

ПСИХОФІЗІОЛОГІЯ

УДК 159.91

DOI <https://doi.org/10.32782/psy-visnyk/2024.4.2>

Тиньков О. М.

*кандидат психологічних наук, доцент,
доцент кафедри психології*

*Національного аерокосмічного університету імені М. Є. Жуковського
«Харківський авіаційний інститут»*

Кузьміна С. В.

старший викладач кафедри психології

*Національного аерокосмічного університету імені М. Є. Жуковського
«Харківський авіаційний інститут»*

ПРОБЛЕМИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ, ЕТИЧНИЙ АСПЕКТ

PROBLEMS OF ARTIFICIAL INTELLIGENCE APPLICATION, ETHICAL ASPECT

Ця стаття узагальнювального характеру спрямована на аналіз знань сучасної нейропсихології, нейрофізіології. Розглянуто результати спільного дослідження міжнародної аудиторської фірми KPMG і університету Квінсленда (Австралія), присвяченого темі довіри до ШІ (довіра до штучного інтелекту). Показано проблеми, з якими стикаються розробники та дослідники ШІ, які можуть бути настільки складними, що їх неможливо алгоритмізувати за допомогою строгих математичних методів, а рішення не доступні. Розглянуто результати першого в Україні соціологічного дослідження «Штучний інтелект: український вимір» – ставлення населення до нових технологій, яке було проведено у вересні 2018 року з ініціативи інституту Горшеніна і групи компаній Everest методом інтерв'ю face-to-face. Розглянуто доповідь «Про етику робототехніки», Всесвітньої комісії ЮНЕСКО з етики наукових знань і технологій (КОМЕСТ), в якій були розроблені рамки для формулювання рекомендацій у сфері етики роботів з урахуванням відмінностей між детермінованими і когнітивними роботами. КОМЕСТ також надає низку конкретних рекомендацій щодо використання робототехніки, включаючи розробку етичного кодексу для розробників і застереження щодо розробки і використання автономних систем озброєнь. У статті розглянуто дослідження: «Довіра до штучного інтелекту: глобальне дослідження». У дослідженні проаналізовано довіру до використання ШІ, ставлення громадськості та очікування від управління ШІ в 17 країнах світу. Звіт за цим дослідженням містить вичерпну глобальну інформацію про довіру та сприйняття систем штучного інтелекту, що охоплює переваги та ризики штучного інтелекту, очікування громадськості, а також регулювання та управління у сфері штучного інтелекту. Він показує, як люди ставляться до використання ШІ у своїй роботі, наскільки широка громадська обізнана про ШІ і як ставлення до ШІ змінюється з часом. Мета статті – здійснити аналіз актуальних питань пов'язаних з недоліками та загрозами впровадження ШІ.

Ключові слова: штучний інтелект, етичні стандарти робототехніки, правові гарантії, загрози, впровадження.

This generalizing article is aimed at analyzing the knowledge of modern neuropsychology, neurophysiology. The results of a joint study by the international auditing firm KPMG and the University of Queensland (Australia) on the topic of trust in AI (trust in artificial intelligence) are considered. It shows the problems faced by AI developers and researchers, which can be so complex that they cannot be algorithmized using rigorous mathematical methods, and solutions are not readily available. The results of the first sociological study in Ukraine "Artificial Intelligence: Ukrainian Dimension" – the attitude of the population to new technologies, which was conducted in September 2018 at the initiative of the Gorshenin Institute and the Everest Group of Companies by the face-to-face interview method, were considered. The report "On the ethics of robotics" by the UNESCO World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) was considered, in which a framework was developed for formulating recommendations in the field of robot ethics, taking into account the differences between deterministic and cognitive robots. COMEST also provides a number of specific recommendations on the use of robotics, including the development of a code of ethics for developers and caveats on the development and use of autonomous weapons systems. The article discusses the study: "Trust in artificial intelligence: a global study." The study analyzed trust in the use of AI, public attitudes and expectations of AI governance in 17 countries around the world. This research report provides comprehensive global information on trust and perception of AI systems, covering AI benefits and risks, public expectations, and AI regulation and governance. It shows how people feel about using AI in their work, how much the general public is aware of AI, and how attitudes towards AI change over time. The purpose of the article is to analyze current issues related to the shortcomings and threats of AI implementation.

Key words: artificial intelligence, ethical standards of robotics, legal guarantees, threats, implementation.

Вступ. У 2017 році Всесвітня комісія ЮНЕСКО з етики наукових знань і технологій (КОМЕСТ) опублікувала звіт про етику робототехніки. У лютому 2017 року «Парламент ЄС» ухвалив цивільно-правовий регламент щодо робототехніки. У 2019 році було

створено спеціальний комітет з питань штучного інтелекту (ШІ).

