

ФІЛОСОФІЯ.

Соціальна філософія та філософія історії.

Філософія науки та техніки

УДК 316.4

Штучний інтелект як джерело відчуження та екзистенційних загроз

ARTIFICIAL INTELLIGENCE AS A SOURCE OF ALIENATION AND EXISTENTIAL THREATS

ДЗЬОБАНЬ Олександр Петрович – доктор філософських наук, професор, головний науковий співробітник Державної наукової установи «Інститут інформації, безпеки і права Національної академії правових наук України», вул. Григорія Сковороди, 77, м. Харків, 61024, Україна

DZOBAN Oleksandr Petrovych – 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», 77 Hryhoriia Skovoroda Str., Kharkiv, 61024, Ukraine

ORCID <https://orcid.org/0000-0002-2075-7508>

DOI <https://doi.org/10.54891/2786-7013/2026-1-1>

Анотація. У статті здійснена спроба комплексного соціально-філософського осмислення впливу систем штучного інтелекту (ШІ) на сучасну людину та суспільство. Обґрунтовується теза, що ШІ є багатоплановим явищем, яке виступає джерелом масштабних соціокультурних трансформацій, несучи в собі як можливості для самореалізації індивіда, так і глибокі екзистенційні загрози. Підкреслюється, що темпи технологічного прогресу значно випереджають філософську рефлексію та теоретичне осмислення наслідків впровадження цих технологій. Розглядаються фундаментальні загрози, які ШІ становить для людської раси: трансформація ринку праці та втрата сенсу буття (проблема масового безробіття, спричиненого здатністю ШІ виконувати не лише рутинні, а й складні когнітивні завдання у творчих та інтелектуальних сферах), соціальна несправедливість та алгоритмічна упередженість (феномен «чорного ящика», через який логіка прийняття рішень ШІ залишається непрозорою, що призводить до дискримінації за етнічними, гендерними чи релігійними ознаками у таких сферах, як правопорядок, освіта та найм персоналу), порушення автономії особистості (зазіхання систем ШІ на право людини на самовизначення через маніпуляції, тотальне стеження та обмеження інформаційного простору). У висновках наголошується на необхідності розробки філософських принципів і нормативно-правових актів, які забезпечать людиноцентризований підхід до розвитку ШІ. Акцентується увага на тому, що лише за умови збереження центрального положення людини у світобудові та встановлення чітких етичних меж використання інтелектуальних систем можливо мінімізувати негативні наслідки та спрямувати потенціал ШІ на благо цивілізації. Звертається особлива увага на те, що екзистенційна загроза, що сформувалася внаслідок фундаментального вбудовування систем ШІ в життя соціуму, вимагає розробки та прийняття цілого комплексу документів, що існують у формі національних стратегій, нормативно-правових актів, етичних керівництв, рекомендацій і стандартів, виданих державними інститутами, комерційними структурами, міжнародними та неурядовими організаціями і регулюють етико-правові аспекти використання ШІ.

Ключові слова: штучний інтелект, екзистенційні ризики, відчуження, соціальна несправедливість, дискримінація, автономія людини, сенс життя.

© ДЗЬОБАНЬ Олександр Петрович

Summary. *This article attempts a comprehensive socio-philosophical reflection on the impact of artificial intelligence (AI) systems on modern individuals and society. It substantiates the thesis that AI is a multifaceted phenomenon acting as a source of large-scale socio-cultural transformations offering both opportunities for self-actualization and profound existential risks. The study emphasizes that the pace of technological progress significantly outstrips philosophical reflection and the theoretical understanding of the consequences of implementing these technologies. The fundamental threats that AI poses to the human race are considered: the transformation of the labour market and the loss of meaning in life (the problem of mass unemployment caused by AI's ability to perform not only routine but also complex cognitive tasks in creative and intellectual fields), social injustice and algorithmic bias (the 'black box' phenomenon, through which the logic of AI decision-making remains opaque, leading to discrimination on ethnic, gender or religious grounds in areas such as law enforcement, education and recruitment), violation of personal autonomy (AI systems infringing on the human right to self-determination through manipulation, total surveillance and restrictions on the information space). The conclusions stress the necessity of developing philosophical principles and regulatory frameworks to ensure a human-centered approach to AI development. It is emphasized that only by maintaining the central position of the human being in the world and establishing clear ethical boundaries for the use of intelligent systems can negative consequences be minimized and the potential of AI be directed toward the benefit of civilization. Particular attention is paid to the fact that the existential threat formed by the fundamental integration of AI into social life requires the development and adoption of a comprehensive set of documents national strategies, legislative acts, ethical guidelines, and standards issued by state institutions, commercial structures, and international organizations to regulate the ethico-legal aspects of AI utilization.*

Key words: *artificial intelligence, existential risks, alienation, social injustice, discrimination, human autonomy, meaning of life.*

Вступ. В останні роки штучний інтелект (ШІ) міцно входить у наше життя, відкриваючи перед людиною різноманітні можливості, покращуючи якість людського життя, розширюючи горизонти її самореалізації. Однак ШІ приховує у собі й нові небезпеки, здійснюючи значний вплив на людину, суспільство, навколишнє середовище.

