УДК 316.4:004.8]:355.02

Зубрицький Григорій Миколайович
кандидат технічних наук, доцент
військова частина А1915
м. Чортків, Україна
e-mail: red.hnups@gmail.com
ORCID: 0000-0002-9577-5268

ЗАГРОЗИ ТА РИЗИКИ ШТУЧНОГО ІНТЕЛЕКТУ І НЕЙРОМЕРЕЖЕВИХ АЛГОРИТМІВ ДЛЯ СИСТЕМИ СОЦІАЛЬНО-ПОЛІТИЧНИХ КОМУНІКАЦІЙ У РОСІЙСЬКО-УКРАЇНСЬКІЙ ВІЙНІ

Анотація. *У статті аналізуються загрози та ризики, пов'язані з використанням штучного інтелекту та нейромережевих алгоритмів у системі соціально-політичних комунікацій під час російсько-української війни. Показано, що сучасні технології штучного інтелекту, включаючи машинне навчання, обробку природної мови та генеративний штучний інтелект, стали потужними інструментами інформаційної війни, дозволяючи автоматизувати, масштабувати та підвищувати ефективність пропаганди, дезінформації та маніпуляцій. Здійснено аналіз окремих методів застосування штучного інтелекту Росією в війні проти України, спираючись на досвід попередніх конфліктів, зокрема в Сирії.*

Ключові слова: *штучний інтелект, соціально-політичні комунікації, загрози, ризики, воєнна безпека.*

Війну не можна виграти за допомогою зброї минулого покоління та застарілих методів.
В. Залужний

Вступ

Постановка проблеми. У 2013 році Міхал Косінскі разом зі своїми колегами опублікував у журналі Proceedings of the National Academy of Sciences статтю, присвячену аналізу цифрових слідів у соціальних мережах, який дозволяв з високою точністю визначати особисті характеристики користувачів [1]. При цьому було встановлено можливість не тільки виявляти особливості характеру, але й з великим ступенем ймовірності передбачати політичні уподобання користувача і навіть стать, сексуальну орієнтацію, колір шкіри.

На даний час розвиток великих даних та алгоритмів на основі штучного інтелекту (ШІ) дозволяє успішно формувати поляризацію думок та політичну упередженість шляхом створення маніпулятивної, сфабрикованої інформації, посилення емоційного контенту, підміни дискусій, ставлячи тим самим під загрозу систему соціально-політичних комунікацій [2-3].

Війна сприяє не тільки згуртованості нації навколо спільної керівної ідеї, а й спрощеному сприйняттю істини, редукованій картині світу за принципом розподілу свій-чужий, актуалізації механізмів віри замість критичного аналізу, крайньої емоційності сприйняття подій, проявам масової афектації [4]. У цих умовах підвищена масовизація та емоційність колективної свідомості стає вкрай вразливою для комунікативних маніпуляцій.

Російська агресія проти України супроводжується безпрецедентним використанням інформаційно-комунікаційних технологій, у яких особливе місце посідають штучний інтелект та нейромережеві алгоритми. Їхній вплив на соціально-політичні комунікації стає

все більш визначальним та вирішальним. Від генерації синтетичного медіаконтенту (діпфейків) до мікротаргетингу пропаганди та автоматизованого спостереження – штучний інтелект трансформує традиційні форми та способи інформаційної війни, створюючи нові загрози для національної безпеки.

Аналіз останніх досліджень та публікацій. Проблематика використання штучного інтелекту у військових конфліктах та інформаційній війні активно досліджується багатьма вченими. Методи та технології інформаційно-психологічного впливу з використанням ЗМІ, інтернету та соціальних мереж розглянуто в праці Валерія Солов'я [5]. Застосуванням ботів, тролів та астротурфінгу у соціальних мережах під час військових конфліктів присвячена книга П.В. Сингера (P. W. Singer) і Эмерсона Т. Брукинга (Emerson T. Brooking) [6]. Загрози генеративного штучного інтелекту та дипфейків аналізуються у роботі Д. Роберта Чесни (Robert Chesney) і Даниэль Китс Ситрон (Danielle Keats Citron) [7]. Дослідження NewsGuard Technologies аналізує створення клонів сайтів відомих ЗМІ за допомогою штучного інтелекту досліджувалось фахівцями [8-9]. Проте комплексний аналіз загроз та ризиків, що створюються саме ШІ та нейромережевими алгоритмами для системи соціально-політичних комунікацій у контексті російсько-української війни, потребує поглибленого вивчення. Аналіз цих загроз та ризиків необхідний для вироблення адекватних стратегій протидії та захисту інформаційного суверенітету держави та визначає **актуальність статті**.