Ще у 2007 році уряд Південної Кореї розпочав роботу над Хартією етичних стандартів робототехніки. Основні положення Хартії подібні до «Законів

робототехніки», сформульованих Айзеком Азімовим, а наприкінці 2008 року Республіка Корея прийняла «Закон про створення та розповсюдження розумних роботів».

Аналіз останніх досліджень і публікацій. Основою для формулювання «Етичного кодексу ШІ» є класичні «Три закони робототехніки», сформульовані Айзеком Азімовим у 1950-х роках (пізніше доповнені «Законом нуля»), які стверджують, що: роботи не повинні завдавати шкоди людству; робот повинен підкорятися всім командам людини, за винятком тих, що суперечать закону нуля і першому закону; робот повинен дбати про власну безпеку в тій мірі, в якій це не суперечить першому і другому законам.

Комітет, скликаний у Великобританії в 2010 році, вніс поправки до «Закону Азімова», уточнивши, що відповідальність за ШІ несе виробник або власник. Спільнота штучного інтелекту і робототехніки, представлена своїми професійними членами, зобов'язана зайняти тверду позицію.

Основні поняття (міжнародні позначення) ШІ включають: штучний інтелект (ШІ), машинне навчання (МН), яке є частиною ШІ, і наука про дані.

Термін «штучний інтелект» ввів американський кібернетик Джон Маккарті, який на семінарі в Дартмутському коледжі в 1956 році поставив перед десятьма вченими завдання. Завдання полягало в тому, щоб змусити машини використовувати природну мову, формувати абстракції та поняття, вирішувати проблеми, які під силу лише людині, і розуміти, як самовдосконалюватися. Цей двомісячний мозковий штурм призвів до створення наукових лабораторій у різних країнах для розробки ШІ, зокрема нейронних мереж [8].

Потім ШІ розглядався на різних рівнях. Найнижчий рівень ієрархії – слабкий ШІ, де нейронні мережі, генетичні алгоритми та інші форми обчислень, що розвиваються, використовуються для розуміння процесів адаптації, сприйняття, виконання та взаємодії з фізичним світом навколо нас.

Після частково зрозумілого процесу розв'язання, цей набір обчислювальних схем трансформується в більш визначену схему другого рівня – сильний ШІ, моделі логічного мислення. Там розглядаються і вивчаються дедукція, абдукція, індукція, підтримка істини та багато інших моделей і принципів міркувань. На цьому високому рівні абстракції фахівці прагнуть втілити в ШІ закони соціальних процесів, передачу та поглиблення знань. Це передбачає розробку експертних систем, інтелектуальних агентів і систем розуміння природної мови.

Найвищим рівнем штучного інтелекту (ШІ), потужного штучного інтелекту, є сам розум, який перевершує людський інтелект. У той час як ШІ в тому вигляді, в якому він створюється сьогодні, відноситься до будь-якої комп'ютерної діяльності, що імітує людський інтелект, сильний ШІ відноситься тільки до тих, які претендують на універсальну загальність, подібну до способу мислення людини.

Термін «сильний штучний інтелект» був введений Джоном Серлом, чий підхід характеризують

його власні слова: «Більше того, така програма буде не просто моделлю розуму, а буквально самим розумом, в тому ж сенсі, в якому людський розум є розумом». Цей автор відомий у філософських колах своєю жорсткою критикою ідей штучного інтелекту та когнітивної психології [6].

Штучний суперінтелект (ASI). Нік Бостром, філософ з Оксфордського університету і провідний мислитель у сфері ШІ, визначає ШІ як «інтелект, який набагато розумніший за найкращі людські мізки майже у всіх сферах, включаючи наукову творчість і соціальні навички. «Ми вже програли. Наша цивілізація закінчилася. Тому що штучні пристрої, безумовно, роблять це краще за нас» [2].

Мисленнєвий експеримент Джона Серла «Китайська кімната» є аргументом на користь того, що проходження тесту Тюрінга не є підставою (критерієм) для того, щоб машина мала справжні мисленнєві процеси [6].

За останнє десятиліття було проведено багато досліджень щодо рівня довіри до ШІ. Це дає змогу проаналізувати взаємодію людини і ШІ, визначити роль етики у прийнятті рішень, зрозуміти вплив різних рівнів ШІ на довіру та дослідити взаємозв'язок між довірою та рівнем використання ШІ. У роботах [1; 3; 5; 10; 13] проаналізовано публікації про довіру до ШІ, що відкриває нові можливості для майбутніх досліджень розвитку ШІ, його впливу на широку громадськість та шляхів підвищення довіри.

