

ВИКЛИКИ ШТУЧНОГО ІНТЕЛЕКТУ ДЕМОКРАТІЇ ТА ДЕМОКРАТИЧНИМ ПРОЦЕСАМ ПРИЙНЯТТЯ РІШЕНЬ

Виділено та проаналізовано основні загрози з боку штучного інтелекту демократії та процесам прийняття рішень. Це насамперед такі основні засади демократичного режиму, як представництво, підзвітність та довіра. Показано, що основна небезпека з боку ШІ полягає у його можливості генерувати неправдиву інформацію та посилювати упередженість у ході прийняття рішень, що може призвести до виключення з демократичних процесів окремих осіб або груп громадян.

Відзначено, що за умови впровадження відповідних запобіжних заходів ця технологія може сприяти автономії особистості; прозорості та нагляду; підзвітності та відповідальності; рівності статі у нагоді демократичним інститутам та значно покращити процес прийняття рішень. Ці заходи мають ґрунтуватися на принципах поваги до людської гідності та недискримінації; конфіденційності та захисту персональних даних; надійності та безпечності.

Ключові слова: штучний інтелект, демократія, участь, прийняття рішень, дезінформація, довіра, інклюзивність.

Vitushko Maksym. Challenges of artificial intelligence to democracy and democratic decision-making processes

The article identifies and analyses the main threats posed by artificial intelligence to democracy and decision-making processes. First of all, it concerns such basic principles of the democratic regime as representation, accountability and trust. The main danger posed by AI is its ability to generate false information and increase bias in decision-making, which can lead to the exclusion of individuals or groups of citizens from democratic processes.

It is noted that, subject to the implementation of appropriate safeguards, this technology can help democratic institutions and significantly improve the decision-making process, including individual autonomy; transparency and oversight; accountability and responsibility; and equality. These measures should be based on the principles of respect for human dignity and non-

discrimination; confidentiality and protection of personal data; and reliability and security.

Key words: *artificial intelligence, democracy, participation, decision-making, disinformation, trust, inclusiveness.*

Технології штучного інтелекту (ШІ) мають значний потенціал для зміцнення демократичних цінностей, інститутів і процесів, удосконалення процесів прийняття рішень. Потенційний вплив включає розвиток розуміння політики серед громадян, що сприяє розширенню участі в демократичних дебатах та глибшій залученості до процесів прийняття рішень. Аналогічно політичні актори, державні службовці або представники громадськості можуть встановити тісніші зв'язки з окремими особами, що в кінцевому підсумку підвищує їх здатність ефективніше представляти інтереси громадськості. Така координація має потенціал трансформувати виборчі кампанії та значно покращити процес прийняття рішень, сприяючи більшій інклюзивності, прозорості та ефективності.

Побоювання щодо зловживання сучасними інформаційно-комунікаційними технологіями, і зокрема ШІ, у політиці існують уже давно, але ті, що безпосередньо пов'язані з демократією та виборами, посилилися в останнє десятиліття. Нещодавні приклади використання ШІ свідчать про те, що вони можуть становити численні загрози демократії та правам людини, слугуючи потужним інструментом фрагментації публічної сфери, підриваючи громадянську участь і довіру до демократичних інститутів. Такі інструменти можуть дозволити користувачам, у тому числі зловмисникам, поширювати дезінформацію, яка може ослабити довіру до інформації (в тому числі через використання контенту, створеного ШІ, або маніпулювання за його допомогою автентичним контентом), ускладнити доступ до інформації; ухвалювати упереджені рішення щодо окремих осіб, що може призвести до дискримінаційних практик; впливати на судові рішення, що може мати потенційні наслідки для системи правосуддя, а також здійснювати незаконне або довільне стеження, яке може призвести до обмежень свободи зібрань, свободи слова, а також недоторканності приватного життя.

