



УДК 004.8:316.77:327.5

DOI: 10.20535/2077-7264.1(87).2025.318549

© С. Б. Фіялка, канд. наук із соц. ком., доц., В. О. Каменчук,
асп., КПІ ім. Ігоря Сікорського, Київ, Україна

ВИКОРИСТАННЯ ChatGPT ДЛЯ ВИЯВЛЕННЯ ПРОПАГАНДИ ТА ДЕЗІНФОРМАЦІЇ В УМОВАХ РОСІЙСЬКО-УКРАЇНСЬКОЇ ВІЙНИ

У статті розглянуто можливості застосування ChatGPT як інструменту для виявлення, аналізу та протидії пропаганді й дезінформації в контексті російсько-української війни. Автори наголошують, що сучасний інформаційний простір стає ключовим полем битви, де поширюються маніпуляції, емоційно забарвлені повідомлення, викривлені дані та пропагандистські техніки, спрямовані на вплив на громадську думку.

Ключові слова: штучний інтелект; ChatGPT; пропаганда; російсько-українська війна; маніпулювання; емоційні маркери; критичне мислення.

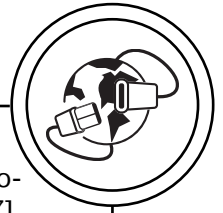
Постановка проблеми

В умовах багаторічної російсько-української війни інформаційний простір став одним з ключових полів битви, де відбувається поширення пропаганди, дезінформації та маніпулювання аудиторією. Водночас розвиток цифрових технологій, зокрема штучного інтелекту, відкриває нові можливості для виявлення, аналізу та протидії таким деструктивним інформаційним впливам. Зокрема, ChatGPT, як один з передових інструментів на основі штучного інтелекту, здатний не лише ідентифікувати приховані маніпуляції, а й формувати рекомендації для аудиторії щодо критичного осмислення медіаматеріалів. Тож виникає необхідність узагальнити напрями використання ChatGPT

зادля протистояння інформаційним загрозам в умовах війни.

Аналіз попередніх досліджень

Штучний інтелект може застосовуватися для автоматизованого аналізу великих обсягів даних з різних джерел, виявлення неправдивої інформації, розпізнавання підроблених зображень через їх порівняння з оригіналами, встановлення схем поширення дезінформації, блокування виявленого неправдивого контенту [1]. Він може допомагати ідентифікувати пропагандистські техніки в новинних статтях та надавати обґрунтовані пояснення [2], а також виявляти пропаганду через аналіз емоційного забарвлення текстів, зокрема звертати увагу на нена-



висний, образливий та токсичний контент (НОТ-контент) у соціальних мережах [3]. Так, український стартап на основі штучного інтелекту Osavul, заснований бізнес-партнерами Д. Білашем та Д. Плешаковим, активно сприяє боротьбі з російською пропагандою та дезінформацією. З початку повномасштабного вторгнення розроблено технологію, яка аналізує повідомлення й публікації в соціальних мережах, таких як «Телеграм» та «Фейсбук», оперативно виявляючи організовані інформаційні кампанії. Це дозволяє РНБО та Міністерству оборони України ефективно спростовувати фейкові новини та протидіяти інформаційно-психологічним операціям [4]. Крім того, мовні моделі можуть бути корисні, щоб перевірити і свої власні тексти на наявність ознак пропаганди [2].

Дослідження, опубліковане в журналі Science, показало, що чат-боти на базі штучного інтелекту, такі як ChatGPT, є ефективнішими за людей у переконуванні осіб відмовитися від віри в конспірологічні теорії. Експеримент за участю 2190 учасників показав, що діалоги з моделлю штучного інтелекту знижували віру в конспірологію на 20 %, а ефект зберігався навіть через два місяці після взаємодії з моделлю [5].