У 2023 році було опубліковано дослідження «Довіра до штучного інтелекту: глобальне дослідження» [9]. У дослідженні проаналізовано довіру до використання ШІ, ставлення громадськості та очікування від управління ШІ в 17 країнах світу. Звіт містить вичерпну глобальну інформацію про довіру та сприйняття систем штучного інтелекту, що охоплює переваги та ризики штучного інтелекту, очікування громадськості, а також регулювання та управління у сфері штучного інтелекту. Він показує, як люди ставляться до використання ШІ у своїй роботі, наскільки широка громадськість обізнана про ШІ і як ставлення до ШІ змінюється з часом. Загалом, результати дослідження вказують на шляхи безпечного та відповідального використання систем штучного інтелекту, а також на шляхи впровадження ШІ в суспільство.

Загалом, результати дослідження пропонують шляхи безпечного та відповідального використання систем ШІ, а також шляхи впровадження ШІ в економіку та суспільство. Ці висновки є важливими для формування політики та побудови стратегій щодо ШІ в бізнесі, уряді та неурядових організаціях, а також для формування стандартів ШІ на національному та міжнародному рівнях. Однак динамічність змін у цій сфері та комплексний і міждисциплінарний характер проблеми вимагають нових досліджень.

Незважаючи на те, що комп'ютерні науки є основою ШІ, експерти зазначають, що довіра людини до ШІ також залежить від виконання умов, визначених гуманітаріями, такими як психологи, соціологи та філософи.

Мета статті – здійснити аналіз актуальних питань пов'язаних з недоліками та загрозами впровадження ШІ.

Виклад основного матеріалу. Перелік завдань, в рамках «штучного інтелекту», дуже широкий і дуже розпливчастий. До нього входять базові задачі з використанням автоматичних рішень, задачі «розуміння» і перекладу мов, доведення теорем, розпізнавання зорових і звукових образів (мовлення).

Це коло «інтелектуальних» задач, які передбачається розв'язувати автоматичними методами шляхом створення і використання комп'ютерних програм. Наприклад, розв'язання складних обчислювальних задач є, можливо, інтелектуальною діяльністю людини. Практично важливі та цікаві завдання, що вимагають від людини інтелектуальних здібностей, часто нагадують геометрію, комбінаторний аналіз і формальну логіку.

Хоча поки що не йдеться про розробку чи створення теорії штучного інтелекту, вже сформувався перелік наукових дисциплін, які складають ядро цієї галузі. Неважко уявити, як інтелектуальні комп'ютери застосовуються у світі майбутнього, схожому на науково-фантастичний: математична логіка, структурна лінгвістика, теорія обчислень, теорія інформаційних структур, теорія управління, теорія статистичної класифікації, теорія графів, теорія евристичного пошуку.

Пов'язані з цим труднощі зазвичай виникають через складність формального опису таких завдань. Людський мозок легко розрізняє котів і собак, але важко формально описати процедуру розпізнавання. Багато завдань (ігрових, наукових, практичних) в принципі можна вирішити шляхом вибору, в певному сенсі, найкращої альтернативи з кінцевого числа альтернатив. Вирішення проблем шляхом евристично керованих спроб і помилок у просторі можливих рішень є домінуючою темою в галузі ШІ.

ШІ вже здатний робити багато речей, які вимагають певного рівня інтелекту. Він може вирішувати багато завдань експертного рівня, міркувати геометричною мовою, розв'язувати математичні задачі, вивчати прості поняття, розуміти прості малюнки, здійснювати прості взаємодії та виконувати корисні завдання. Основні поняття проблем ШІ включають об'єкти, стани, простори станів, оператори, які перетворюють об'єкти, простори станів та обмеження, що накладаються на оператори.

Проблеми, з якими стикаються розробники та дослідники ШІ, численні і можуть бути настільки складними, що їх неможливо алгоритмізувати за допомогою строгих математичних методів, а рішення не доступні. У таких ситуаціях використовують евристику. Евристика – це техніка пошуку істини. Евристики – це правила здорового глузду, які спрямовують інтелектуальні системи до мети. Іноді вони нагадують методи управління. Ці методи не завжди працюють, але часто використовуються.

Недоліки та загрози поширення ШІ. У розробку систем ШІ вкладається набагато більше грошей, ніж у дослідження їхньої безпеки. Недоліком є те, що найбільш оптимізовані моделі (наприклад, нейронні мережі) часто функціонують як чорний ящик, що унеможливає розу-

міння або вивчення мотивацій і причин вибору того чи іншого рішення. Цей факт може призвести до етичних проблем, пов'язаних з інформаційною прозорістю.

