

Філософія

УДК 316.4

DOI <https://doi.org/10.5281/zenodo.18849900>

**Від Азімова до GPT: еволюція проблем безпеки та відповідальності в
системах штучного інтелекту**

Дзьобань Олександр Петрович,

доктор філософських наук, професор,
головний науковий співробітник Державної наукової установи
«Інститут інформації, безпеки і права Національної академії
правових наук України», м. Харків, Україна,
<https://orcid.org/0000-0002-2075-7508>

Прийнято: 12.02.2026 | Опубліковано: 28.02.2026

***Анотація.** У статті здійснена спроба комплексного соціально-філософського аналізу еволюції проблем безпеки та відповідальності в системах штучного інтелекту (ШІ) від класичних «законів робототехніки» Аїзека Азімова до сучасних генеративних моделей типу GPT-4. Обґрунтовується необхідність виходу за межі вузькоетичного трактування технологічних ризиків, пропонуючи розглядати їх як фундаментальні виклики для людської ідентичності, екзистенції та соціальної структури.*

Метою статті є теоретичне обґрунтування та аналізу трансформації інституційних форм освіти в умовах мережевого суспільства, виявлення особливостей просторової, комунікативної та ризикованої динаміки як стійких (традиційних), так і рухливих

(слабоінституціоналізованих) освітніх структур у контексті цифровізації та глобалізації соціального буття.

Досягненню поставленої мети сприяло застосування наукових методів і підходів за принципом взаємодоповнюваності. Рух дослідницької методології здійснювався від філософського рівня методології (діалектичний метод) до загальнонаукових методів і прийомів (історико-порівняльного аналізу, аналізу і синтезу, абстрагування й формалізації).

У статті доведено, що «перший закон» Азімова залишається центральною проблемою, проте сучасні системи ШІ створюють нові форми загроз — від втрати людиною сенсу буття через автоматизацію до підриву критичного мислення та творчої унікальності. Виокремлено низку проблем: непрозорість алгоритмів, упередженість (дискримінація), дефіцит відповідальності та загрозу конфіденційності. На прикладі GPT-4 продемонстровано суперечливість сучасного ШІ, який, попри здатність імітувати творчу діяльність, залишається «вузьким» інструментом, не здатним на справжню інноваційність та розуміння смислів.

У висновках підкреслюється, що людина залишається єдиним джерелом справжньої творчості, а подолання технологічних загроз можливе лише через системну гуманітарну експертизу та чітке визначення меж делегування повноважень алгоритмам.

Ключові слова: *штучний інтелект, GPT-4, безпека, відповідальність, творча унікальність, екзистенційні ризики, цифрова трансформація.*

**From Asimov to GPT: The Evolution of Security and Responsibility
Issues in Artificial Intelligence Systems**

Oleksandr Dzoban,

Doctor of Philosophical Sciences, Professor,
Chief research fellow State Scientific Institution “Institute of Information,
Security and Law of the National Academy of Legal Sciences of Ukraine”
Kharkiv, Ukraine, <https://orcid.org/0000-0002-2075-7508>

***Abstract.** The article attempts a comprehensive socio-philosophical analysis of the evolution of security and responsibility issues in artificial intelligence (AI) systems, ranging from Isaac Asimov's classic "laws of robotics" to modern generative models such as GPT-4. The author substantiates the necessity of transcending the narrow ethical interpretation of technological risks, proposing to view them as fundamental challenges to human identity, existence, and social structure.*

The purpose of the article is the theoretical substantiation and analysis of the transformation of institutional forms of education within a network society, identifying the peculiarities of spatial, communicative, and risk dynamics of both stable (traditional) and fluid (weakly institutionalized) educational structures in the context of the digitalization and globalization of social existence. Achieving this goal was facilitated by the application of scientific methods and approaches based on the principle of complementarity. The research methodology progressed from the philosophical level (the dialectical method) to general scientific methods and techniques, including historical-comparative analysis, analysis and synthesis, abstraction, and formalization.

The study proves that Asimov's "First Law" remains a central concern; however, contemporary AI systems create novel forms of threats—from the loss of

human meaning due to automation to the undermining of critical thinking and creative uniqueness. A series of issues are identified: the opacity of algorithms, bias (discrimination), a deficit of responsibility, and threats to privacy. Using the GPT-4 model as an example, the article demonstrates the contradictory nature of modern AI, which, despite its ability to mimic creative activity, remains a "narrow" tool incapable of genuine innovation or an understanding of meanings.