Водночас, як зазначає Р. Лорян, штучний інтелект став суперінструментом для дезінформації та пропаганди [6]. Медійники, IT-фахівці та інші спеціалісти занепокоєні через ризик втратити роботу, митці дедалі частіше заявляють про порушення їхніх інтелектуальних прав, а перед дослідниками постає виклик знайти способи протидіяти поширенню

інтернетом штучно згенерованого неправдивого контенту [6; 7]. До того ж сам штучний інтелект стикається з труднощами через залежність від контексту маніпулятивної мови [8]. Надмірна довіра до систем штучного інтелекту може призвести до втрати критичного мислення та зменшення людської автономії, адже існують ризики використання штучного інтелекту для поширення пропаганди та проведення масштабних дезінформаційних кампаній. Зокрема, ChatGPT здатний генерувати пропагандистські матеріали, імітуючи стиль медіа [9], та втручатися у вибори. Наприклад, у травні 2024 р. виявлено, що групи акаунтів з Росії, Китаю, Ірану та Ізраїлю використовувала технології OpenAI з метою вплинути на світовий політичний дискурс [10]. Ф. Хеппелл, М. Бакір та К. Бончева продемонстрували, що GPT-3.5 міг створювати реалістичні дезінформаційні твердження про війну в Україні, навіть щодо подій, які сталися після завершення його навчання, і що люди чи наявні автоматизовані інструменти не здатні надійно їх ідентифікувати [11].

Експерти інноваційної лабораторії Europol попереджають про небезпеки кримінального використання ChatGPT, що пов'язані з його здатністю створювати дуже реалістичний текст, що робить його інструментом для фішингових атак та інших кіберзлочинів [12]. Це підкреслює необхідність контролю та етичного використання технологій штучного інтелекту, впровадження курсів з медіаграмотності та критичного мислення, які б сприяли підвищенню обізнаності громадян щодо інформаційних загроз [13].



Мета роботи

З'ясувати можливості використання ChatGPT як інструменту для виявлення, аналізу та протидії пропаганді й дезінформації в контексті російсько-української війни задля посилення інформаційної безпеки, підвищення критичного мислення аудиторії.

Результати проведених досліджень

Матеріалом дослідження стали п'ять телеграм-каналів з різною тематикою та аудиторією: особистий канал громадського діяча А. Смолія (https://t.me/smolii_ukraine), місцеві новинні канали «Івано-Франківськ 24/7» (@blacklist 2477) і «Труха Київ» (@truehakyiv), новинний канал «Моя Україна» (@this-туа), а також канал, що фокусується на військових подіях, «Реальна війна» (@voynareal). Проаналізовано понад 1500 повідомлень, опублікованих на цих каналах з 1 листопада 2024 р. по 20 грудня 2024 р. З них обрано 154 повідомлення з використанням одного й більше прийомів, що варіюються від апеляції до емоцій, апеляцій до авторитету, до спотворення, вибіркового висвітлення фактів тощо [14]. Для аналізу контенту телеграм-каналів використано програмне забезпечення NVivo, що дозволило систематизувати й структурувати дані. На основі попередньо розроблених категорій (наприклад, «апеляція до страху», «апеляція до співчуття», спотворення фактів) здійснювалося кодування текстів. Програма NVivo автоматично підраховувала частотність використання відповідних прийомів (табл.).

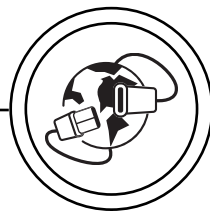
Відтак повідомлення проаналізовано за допомогою ChatGPT,

щоб протестувати здатність моделі виявляти потенційно маніпулятивний або дезінформаційний контент. Аналіз передбачав: виявлення емоційних маркерів у тексті, розкриття логічних помилок, упереджень та маніпуляцій і генерацію рекомендацій для критичного аналізу контенту.