ШІ має здатність впливати на культуру, істотно трансформуючи її базові складові. Можна з повною впевненістю вже сьогодні вести мову про те, що ШІ – це складне і багатопланове явище, джерело масштабних соціокультурних змін як позитивних, так і таких, що негативно впливають на людину й людську цивілізацію.

У цих умовах філософія, будучи раціональним типом світогляду, обґрунтовано ставить перед собою завдання виробити аргументоване й виважене ставлення до впровадження систем штучного інтелекту, розглянувши, які ризики супроводжують цей процес, який саме вплив вони мають на світорозуміння сучасного індивіда, на його ставлення до навколишнього світу.

Будь-які досягнення у сфері ШІ мають сенс тільки у тому випадку, якщо вони співвідносяться з ідеалами процвітання людини, людської цивілізації. Надмірне делегування повноважень від людини системам ШІ, навпаки, здатне негативно вплинути на екзистенційні основи буття людини, з особливою гостротою поставивши питання про сенс її існування і життєве призначення.

Таким чином, існує необхідність більш глибокого й адекватного філософського рефлексування сучасних реалій впливу ШІ на суспільство й людину, потреба виявлення змісту базових принципів взаємодії людини з системами ШІ; виявлення можливих соціальних, етичних наслідків і перспектив взаємодії людини зі ШІ для досягнення цілей гуманістичного, безпечного та ефективного розвитку сучасного соціуму.

Аналіз останніх досліджень. Результати аналізу джерел і публікацій свідчать про те, що філософський дискурс не раз звертався до теми ШІ, проте теоретичне осмислення наслідків

впровадження цих технологій все ще суттєво відстає від прогресу самих розробок. Сучасні дослідження можна структурувати за такими ключовими напрямками:

- увага до глобальних екзистенційних ризиків – С. Хокінг ще у 2014 році застерігав, що розробка повноцінного штучного інтелекту може означати кінець людської раси, якщо людина не навчиться уникати ризиків, пов'язаних із втратою контролю над системою [8];

- акцентування на трансформації праці та втраті сенсу буття – А. Кларк і А. Освальд довели, що безробіття призводить до втрати навичок, соціальної ізоляції та відсутності мети в житті [5, р. 648–659]. Т. Кім та А. Шеллер-Вольф підкреслюють, що автоматизація несе екзистенційну загрозу втрати сенсу життя навіть у разі фінансової компенсації втраченого доходу [12, р. 319–337];

- виділення проблеми соціальної несправедливості та алгоритмічної упередженості – Дж. Ангвін, Дж. Ларсон та С. Матту виявили упередженість ШІ в автоматичних оцінках рецидивізму [2]. М. Боген і А. Рік зафіксували ризики систематичних помилок на всіх етапах найму персоналу через ШІ [4];

- звернення уваги до правової та когнітивної прозорості – С. Вахтер, Б. Мітльштат та С. Рассел зазначають, що через швидкість і складність алгоритмів вкрай важко визначити факт порушення прав людини, оскільки логіка ШІ не піддається людському розумінню [18, р. 105–567];

- спроби осмислення порушення автономії особистості – Л. Флориді обґрунтовує, що поступове розширення делегування повноважень системам ШІ неминуче знижує людську автономію [7].

Водночас теоретико-методологічні та соціально-філософські засади означеної проблеми як світоглядне підґрунтя планування процесу розвитку і впровадження систем ШІ сьогодні розроблені все ще недостатньо.

Мета статті – спроба соціально-філософського осмислення впливу систем ШІ на людину й суспільство, виявлення та аналізу основних негативних проявів і ризиків, що виникають при впровадженні ШІ.

Виклад основного матеріалу. Системи ШІ створюються людиною для вирішення важливих для неї завдань, досягнення поставлених нею цілей, але 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 й майбутнього. Навпаки, вони покликані полегшити життя людини, звільнивши її від рутинних і нецікавих, або надзвичайно трудомістких операцій. Проте, у процесі проектування, впровадження й 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» проблеми, однак, на наш погляд, вони далеко виходять за межі власне етичної проблематики. Їх слід характеризувати як соціально-філософські, оскільки вони охоплюють широкий комплекс питань, що торкаються в цілому взаємодії людини зі ШІ в різних сферах суспільного життя, породжуючи крім власне етичних, моральних проблем, екзистенційні, аксіологічні та інші питання. Крім того, їх дія погіршує стан і становище людини у світі. Розглянемо найбільш очевидні з таких проблем.