Вплив штучного інтелекту на суспільно-політичні процеси активно досліджується зарубіжними дослідниками, зокрема А. Агаравлом [1], Ф. Паскуале [2], С. Расселлом [3], С. Зубофф [4]

та низкою інших. Тема штучного інтелекту ще мало розроблена вітчизняними вченими, особливо гуманітаріями. У 2023 р. вийшла друком колективна монографія, присвячена використанню ШІ в у сфері безпеки та оборони, науковій діяльності та освіті; медицині; промисловості та енергетиці, телекомунікаційній галузі, транспорті та інфраструктурі, сільському господарстві та екології [5]. Однак вплив ШІ на політичні процеси в Україні практично не досліджений, що і зумовлює актуальність даного дослідження.

Метою статті є аналіз потенційних небезпек використання штучного інтелекту для демократичних інститутів та процесів прийняття рішень.

Генеративний ШІ загрожує трьома основним стовпам демократичного режиму: представництву, підзвітності та довірі [6].

1. Загроза представництву. Для ефективного реагування керівництва держави на запити громадян вони повинні чітко їх розуміти. Зазвичай вподобання виборців з тих чи інших питань можна дізнатися через соціологічні опитування, проведення різноманітних громадських обговорень та консультацій, аналіз соцмереж та інших інструментів. Однак широке використання ШІ може спотворювати уявлення про думку громадян шляхом генерування фальшивих електронних листів, дописів у соцмережах із наперед визначеною позицією. Доступність масштабної розсилки сфальшованих ШІ повідомлень нівелює важливість громадського обговорення як одного з методів представництва інтересів громадян шляхом виявлення важливих проблем і пошуку способів їх розв'язання.

2. Загроза підзвітності. Функціонуюча демократія також передбачає наявність механізмів підзвітності – можливість громадян вимагати від чиновників відповідати за свою діяльність насамперед через вільні та чесні вибори [7]. З допомогою ШІ можна максимально деталізувати профіль виборця, виокремити групу людей із нестійкими політичними уподобаннями і за допомогою таргетованої рекламної кампанії чи кастомізованого контенту в мережі запропонувати саме ту виборчу програму кандидата чи політичної партії, що найкраще відповідає їх потребам [8, с. 686]. В умовах демократії поразка на виборах є результатом неправильної оцінки настроїв виборців, а використання ШІ дасть змогу мінімізувати цю помилку. До того ж агрегація потреб виборців у реальному часі

з широким використанням ІКТ перетворить політичні партії на рудименти політичної системи.

У поєднанні з новими можливостями мікротаргетингу використання ІІІ дає потужний імпульс дезінформаційним кампаніям, зробивши їх набагато ефективнішими, ніж спроби вплинути на вибори більш традиційними методами – рекламними кампаніями, агітаційними турами кандидатів тощо [9].

3. Загроза довірі. Сучасний стан розвитку ІІІ дозволяє зловмисникам продукувати дезінформацію, в тому числі контент, мікротаргетований на певні демографічні групи і навіть окремих осіб, у великих масштабах. Популярність соціальних мереж дає змогу без особливих зусиль поширювати дезінформацію, в тому числі ефективно спрямовувати її на конкретні аудиторії.