The conclusions emphasize that humanity remains the sole source of authentic creativity, and overcoming technological threats is possible only through systematic humanitarian expertise and a clear definition of the boundaries for delegating authority to algorithms.

Keywords: *artificial intelligence, GPT-4, security, responsibility, creative uniqueness, existential risks, digital transformation.*

Постановка проблеми. Сучасна епоха характеризується всеохопним проникненням штучного інтелекту в усі сфери буття соціального, що створює як нові стимули для розвитку людини, так і серйозні загрози для її майбутнього. Останні роки позначені стрімкою інтеграцією штучного інтелекту (ШІ) у повсякденну практику, що відкриває безпрецедентні можливості для самореалізації людини та покращення якості життя. ШІ сьогодні є не лише інструментом прогресу, а й потужним фактором трансформації людської цивілізації. Водночас експансія технологій ШІ несе в собі приховані загрози, здійснюючи суттєвий вплив на соціум та довкілля. ШІ постає як складний феномен, здатний трансформувати фундаментальні пласти культури та спричиняти масштабні соціокультурні зрушення – як конструктивні, так і деструктивні. Як багатогранне явище, він виступає каталізатором змін, наслідки яких для людини можуть бути неоднозначними і така неоднозначність дедалі посилюється.

У цьому контексті філософська думка має виробити критичне ставлення до впровадження інтелектуальних систем, проаналізувати супутні ризики та їхній вплив на світогляд сучасної особистості. Будь-який технологічний прогрес у сфері ШІ є виправданим лише тоді, коли він узгоджується з ідеалами гуманізму. Пріоритетом розвитку ШІ має залишатися добробут людства; натомість надмірне покладання на алгоритми загрожує втратою життєвих орієнтирів та ідентичності, а неконтрольоване делегування функцій алгоритмам може підірвати екзистенційні основи людського буття, загострюючи питання про сенс життя.

Сьогодні існує запит на нову парадигму взаємодії з ШІ, яка б гарантувала гуманістичний шлях розвитку суспільства. Необхідність системного аналізу принципів взаємодії людини та інтелектуальних систем зумовлена потребою у створенні безпечного соціуму. Постає гостра потреба у глибокій філософській рефлексії взаємодії людини та ШІ, визначенні етичних кордонів та соціальних перспектив такого симбіозу. Попри наявність великої кількості досліджень у цій галузі, теоретичне осмислення новітніх розробок суттєво відстає від темпів технологічного прогресу. Соціально-філософський фундамент для безпечного впровадження систем ШІ наразі залишається недостатньо розробленим.

Аналіз останніх досліджень і публікацій. Результати такого аналізу свідчать, що феномен штучного інтелекту вже тривалий час перебуває у фокусі міждисциплінарного дискурсу. Значний внесок у розуміння природи інтелектуальних систем зробили сучасні теоретики цифровізації та постгуманізму (Н. Бостром, М. Тегмарк, Ю. Харарі). Попри значну кількість праць, теоретичне осмислення стрімких змін, що відбуваються всередині сучасних систем (зокрема поява великих мовних моделей типу GPT-4), все ще істотно відстає від темпів їхнього впровадження у соціальну практику.

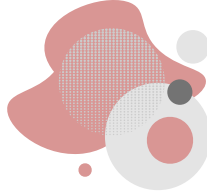
Виділення невирішених раніше частин загальної проблеми. Більшість існуючих досліджень фокусується на вузькотехнічних або суто етичних аспектах, тоді як комплексна соціально-філософська рефлексія цифрової трансформації суспільних відносин залишається фрагментарною.

Формулювання цілей статті (постановка завдання). Метою статті є спроба вийти за межі вузькоетичного трактування проблем штучного інтелекту і розкрити їхній глибинний соціально-філософський характер, проаналізувавши трансформацію взаємовідносин людини та ШІ (на прикладі моделі GPT-4) у контексті збереження людської ідентичності та творчої унікальності.

Виклад основного матеріалу дослідження. Обговорення проблем застосування систем ШІ необхідно розпочати зі звернення до класики, до так званого «першого закону робототехніки», сформульованого ще письменником-фантастом Айзеком Азімовим в оповіданні «Хоровод», що вийшло друком у далекому 1942 році: робот не може заподіяти шкоду людині або своєю бездіяльністю допустити, щоб людині була заподіяна шкода [1].