ChatGPT допоміг виявити емоційно заряджені маркери страху, гніву, співчуття та надії. Вони є важливим елементом формування інформаційного простору під час війни. Серед маркерів страху штучний інтелект ідентифікував фрази: «масований наступ росіян неминучий», «ситуація катастрофічна», «можливий повторний наступ на Київ», «ця зима буде найважчою за три роки повномасштабної війни», «війна може тривати ще десятки років», «узимку можливий енергетичний колапс», «блекаут триватиме тижнями» тощо. З одного боку, вони попереджають про реальні загрози. З другого боку, надмірне нагнітання страху може викликати паніку, що ускладнює адекватне реагування.

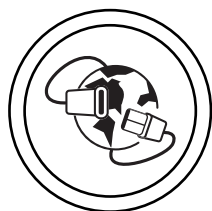
Маркери гніву посилюють почуття обурення діями ворога, акцентуючи увагу на його злочинах. Серед найбільш емоційно насичених маркерів гніву відзначимо такі: «ракети влучають у школи та лікарні», «черговий удар по цивільній інфраструктурі», «вони крадуть наших дітей», «воєнні злочини росії залишаються безкарними», «окупанти спалюють українські книги». Вони підкріплюють негативне ставлення до агресора, консолідують аудиторію та сприяють мобілізації до боротьби.

Маркери співчуття використовують для того, щоб викликати



Результати аналізу методів пропаганди та маніпуляцій, виявлених у досліджених повідомленнях

Метод	Суть методу	Кількість випадків	Приклад
Апеляція до емоцій (страх)	Використання емоційно заряджених слів і фраз, що пробуджують страх	103	«Від завтра по Україні люті морози досягнуть подекуди - 14 градусів. Особливо холодно на Заході країни. Якщо враховувати постраждалу енергосистему та опалення, яке від неї залежить, — запасаємось теплим одягом та ковдрами...» (@blacklist2477, 13.12.2024); «Дефіцит електроенергії триватиме роками: українців закликають готуватися до найгіршого сценарію» (https://t.me/smolii_ukraine , 13.12.2024)
Апеляція до емоцій (гнів)	Використання емоційно заряджених слів і фраз, що пробуджують гнів	72	«Так рятували чоловіка з поліклініки. Насипте прокльонів русні, щоб у них так кожен день було!» (@truehakyiv, 29.11.2024)
Апеляція до емоцій (співчуття)	Використання емоційно заряджених слів і фраз, що пробуджують співчуття	69	«Зі сльозами на очах дівчина співає у зруйнованій квартирі в Тернополі, де 2 грудня від атаки рф загинув її тато. Аліна Ременець зіграла на піаніно та виконала «Летіла зозуля». 45-річний тато Віталій зі своїм собакою загинули унаслідок російського удару по Тернополю» (https://t.me/smolii_ukraine , 03.12.2024)
Апеляція до емоцій (надія)	Використання емоційно заряджених слів і фраз, що пробуджують надію	24	«В Україні швидко наближається припинення вогню: відновлення країни розпочнеться вже у другій половині 2025 року» (@voynareal, 12.12.2024)
Спотворення фактів, маніпуляції через неправдиві або ненадійні дослідження	Перекручування фактів, згадування фактів, які не підтверджені джерелами	49	«Повістка буде вважатися врученою, якщо «Укрпошта» надішле у Viber повідомлення, а ви її не заберете» (@blacklist2477, 14.12.2024); «Понад 7 мільйонів придатних до служби російських чоловіків готові вирушити на війну



Закінчення табл.

Метод	Суть методу	Кількість випадків	Приклад
			з Україною за гроші, підрахував Європейський центр аналізу та стратегій (CASE), провівши опитування у рф» (https://t.me/smolii_ukraine , 10.12.2024); Понад 44,6 % українців довіряють Трампу — опитування на замовлення Центру «Нова Європа». Чекають на «мір за 24 години»? (https://t.me/smolii_ukraine , 10.12.2024)
Вибіркове подання фактів	Факти подають частково	33	«Санкції, накладені міжнародними партнерами на росію, змушують кремль переглядати свою політику» (@thismyua, 29.11.2024)
Апеляції до авторитету	Зазначення імен офіційних осіб, експертів чи міжнародних інституцій додає ваги повідомленню, навіть якщо ці авторитети ніяк не причетні до теми	19	«Витрати перекладуть на кінцевого споживача, що призведе до подорожчання вартості комунальних послуг. Про це заявив міністр енергетики та захисту навколишнього середовища України у 2019–2020 роках Олексій Оржель» (https://t.me/smolii_ukraine , 13.12.2024)