Ризик також полягає в тому, що з поширенням неавтентичного аудіовізуального контенту в Інтернеті люди можуть просто не знати, чому вірити, і тому не довірятимуть усій інформаційній екосистемі. Довіра до медіа вже є вкрай низькою, а поширення інструментів, які можуть генерувати неавтентичний контент, ще більше ослабить цю довіру. Це, своєю чергою, може ще більше знизити і без того досить низький рівень довіри до уряду. У листопаді 2023 р. у країнах ОЕСР більша частка людей (44%) мали низьку або нульову довіру до свого уряду, ніж високий або помірно високий рівень довіри (39%) [10, р. 23]. Однак існують значні відмінності між країнами та групами. Порівняно з 2021 р. спостерігається помірне зростання частки тих, хто не довіряє, та зменшення частки тих, хто довіряє уряду на високому або помірно високому рівні. Для 18 країн, у яких рівень довіри до національного уряду вимірювався у 2021 та 2023 роках, частка тих, хто має високу або помірно високу довіру, зменшилася з 43 до 41%. Зростання частки тих, хто не довіряє або довіряє дуже мало, майже рівнозначне – з 40 до 43%. Однак ці дані приховують важливі зміни в рівнях довіри в окремих країнах. Зокрема, частка з високим або помірно високим рівнем довіри суттєво зросла в Бельгії, Колумбії, Австралії, Канаді, Франції, Латвії та Швеції. У Фінляндії та Норвегії ця частка суттєво знизилася. У кількох країнах (Люксембург, Мексика та Швейцарія) більшість людей мають високий або помірно високий рівень довіри до національного уряду, тоді як приблизно в третині країн – менше, ніж кожен третій.

Соціальна довіра – це важливий клей, який скріплює демократичні суспільства. Вона стимулює громадянську активність і політичну участь, зміцнює довіру до політичних інститутів і сприяє повазі до демократичних цінностей, що є важливим запобіжником відступу від демократії та авторитаризму [11]. Високий рівень довіри громадян до органів публічної влади може мотивувати їх до політичної участі та залучення, а також заохочувати до протидії загрозам демократичним інститутам і практикам.

Своєю чергою, дезінформація є серйозною загрозою для суспільства і демократії загалом, оскільки вона ефективно змінює і маніпулює фактами, щоб створити соціальний зворотний зв'язок, який підриває будь-яке відчуття об'єктивної істини [8, с. 686]. Дебати про те, що є реальним, швидко переростають в обговорення того, хто має право вирішувати, що є реальним, що призводить до перегляду владних структур, які часто слугують укоріненим інтересам. Водночас різноманітні процедури фактчекінгу здійснюються із застосуванням ШІ, що зумовлює необхідність його введення у правові рамки.

У відповідь на ці виклики демократії з боку ШІ 17 травня 2024 р. Рада Європи прийняла перший в історії міжнародний юридично обов'язковий договір у сфері ШІ – Рамкову конвенцію про штучний інтелект і права людини, демократію та верховенство права [12]. Вона спрямована на забезпечення відповідності діяльності систем ШІ правам людини, демократії та верховенству права, а також сприяла технологічному прогресу та інноваціям. Ця Рамкова конвенція зосереджена на захисті та просуванні прав людини, демократії та верховенства права і прямо не регулює економічні та ринкові аспекти систем ШІ. Вона забезпечує загальну правову основу на глобальному рівні для дотримання чинних міжнародних і національних правових зобов'язань, які застосовуються державами-учасниками у сфері прав людини, демократії та верховенства права. і спрямована на забезпечення того, щоб діяльність систем ШІ будь-якої форми власності відповідала цим обов'язкам, стандартам і зобов'язанням.

Рамкова конвенція визначає такі принципи використання ШІ.

1. Повага до людської гідності та автономії особистості (ст. 7): використання ШІ не повинно призводити до дегуманізації людей, підривати їхню самостійність або зводити їх до набору даних, а

також антропоморфізувати системи ШІ. Людська гідність вимагає визнання складності та багатства людської ідентичності, досвіду, цінностей та емоцій. Відстоювання людської гідності передбачає повагу до невідчужуваної цінності та вартості кожної людини, незалежно від її походження, особливостей чи обставин, і стосується, зокрема, способу, в який слід ставитися до всіх людей.

Автономія особистості є одним із важливих аспектів людської гідності й стосується здатності людей до самовизначення, тобто їхньої здатності робити вибір і приймати рішення, в тому числі без примусу, і вільно жити своїм життям. У контексті ШІ незалежність особи передбачає наявність у людей контролю над використанням і впливом технологій ШІ на своє життя, при цьому їхня свобода дій і автономія не зменшується. Антропоцентричне регулювання визнає важливість надання людям можливості формувати свій досвід роботи зі ШІ, гарантуючи, що ці технології зміцнюють, а не ослаблюють їхню автономію, що особливо важливо з огляду на здатність ШІ до імітації та маніпулювання.