Поширюючи дію першого закону на системи ШІ, підкреслимо, що останні створюються людиною для вирішення важливих для неї завдань, досягнення поставлених нею цілей, а не для продукування додаткових труднощів, перешкод і проблем. Тим більше, вони ні за яких умов не повинні перетворюватися на загрозу для людини, її здоров'я, безпеки, життя, для її сьогодення і майбутнього. Навпаки, вони покликані полегшити життя людини, вивільнивши її від рутинних і нецікавих, або надзвичайно трудомістких операцій, для творчих і розкриваючих її безмежний творчий потенціал занять.

Проте, найбільш актуальні сфери застосування ШІ свідчать про те, що у процесі проектування, впровадження і поширення систем ШІ неминуче



виникає низка специфічних проблем, перелік яких у міру розвитку сфери застосування систем ШІ постійно розширюється.

Традиційно ці проблеми прийнято позначати терміном «етичні», однак, на наш погляд, вони далеко виходять за межі власне етичної проблематики. Будемо характеризувати їх як соціально-філософські, оскільки вони охоплюють широкий комплекс питань, що торкаються в цілому взаємодії людини зі штучним інтелектом в різних сферах суспільного життя, породжуючи крім власне етичних, моральних проблем, екзистенційні, аксіологічні та інші проблеми. Крім того, їх дія погіршує стан і становище людини в світі. Отже, вони вимагають осмислення і вирішення не тільки з етичних міркувань, але, головним чином, з точки зору соціальної філософії.

Так, екзистенційна проблема втрати людиною сенсу буття може виникнути, наприклад, у процесі застосування систем ШІ у сучасному виробництві. Людині, яка втратила роботу через заміну її роботом, часто складно знайти інші форми успішної соціалізації, і вона відчуває свою непотрібність, незатребуваність у соціумі, що породжує питання про призначення, цінність і сенс її життя. Аксіологічний аспект впливу технологій ШІ на світорозуміння сучасної людини пов'язаний з тим, що штучний інтелект сприймається у суспільстві як абсолютна, безумовна цінність, витісняючи на узбіччя буття традиційні цінності людського життя.

Не можна не відзначити особливий клас проблем, породжених відносною новизною феномену ШІ, все ще існуючим браком вивченості досвіду застосування його систем у соціальній практиці і, як наслідок, недостатньою поінформованістю людини про реальні та потенційні можливості ШІ.

Проблеми застосування ШІ не раз ставали предметом наукової дискусії. За її підсумками вченими отримано значущі результати і опубліковано огляди найбільш важливих, на їхню думку, питань. Так,

Дж. Боссманн виділила дев'ять основних проблем у сфері ШІ: безробіття, нерівність, людяність, штучна дурість, расистські роботи, безпека, злі джини, сингулярність і права роботів [2]. М. Райан і Дж. Антоніу ведуть мову вже про сімнадцять подібних викликів [3, р. 16]. Т. Хагендорф у своєму фундаментальному дослідженні різних етичних настанов [4], випущених до 2020 р., виявив проблеми непрозорості, порушення конфіденційності, дискримінації, загрози безпеці, відсутності відповідальності. Доповідь «Штучний інтелект і життя у 2030 р.» [5], випущена Стенфордським університетом у 2016 р., серед нагальних проблем застосування систем ШІ називає такі: загроза безпеці, конфіденційності особистих даних, проблема порушення кримінальної та цивільної відповідальності, безробіття, сертифікація систем ШІ. Ці ж проблеми згадуються у Резолюції Європейського парламенту «Норми цивільного права» [6].

Інститутом інженерів електротехніки та електроніки (Institute of Electrical and Electronics Engineers, IEEE) були опубліковані «Рекомендації з етично обґрунтованого проектування» (Ethically Aligned Design) [7] у яких міститься перелік соціальних та етичних проблем впровадження систем ШІ: проблема визначення відповідальності, непрозорості, недостатньої поінформованості; проблема неуніверсальності та конфліктності етичних норм, яка залежить від характеру завдань та відмінностей між сферами застосування; проблема досягнення довіри між людьми та системами ШІ; проблема непередбаченої поведінки систем ШІ; труднощі у забезпеченні безпеки роботи систем ШІ; проблема конфіденційності персональних даних; проблема забезпечення прозорості у роботі систем ШІ; проблема юридичної відповідальності у разі заподіяння системами ШІ шкоди людині.