Заподіяння шкоди. Проблема заподіяння шкоди викликає занепокоєння й найсерйознішу стурбованість вчених, громадських діячів, світової спільноти в цілому. Британський фізик С. Хокінг ще у 2014 році висловлював думки про те, що успішне створення ШІ стане найбільшою подією в історії людства, але, на жаль, воно може виявитися останньою, якщо ми не навчимося уникати ризиків. На його думку, коли штучний інтелект почне керувати фінансовими ринками,

науковими дослідженнями, людьми і розробкою зброї, це будуть речі, недоступні нашому розумінню. І якщо короткостроковий ефект штучного інтелекту залежить від того, хто ним керує, то довгостроковий ефект – від того, чи можна буде ним керувати взагалі «Розробка повноцінного штучного інтелекту може означати кінець людської раси» [7].

Обговорювана проблема часто зводиться до навмисного використання ШІ в цілях, що становлять загрозу для життя, здоров'я і благополуччя людини й суспільства. Деякі системи ШІ використовуються в розробці автономної зброї, спеціально створюються для вирішення військових завдань. Регулярно на міжнародних конференціях зі штучного інтелекту дослідники в галузі робототехніки та ШІ, громадські діячі, вчені закликають припинити подальший розвиток озброєння, заснованого на ШІ, яке може діяти без суворого людського контролю [1].

Проте, сьогодні існують різні види автономного озброєння на основі ШІ. Крім тих з них, які підконтрольні людині й надають їй можливість скасування рішення системи, обговорюється ідея створення повністю автономної зброї, принципово невіддільної людському контролю [5]. Автономне озброєння, таким чином, видається ще більш небезпечним у порівнянні з традиційним, оскільки його застосування істотно скорочує час на прийняття й осмислення важливих стратегічних рішень, мінімізує контроль над ситуацією з боку людини і позбавляє людину можливості самостійно приймати рішення, скорочуючи її автономію, обмежуючи її свободу.

Загрози поступово виникають у всіх сферах: у медицині – неналежна діяльність автономного медичного обладнання; в агросекторі – руйнування, спричинені безпіотною сільськогосподарською технікою; в обороні – помилка в системі «свій – чужий» військових пристроїв для придушення противника.

На жаль, застосування систем ШІ у військових і злочинних цілях не вичерпує весь потенційний шкоди, що завдається людині. Проблема заподіяння шкоди має більш складний, комплексний характер, що виявляється через безліч соціально-філософських проблем, які розкривають окремі грані шкоди, що завдається людині, і причини, що їх породили.

До недавнього часу вважалося, що технології ШІ замінять людину у виконанні відносно простих, рутинних робіт, але насправді виявилось, що вони можуть замінити людину, виконуючи практично будь-яку роботу, алгоритм якої можна обчислити і передбачити.

Не без причини ще на зорі технологій ШІ Н. Вінер застерігав, що цифрові обчислювальні машини – це щось, що докорінно відрізняється від усіх попередніх механічних засобів і здатне зруйнувати звичний соціальний порядок. Звичайно, слід враховувати, що різні технології, що автоматизують роботу, мають різний потенційний ефект. Проте, деякі з них здатні призвести до втрати людиною роботи, яку вона цінує і вважає важливою і значущою, позбавити її робочого місця, законного заробітку і засобів забезпечити себе і своїх близьких усім необхідним. Більше того, технології ШІ можуть зробити це дуже тонко, прибравши певні завдання в алгоритмах дій працівників, зробивши участь людини в них необов'язковою, не вигідною і зайвою для роботодавця.

Унікальна особливість ШІ як технології полягає в тому, що він володіє рівнем інтелекту, безпрецедентним у порівнянні з іншими попередніми технологіями в історії розвитку (інформаційними та комунікаційними), що проявляється в здатності ШІ приймати рішення, вирішувати проблеми і навіть думати в вузькоспеціалізованому розумінні цього слова. Системи ШІ, детально вивчаючи опис наших дій і повторюючи за нами, можуть опанувати необхідні навички, якісно виконуючи набагато більше когнітивних завдань, аніж середньостатистичний працівник-людина. Тим самим системи ШІ отримують можливість впливати на характер змін, що відбуваються, породжуючи нові, несподівані, суперечливі і негативні наслідки для зайнятості людей, для їх самооцінки, самореалізації, розуміння свого місця і ролі у світі.

Цілком очевидно, що якщо системи ШІ беруть на себе більш складні й важливі завдання, то наслідки можуть бути непередбачуваними. Приклад з GPT 4 демонструє, що професії, які раніше вважалися прерогативою висококваліфікованих кадрів, стають тепер вразливими для ШІ і створюють ризики працевлаштування, професійної затребуваності для фахівців, зайнятих у творчих та інтелектуальних видах діяльності.

Системи ШІ активно впроваджуються навіть у галузі живопису та музики. Відомі приклади, коли ШІ склав цілий музичний твір [9]. Тоді як ставитися до того творчого продукту, який здійснюється системами ШІ? Чи можна результат їхньої «творчості» сприймати як щось, що заслуговує на увагу? Набагато більше побоювань викликає те, що подібні програмні продукти ґрунтуються на генетичному програмуванні, яке націлене на постійне навчання доти, доки не почне вирішувати завдання близько до його ідеального варіанту виконання. Тоді виникає питання, чи можливо, що алгоритми навчатися в найближчому майбутньому створювати за нас, наприклад, правові концепції або вирішувати завдання в галузі державного управління?