2. Прозорість та нагляд (ст. 8): процеси прийняття рішень і загальна робота систем ШІ мають бути зрозумілими та доступними для відповідних суб'єктів, а також за певних умов для відповідних зацікавлених сторін. У деяких випадках це також може стосуватися надання додаткової інформації, наприклад, про алгоритми, що використовуються, з урахуванням міркувань безпеки, комерційної, інтелектуальної власності та інших міркувань. Засоби забезпечення прозорості залежать від низки чинників, таких, як тип системи ШІ, контекст її використання або її роль, а також досвід відповідного суб'єкта або зацікавленої сторони.

Цей принцип також передбачає заходи щодо ідентифікації контенту, створеного ШІ для уникнення ризику обману та уможливлення розрізнення автентичного контенту, створеного людиною, і штучного генерованого контенту, оскільки людям стає дедалі важче його ідентифікувати. Такі заходи можуть включати такі методи, як маркування та водяні знаки, що передбачають вбудовування розпізнаваного підпису у вихідні дані системи ШІ, за умови доступності цих технологій та їхньої доведеної ефективності, загально-визнаного рівня розвитку техніки та особливостей різних типів контенту. Сприяння використанню технічних стандартів, ліцензій із відкритим вихідним кодом і співпраці дослідників і розробників

підтримує розвиток прозоріших систем штучного інтелекту в довгостроковій перспективі.

Нагляд стосується впровадження різних механізмів, процесів і структур, призначених для моніторингу, оцінки та керівництва діяльністю протягом життєвого циклу систем штучного інтелекту. Потенційно вони можуть складатися з правових, політичних і нормативних рамок, рекомендацій, етичних принципів, кодексів практики, програм аудиту та сертифікації, інструментів виявлення та пом'якшення упередженості.

3. Підзвітність та відповідальність (ст. 9): необхідність забезпечення механізмів, які дозволять особам, організаціям або установам, відповідальним за діяльність систем ШІ, нести відповідальність за негативний вплив на права людини, демократію або верховенство права, спричинений їх діяльністю. Зокрема, це вимагає створення нових або удосконалення чинних рамок і механізмів, які можуть бути застосовані до діяльності систем ШІ для виконання цієї вимоги. Це також може включати судові та адміністративні заходи, режими цивільної, кримінальної та іншої відповідальності, а в державному секторі – адміністративні та інші процедури, що дозволяють оскаржувати рішення, або покладання конкретних обов'язків і зобов'язань на операторів.

Цей принцип підкреслює необхідність чіткого розподілу відповідальності та можливості простежити дії та рішення до конкретних фізичних або юридичних осіб таким чином, щоб визнати різноманітність відповідних суб'єктів, їхніх ролей та обов'язків. Це важливо для запровадження механізму виявлення наслідків негативного впливу ШІ на права людини, демократію або верховенство права та належного розподілу відповідальності. Тобто незалежно від форми власності (державна чи приватна) всі суб'єкти, відповідальні за діяльність протягом життєвого циклу систем ШІ, повинні підпорядковуватися чинній системі правил, правових норм, щоб уможливити ефективне визначення відповідальності.

4. Рівність і недискримінація (ст. 10): передбачає недопущення технічної та соціальної упередженості при використанні ШІ. Технічна упередженість виникає в ході застосування машинного навчання, що призводить до додаткових упереджень, які не присутні в даних, що використовуються для навчання системи або прийняття рішень. Соціальна упередженість – нездатність належним чином

врахувати історичну або поточну нерівність у суспільстві в діяльності ІІІ, зокрема у розробці та навчанні моделей. До таких нерівностей належать, наприклад, історичні та структурні перешкоди на шляху до гендерної рівності та справедливого ставлення до осіб, що належать до груп, які були або досі залишаються частково недостатньо забезпеченими послугами, зазнають дискримінації або інших форм нерівності.