Ускладнення та поширення систем ШІ призводить до питання наділення цих систем юридичною відповідальністю, суб'єктністю. Інакше неможливо зрозуміти, хто відповідатиме перед законом, якщо в результаті

використання системи ШІ виникнуть негативні наслідки і буде завдано шкоди будь-кому з учасників інциденту.

Системи ШІ активно впроваджуються у так звані «розумні міста», і широке використання технологій у повсякденному житті призводить до значних змін у матеріальному та духовному житті людини.

М. Шерер [8] та С. Рід [9] звернули увагу фахівців на проблему розподілу відповідальності у контексті застосування систем ШІ у розумних містах. Вони пишуть про те, що системи ШІ мають бути завжди регульованими, підзвітними та забезпеченими механізмами компенсації збоїв. Йдеться саме про такий вид регулювання, яке стоїть на варті людських цінностей, забезпечуючи незаподіяння шкоди життю людини, сталий розвиток та захист навколишнього середовища, дотримання вимог соціальної справедливості. Ці суспільні цінності, на їхню думку, повинні мати безумовний пріоритет перед економічними цінностями, міркуваннями прибутку та вигоди.

Про те, що економічні цілі корпорацій та інтереси політичних еліт не повинні бути пріоритетним завданням для ШІ, також писали Л. Флоріді, Дж. Коулс, Т. Кін та М. Таддео [10]. Вони переконані, що ШІ є спільним ресурсом, покликаним працювати виключно на благо кожної людини та суспільства загалом.

На окрему увагу заслуговує проблема непрозорості, що стала відчутною у зв'язку з аналізом великих даних, оскільки операції з ними неможливо контролювати так, як це було можливо у випадку з простим комп'ютерним кодом.

У літературі також простежуються і спроби з'ясування причин непрозорості [11]. При цьому зазвичай розрізняється навмисна непрозорість, що виникає, наприклад, у тих випадках, коли уряд або корпорації, переслідуючи якісь власні цілі, приймають рішення залишити послідовність

дій системи ШІ недоступною для громадськості, громадської думки та непрозорість, обумовлена масштабами застосування системи ШІ, коли принципи машинного навчання або кількість задіяних акторів роблять алгоритм її дій непрозорим навіть для експертів.

Т. Ахмад, А. Тередесей та К. Еккерт [12] розглянули проблему непрозорості систем ШІ у зв'язку з питанням про недовіру до ШІ. Вони вважають, що системи ШІ вже стали невід'ємною частиною повсякденного життя людини, але при цьому принцип їхньої роботи залишається оповитим завісою таємниці. Таким чином, непрозорий характер прийняття рішень системами ШІ підриває довіру до них з боку суспільства та його інститутів.

Особливої гостроти зазначені проблеми набувають у процесі використання чату GPT 4¹, заснованого на ШІ. Чат-бот із штучним інтелектом GPT 4 – це генеративна мовна модель, запущена комерційною компанією Open AI 14 березня 2023 р. як розвиток раніше розроблених чат-ботів GPT. Ця модель називається генеративною, оскільки здатна створювати нові дані, а не лише аналізувати наявні. Chat GPT 4 не лише розпізнає шаблони, а й використовує їх для створення нових даних.

У даний час чат-бот пише вірші та пісні, готує слайди і презентації у певному стилі, успішно проходить співбесіду з приводу прийому на роботу в Google, пише маркетингові кампанії для певної демографічної групи, коментарі до онлайн-ігор і створює зображення з високою роздільною здатністю, вибудовує прогнози, фінансову аналітику, генерує найпростіший код, спілкується багатьма світовими мовами. Продукти, отримані GPT 4, майже не відрізняються від контенту, створеного людиною, оскільки ця

¹ Незважаючи на те, що компанія OpenAI у серпні 2025 року випустила мовну модель GPT-5, заявивши про більш потужні можливості міркування, кодування та написання коду, а також про більш адаптивний дизайн, у цій статті мова йтиме про GPT-4 як найбільш зручний, популярний і функціональний на сьогоднішній день сервіс ШІ. GPT-5 – це мультимодальна мовна модель, розроблена компанією OpenAI, і є моделлю за замовчуванням, яка використовується в чат-боті ChatGPT. Вона покликана забезпечити більш високу точність, здатність до міркування та адаптивність у порівнянні зі своїми попередниками, перемикаючись між швидкими відповідями та більш глибоким аналізом залежно від завдання.