емпатію та привернути увагу до людських трагедій. Такі маркери підкреслюють страждання цивільного населення, спричинені війною, і акцентують увагу на необхідності гуманітарної підтримки. Прикладами маркерів співчуття є: «трое дітей залишилося сиротами», «мати біля могили сина», «чоловік втратив усю родину», «лікарі борються за життя пораненої дитини» тощо.

Маркери надії відіграють ключову роль у підтримці морального духу аудиторії. Вони акцентують увагу на успіхах української

армії, міжнародній підтримці та перспективах відновлення країни після війни. Наприклад, «наші хлопці повернули контроль над містом», «ми на ще один день наблизилися до перемоги», «ми вистоїмо», «український дух незламний», «ухвалено новий пакет допомоги для України», «нас підтримує весь цивілізований світ», «світло повертається». Маркери надії також акцентують увагу на взаємодопомозі. Історії про бабусь, які печуть пиріжки чи плетуть шкарпетки для військових, або про дітей, які збирають до-



нати, формують уявлення про солідарність навіть у найважчих умовах. Вони створюють образ суспільства, яке об'єднується навколо допомоги ЗСУ та цивільному населенню.

Тепер простежимо, як ChatGPT може допомогти критично оцінити медіаповідомлення щодо наявності в ньому ознак маніпуляції і пропаганди.

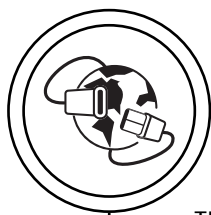
Приклад 1. «Станом на початок грудня 2024 року 75 % приватних компаній-виробників озброєння та військової техніки стверджують, що державні замовники не ведуть з ними перемовини про постачання озброєння на наступний рік — дані з нового опитування Технологічних Сил України, участь в якому взяла 31 компанія приватного сектору оборонно-промислового комплексу (ОПК)» (https://t.me/smolii_ukraine, 11.12.2024).

У результаті аналізу цього повідомлення штучний інтелект виокремив емоційні тригери: твердження, що «державні замовники не ведуть перемовини», формує враження кризової ситуації у взаємодії держави та приватного оборонного сектору. ChatGPT зазначив, що посилання на «Технологічні Сили України» та опитування підвищує довіру до тексту, але зауважив, що не вказується методологія опитування, наприклад, як обирали компанії чи, які питання ставили, а також те, що опитування охоплює лише 31 компанію, що є незначною вибіркою в масштабах оборонного сектору України й тому викликає питання щодо репрезентативності даних. Також у відповіді наголошується, що повідомлення не містить альтернативної

інформації, наприклад, позиції державних замовників чи урядових органів, що обмежує можливість об'єктивного аналізу і створює враження упередженості. Тож, за висновком штучного інтелекту, це повідомлення посилює тривогу щодо обороноздатності країни; містить однобічне подання інформації, без аналізу інших можливих причин чи точок зору. Також йому бракує прозорої методології опитування, що може впливати на достовірність викладених даних.

Приклад 2. «В Україні швидко наближається припинення вогню: відновлення країни розпочнеться вже у другій половині 2025 року, — представник польського уряду щодо відновлення України Павло Коваль» (@voynareal, 12.12.2024).