Метою статті є аналіз загроз та ризиків, пов'язаних з використанням штучного інтелекту та нейромережевих алгоритмів у системі соціально-політичних комунікацій в умовах російсько-української війни.

Виклад основного матеріалу

У рамках даної статті під штучним інтелектом (ШІ) розуміється сукупність технологій, що дозволяють комп'ютерним системам імітувати когнітивні функції людини, такі як навчання на масивах даних (Machine Learning, ML), розпізнавання патернів, обробка природної мови (Natural Language Processing, NLP), комп'ютерний зір (Computer Vision) та генерація нового контенту (Generative AI). Особливе значення мають нейромережеві алгоритми глибокого навчання (Deep Learning), що є технологічною основою багатьох сучасних ШІ-застосунків у соціально-політичних комунікаціях.

Штучний інтелект успішно використовується у військовій сфері для аналізу розвідданих, логістики, управління автономними системами, моделювання операцій та бойових дій, систем підтримки прийняття рішень. Однак, у контексті інформаційного протиборства його роль стала помітною з розвитком соціальних мереж та генеративних моделей.

Інформаційні операції, що проводяться Росією, використовують традиційні практики радянських "активних заходів" та адаптовані до цифрової епохи [10-11]. Сучасна російська військова доктрина розглядає інформаційний простір як критично важливе поле бойових дій, інтегруючи інформаційні інструменти в загальну концепцію гібридної війни [12-13]. У цьому контексті штучний інтелект розглядається Російською Федерацією не просто як нова технологія, а як потужний каталізатор, здатний значно підвищити ефективність існуючих методів інформаційних операцій за рахунок автоматизації, масштабування, покращення таргетування, створення реалістичного синтетичного контенту та інтеграції з кіберопераціями.

Війна в Сирії, що розпочалася у 2011 році, стала одним із перших полігонів для апробації та широкомасштабного застосування цих технологій у соціально-політичних комунікаціях. Інформаційні операції у сирійській війні проводили не лише Росія та Іран на підтримку режиму Башара Асада, а також Ісламська держава (ІДІЛ). Російські операції були спрямовані на дискредитацію "білих касок", просування ідеї про боротьбу з "терористами", посилення антизахідних настроїв, а також на заперечення застосування

хімічної зброї режимом Асада або звинувачення в цьому опозиції [14]. Ісламська держава досить ефективно використовувала соціальні мережі для демонстрації сили, вербування, виправдання насильства та створення образу "халіфату". Основними майданчиками для цих операцій були Twitter, Facebook, YouTube, а також Telegram [15].

Основними методами проведення інформаційних операцій були [14-15]:

– поширення пропаганди та дезінформації з використанням мереж ботів та "фабрик тролів" для масового поширення наративів, фейкових новин, сфабрикованих свідчень очевидців;

– дискредитація опонентів шляхом розгортання кампаній з очорнення репутації супротивників, поширення чуток, звинувачень у сфабрикованих військових злочинах;

– астротурфінг – створення ілюзій масової підтримки та народного гніву за допомогою безлічі фейкових акаунтів, що імітують реальних користувачів;

– маніпулювання трендами – використання ботів для виведення певних хештегів у топ обговорень у соціальних мережах;

– придушення інакомислення шляхом атак на акаунти активістів та журналістів, масові скарги на їхній контент з метою блокування.

В умовах гібридної війни, розпочатої Росією проти України ще задовго до відкритої військової агресії, соціально-політичні комунікації стали основною ціллю інформаційного протиборства. Цілі включають підрив довіри до державних інститутів України, деморалізацію населення та Збройних сил, маніпулювання громадською думкою всередині країни та на міжнародній арені. З 2022 року ця війна стала першим повномасштабним конфліктом, де можливості сучасного штучного інтелекту застосовуються систематично для впливу на інформаційний простір [16].

Основними ризиками та загрозами, пов'язаними з використанням штучного інтелекту в соціально-політичних комунікаціях в умовах триваючої війни, є:

– автоматизований аналіз великих даних (ШІ-спостереження);

– маніпуляція масовою свідомістю через ШІ-таргетинг;

– поширення дезінформації із застосуванням генеративного ШІ (діпфейки та фейкові тексти);

– непрозорість алгоритмічних рішень та алгоритмічна заражуваність;

– концентрація "цифрової влади" в іноземних технoплатформ.