5. Конфіденційність та захист персональних даних (ст. 11). Збір персональних даних вже є повсюдним не лише як основа бізнес-моделей у багатьох галузях, але й як один із ключових видів діяльності органів державної влади, зокрема правоохоронних, які використовують різноманітні технології та автоматизовані системи, що збирають, обробляють та генерують персональні дані у процесах прийняття рішень. Оскільки системи ІІІ в основному керуються персональними даними, то за відсутності належних запобіжних заходів це може становити серйозні ризики для приватного життя людей.

6. Надійність (ст. 12): впровадження стандартів, технічних специфікацій, методів забезпечення і схем відповідності в оцінці та перевірці надійності систем ІІІ. Стандарти можуть забезпечити надійну основу для обміну спільними очікуваннями щодо певних аспектів продукту, процесу, системи або послуги з метою побудови обґрунтованої впевненості в надійності системи ІІІ, якщо її розробка і використання відповідають цим стандартам. Цей принцип підкреслює важливість запровадження заходів, спрямованих на забезпечення надійності систем ІІІ за допомогою заходів, що стосуються ключових аспектів функціонування (надійність, безпека, захищеність, точність і продуктивність), а також функціональних передумов, таких, як якість і точність даних, цілісність даних, безпека даних і кібербезпека. Відповідні стандарти, вимоги, схеми забезпечення та дотримання вимог є передумовою зміцнення довіри громадськості до технологій ІІІ.

7. Безпечність інновації (ст. 13): заохочення і сприяння інноваціям відповідно до прав людини, демократії та верховенства права. Одним зі способів стимулювання є надання органам влади у відповідному секторі діяльності можливості створювати «контрольоване середовище» або «рамки», що дозволяють розробляти, навчати, експериментувати й тестувати інновації під безпосереднім нагля-

дом компетентних органів, зокрема для заохочення врахування на ранніх стадіях питань якості, недоторканності приватного життя та інших прав людини, а також питань безпеки і захисту інформації. Це особливо важливо, оскільки певні ризики, пов'язані з ШІ, можуть бути ефективно вирішені лише на етапі проєктування.

Отже, використання технологій ШІ може призвести до ескалації напруженості або підриву суспільної довіри, яка є основним елементом ефективного демократичного уряду та запорукою ефективного процесу прийняття рішень. ШІ може генерувати неправдиву інформацію або призвести до виключення з демократичних процесів окремих осіб або тих, хто може бути недостатньо представлений або перебувати у вразливому становищі. Отже, попри свої позитивні аспекти ШІ несе в собі значний ризик негативного впливу на демократичний процес, прийняття рішень органами публічної влади та реалізацію відповідних прав людини. У найбільш узагальненому вигляді ШІ загрожує трьом основним засадам демократичного режиму: представництву, підзвітності та довірі. Однак за умови впровадження відповідних запобіжних заходів ці технології можуть виявитися корисними для демократії. Ці заходи мають ґрунтуватися на принципах поваги до людської гідності та автономії особистості; прозорості та нагляду; підзвітності та відповідальності; рівності й недискримінації; конфіденційності та захисту персональних даних; надійності та безпечності.