На думку експертів, на планеті з часом з'явиться надпотужний штучний інтелект (ШІ). За прогнозами середньостатистичних експертів, це станеться до 2060 року, Рей Курцвейл вважає, що це станеться до 2045 року, а Нік Бостром вважає, що це може статися в будь-який час в наступному десятилітті або в кінці століття, але коли це станеться, стрімке зростання ШІ застане нас зненацька [2; 7].

Перед обличчям інтелектуального вибуху людство схоже на дитину, яка грається з бомбою. Потужність наших іграшок занадто далека від незрілості нашої поведінки. Суперінтелект – це випробування, до якого ми ще не готові і не будемо готові ще довго. Ми не знаємо, коли пристрій вибухне, але якщо піднести його до вуха, то почуємо тихе цокання.

До інших недоліків (загроз) впровадження ШІ на національному та міжнародному рівнях можна віднести:

- економічна нестабільність.
- посилення нерівності між тими, хто має доступ до нової технології, і тими, хто його не має
- прийняття неправильних рішень (наприклад, газопровід Канада-США);
- зменшення різноманітності в культурних індустріях;
- безробіття; нестабільність на ринку праці.

У березні 2021 р. світ побачили результати спільного дослідження міжнародної аудиторської фірми KPMG і університету Квінсленда (Австралія), присвяченого темі довіри до ШІ (довіра до штучного інтелекту): результати дослідження в п'яти розвинутих країнах. (Trust in Artificial Intelligence: A five country study) [11].

В ході дослідження було опитано понад 4 500 осіб з країн з розвинутою економікою і високим рівнем життя: з США, Канади, Великої Британії, Німеччини та Австралії. Люди не довіряють ШІ, але допускають його використання. більшість опитаних вказало, що ставляться терпимо до ШІ (42%) або нейтрально (28%). меншість опитаних схвалюють використання ШІ (15%), підтримують (6%). 9% опитаних ставляться до ШІ негативно. Поточні правові гарантії бачаться недостатніми для забезпечення безпеки використання ШІ. Незважаючи на впевненість в необхідності правового регулювання ШІ, більшість опитаних не згодні або утрималися від оцінки поточного правового регулювання як достатнього для забезпечення безпеки ШІ – технологій [10].

Перше в Україні соціологічне дослідження «Штучний інтелект: український вимір» – ставлення населення до нових технологій було проведено у вересні 2018 року з ініціативи інституту Горшеніна і групи компаній Everest методом інтерв'ю face-to-face за місцем проживання респондентів. Було опитано 1000 респондентів віком від 16 років до 65 років. Українці готові покластися на нові технології не тільки в питанні виборів, але і в боротьбі з корупцією. 42,3% українців не проти довірити контроль

за державними видатками ШІ, респонденти відкриті до того, щоб задіяти роботів у будівництві (52,0%), промисловості (46,0%) та освіти (39,1%). Незважаючи на таку відкритість, серед найбільш неприйнятних для використання ШІ українці назвали політику і державне управління (40,1%), мистецтво (37,1%), юриспруденцію (34,7%), спорт (33,1%) і журналістику (28,3%). Медицина є проблемною точкою, що поляризує суспільство, адже вона лідирує і в позитивному (51,4%), і в негативному ставленні (28,3%) з впровадження нових технологій. [1; 4].

Всесвітня комісія ЮНЕСКО з етики наукових знань і технологій (КОМЕСТ) опублікувала доповідь «Про етику робототехніки», в якій розробила рамки для формулювання рекомендацій у сфері етики роботів з урахуванням відмінностей між детермінованими і когнітивними роботами. Сучасна робототехніка, відома як когнітивна робототехніка, базується на технології штучного інтелекту (ШІ). Такі роботи можуть взаємодіяти з людьми та іншими роботами, вирішувати проблеми, навчатися і навіть проявляти творчість, розпізнавати просторове середовище і використовувати мову, але вони відрізняються від так званих детермінованих роботів [12].

Рішення, які приймають роботи, непередбачувані, а їхня поведінка залежить від досвіду та стохастичних (випадкових) умов. Все це піднімає питання безпеки, приватності та захисту людської гідності.