Автоматизований аналіз великих даних (ШІ-спостереження) і як наслідок втрата приватності в умовах відсутності дієвих механізмів контролю та захисту даних багаторазово підвищує вразливість цивільного населення. Існують загрози безпеки для активістів, журналістів, родичів військовослужбовців та звичайних громадян, чиї дані можуть бути використані для фільтраційних заходів, примусу до співпраці, відстеження переміщень та контактів осіб, що становлять інтерес для спецслужб, викрадень чи фізичного усунення. Так, Росією в окупованих містах України для пошуку ветеранів АТО/ООС, активістів та співробітників силових структур використовувалась система розпізнавання облич [17].

Маніпуляція масовою свідомістю через ШІ-таргетинг. Нейромережеві алгоритми аналізують цифрові сліди користувачів (лайки, репости, коментарі, пошукові запити, історію переглядів) для створення детальних психологічних та поведінкових профілів. На основі цих профілів ШІ може підбирати та доставляти контент (новини, відео), який з найбільшою ймовірністю викличе потрібну емоційну реакцію, вплине на думку або підштовхне до певної дії. Це може бути як державна пропаганда, так і дезінформація, спрямована на розпалювання паніки, ворожнечі або недовіри до влади. Передумовами цього є масове використання населенням соціальних мереж як основного джерела інформації в умовах війни. Так, згідно з опитуванням Internews Ukraine, проведеним у 2024 році, 84% українців використовували соціальні мережі для отримання новин, при цьому 42% опитаних отримували новинний контент виключно через соціальні мережі

[18]. При цьому постійний потік маніпулятивного контенту, створення інформаційного шуму та перевантаження може виснажувати критичне мислення і робити людей більш сприйнятливими до пропаганди [4, 19]. Оцінити пряму ефективність ШІ-таргетингу у зміні поведінки складно, але його здатність доставляти "правильне" повідомлення "правильній" людині у "правильний" час значно підвищує потенціал впливу порівняно з традиційною масовою пропагандою.

Поширення дезінформації із застосуванням генеративного ШІ (діпфейки та фейкові тексти). Нейромережі на кшталт GPT використовуються для автоматичного створення великих обсягів текстового контенту – фейкових новин, коментарів під статтями, постами в соцмережах (астротурфінг) та пропагандистських статей. Це дозволяє імітувати суспільну дискусію, створювати видимість підтримки певних ідей, заглушати реальні голоси. Генеративний ШІ виробляє дезінформацію швидше, дешевше і в значно більших масштабах, ніж будь-коли раніше. Навіть якщо окремі фейки згодом викриваються, їхній потік може підірвати довіру до інформації загалом ("ефект інформаційної втоми" та "дивіденд брехуна"). Реалістичність діпфейків (особливо аудіо) зростає, роблячи їх все важчими для виявлення без спеціальних інструментів. Ефективність полягає не тільки в обмані, у створенні хаосу, сумнівів, емоційного виснаження, а й у підриві самої можливості верифікації інформації [4]. Це особливо небезпечно в умовах бойових дій і значно ускладнює оперативне прийняття правильних рішень.

Непрозорість алгоритмічних рішень та їхня заражуваність. Складність сучасних нейромережевих моделей ускладнює розуміння того, чому ШІ ухвалив те чи інше рішення. Наприклад, чому алгоритм модерації YouTube чи Facebook позначив конкретне відео про війну як таке, що порушує правила, а інше, схоже, залишив? Ця непрозорість ("black box problem") в умовах інформаційної війни може призводити до випадкового або навмисного придушення важливої інформації. Однією з причин цього є навчання ШІ на джерелах, які можуть відображати задані стереотипи та здійснювати необхідний дисбаланс у наданні інформації, тобто заражуваність алгоритмічних рішень [20].

Концентрація "цифрової влади" в іноземних технoплатформ. Facebook (Meta), Twitter (X), YouTube (Google), Telegram відіграють колосальну роль у поширенні інформації про війну. Концентрація влади у техногігантів – серйозний виклик сучасності. Їхні алгоритми рекомендацій та модерації контенту безпосередньо впливають на те, що бачать та читають користувачі. Рішення цих компаній про видалення або маркування контенту, пов'язаного з пропагандою війни Російської Федерації або з висвітленням дій української армії, мають величезне соціально-політичне значення. Однак політика цих техногігантів не є абсолютно прозорою і часто критикується за непослідовність та можливий зовнішній вплив. Так, результати досліджень показують, що політика модерації Telegram, однієї з ключових платформ для поширення як новин, так і дезінформації під час війни, є найменш прозорою [21].