Подібна ситуація створює значні ризики для здатності цих фахівців використовувати наявні у них знання та вміння, демонструвати свою майстерність і використовувати наявні у них навички, а значить, бути і відчувати себе професіоналами, потрібними, затребуваними в суспільстві людьми, які слідує своєму призначенню, своїй місії.

Проблема декваліфікації актуалізує соціально-політичний аспект застосування систем ШІ, питання про те, яким чином суспільство і держава повинні реагувати, якщо великі маси людей раптово опиняться позбавленими джерела доходу без перспективи знайти собі інше заняття, щоб задовольняти свої базові потреби, мати можливості й умови для особистої та соціальної реалізації. У даний час політичне регулювання не здатне ефективно мінімізувати наслідки структурних змін, викликаних впровадженням цифрових технологій, зменшити зростаючу нерівність в контексті скорочення частки праці в національному доході. З кожним днем залишається все менше професій, де людину не замінила б машина.

Крім очевидної економічної загрози матеріальному благополуччю людей, існує ще одне не менш важливе питання, що пов'язує втрату роботи з екзистенційною загрозою втрати сенсу життя, про яке, зокрема, пишуть Т. Кім і А. Шеллер-Вольф [11]. Більше того, зафіксувавши даний взаємозв'язок, автори вказують на збереження актуальності проблеми втрати сенсу навіть при компенсації втрати доходу, що виникла внаслідок автоматизації. Цю думку свого часу висловлювали ще А. Кларк і А. Освальд, показуючи, що безробітні страждають від втрати навичок, соціальної ізоляції та відсутності мети в житті, втрата роботи має довгостроковий негативний вплив на їхнє соціальне самопочуття [5]. У пізніших дослідженнях з'явився такий феномен як «смерть від відчаю», що виникає у зв'язку з відсутністю оплачуваної роботи. Це означає, що автоматизація може нести потенційну екзистенційну загрозу для людства, оскільки без роботи набагато більше людей мають шанс зіткнутися з хронічними захворюваннями і, в крайньому випадку, втративши надію, зробити непоправний і неправильний вибір проти самого життя.

Якби системи ШІ співвідносилися з механічним ткацьким верстатом або паровою машиною, то економічне становище людини не піддавалося б такій серйозній загрозі. Якщо раніше робітник турбувався про втрату роботи в своєму професійному секторі, то у нього завжди була можливість перенавчитися і знову знайти свою безумовну цінність з економічної точки зору. Використання систем ШІ у сучасному світі надзвичайно ускладнює, проблематизує цю можливість, стрімко скорочуючи перелік сфер діяльності людини, що не підлягають автоматизації.

Таким чином, можна зробити висновок, що системи ШІ, на відміну від інших технологій, у перспективі можуть стати необмеженими у своїх можливостях і здатними навчатися та здійснювати виробничий процес у всіх сферах життя соціуму. Відповідно, виникає ризик відчуження людини від самої себе внаслідок її взаємодії з системами ШІ, наділеними людиною її найістотнішою людською особливістю, здатністю до інтелектуальної діяльності та побудованою на її основі цілеспрямованою матеріальною діяльністю, націленою на зміну, перетворення навколишньої реальності, виходячи з цілей і завдань людини, людської цивілізації. У короткостроковій перспективі застосування систем ШІ загрожує безробіттям, а в довгостроковій перспективі виникає загроза не тільки повного безробіття, але й нагальної необхідності захищати особливий статус людини у світі.

Стосовно проблеми заподіяння шкоди всі інші труднощі виступають у ролі вторинних, підлеглих, що розкривають і уточнюють зміст головної проблеми. Вони покликані виявити джерело, причини виникнення небажаних для людини наслідків, а також описати різноманітні прояви можливої шкоди для людини і наслідки, що виникають у разі некоректного впровадження, застосування систем ШІ або помилок, допущених творцями таких систем у процесі їх розробки.

Соціальна несправедливість. Крім зазначених вище проблем, системи ШІ можуть бути упередженими щодо певних індивідів або груп, піддаючи їх несправедливій дискримінації за етнічними, гендерними або релігійними ознаками, що дозволяє зробити висновок про наявність проблеми соціальної несправедливості.

Табір дослідників на тему справедливості ділиться на тих, хто вважає, що справедливість можливо алгоритмізувати і ввести в системи як вимірюваний показник [14], і тих, хто впевнений, що справедливість не можна тлумачити з точки зору чітко визначених кількісних показників [9].