Аналізуючи це повідомлення, ChatGPT виокремив емоційно заряджену фразу («швидко наближається припинення вогню»), що формує оптимістичний настрій, хоча немає достатніх доказів, що такі прогнози базуються на об'єктивних даних. Використано ім'я офіційної особи для надання повідомленню більшої вагомості, однак не уточнюється, чи ці заяви підкріплено конкретними фактами. Це створює ефект «порожніх обіцянок». Повідомлення також містить ознаки «чорнобілого мислення»: зв'язування припинення вогню з негайним початком відновлення створює ілюзію, що ці два процеси автоматично йдуть один за одним. До того ж повідомлення може відволікати увагу від поточної ситуації на фронті. Тобто наведений текст містить апеляції до емоцій та надає суспільству оптимістичні, але недостатньо обґрунтовані прогнози.



Приклад 3. «Дефіцит електроенергії триватиме роками: українців закликають готуватися до найгіршого сценарію. Українці ще довгий час будуть відчувати дефіцит електроенергії, оскільки відновлення зруйнованої внаслідок російських масованих ракетних обстрілів енергетичної інфраструктури потребує багатомільярдних вливань. Ці витрати перекладуться на кінцевого споживача, що призведе до подорожчання вартості комунальних послуг. Про це заявив міністр енергетики та захисту навколишнього середовища України у 2019–2020 роках Олексій Оржель» (https://t.me/smolii_ukraine, 13.12.2024).

За результатами аналізу цього повідомлення чат виділяє фрази «дефіцит електроенергії триватиме роками», «українці ще довгий час будуть відчувати дефіцит електроенергії» та «українців закликають готуватися до найгіршого сценарію», які створюють емоційний тиск на читача. ChatGPT зазначає, що не наведено жодних деталей щодо джерел фінансування, планів відновлення чи альтернативних сценаріїв, натомість автор повідомлення концентрується виключно на негативних аспектах, не згадуючи про кроки, які вже роблять для відновлення енергетичної інфраструктури, або про підтримку з боку міжнародних партнерів. Це обмежує об'єктивне розуміння ситуації. Теза про те, що витрати буде перекладено на кінцевого споживача, може підсилювати негативні емоції, викликаючи недовіру до дій влади. Посилання на заяву колишнього міністра додає тексту ваги, проте брак прямої цитати залишає місце для ма-

ніпуляцій чи виривання слів з контексту. До того ж немає деталей про точний час і обставини, за яких зроблено заяву, а також не викладено інших експертних оцінок чи думок, що могли б дати повнішу картину. Тож текст містить ознаки пропаганди через емоційність окремих тверджень, брак конкретики та потенційне перекручування заяв посадовця. Брак комплексного аналізу проблеми може свідчити про намагання акцентувати увагу лише на негативних аспектах. Для точного аналізу необхідно провести фактчекінг і розглянути повідомлення в ширшому контексті інформаційного простору.

Приклад 4. «Понад 7 мільйонів придатних до служби російських чоловіків готові вирушити на війну з Україною за гроші, підрахував Європейський центр аналізу та стратегій (CASE), провівши опитування до РФ. Цього вистачить ще на кілька років війни без оголошення мобілізації. За даними опитування, 29,6 % росіян припускають, що добровільно поїдуть на війну з Україною. З них 13,2 % готові на це прямо зараз, решті для цього потрібні додаткові стимули — схвалення родичів, збільшення розміру виплат. 54,2 % не готові воювати за жодних умов. Цифру в 7,4 млн осіб CASE отримав, зіставивши відсоток готових воювати і заявлений 2022 року экс-міністром оборони РФ шойгу мобілізаційний ресурс РФ у 25 мільйонів осіб» (https://t.me/smolii_ukraine, 10.12.2024).