1. Agarwal A., Gans J.S., Goldfarb A. Prediction machines: The simple economics of artificial intelligence. Harvard: Harvard Business Review Press, 2018. 2. Pasquale F. The black box society: The secret algorithms that control money and information. Harvard: Harvard University Press, 2015. 3. Russell S.J. Human compatible: Artificial intelligence and the problem of control. New York: Penguin Press, 2019. 4. Zuboff S. The age of surveillance capitalism: The fight for a human future at the new frontier of power. New York: Profile Books, 2019. 5. Стратегія розвитку штучного інтелекту в Україні: монографія / за заг. ред. А.І. Шевченка. Київ: ІПШІ, 2023. 305 с. 6. Kreps S., Kriner D. How AI Threatens Democracy. *Journal of democracy*. 2023. Vol. 34. №4. P. 122-131. URL: <https://www.journalofdemocracy.org/articles/how-ai-threatens-democracy/> 7. Przeworski A., Stokes S.C., Manin B. Democracy, Accountability, and Representation. New York: Cambridge University Press, 1999. 8. Стойко О.М. Політико-правові наслідки використання штучного інтелекту. *Правова держава*. 2023. Вип. 34. С. 684-693. 9. Kreps S. The Role of Technology in Online Misinformation. Brookings Institution, June 2020. URL:

www.brookings.edu/articles/the-role-of-technology-in-online-misinformation.
10. OECD Survey on Drivers of Trust in Public Institutions – 2024 Results. URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/07/oecd-survey-on-drivers-of-trust-in-public-institutions-2024-results_eeb36452/9a20554b-en.pdf **11.** Levitsky S., Ziblatt D. *How Democracies Die*. New York: Crown, 2019. **12.** The Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law. URL: <https://rm.coe.int/1680afae3c>

***Vitushko Maksym.* Challenges of artificial intelligence to democracy and democratic decision-making processes**

Artificial intelligence (AI) technologies have significant potential to strengthen democratic values, institutions and processes, and improve decision-making. Potential impacts include the development of a deeper understanding of politics among citizens, which contributes to increased participation in democratic debate and deeper engagement in decision-making processes. Likewise, political actors, civil servants or members of CSO can establish closer ties with individuals, ultimately enhancing their ability to represent the public interest more effectively. Such coordination has the potential to transform electoral campaigns and significantly improve decision-making, contributing to greater inclusiveness, transparency and efficiency.

The article identifies and analyses the threats posed by AI to the three main pillars of a democratic regime: representation, accountability and trust. 1. Threat to representation. For the government to respond effectively to citizens' requests, they must clearly understand them. Usually, voters' preferences on certain issues can be found out through opinion polls, various public discussions and consultations, analysis of social media and other tools. However, the widespread use of AI can distort the view of citizens' opinions by generating fake emails and social media posts with a predetermined position.

Threat to accountability. A functioning democracy also implies the existence of accountability mechanisms – the ability of citizens to hold officials accountable for their activities primarily through free and fair elections. In a democracy, electoral defeat is the result of an incorrect assessment of voter sentiment, and the use of AI will help minimise this error. Combined with new microtargeting capabilities, AI gives a powerful boost to disinformation campaigns, making them much more effective than attempts to influence elections using more traditional methods such as advertising campaigns, candidate campaign tours, etc.

Threat to trust. The current state of AI development allows attackers to produce disinformation, including content microtargeted at certain demographic groups and even individuals, on a large scale. The popularity of social media makes it easy to spread disinformation, including effectively targeting specific audiences.

In response to these challenges to democracy posed by AI, on 17 May 2024, the Council of Europe adopted the first ever international legally binding treaty in the field of AI – The Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law. It aims to ensure that AI systems comply with human rights, democracy and the rule of law, and to promote technological progress and innovation. This Framework Convention focuses on the protection and promotion of human rights, democracy and the rule of law and does not explicitly address the economic and market aspects of AI systems. It provides a common legal framework at the global level for the application of existing international and national legal obligations applicable to States Parties in the areas of human rights, democracy and the rule of law and aims to ensure that the activities of AI systems of any form of ownership comply with these duties, standards and obligations. If appropriate safeguards are in place, the use of AI can be beneficial to democracy. These measures should be based on the principles of respect for human dignity and individual autonomy; transparency and oversight; accountability and responsibility; equality and non-discrimination; privacy and personal data protection; and reliability and security.

Key words: artificial intelligence, democracy, participation, decision-making, disinformation, trust, inclusiveness.