УДК / UDC: 1:004.8
DOI: https://doi.org/10.37750/2616-6798.2026.2(57).364303

Олександр Петрович Дзьобань
Державна наукова установа "Інститут інформації, безпеки і права Національної академії правових наук України"
Київ, Україна
ORCID: http://orcid.org/0000-0002-2075-7508
Максим Вікторович Жушман
Національний юридичний університет імені Ярослава Мудрого
Київ, Україна
ORCID: https://orcid.org/0000-0003-1235-6189

КОНЦЕПТУАЛЬНІ МЕЖІ ОСНОВНИХ СТРАТЕГІЙ ОСМИСЛЕННЯ ПРИРОДИ ШТУЧНОГО ІНТЕЛЕКТУ: КРІЗЬ ПРИЗМУ УЯВНИХ ЕКСПЕРИМЕНТІВ

Анотація. *У статті представлено результати комплексного філософського аналізу основних напрямів філософії штучного інтелекту (ШІ), таких як фізикалізм, функціоналізм і когнітивний плюралізм. Досліджуються теоретичні труднощі, з якими стикаються ці підходи при спробі пояснити феномен свідомості та суб'єктивного досвіду ("кваліа"). Особлива увага приділяється аналізу ключових мисленнєвих експериментів ("Кажан" Т. Нагеля, "Китайська кімната" Дж. Серла, "Болотна людина" Д. Девідсона, "Китайська нація" Н. Блока і "Філософські зомбі" Д. Чалмерса), які демонструють обмеження редукціоністських і чисто технічних підходів до розуміння інтелекту. Обґрунтовується необхідність переходу до трансдисциплінарної парадигми, що об'єднує природничо-наукове та соціогуманітарне знання для глибокого осмислення природи свідомості та оцінювання перспектив створення "сильного" ШІ.*

Ключові слова: *штучний інтелект, фізикалізм, функціоналізм, кваліа, мисленнєвий експеримент, трансдисциплінарність.*

Oleksandr P. Dzioban
State Scientific Institution "Institute of Information, Security and Law of the National Academy of Legal Sciences of Ukraine"
Kyiv, Ukraine
ORCID: http://orcid.org/0000-0002-2075-7508
Maksym V. Zhushman
Yaroslav the Wise National Law University
Kharkiv, Ukraine
ORCID: https://orcid.org/0000-0003-1235-6189

CONCEPTUAL LIMITS OF THE BASIC STRATEGIES FOR UNDERSTANDING THE NATURE OF ARTIFICIAL INTELLIGENCE: THROUGH THE PRISM OF IMAGINARY EXPERIMENTS

Summary. *The article provides a comprehensive socio–philosophical analysis of the main directions in the philosophy of artificial intelligence (AI), such as physicalism, functionalism, and cognitive pluralism. The authors examine key thought experiments ("What Is It Like to Be a Bat?",*

“Chinese Room”, “Swampman”, “Chinese Nation”, and “Philosophical Zombies”) that demonstrate the limitations of purely technical and reductionist approaches to understanding consciousness. Special attention is paid to the problem of “qualia” and the fundamental distinction between syntax and semantics in AI operations. The paper substantiates the need for a transition to a transdisciplinary paradigm that combines natural science and socio–humanitarian knowledge for a deep understanding of the nature of intelligence and the prospects for creating “strong” AI.

Keywords: *artificial intelligence, physicalism, functionalism, qualia, thought experiment, transdisciplinarity.*

Постановка проблеми. Незважаючи на наявність широкої понятійної бази, сучасний етап розвитку цифрових технологій вимагає комплексного, трансдисциплінарного осмислення напрацювань різних наукових напрямків. Результати нейронаук і когнітивної нейробіології не дозволяють сформувати цілісну картину пізнання, оскільки воно розподілене у “спільноті знання” і не може бути повністю зведене до активності нейромереж. Існуючі підходи стикаються з серйозними теоретичними труднощами при спробі пояснити феномен свідомості і “кваліа” (суб’єктивних переживань). Без вирішення питання про те, чи може машина володіти розумінням і семантикою, а не просто маніпулювати символами, неможливо адекватно оцінити потенціал створення “сильного” штучного інтелекту. Крім того, сьогодні чітко простежується необхідність соціокультурної інтеграції: існує запит на посилення взаємодії між розробниками ШІ та представниками соціогуманітарних наук для уточнення стратегій імплементації технологій у сучасне суспільство.

Результати ***аналізу наукових джерел і публікацій*** свідчать про те, що на сьогоднішній день в арсеналі науки є досить широка понятійна база проблематики штучного інтелекту. Перш за все, в контексті даної статті особливий інтерес представляють твори другої половини XX ст. з філософії штучного інтелекту та філософії свідомості, пов’язані з такими напрямками, як фізикалізм, функціоналізм і когнітивний плюралізм, а також з суміжними підходами енактивізму, трансгуманізму, критики “сильного” штучного інтелекту та основ штучних нейронних мереж (доробки А. Тьюринга [1], Дж. МакКарті [2], Дж. Серла [3–4], Т. Нагеля [5], Д. Девідсона [6], Х. Дрейфуса [7], Ф. Варели [8], Н. Бострома [9], Д. Чалмерса [10] та інших).