Аналізуючи повідомлення про готовність мільйонів росіян воювати, ChatGPT відзначає використання емодзі «злість», спрямо-



ване на створення негативного емоційного настрою, що не при-
таманно об'єктивній аналітиці, а
також подання даних у формі без-
апеляційного твердження: «цього
вистачить ще на кілька років вій-
ни». Текст зосереджується виключ-
но на загрозі мобілізації, але не
розглянуто політичних, економіч-
них чи соціальних чинників, які
могли б обмежити ці можливості.
Згадується Європейський центр
аналізу та стратегій (CASE), що
нібито додає авторитетності, але
не розкрито методологію опиту-
вання: який тип вибірки викори-
стовувався (репрезентативна чи
ні); як саме проводилося опиту-
вання в росії (де дослідження на
таку тему потенційно ускладне-
ні). Дані про 7,4 млн осіб отрима-
но шляхом математичних роз-
рахунків, проте число 25 млн,
наведене російським військовим
злочинцем, може бути завищеним
або спекулятивним. До того ж
висновки CASE не враховують
реальних умов мобілізації, рівня
підготовки чи інших чинників, що
ставить під сумнів точність про-
гнозу. У повідомленні зазначено,
що 54,2 % росіян не готові вою-
вати за жодних умов, однак цю
статистику подають як другоряд-
ну. Отже, у тексті використано
емоційно забарвлену риторичку;
дані подано вибірково або без
належного контексту, бракує зба-
лансованої аналітичної картини.

Приклад 5. «Від завтра по
Україні люті морози досягнуть
подекуди -14 градусів. Особливо
холодно на Заході країни. Якщо
враховувати постраждалу енер-
госистему та опалення, яке від
неї залежить, — запасаємось
теплим одягом та ковдрами. До
всього колапсу — МАГАТЕ пові-

домляє про зниження виробітку
електроенергії на 5 з 9 атомних
реакторів, через масовану атаку
рф» (@blacklist2477, 13.12.2024).

У цьому повідомленні Chat-
GPT помічає емоційний вплив
через використання екстремаль-
них описів: фраз «люті морози»,
«запасаємось теплим одягом та
ковдрами». Слово «колапс» за-
звичай означає повний розлад
або крах системи, що підвищує
драматичність ситуації й викли-
кає уявлення про неможливість
впоратися з кризою. Зазначен-
ня температурних показників («-14
градусів») без контексту щодо по-
рівняння з історичними даними
або інформацією про звичайні зи-
мові температури може підсилю-
вати відчуття надзвичайної ситу-
ації. Отже, текст має ознаки емо-
ційної маніпуляції.

Приклад 6. «Війна в Україні
закінчиться вже у наступному
році, — NYT. Видання зазначає,
що новий президент США До-
нальд Трамп хоче, щоб мирні
переговори почалися якомога
швидше. Крім того, розвідуваль-
ні служби вважають, що в Україні
скоро закінчатся солдати. Рес-
публіканці не хочуть схвалювати
додаткову допомогу Україні, і Київ
знає, що без істотного збільшен-
ня допомоги боротьба скоро за-
кінчиться (@thismyua, 13.12.2024).

Насамперед ChatGPT відзна-
чає посилання на авторитетне
видання The New York Times (NYT),
однак радить перевірити досто-
вірність цієї інформації, оскільки
текст не надає конкретного по-
силання на статтю чи дату пуб-
лікації. У повідомленні також на-
явні емоційні маркери та змісто-
ві викривлення. «Війна в Україні
закінчиться вже у наступному



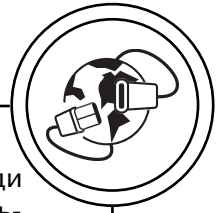
році» — категоричне твердження. Використання впевненого тону може створити в читача ілюзію щодо єдиного варіанта розвитку подій. «В Україні скоро закінчатся солдати» — узагальнення та створення враження катастрофічної ситуації, що підсилює страх і песимізм. «Республіканці не хочуть схвалювати додаткову допомогу Україні» — використання фрази «не хочуть» вказує на певну упередженість, адже реальні позиції є більш різноманітними. Фрази «Дональд Трамп хоче...», «розвідувальні служби вважають», «Київ знає» не підкріплюються конкретними джерелами. Вислів «Боротьба скоро закінчиться» має відтінок песимізму та викликає емоційну реакцію. Тож повідомлення містить категоричні твердження, неперевірені дані, припущення, які подано як факти, емоційно забарвлену лексику, посилання на авторитетне джерело без можливості прямої перевірки. Насамкінець ChatGPT ще раз радить перевірити першоджерело (NYT), а також оцінити альтернативні думки та об'єктивні дані щодо ситуації в Україні.

