, Find, Trace) для швидкої оцінки джерел [7]. Поєднання ігрових форм навчання з опануванням OSINT-інструментів дозволяє сформувати у молоді стійкий імунітет до когнітивних атак, згенерованих ШІ.

ВИСНОВОК

Штучний інтелект є технологією подвійного призначення. Можливості таких моделей, як GPT-4 чи Midjourney, дозволяють створювати контент, який обходить критичні фільтри людської свідомості. Маніпуляція стає автоматизованою, дешевою та персоналізованою. Для протидії цьому необхідно не лише вдосконалювати технічні засоби детекції ШІ-контенту, а й розвивати медіаграмотність суспільства, навчаючи людей критично оцінювати інформацію в епоху «постправди».

В той час ефективна боротьба з викликами впливу штучного інтелекту можлива лише за умови комплексної взаємодії технічних інструментів розпізнавання контенту, суворих правових рамок та високого рівня інформаційної культури загалу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Що таке штучний інтелект і чого всім раптом стало так цікаво. URL: https://ailaboratory.wixsite.com/shi-ua/post/shcho-take-shtuchnyj-intelect (дата звернення 20.12.25)

2. OpenAI. GPT-4 Technical Report. arXiv preprint. 2023. URL: https://arxiv.org/abs/2303.08774 (дата звернення 20.12.25)
3. Штучний інтелект на службі у шахраїв: аферисти вигадали нову шокувальну схему. URL: https://tsn.ua/exclusive/shtuchniy-intelekt-na-sluzhbi-u-shahrayiv-aferisti-vigadali-novu-shokuyuchu-shemu-2734428.html (дата звернення 20.12.25)
4. Fake Biden robocall tells New Hampshire Democrats not to vote. *NBC News*. 2024. URL: https://www.nbcnews.com/politics/2024-election/fake-joe-biden-robocall-tells-new-hampshire-democrats-not-vote-tuesday-rcna134984 (дата звернення 20.12.25)
5. Fake AI image of Pentagon explosion briefly sends US stocks dipping. *The Guardian*. 2023. URL: https://www.theguardian.com/technology/2023/may/22/pentagon-ai-generated-image-explosion (дата звернення 20.12.25)
6. Вакцинація від фейків: як працює гра Get Bad News. URL: https://ms.detector.media/manipulyatsii/post/21663/2018-08-21-vaktsyna-vid-feykiv-yak-pratsyuie-gra-get-bad-news/ (дата звернення 25.12.25)
7. Метод: SIFT. Останнє оновлення 27.10.22. URL: https://ukrayinska.libretexts.org/%D0%93%D1%83%D0%BC%D0%B0%D0%BD%D1%96%D1%82%D0%B0%D1%80%D0%BD%D1%96_%D0%BD%D0%B0%D1%83%D0%BA%D0%B8/%D0%94%D0%BE%D1%81%D0%BB%D1%96%D0%B4%D0%B6%D0%B5%D0%BD%D0%BD%D1%8F_%D1%82%D0%B0_%D1%96%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D1%96%D0%B9%D0%BD%D0%B0_%D0%B3%D1%80%D0%B0%D0%BC%D0%BE%D1%82%D0%BD%D1%96%D1%81%D1%82%D1%8C/%D0%92%D1%81%D1%82%D1%83%D0%BF_%D0%B4%D0%BE_%D0%B4%D0%BE%D1%81%D0%BB%D1%96%D0%B4%D0%B6%D0%B5%D0%BD%D1%8C_%D0%BA%D0%BE%D0%BB%D0%B5%D0%B4%D0%B6%D1%83_(Butler%2C_Sargent_%D1%96_Smith)/04%3A_%D0%9F%D0%B5%D1%80%D0%B5%D0%B2%D1%96%D1%80%D0%BA%D0%B0_%D1%84%D0%B0%D0%BA%D1%82%D1%96%D0%B2/4.02%3A_%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_SIFT

Павловська Анастасія Вячеславівна – студентка групу 1БС-24б, факультету інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, nastyapav2006@gmail.com

Науковий керівник: ***Теклюк Анатолій Іванович*** – канд. філософських наук, доцент кафедри філософії та гуманітарних наук, Вінницький національний технічний університет, Вінниця, teklyuk.a.i@vntu.edu.ua

Pavlovska Anastasia Vyacheslavovna – student group 1SS-24b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, nastyapav2006@gmail.com

Supervisor: ***Tekliuk Anatolii I.*** — PhD in Philosophy, Associate Professor, Vinnytsia National Technical University, Vinnytsia, teklyuk.a.i@vntu.edu.ua