Наукові доробки зазначених мислителів дійсно окреслюють фундамент, на якому стоїть сучасна філософія штучного інтелекту. Проте, попри значний прогрес у функціоналізмі та нейронних мережах, перехід від “обчислювальної потужності” до “свідомого досвіду” залишається головним каменем спотикання. Основна прогалина лежить на перетині технічного виконання та онтологічного статусу. Ми навчилися будувати нейронні мережі, які перемагають людину в іграх та мистецтві, але ми досі не маємо теорії, яка б пояснила, як математична функція може стати “кимось”.

Отже, сучасний етап розвитку інформаційних і цифрових технологій потребує комплексного трансдисциплінарного філософського осмислення напрацювань різних наукових напрямків, присвячених проблематиці штучного інтелекту.

Мета статті – оцінити рефлексивний потенціал наявних точок зору та підходів до розуміння основ штучного інтелекту для уточнення стратегій його розуміння та імплементації в соціокультурне середовище сучасного суспільства.

Виклад основного матеріалу. Філософія штучного інтелекту є невід’ємною частиною сучасної онтології, філософії свідомості, і намагається випередити онтологічний статус свідомості, її ставлення до мозку як до вищої форми організації матерії і до фізичної реальності. Існує величезна кількість різних підходів до рефлексії

даного питання, проте найбільш впливовими напрямками можна назвати фізикалізм, функціоналізм і когнітивний плюралізм.

У загальному сенсі під фізикалізмом у філософії розуміється точка зору, згідно з якою все існуюче у світі (у тому числі ментальне) є похідним від фізичних процесів. Фізикалізм безпосередньо пов'язаний з природничими науками (у першу чергу з фізикою) і спирається на їх авторитет як онтологічно, так і епістемологічно. Деякі дослідники (наприклад, П. Петті) виділяють наступні ключові тези фізикалізму [11]:

• існують мікрофізичні сутності: є емпіричний світ, подібний до того, який описується за допомогою фізичних законів, і все в цьому емпіричному світі розділено на атомарні та субатомарні рівні, що допускаються фізичною наукою;

• все складається з мікрофізичних сутностей: будь-яка річ в емпіричному світі без залишку складена з субатомних частинок або сама є такою частинкою;

• існують мікрофізичні закономірності: мікрофізичні об'єкти підкоряються певним законам в силу їх властивостей і взаємозв'язків; закони, що діють на мікрорівні, не завжди можуть збігатися з тими, що діють на макрорівні;

• мікрофізичні закономірності регулюють все: макрорівневі закони не є незалежними від субатомних взаємодій.

Однією з ключових проблем фізикалізму можна назвати так звану "дилему теоретика", сформульовану К. Гемпелем [12]. Фізикалізм спирається на фізичну науку, але що під нею розуміється? З одного боку, є сенс ґрунтуватися на сучасних наукових досягненнях, але навіть найпередовіші напрямки, наприклад, квантова фізика, є неповними і суперечливими. З іншого боку, якщо спиратися на якусь ідеальну і завершену науку майбутнього, то фізикалізм стає досить розпливчастою і невизначеною концепцією, оскільки люди зараз не мають ні найменшого уявлення про те, яким може стати наукове знання через десятиліття.

Однак, незважаючи на "дилему теоретика", а також очевидні складнощі, які відчуває сучасна наука при спробах описати і зрозуміти функціонування свідомості, фізикалізм підтримується великою кількістю відомих філософів, серед яких У. Куайн, Р. Рорті, Д. Деннет, П. Черчленд, Р. Пенроуз [13–16] та ін.

Функціоналізм також є досить популярним напрямком у філософії свідомості, причому в силу досить широких поглядів, що лежать в основі, його можна вважати як похідним від фізикалізму, так і продовженням біхевіоризму або самостійною концепцією. Функціоналізм постулює, що ментальний стан залежить не від внутрішньої структури об'єкта, а від способу функціонування. Проводячи аналогію з комп'ютерами, тіло (або мозок) – це процесор, а свідомість – програмне забезпечення. Головні тези функціоналізму – "нейтральний характер функцій" (природа ментальних станів не матеріальна і не ідеальна) і "множинність реалізації" (ментальні процеси можуть бути реалізовані на різних носіях: на білковому мозку, комп'ютерному процесорі або кремнієвому субстраті).

Однією з ключових проблем, пов'язаних з функціоналізмом, є питання співвідношення даної концепції з важливим поняттям аналітичної філософії свідомості – з кваліа, тобто зі здатністю мати "психічні стани", з тим, як "речі виглядають, звучать і пахнуть", з тим, "як відчувається біль" [17]. Оскільки функціоналісти часто мають на увазі під психічними станами тільки функції, що перетворюють вхідні сенсорні дані в поведінковий "вихід", то відчуття самі по собі втрачають свою значущість. Однак противники такого підходу стверджують: у феноменальній свідомості є дещо, що виходить за межі функцій, а відчуття, наприклад, болю – це щось більше, аніж

“причинно-наслідковий зв’язок між дотиком до гарячої праски і відсмикуванням руки” [18].