Несправедливість у контексті застосування ШІ прийнято позначати в літературі як алгоритмічну упередженість [17]. Справедливість виражається головним чином у запобіганні або пом'якшенні наслідків небажаної упередженості та дискримінації. Джерела, як правило, визначають несправедливість систем ШІ як відсутність різноманітності, інтеграції та рівності.

Прийнято вважати, що системи ШІ можуть допомогти при справедливому прийнятті рішень і сприяти запобіганню упередженості [13]. Однак практика показала, що прийняття рішень системами ШІ часто відтворює існуючі в суспільстві системні, інституційні та соціальні упередження. Такі упередження можуть призвести до дискримінації, несправедливості та проблем порушення конфіденційності, порушення прав людини.

Так, у більшості європейських держав у сфері забезпечення правопорядку застосовуються системи ШІ для профілювання людей, спроб передбачити ймовірну злочинну поведінку людей або місця майбутніх злочинів, а також для оцінювання передбачуваного ризику причетності окремих осіб до скоєння злочинів (наприклад, список злочинців 600 (Нідерланди)) [13]. Результати цих прогнозів використовуються для стеження, проведення обшуків або допитів осіб, яким системою ШІ був присвоєний високий ризиковий рейтинг.

Оцінюючи роботу подібних систем, фахівці дійшли до несподіваних результатів. Так, Дж. Ангвін, Дж. Ларсон, С. Матту у своєму дослідженні з'ясували, що системи ШІ упереджені в автоматичних оцінках можливості повторення злочину засудженими [2].

Системи штучного інтелекту застосовуються також у сфері освіти та працевлаштування для відбору кандидатів при вступі до ЗВО, для найму працівників, наприклад, з переглядом резюме та для цільової реклами вакансій. М. Боген і А. Рік зауважили, що ризик систематичної помилки був зафіксований на кожному з цих етапів процесу найму [4]. Застосування систем ШІ в зазначеному напрямку, безумовно, має переваги у вигляді автоматизації рутинної роботи

і скорочення тимчасових витрат. Однак виникають побоювання з приводу того, наскільки адекватною є оцінка системами ШІ навичок претендентів, на яких підставах і фактах приймається рішення, чи не піддається загрозі безпека соціуму і окремої людини.

Рекомендаційні системи у пошукових системах, засновані на ШІ, допомагають користувачам орієнтуватися в мережі, пропонують інформацію, що їх цікавить. Встановлено, що рекомендаційні системи посилюють різні види упередженості, наполегливо пропонуючи певні матеріали або представляючи як найкращі тільки певні підприємства та маркетплейси. Це збільшує дисбаланс сил між домінуючими на ринку великими платформами і більш дрібними, які не мають доступу до рівних обсягів високоякісних споживчих даних, які життєво важливі для виходу на ринок.

Особливо очевидною стає соціальна несправедливість у контексті європейського законодавства, згідно з яким чоловіки і жінки повинні користуватися рівними правами на ринку праці і мати рівний доступ у сфері пропозиції товарів і послуг. Також має бути гарантовано рівне ставлення до людей, незалежно від їхнього расового чи етнічного походження. Аналогічним чином має дотримуватися рівність у сфері праці та зайнятості, якщо йдеться про людей з інвалідністю, про людей з різними релігійними уподобаннями, переконаннями, про людей різних вікових груп і сексуальної орієнтації.

С. Вахтер, Б. Мітльштат, С. Рассел, вивчивши цілу низку європейських документів, дійшли висновку, що стосовно ШІ надзвичайно складно визначити, чи дійсно були порушені права людей в тому чи іншому випадку, оскільки алгоритми ШІ здійснюють свою діяльність з такою швидкістю, масштабом і рівнем складності, які в принципі не можна порівняти з рівнем людського інтелекту, не піддаються людському розумінню [18]. Наприклад, споживачам складно оцінити, чи була їм запропонована найкраща можлива ціна. Вони навіть не мають можливості дізнатися про те, що певні рекламні оголошення їм взагалі не були запропоновані. Масштаби та ефекти автоматизованого прийняття рішень не дозволяють людям зрозуміти, що вони опинилися в обділеному, ущербному становищі, і, як результат, у них не буде явних підстав для подання позову відповідно до закону про недискримінацію. Використання систем ШІ, таким чином, піднімає питання про справедливість на фундаментально новий рівень, що передбачає виконання вимог про методи виявлення можливої дискримінації та обов'язкове подання доказів, що свідчать про те, що права людей не були ущемлені.

Ще один аспект розглянутої проблеми виникає через складний характер сучасних систем машинного навчання, дії яких здійснюються за принципом «чорного ящика», коли простежити за цілим рядом факторів, що складають алгоритм діяльності ШІ, не представляється можливим.

Проблема полягає в тому, що людина не знає, яким чином системи ШІ, засновані на нейронних мережах, приходять до тих чи інших результатів. Наприклад, алгоритм COMPAS, що оцінює ймовірність вчинення насильницького злочину, як виявилось, дискримінує злочинців з темним кольором шкіри. Однак, як показало дослідження, система насправді не враховувала расові ознаки як входні дані. Натомість вона отримала конфіденційні дані про расову приналежність як наслідок з іншої доступної їй інформації про місце проживання потенційного злочинця.