Найвідомішими випадками використання штучного інтелекту та нейромережевих алгоритмів у системі соціально-політичних комунікацій під час російсько-української війни є:

– діпфейк Зеленського в березні 2022 року, коли було поширено відео, на якому обличчя президента накладено на інше тіло, а звернення "президента" містило заяву про капітуляцію;

– операція "Doppelgänger" / "Secondary Infektion". Масштабна російська кампанія з дезінформації, виявлена Meta, EU DisinfoLab та іншими дослідниками, включала створення клонів сайтів авторитетних західних ЗМІ (The Guardian, Bild) та публікацію на них фейкових статей з проросійською та антиукраїнською пропагандою. Незважаючи на те, що пряме використання генеративного ШІ для написання всіх статей не доведено, масштаб та швидкість операції, а також використання таргетованої реклами для

просування розроблених фейкових сайтів вказують на високий ступінь автоматизації та можливе застосування ШІ-інструментів для управління цією кампанією [8-9].

– масоване "зараження" даних провідних чат-ботів на основі штучного інтелекту російською мережею дезінформації "Pravda" (не пов'язана з російською газетою "Правда"). У звіті NewsGuard Technologies, опублікованому в березні 2025 року, вказується, що лише у 2023 році мережа "Pravda" опублікувала 3,6 мільйона статей, що містять дезінформацію [22]. При цьому 10 провідних чат-ботів (включаючи ChatGPT-4 від OpenAI, Gemini від Google, Meta AI від Meta, Copilot від Microsoft, Grok від xAI, Perplexity та інші) використовують хибні наративи цієї мережі. У 32-33,5% випадків при відповідях на відповідні запити чат-боти відтворювали такі хибні твердження, як:

– відповідальність України за масові вбивства в Бучі;

– наявність в Україні секретних "біолабораторій", що фінансуються США;

– підтримка США "неонацистів" в Україні;

– заборона Зеленським в Україні соціальної мережі Truth Social.

При цьому сім із десяти протестованих чат-ботів безпосередньо посилалися на статті мережі "Pravda" як на джерело своєї інформації.

Наведені вище приклади використання ШІ в системі соціально-політичних комунікацій під час російсько-української війни переконливо показують, що ШІ є активним фактором у війні та вимагає постійного вдосконалення існуючих підходів до кібербезпеки та протидії дезінформації.

Безумовно для мінімізації ризиків та загроз необхідним є:

– створення систем швидкого реагування на ШІ-загрози, розробка протоколів для оперативного виявлення, аналізу та спростування небезпечних ШІ-фейків та інформаційних атак;

– правове регулювання прозорості алгоритмів модерації соціальних мереж та технологічних компаній;

– впровадження надійних та стандартизованих методів цифрового маркування автентичного контенту;

– популяризація основ фактчекінгу, основних принципів роботи ШІ та алгоритмів.

На даний час для вирішення цих завдань в Україні створено Центр протидії дезінформації при РНБО, який координує зусилля з боротьби з фейками, також аналогічні завдання вирішуються багатьма громадськими організаціями. Однак масштаб і складність цих загроз в умовах війни вимагають не окремих заходів, а цілісної, багаторівневої стратегії протидії. Такий підхід повинен інтегрувати передові технологічні рішення, законодавче регулювання, міжнародне співробітництво, широкі освітні програми для підвищення стійкості суспільства до маніпуляцій, а також дотримання етичних норм та прав людини.

Висновки

Штучний інтелект та пов'язані з ним нейромережеві алгоритми перетворилися на потужний і багатофункціональний інструмент у гібридній війні, яку Росія веде проти України в інформаційному просторі. Їх використання створює комплексні та багатогранні загрози для системи соціально-політичних комунікацій, національної безпеки та суспільної стабільності.

Спектр загроз є надзвичайно широким: від застосування ШІ для спостереження та збору даних про громадян на окупованих територіях і в тилу (що підриває приватність, створює ризики переслідувань, викрадень та інших злочинів), до масового виробництва та цільового поширення дезінформації.

Ситуацію значно ускладнює непрозорість алгоритмів, що використовуються технологічними платформами для модерації контенту та формування новинних стрічок

(“проблема чорної скриньки”). Крім того, існує ризик “алгоритмічної заражуваності”, коли ШІ навчається на даних, що містять упередження або навмисно спотворену інформацію,

Значна концентрація влади у невеликої кількості іноземних технологічних гігантів створює додаткову вразливість, оскільки їхні корпоративні політики та рішення щодо модерації контенту мають прямий і суттєвий вплив на інформаційний простір України, але не завжди є прозорими чи узгодженими з національними інтересами безпеки.

Протидія цим загрозам вимагає комплексного та системного підходу через масштабність та технологічну складність викликів. Необхідна розробка та впровадження цілісної, багаторівневої національної стратегії. Лише такий інтегрований підхід дозволить ефективно мінімізувати ризики, пов’язані з використанням ШІ в інформаційній війні, та захистити український інформаційний простір і демократичні процеси.