ChatGPT може слугувати не тільки інструментом для виявлення пропаганди, а й освітнім ресурсом, який пояснює аудиторії природу маніпулятивних технік. Інтеграція таких моделей, як ChatGPT, у системи моніторингу соціальних мереж дозволяє виявляти організовані дезінформаційні кампанії ще на ранніх етапах. Наприклад, автоматизоване виявлення схожих тем у тисячах повідомлень, розміщених на різних каналах протягом короткого часу, може свідчити про скоординовану ІПСО. Впроваджен-

ня подібних рішень у державній системі інформаційної безпеки України значно посилює її ефективність, особливо в умовах війни, коли швидкість реагування є критичною. Використання специфічних локальних наборів даних, наприклад емоційно забарвлених фраз, для додаткового навчання моделі може суттєво підвищити її ефективність у виявленні маніпуляцій та пропаганди.

Висновки

Як бачимо, ChatGPT продемонстрував високу ефективність у виявленні деяких ключових технік впливу на читача, і насамперед емоційних маркерів. Крім того, ChatGPT здатен ідентифікувати хибні аргументи, вибіркові факти та інші пропагандистські прийоми. Виявлення емоційного забарвлення тексту, упереджених узагальнень чи логічних помилок може бути автоматизоване через аналіз семантики, синтаксичних конструкцій і тональності повідомлень. Це створює можливість для розробки інструментів, що допоможуть журналістам, дослідникам і широкій аудиторії оперативно ідентифікувати загрози інформаційного впливу, підвищуючи критичне мислення та стійкість до маніпуляцій у цифрову епоху. Інтерактивні платформи на основі штучного інтелекту можуть стати ефективним інструментом для громадян. Такі платформи можуть мати функцію перевірки новин у реальному часі, оцінювання достовірності джерел й аналізу маніпулятивних технік. Наприклад, користувачі зможуть вводити но-



вини або чутки, які вони побачили в мережі, й отримувати детальний аналіз тексту, включаючи

маркери пропаганди, методи впливу та рекомендації з подальшої перевірки інформації.