На нашу думку, впровадження та використання систем ШІ в усіх сферах суспільного життя має ґрунтуватися на справедливому, рівному, неупередженому ставленні до всіх учасників різноманітних процесів соціальної взаємодії. Подібні вимоги висувують сьогодні численні міжнародні організації, комітети з етики, наукові інститути, що розробляють власні етичні кодекси, а також вчені, які беруть участь у роботі наукових конференцій. У міру того, як все більше і більше рішень делегується людиною системам ШІ, суспільство, його

інститути повинні забезпечити, щоб ці рішення були вільні від будь-якої упередженості та дискримінації.

Порушення автономії. Дана проблема, спеціально виділена дослідниками систем ШІ, полягає в тому, що останні можуть посягати на автономію людини, обмежуючи свободу прийняття нею рішення і таким чином утискаючи людську гідність. У людей забирають саму можливість самостійно, без допомоги ШІ, приймати усвідомлені та незалежні рішення, що можна трактувати як втрату людиною її права на самовизначення. Очевидно, що ШІ та подібні до нього технології здатні істотно впливати на свободу людини, породжуючи нові залежності та обмеження. Системи ШІ, наприклад, можуть прямо або опосередковано нав'язувати людям певний спосіб життя, застосовуючи репресивні механізми спостереження або стимулювання. У результаті страждає процес цілепокладання, втрачаються навички самостійності, здійснюється менш автентичний вибір, принцип людської автономії фактично порушується. Привабливі на перший погляд перспективи використання систем ШІ (розширення когнітивних здібностей людини, звільнення людини від монотонної праці, підвищення якості її життя та обслуговування тощо) супроводжуються прокламаціями про неминучі загрози та ризики для самостійності та автономії людини, що неминуче виникають внаслідок впровадження ШІ в життя суспільства.

Проблема порушення автономії особливо наочна у сфері медицини, де системи ШІ перетворюють базові структури лікування, змінюють соціальний контекст надання медичної допомоги. Ці перетворення мають системний характер, оскільки кардинально зачіпають усі аспекти діяльності лікаря, змінюючи діагностичні процедури, практики прийняття лікарських рішень, порядок проведення необхідних маніпуляцій (наприклад, хірургічних операцій, післяопераційного догляду).

Системи ШІ в галузі медицини вже сьогодні претендують на роль автономних, вільно діючих суб'єктів, виконуючи асистуючі та консультуючі функції. Роботи-асистенти в деяких клініках займаються доглядом за хворими після важких операцій, а програми на основі ШІ консультують лікарів, допомагаючи їм у прийнятті рішень [3]. У сфері інноваційної медицини вже ведуться дискусії про те, як розробити і впровадити повністю автономну систему ШІ.

Проблема порушення автономії, в свою чергу, породжує проблему декваліфікації співробітників і проблему довіри. У лікарів, які все більше спираються на асистуючі системи ШІ, з часом змінюються методи прийняття рішень (мається на увазі обґрунтоване прийняття рішення), стиль навчання, знижується якість медичних знань, виробляється стереотипне, недбале, знеособлене ставлення до пацієнтів. Чим значнішим стає вплив ШІ на людську природу і навколишнє середовище, тим важче його контролювати, тим більша небезпека для людини. Роботизація і заміщення людини призведуть до значної трансформації ціннісних пріоритетів людини, зміняться її погляди на обов'язок, справедливість, необхідність тощо. До того ж, системи на основі глибокого навчання не є цілком прозорими, і їх дії принципово незрозумілі для лікаря, не кажучи вже про те, що вони ще більш незрозумілі пацієнтам. Можна сказати, що подібні системи в певному сенсі мають суб'єктність, оскільки їм делегується частина лікарських повноважень. Це, в свою чергу, знову піднімає питання про те, хто несе відповідальність: людина чи машина?

Як правило, високий ступінь самостійності ШІ обмежує можливості контролю і впливу на нього з боку людини, а значить і автономію людини. Тому розробникам систем ШІ слід передбачити, щоб дії людини завжди мали пріоритет при їх використанні, тобто щоб людина завжди була в центрі додатка. Найкращий спосіб забезпечити необхідний рівень автономії людини – залучити майбутніх користувачів, а також експертів у предметній сфері до самого процесу розробки системи ШІ. Ступінь автоматизації повинен відповідати

контексту додатка і надавати користувачам всі необхідні можливості управління. У результаті, це дасть можливість створення ШІ, орієнтованого на людину. Системи ШІ вже сьогодні широко використовуються в багатьох додатках, пов'язаних з безпекою, причому в таких галузях діяльності, які пов'язані з надзвичайно високим рівнем ризику для людини, її життя і здоров'я (наприклад, авіація або робота атомних електростанцій). Тут особливо важливо забезпечити, щоб засоби управління системою були зрозумілі і підконтрольні людям, поводитися в роботі так само, як на етапі проектування.