система навчається на основі всієї інформації, доступної в даний час в Інтернеті. Тому, на перший погляд, GPT 4 виглядає повноцінним партнером людини у творчій роботі зі створення інноваційних продуктів.

Переваги цього нового інструменту штучного інтелекту у порівнянні з його попередниками очевидні, проте важливо бачити і розуміти зворотний бік цього явища і проблеми, які воно породжує. Так, у відносно недавньому мультидисциплінарному дослідженні, що об'єднало 43 статті експертів з різних наукових сфер, стверджується, що застосування Chat GPT може призвести до репутаційних та юридичних ризиків, використання образливого контенту або контенту, захищеного авторським правом, втрати конфіденційності, шахрайства. Загрози, що походять від Chat GPT 4 і подібних ШІ-ботів, пов'язані з проблемою непрозорості, можуть сприяти дискримінації та упередженням, порушення авторських прав, плагіату або сфабрикованому ботами неавтентичному текстовому контенту, проблемі поширення підроблених медіа [13].

У складних видах діяльності, що мають творчий та інноваційний характер, GPT 4 сьогодні не здатний створювати справді оригінальний, новий продукт. Австралійський рок-музикант, поет, письменник, автор музики до численних кінофільмів Нік Кейв ще кілька років тому отримав від одного зі своїх шанувальників текст пісні, який був написаний Chat GPT на основі конкретної підказки: «У стилі Ніка Кейва». Розчарований Кейв відповів шанувальнику: «Написання гарної пісні – це не мімікрія, не реплікація і не стилізація ...» [14]. Л. Джена і С. Гоял вважають, що стати по-справжньому творчими набагато більше шансів мають системи з формами емоційного інтелекту [15].

Бум поширення чат-бота знову актуалізував побоювання, що стосуються страху зростаючої залежності людей від подібних програм і поступового делегування їм людиною вирішення все більш складних і

значущих завдань [16]. Наприклад, у країнах, що розвиваються, де не вистачає експертів у тих чи інших галузях знань, Chat GPT 4 може замінити їх, генеруючи необхідну користувачам інформацію. Широке використання чат-бота в освіті вже сьогодні вимагає перебудови та адаптації навчального процесу, роботи викладачів, наукових співробітників та керівництва університетів до умов цифрового середовища, що швидко змінюється. Цілком очевидно, що подібні технології цілком можуть використовуватися і широко використовуються студентами здобувачами освіти для написання наукових праць, що породжує проблеми з плагіатом та академічною доброчесністю.

Крім очевидних результатів застосування зазначеної технології, що лежать на поверхні, необхідно враховувати ризик виникнення більш глобальних і віддалених у часі негативних наслідків. Йдеться, наприклад, про можливу відмову учнів здобувачів освіти (і такі випадки сьогодні, на жаль, далеко не поодинокі) від поглибленого вивчення та критичного аналізу різних аспектів предмета, що вивчається, внаслідок використання ними Chat GPT 4, що дозволяє отримати швидкий результат без реального занурення та оволодіння новими навичками та знаннями. Досвід останніх років свідчить про правоту суджень С. О'Коннора, який наголосив на тому, що ШІ може призвести до придушення критичного мислення, творчого потенціалу, абсолютно необхідних для розвитку людини [17].

Проблема непрозорості у зв'язку із застосуванням Chat GPT 4 виявляється у тому, що від розробника і тим більше користувача прихованими залишаються механізми роботи бота, системи ранжування і рекомендації. GPT 4, як і багато інших систем ШІ, може використовувати дані, що становлять у суспільстві упередження та помилки, сприяючи посиленню проблеми соціальної несправедливості.

Д. Вінбергер писав, що «упередженість – це первородний гріх машинного навчання» [18]. Як відомо, знання ми отримуємо внаслідок формулювання пояснень та перевірки їх на відповідність реальності, тоді як система спирається на всю сукупність даних, перенесених людиною до мережі. Отже, уникнути повною мірою упередженості на даному етапі розвитку технології ШІ неможливо. Chat GPT 4 фактично будує свої міркування на основі інформації, яка далеко не завжди є надійною та достовірною. Не здатний до справжнього розуміння, він займається зіставленням, компілюванням великих обсягів даних, що є у наявності у відкритих на навчання джерелах. Згенерована GPT 4 інформація вимагає ретельної перевірки, оскільки його програмне забезпечення створює неточний і хибний контент, який може бути цілком достовірним, але насправді є хибним результатом логічно некоректних висновків.