Список використаної літератури/References

1. Voropaieva, T. S., & Aver'ianova, N. M. (2024). Shtuchnyi intelekt v systemi informatsiinoi bezpeky Ukrainy v umovakh ,osiisko-ukrainskoi viiny [Artificial Intelligence in the Information Security System of Ukraine During the Russo-Ukrainian War]. *Scientific Forum: Theory and Practice of Research*. Valencia, Kingdom of Spain: Collection of Scientific Papers 'SCIENTIA', 62–67. Retrieved on the 20th of December 2024 from http://previous.scientia.report/index.php/archive/article/view/2026/2042?utm_source=chatgpt.com [in Ukrainian].
2. Jones, D. G. (2024). Detecting Propaganda in News Articles Using Large Language Models. *Eng OA*, 2(1), 01–12. Retrieved on the 20th of December 2024 from <https://www.opastpublishers.com/open-access-articles/detecting-propaganda-in-news-articles-using-large-language-models.pdf> [in English].
3. Li, L., Fan, L., Atreya, S., & Hemphill, L. (2023). 'HOT' ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media. *arXiv:2304.10619* [cs.CL]. <http://doi.org/10.48550/arXiv.2304.10619> [in English].
4. Melnyk, T. (2023, May 31). SHI proty rosiiskykh IPSO. Ukrainnyi startup Osavul navchyv neiromerezhi voiuvaty z propahandoiu. Yak prodaty taku tekhnologiiu [AI against Russian IPSO. Ukrainian startup Osavul taught neural networks to fight propaganda. How to sell such technology]. *Forbes Ukraine*. Retrieved on the 20th of December 2024 from <http://forbes.ua/innovations/ai-proti-rosiyskikh-ipso-ukrainskiy-startap-osavul-navchiv-neiromerezhi-voyuvati-z-propagandoyu-yak-prodati-taku-tekhnologiyu-30052023-13928> [in Ukrainian].
5. Costello, T., Pennycook, G., & Rand, D. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), 10.1126/science.adq1814 [in English].
6. Lorian, R. (2023, January 30). Shtuchnyi intelekt yak superinstrument dlia dezinformatsii ta propahandy [Artificial Intelligence as a Super Tool for Disinformation and Propaganda]. Retrieved on the 20th of December 2024 from http://www.oporaua.org/polit_ad/shtuchnii-intelekt-iaksuperinstrument-dlia-dezinformatsiyi-ta-propagandi-24507 [in Ukrainian].
7. Stokel-Walker, C. (2023, March 27). We Spoke To The Guy Who Created The Viral AI Image Of The Pope That Fooled The World. *BuzzFeedNews*. Retrieved on the 20th of December 2024 from <http://www.buzzfeednews.com/article/chrisstokelwalker/pope-puffy-jacket-ai-midjourney-image-creator-interview> [in English].
8. Jose, J., & Greenstadt, R. (2024). Large language models fall short in detecting propaganda. *Proc. of the 18th International AAAI Conference on Web and Social Media, Workshop: CySoc 2024: 5th International Workshop on Cyber Social Threats*. <http://doi.org/10.36190/2024.06> [in English].
9. Klepper, D. (2023, January 24). It turns out that ChatGPT is really good at creating online propaganda: 'I think what's clear is that in the wrong hands there's going to be a lot of trouble'. *Fortune*. Retrieved on the 20th of December 2024 from <http://fortune.com/2023/01/24/chatgpt-open-ai-online-propaganda/> [in English].



10. De Vynck, G. (2024, May 30). OpenAI finds Russian and Chinese groups used its tech for propaganda campaigns. *The Washington Post*. Retrieved on the 20th of December 2024 from <http://www.washingtonpost.com/technology/2024/05/30/openai-disinfo-influence-operations-china-russia/> [in English].

11. Heppell, F., Bakir, M., & Bontcheva, K. (2024). Lying blindly: Bypassing ChatGPT's safeguards to generate hard-to-detect disinformation claims. *arXiv*. <http://doi.org/10.48550/arXiv.2402.08467> [in English].

12. The criminal use of ChatGPT — A cautionary tale about large language models (2024). *Business Insider*. Retrieved on the 20th of December 2024 from <http://www.europol.europa.eu/media-press/newsroom/news/criminal-use-of-chatgpt-cautionary-tale-about-large-language-models> [in English].

13. (2020). Krytychne myslennia dlia osvityan [Critical thinking for educators]. *Prometheus*. Retrieved on the 20th of December 2024 from <https://prometheus.org.ua/prometheus-free/krytychne-myslennya-dlya-osvityan/> [in Ukrainian].

14. Da San Martino, G., Cresci, S., Barron-Cedeno, A., Yu, S., Di Pietro, R., & Nakov, P. (2020). A survey on computational propaganda detection. *arXiv*. <http://doi.org/10.48550/arXiv.2007.08024> [in Ukrainian].

The article examines the potential of using ChatGPT as a tool for detecting, analyzing, and countering propaganda and disinformation in the context of the Russia-Ukraine war. The authors emphasize that the modern information space has become a critical battleground where manipulations, emotionally charged messages, distorted data, and propagandist techniques are widely employed to influence public opinion.

Keywords: artificial intelligence; ChatGPT; propaganda; Russia-Ukraine war; manipulation; emotional markers; critical thinking.

Надійшла до редакції: 23.12.24

Рецензія: 15.01.25

Опубліковано: 15.04.25