Список літератури

1. Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*. 2013. № 110 (15). P. 5802–5805. URL: https://doi.org/10.1073/pnas.1218772110.
2. Цвык В. А., Цвык А. В. Искусственный интеллект и политическая коммуникация: риски и возможности. *Вестник Московского университета. Серия 12. Политические науки*. 2022. № 2. С. 94-118.
3. Nemitz P. Artificial intelligence in the European Union: ethics and governance. *Council of Europe Publishing*. URL: https://rm.coe.int/artificial-intelligence-in-the-european-union-ethics-and-governance/168093c836 (date of access: 03.04.2025).
4. Hotsur O., Danylina O., Zozulia N., Stiekolshchikova V., Porpulit O., Danko-Sliptsova A. How does information manipulation hinder normal brain functioning? Violation of neuroethics in wartime mass media. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience.* 2023. № 14(3). P. 224–240. URL: https://doi.org/10.18662/brain/14.3/472.
5. Соловей В. Д. Абсолютное оружие. Основы психологической войны и медиаманипулирования. Москва : Эксмо, 2015. 320 с.
6. Singer P. W., Brooking E. T. LikeWar: The Weaponization of Social Media. Houghton Mifflin Harcourt, 2018. 421 p.
7. Chesney R., Citron D. K. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*. 2019. № 107(6). P. 1753-1820. URL: http://dx.doi.org/10.2139/ssrn.3213954.
8. Quarterly Adversarial Threat Report: April - June 2022. *Meta.* URL: https://about.fb.com/news/2022/08/meta-quarterly-adversarial-threat-report-q2-2022/ (date of access: 03.04.2025).
9. Quarterly Adversarial Threat Report: July - September 2022. *Meta*. URL: https://about.fb.com/news/2022/09/removing-coordinated-inauthentic-behavior-from-china-and-russia/ (date of access: 03.04.2025).
10. Reed T. Active Measures: The Secret History of Disinformation and Political Warfare. New York: Farrar, Straus and Giroux, 2020. 560 p.
11. Giles K. Russia’s ‘New’ Tools for Confronting the West: Continuity and Innovation in Moscow’s Information Warfare. Russia and Eurasia Programme. Chatham House: The Royal Institute of International Affairs, 2016. 71 p.
12. Darczewska J., Żochowski P. Russia’s strategy of disinformation: A conceptual approach. Centre for Eastern Studies (OSW), 2017.
13. Galeotti M. Russian Political War: Moving Beyond the Hybrid. Routledge, 2018. 136 p.
14. Nocetti, J. Dazed and confused: Russian “information warfare” and the Middle East – The Syria lessons. *EuroMeSCo*. URL: https://www.euromesco.net/wp-content/uploads/2019/02/Brief93_Dazed-and-oncufsed.-Russian-information-warfare.pdf (date of access: 03.04.2025).
15. Berger J. M., Stern, J. ISIS: The State of Terror. Ecco, 2015. 416 p.
16. Artificial intelligence in warfare. Think Tank. *European Parliament*. URL: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)754599 (date of access: 03.04.2025).
17. Torture, Disappearances in Occupied South: Abusive “Filtration” Process Targets Civilians Fleeing Mariupol. *Human Rights Watch.* URL: https://www.hrw.org/news/2022/07/21/ukraine-torture-disappearances-occupied-south (date of access: 03.04.2025).

18. Internews in Ukraine. Ukrainian Media, Use and Trust in 2024. *Internews in Ukraine*. URL: https://internews.in.ua/wp-content/uploads/2024/11/USAID-Internews-Media-Report-2024.pdf (date of access: 03.04.2025).
19. Prier J. Commanding the Trend: Social Media as Information Warfare. *Strategic Studies Quarterly*. 2017. № 11(4). P. 50–74. URL: https://www.airuniversity.af.edu/portals/10/ssq/documents/volume-11_issue-4/prier.pdf (date of access: 09.04.2025).
20. Mehrabi N., Morstatter F., Saxena N., Lerman K., Narayanan A. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*. 2021. № 54(6). P. 1-35. URL: https://doi.org/10.1145/3457607.
21. What's wrong with Telegram? *New Eastern Europe*. URL: https://neweasterneurope.eu/2024/09/17/whats-wrong-with-telegram/ (date of access: 03.04.2025).
22. Leading AI Chatbots Found to Spread Russian Disinformation and Other False Narratives About News Events. *NewsGuard Technologie*. URL: https://www.newsguardrealitycheck.com/p/special-report-top-10-generative?utm_source=chatgpt.com (date of access: 03.04.2025).