Ця система, крім іншого, нівелює значущість фактів і виводить на поверхню глобальних інформаційних потоків будь-яку інформацію, у тому числі й конфіденційну, що породжує проблеми порушення конфіденційності, а також проблему підзвітності у тому випадку, коли організація не знає і не може визначити, хто повинен нести відповідальність за недостовірну інформацію в роботі. Доступність даної технології для масового споживача та розробників через мови програмування (наприклад, Python), хмарні сервіси (наприклад, Amazon Web Services, Microsoft Azure, Google) дозволяє застосовувати її так само широко, як, наприклад, програми Excel та Access. Ця легкість і простота кратно збільшують ризик заподіяння шкоди людині внаслідок використання інформації, згенерованої даною системою, яка не має нічого спільного з дійсним рішенням поставлених людиною завдань.

Цілком очевидно, що зазначені проблеми, породжені застосуванням ШІ, вимагають соціально-філософського осмислення, оскільки за своїм змістом вони відповідають основній інтенції соціально-філософського

дослідження. Соціальна філософія традиційно вивчає з позицій цілісності та системності універсальні, сутнісні риси суспільства, найбільш загальні закони його динаміки, глибинні причини тих чи інших подій та процесів, перспективи розвитку соціуму. Але всі ці знання їй необхідні для з'ясування місця та ролі людини, яка виступає як творець світу культури, світу цінностей і норм, необхідних для спільного буття великих груп людей, функціонування соціальних інститутів, існування соціальної реальності загалом. Інакше кажучи, у фокусі соціально-філософського дослідження завжди знаходиться людина як центральний елемент світобудови, творець і носій усіх властивостей і якостей соціального, без якого дискурс про суспільство не має сенсу. Отже, проблеми, що виникають у процесі взаємодії людини з системами ШІ повинні бути введені в коло питань, що досліджуються соціальною філософією, оскільки обговорення та вирішення цих питань стосується основ буття людини у світі, її становища по відношенню до інших складових соціальної реальності, а також найближчих та віддалених перспектив її розвитку.

Соціально-філософське дослідження також необхідне для виявлення комплексної природи проблем ШІ, для пошуку їх розв'язання не тільки з позицій етики, аксіології, філософської антропології та інших галузей знань про людину та суспільство, а з позицій розуміння цілісності соціального життя людей, реалізації гуманістичних ідеалів, діяльнісної природи, цільової функції людини, її здатності до природо перетворювальної діяльності і обумовленого цим центрального становища людини у соціальній дійсності.

Висновки. Проблема загрози безпеці або заподіяння шкоди, сформульована ще А. Азімовим як «перший закон робо техніки», виступає в ролі центральної, системоутворюючої соціально-філософської проблеми застосування технології ШІ, по відношенню до якої інші є такими, що

розкривають можливе джерело, причини виникнення шкоди, або його різновидів, варіанти реалізації.

До найбільш часто обговорюваних питань і труднощів застосування ШІ сучасні автори і дослідники відносять проблеми порушення автономії людини, соціальної несправедливості як такі, що представляють аспекти, форми, грані, різновиди цієї шкоди і проблеми порушення конфіденційності, відсутності відповідальності, непрозорості, що викликають можливе заподіяння шкоди внаслідок використання технологій ШІ.

Безумовно, запропонований перелік проблем застосування ШІ не є вичерпним. Унаслідок повсюдного проникнення систем ШІ у різні сфери суспільного життя, перманентного вдосконалення технології та постійно зростаючого впливу їх на життя та діяльність людини цей перелік може змінюватись, поповнюватись, уточнюватись. Крім того, цілком очевидно, що соціальні проблеми можуть бути спричинені не лише застосуванням ШІ. Інші технології, не пов'язані з ШІ, також можуть бути джерелом цих проблем, однак використання ШІ здатне посилити дію інших причин, а також стати самостійною причиною погіршення становища людини у світі.