Проблема порушення автономії проявляється і в неможливості для людини вибрати ту роботу, яка сприяє саморозвитку, моральному вдосконаленню і дозволяє дотримуватися власних цінностей, жити відповідно до тих принципів, які вони дійсно поділяють. У цьому контексті Л. Флориді вважає, що поступове розширення делегування повноважень системам ШІ знижує людську автономію [7]. Це означає, що в разі активної участі ШІ в процесі вирішення завдання, останній може обмежувати людину в її прагненні слідувати високим духовним цілям своєї діяльності, що призводить до втрати важливих, істотних людських якостей і обмежує можливості для розвитку людини. ШІ також може скоротити нашу автономію і в тому випадку, якщо доступ до передових технологій буде закритий для більшості членів суспільства. Автономія працівника, який виконує функції оператора систем ШІ, може бути порушена шляхом неправомірного обмеження штучним розумом обсягу інформації, доступної для перегляду та використання [11]. До того ж, якщо робота оператора сама по собі є буденною і нудною, то це змушує людину відчувати себе «рабом машини» і, з цієї причини, також відчувати почуття несвободи.

На нашу думку, набагато більший ризик для автономії людини виникає через стеження і маніпуляції з боку систем ШІ. Коли люди перебувають під наглядом, вони, як правило, відчувають скутість і діють менш вільно. Використання систем ШІ для спостереження за працівниками, як ми припускаємо, матиме аналогічні наслідки і може стати для роботодавців засобом посилення контролю за працівниками. Деякі системи ШІ впроваджуються для моніторингу онлайн-нарад з метою здійснення спостереження за залученістю працівників до процесу обговорення, що може призвести до стресу, депресії та інших негативних наслідків відчуття власної несвободи [15].

Висновки. Сфери застосування ШІ у майбутньому будуть тільки розширюватися, і людині необхідно подбати про те, щоб при розробці інтелектуальних систем будь-якого рівня складності людина не втратила свого центрального положення у світобудові, щоб в основі будь-яких технологій незмінним залишався підхід, орієнтований на людину, на досягнення нею цілей і завдань свого розвитку, на реалізацію її справжнього призначення у світі.

На наш погляд, проблема заподіяння шкоди, що виникає в процесі взаємодії людини з системами ШІ, призводить до ризику відчуження людини від самої себе, від власної сутності і призначення, можливої втрати і спотворення її особливого статусу в світі як єдиної істоти, здатної до інтелектуальної діяльності, націленої на зміну, перетворення навколишньої реальності. Проблеми соціальної несправедливості та порушення автономії покликані уточнити зміст, розкрити окремі грані та описати різноманітні прояви шкоди, що завдається людині, а також наслідки, що виникають через некоректне впровадження, застосування систем ШІ.

Екзистенційна загроза, що сформувалася внаслідок фундаментального вбудовування систем ШІ в життя соціуму, вимагає розробки та прийняття цілого комплексу документів, що існують у формі національних стратегій, нормативно-правових актів, етичних керівництв, рекомендацій і стандартів, виданих державними інститутами, комерційними структурами, міжнародними та неурядовими організаціями і регулюють етико-правові аспекти використання ШІ.