У звіті також представлені етичні цінності та принципи, які можуть допомогти в узгодженій підготовці нормативних документів на всіх рівнях, від кодексів поведінки технічного персоналу до національних законів і міжнародних конвенцій. Особливий акцент зроблено на таких цінностях і принципах, як людська гідність, незалежність, недоторканність приватного життя, безпека, відповідальність, благодійність і справедливність. Всі цінності, що обговорюються в цьому

звіті, об'єднані принципом людської відповідальності.

КОМЕСТ також надає низку конкретних рекомендацій щодо використання робототехніки, включаючи розробку етичного кодексу для розробників і застереження щодо розробки і використання автономних систем озброєнь.

Відповідальність за поведінку когнітивних роботів дуже серйозна. Якщо пацієнт помирає під час операції робота-хірурга, або якщо, наприклад, безпілотний автомобіль збиває людину, хто несе відповідальність за рішення, прийняті роботом: творець ШІ, користувач і, можливо, страхова компанія, хто розділяє відповідальність? Алгоритми прийняття рішень є складною концепцією, і їх нелегко застосувати до існуючих правових систем. Це означає, що спочатку потрібно продумати низку ситуацій, а також те, як ці ситуації повинні бути юридично врегульовані.

Висновки. Ми живемо у світі, де значною мірою домінує бізнес і ринок. А реклама, як відомо, є двигуном бізнесу. Тому заяви про необхідність діяти без оглядки на вигоду виробників ШІ, про високу ефективність додатків зі штучним інтелектом і його повсюдне впровадження рекламодавці навряд чи скажуть у своїй рекламі, навіть якщо вони знають про побічні ефекти.

Насправді ШІ – це потужний інструмент обробки даних, здатний знаходити рішення складних завдань швидше, ніж традиційні алгоритми, написані програмістами. ШІ та методи глибокого навчання можуть допомогти у вирішенні широкого кола проблем.

Закони робототехніки Айзеком Азімовим небезпечні тим, що можуть, наприклад, заохотити ШІ захопити владу на планеті і «захистити» людей від шкоди. Тим не менш, основою для розробки етичного кодексу для ШІ є класичні «Три закони робототехніки», сформульовані Айзеком Азімовим у 1950-х роках і пізніше доповнені «Законом нуля».

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Андрощук Г.О. Рівень довіри і ставлення до штучного інтелекту: аналіз результатів досліджень. *Часопис Київського університету права*. 2021. № 3. С. 195–201.
2. Бостром Нік. Суперінтелект. Стратегії і небезпеки розвитку розумних машин. Київ: Наш формат, 2020. 408 с. ISBN 978-617-7866-31-1.
3. Системи та засоби штучного інтелекту: тези доповідей Міжнародної наукової конференції «Штучний інтелект: досягнення, виклики та ризики». Київ: ІПШ «Наука і освіта», 15–16.03. 2024. 550 с.
4. Штучний інтелект: Український вимір. Звіт за результатами кількісного соціологічного дослідження. Київ, 2018. URL: <http://gorshenin.ua/wp-content/uploads/2018/12/Iskusstvennyj-intellekt.pdf>
5. Bishop, M. J. Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It. *Frontiers in Psychology*. 2021. 11:513474. <https://doi.org/10.3389/fpsyg.2020.513474>
6. John Searle's Philosophy of Language: Force, Meaning, and Mind: Savas L. Tsohatzidis (Ed.). Cambridge University Press, 2007. 297 p.
7. Kurzweil R. *The Singularity is Near*. New York: Viking Books, 2005. 434 p.
8. McCarthy John. Recursive Functions of Symbolic Expressions and Their Computation by Machine. *Communications of the ACM*. 1960. Vol. 3, № 4. P. 184–195.
9. Professor Nicole Gillespie, Dr Steve Lockey, Dr Caitlin Curtis and Dr Javad Pool. Trust in artificial intelligence: A global study (2023). URL: <https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/trust-in-ai-global-insights-2023.pdf>
10. Ramos-Cruz B., Andreu-Perez J., Martínez L. The cybersecurity mesh: A comprehensive survey of involved artificial intelligence methods, cryptographic protocols and challenges for future research. *Neurocomputing*. 2024. Art. 127427. <https://doi.org/10.1016/j.neucom.2024.127427>
11. Trust in Artificial Intelligence: A five country study. URL: <https://home.kpmg/au/en/home/insights/2021/03/artificial-intelligence-five-country-study.html>
12. UNESCO. Режим доступу: <https://on.unesco.org/48XkvEO>
13. Yang G., Ai H., Liu W., Wang Q. Weak signal detection based on variablesituation-potential with time-delay feedback and colored noise. *Chaos, Solitons Fractals*. 2023. Vol. 169. Art. 113250. <https://doi.org/10.1016/j.chaos.2023.113250>