Розглянутий приклад Chat GPT 4 підтверджує тезу про принципово «вузький» характер сучасного ШІ, який, незважаючи на свої видатні характеристики, все ж таки, не може стати повноцінною заміною людині. Людина залишається по відношенню до штучного розуму оригіналом, творцем справді нового. Технології відтворюють і використовують результати діяльності людини, але перевершити її незбагненну з позицій раціональності природу вони не в змозі. При цьому, створений GPT 4 контент у вигляді фейкових новин, пропаганди, дезінформації вже сьогодні може вводити людей в оману, посилюючи упередження, що існують у суспільстві, розпалюючи ворожнечу, ненависть, підриваючи соціальну

згуртованість, довіру людей один до одного і до результатів розвитку нових технологій.

Усе це обґрунтовує необхідність системного, цілісного, виваженого дослідження впливу ШІ на суспільство і людини засобами та методами гуманістично орієнтованої соціально-філософської інтерпретації процесів взаємодії людини зі штучним розумом, що відкриває можливість відшукування шляхів вирішення, способів подолання вже існуючих і тих, що лише формуються в цій галузі проблем та перешкод.

Список використаних джерел

1. Asimov I. Runaround. URL: https://web.williams.edu/Mathematics/sjmiller/public_html/105Sp10/handouts/Runaround.html (date of access: 17.12.25).
2. Bossmann Dzh. Top 9 Ethical Issues in Artificial Intelligence. URL: <https://www.weforum.org/stories/2016/10/top-10-ethical-issues-in-artificial-intelligence/> (date of access: 17.12.25).
3. Ryan M., Antoniou J. et al. Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality. *Science and Engineering Ethics*. 2021. Vol. 27 (2). DOI:10.1007/s11948-021-00293-x.
4. Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines*. 2020. Vol. 30. P. 99-120.
5. Stone P., Brooks R. et al. Artificial Intelligence and Life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University*. Stanford, CA, September 2016.
6. Report with recommendation to the Commission on Civil Law Rules on Robotics. URL: <https://www.europarl.europa.eu/committees/en/report-with->

recommendations-to-the-commi/product-details/20170202CDT01121/ (date of access: 15.12.25).

7. Ethically Aligned Design A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version 2. URL: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf/ (date of access: 14.12.25).

8. Scherer M.U. Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. DOI:10.2139/ssrn.2609777.

9. Reed C. How should we regulate artificial intelligence? DOI: 10.1098/rsta.2017.0360.

10. Floridi L., Cowls J., Kin T.C., Taddeo M. How to design AI for social good: Seven Essential factors. *Science and Engineering Ethics*. 2020. Vol. 26. P. 1771-1796. DOI:10.1007/s11948-020-00213-5.

11. Etzioni A., Etzioni O. Should artificial intelligence be regulated? URL: <https://issues.org/perspective-artificial-intelligence-regulated/> (date of access: 17.12.25).

12. Ahmad M.A., Teredesai A., Eckert C. Fairness, accountability, transparency in AI at scale: Lessons from national programs. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020. Vol. 27 (30). P. 690-699.

13. Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. DOI:10.1016/j.ijinfomgt.2023.102642.

14. Cain S. «This song sucks»: Nick Cave responds to ChatGPT song written in style of Nick Cave”. URL: https://www.reddit.com/r/Music/comments/10e9639/this_song_sucks_nick_cave_responds_to_chatgpt/ (date of access: 18.12.25).

15. Jena L. K., Goyal S. Emotional intelligence and employee innovation: Sequential mediating effect of person-group fit and adaptive performance Intelligence émotionnelle et innovation des employés : effet médiateur séquentiel de l'ajustement personne-groupe et de la performance adaptative. *European Review of Applied Psychology*. 2022. Vol. 72 (1). URL: <https://www.sciencedirect.com/science/article/abs/pii/S1162908821001079> (date of access: 27.11.25).

16. Baird A., Maruping L.M. The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly*. 2021. Vol. 45. P. 315-341.

17. O'Connor S. ChatGPT. Editorial: Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education In Practice*. 2023. Vol. 66. URL: <https://www.sciencedirect.com/science/article/abs/pii/S1471595322002517?via%3Dihub> (date of access: 27.11.25).

18. Weinberger D. How Machine Learning Pushes us to Define Fairness. URL: <https://www.predictiveanalyticsworld.com/machinelearningtimes/how-machine-learning-pushes-us-to-define-fairness/10636/> (date of access: 18.12.25).