Список використаних джерел

1. AI Conf 2025. DOI: <https://dou.ua/calendar/54093/>
2. Angwin J., Larson J. et al. Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks. URL: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (дата звернення: 10.02.2026).
3. Bian Y. et al. Artificial intelligence–assisted system in postoperative follow-up of orthopedic patients: e-ploratory quantitative and qualitative study. URL: <https://pubmed.ncbi.nlm.nih.gov/32452807/> (дата звернення: 18.02.2026).
4. Bogen M., Rieke A. Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias. URL: <https://apo.org.au/node/210071> (дата звернення: 20.02.2026).
5. Clark A., Oswald A. Unhappiness and unemployment. *Economic Journal*. 1994. Vol. 104 (424). P. 648–659.
6. Etzioni A., Etzioni O. Should artificial intelligence be regulated. URL: <https://www.issues.org/334/perspective-should-artificial-intelligence-be-regulated/> (дата звернення: 14.02.2026).
7. Floridi L., Cowls J., Beltrametti M. et al. AI4People – An ethical framework for a good AI society. URL: https://ai4people.org/PDF/AI4People_Ethical_Framework_For_A_Good_AI_Society.pdf (дата звернення: 20.02.2026).
8. Hawking St. Will artificial intelligence outsmart US? URL: <https://ru.scribd.com/document/470378723/Will-Artificial-Intelligence-Outsmart-Us> (дата звернення: 14.02.2026).
9. Holstein K., Wortman J. et al. Improving fairness in machine learning systems: What do industry practitioners need? *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019. P. 1–16.
10. Iamus, a music-making computer, could be the next Mozart. URL: <https://www.vice.com/en/article/pgg8yy/iamus-a-music-making-computer-could-be-the-next-mozart> (дата звернення: 20.02.2026).
11. Kellogg K. C., Valentine M. A., Christin A. Algorithms at work: The new contested terrain of control. *Academy of Management Annals*. 2020. Vol. 14 (1). P. 366–410.
12. Kim T. W., Scheller-Wolf A. Technological unemployment, meaning in life, purpose of business, and the future of stakeholders. *Journal of Business Ethics*. 2019. Vol. 160 (2). P. 319–337.
13. Kleinberg J., Mullainathan S., Manish R. Inherent Trade-Offs in the Fair Determination of Risk Scores. *8th Innovations in Theoretical Computer Science Conference. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik*. 2018. P. 1–23.
14. Michael V., Van Kleek M., Binns R. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *Proceedings of the 2018 chi conference on human factors in computing systems*. 2018. P. 1–14.
15. Pardes A. AI can run your work meetings now. URL: <https://www.wired.com/story/ai-can-run-work-meetings-now-headroom-clockwise/> (дата звернення: 20.02.2026).
16. Pushback against AI policing in Europe heats up over. URL: <https://www.globaltimes.cn/page/202110/1237232.shtml> (дата звернення: 22.02.2026).
17. Ryan M., Antoniou J. et al. Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality. URL: <https://edepot.wur.nl/543861> (дата звернення: 20.02.2026).
18. Wachter S., Mittelstadt B., Russel C. Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI. *Computer Law & Security Review*. 2021. Vol. 41. P. 105–567.

References

1. AI Conf (2025). DOI: <https://dou.ua/calendar/54093/>
2. Angwin J., Larson J. et al. Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks. URL: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (accessed: 20.02.2026).
3. Bian Y. et al. Artificial intelligence–assisted system in postoperative follow-up of orthopedic patients: e-ploratory quantitative and qualitative study. URL: <https://pubmed.ncbi.nlm.nih.gov/32452807/> (accessed: 20.02.2026).

4. Bogen M., Rieke A. Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias. URL: <https://apo.org.au/node/210071> (accessed: 20.02.2026).
5. Clark A., Oswald A. (1994). Unhappiness and unemployment. *Economic Journal*, 104 (424), 648–659.
6. Etzioni A., Etzioni O. Should artificial intelligence be regulated. URL: <https://www.issues.org/334/perspective-should-artificial-intelligence-be-regulated/> (accessed: 20.02.2026).
7. Floridi L., Cowls J., Beltrametti M. et al. AI4People – An ethical framework for a good AI society. URL: https://ai4people.org/PDF/AI4People_Ethical_Framework_For_A_Good_AI_Society.pdf (accessed: 20.02.2026).
8. Hawking St. Will artificial intelligence outsmart US? URL: <https://ru.scribd.com/document/470378723/Will-Artificial-Intelligence-Outsmart-Us> (accessed: 20.02.2026).
9. Holstein K., Wortman J. et al. (2019). Improving fairness in machine learning systems: What do industry practitioners need? Proceedings of the 2019 CHI conference on human factors in computing systems. P. 1–16.
10. Iamus, a music-making computer, could be the next Mozart. URL: <https://www.vice.com/en/article/pgg8yy/iamus-a-music-making-computer-could-be-the-next-mozart> (accessed: 20.02.2026).
11. Kellogg K. C., Valentine M. A., Christin A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14 (1), 366–410.
12. Kim T. W., Scheller-Wolf A. (2019). Technological unemployment, meaning in life, purpose of business, and the future of stakeholders. *Journal of Business Ethics*, 160 (2), 319–337.
13. Kleinberg J., Mullainathan S., Manish R. (2018). Inherent Trade-Offs in the Fair Determination of Risk Scores. 8th Innovations in Theoretical Computer Science Conference. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 1–23.
14. Michael V., Van Kleek M., Binns R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. Proceedings of the 2018 chi conference on human factors in computing systems, 1–14.
15. Pardes A. AI can run your work meetings now. URL: <https://www.wired.com/story/ai-can-run-work-meetings-now-headroom-clockwise/> (accessed: 20.02.2026).
16. Pushback against AI policing in Europe heats up over. URL: <https://www.globaltimes.cn/page/202110/1237232.shtml> (accessed: 22.02.2026).
17. Ryan M., Antoniou J. et al. Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality. URL: <https://edepot.wur.nl/543861> (accessed: 20.02.2026).
18. Wachter S., Mittelstadt B., Russel C. (2021). Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI. *Computer Law & Security Review*, 41, 105–567.

Отримано 14.02.2026.

Отримано в доопрацьованому вигляді 22.02.2026.

Прийнято до друку 10.03.2026.

Опубліковано 08.04.2026.

Received 14.02.2026.

Received in revised form 22.02.2026.

Accepted for publication 10.03.2026.

Published 08.04.2026.