Найвідомішими аргументами проти функціоналізму є “відсутність кваліа” і “перевернуті кваліа”. Сенс “відсутності кваліа” полягає в тому, що функціонального опису розуму недостатньо для вловлювання феноменальних властивостей свідомості. Прикладом цього аргументу може служити уявний експеримент “Китайська нація”, який детальніше буде розглянуто нижче. “Перевернуті кваліа” – це ситуація, коли дві людини функціонально ідентичні, але відчувають однакові відчуття при взаємодії з різними предметами. Оскільки подібний випадок інвертованого сприйняття можливий, значить, функціоналізм – як мінімум обмежений у застосуванні підхід.

Втім, навіть наявність досить серйозних контраргументів не зменшує популярності функціоналізму і подібних поглядів дотримуються Д. Льюїс [19], Т. Хоган [20], С. Шемейкер [21].

Когнітивний плюралізм ґрунтується на ідеї не єдності, а множинності, не зведення свідомості до тіла, а фундаментальності свідомості нарівні з фізичним світом. Поняття “когнітивний плюралізм” було запропоновано С. Хорстом [22]; у сучасній філософії свідомості існує ціла низка немоністичних концепцій, наприклад, містеріанство К. МакГінна [23] і панпсихізм Д. Чалмерса [10]. Напрямки, подібні до когнітивного плюралізму, ймовірно, в даний момент стають найсерйознішими противниками фізикалізму і функціоналізму, проте їх розробка і практичне застосування видаються можливими лише у відносно віддаленому майбутньому за умови перегляду фундаментальних наукових принципів. Проте саме наявність подібних ідей сприяє виробленню аргументації антифізикалістів, причому найцікавіші з них представлені яскравими уявними експериментами, про які йтиметься далі.

“Кажан”. У 1974 р. філософ Т. Нагель опублікував статтю “Як це – бути кажаном?” (“What is it like to be a bat?”), де запропонував наступний уявний експеримент: уявімо, що нам необхідно зрозуміти, як це бути кажаном. Автор говорить про те, що ми обмежені можливостями свого розуму і не можемо уявити, як користуватися ехолокацією, літати в сутінках і ловити пащею комах. У наших силах лише створити якусь досить примітивну схему, що описує існування кажана, адже люди в змозі виділити тільки той досвід, який може бути порівнянний з людським, або про наявність якого нам відомо (голод, зір, ехолокація), проте буде неправильно заперечувати ймовірність того, що у миші є відчуття і переживання, не усвідомлювані людьми в принципі, що виходять за рамки всіх наших знань і уявлень. Як зазначає Нагель, “роздуми про те, як це бути кажаном, приводять нас до висновку, що існують факти, які не ґрунтуються на істинності висловлювань, виражених людською мовою” [5, p. 441]. Даний висновок філософа видається вкрай важливим для визначення рівня розвитку штучного інтелекту, оскільки уможливлюється помилкове заниження здібностей ШІ в тому випадку, коли ці здібності виходять за межі людського знання.

Крім того, Т. Нагель звертає увагу на те, що на перший погляд однаковий для двох людей досвід може сприйматися ними абсолютно по-різному, до того ж, людині вкрай важко побачити свій досвід з позиції третьої особи, з іншої точки зору [5, p. 443]. Наприклад, кожен з нас здатний оцінити те, що відбувається, за допомогою об’єктивних параметрів (зафіксувати за допомогою фізичних приладів розряд блискавки або подивитися результати МРТ випробуваного, який малює коло), проте дві людини, які бачать блискавку або малюють кола, можуть при цьому переживати абсолютно різний досвід. Таким чином, філософ показує суб’єктивність свідомості і неможливість звести

психічні процеси до фізичних і висловлюється проти редукціонізму в питанні "свідомість – тіло".

"Китайська кімната". Одним з найважливіших уявних експериментів у філософії штучного інтелекту і філософії свідомості є "Китайська кімната" Дж. Серла. Цей експеримент був придуманий як відповідь на критерій Тьюринга, згідно з яким можна визнати існування інтелекту у машини в тому випадку, якщо в процесі дистанційного спілкування людина не здатна визначити, що взаємодіє з комп'ютером. Серл виступив категорично проти прирівнювання автоматичної видачі машиною відповідної відповіді до процесів мислення і розуміння і в 1980 р. у статті "Розум, мозок і програми" ("Minds, brains and programs") запропонував наступний експеримент. Необхідно уявити людину – носія англійської мови, яка не говорить китайською, – і замкнути її в кімнаті з текстом китайською мовою, після чого передати їй другий текст китайською та написану англійською інструкцію, що дозволяє знайти співвідношення між двома текстами. Потім принести третій текст китайською мовою і нові інструкції англійською мовою, які дозволять вибудувати зв'язки з першими двома текстами і написати послідовність китайських символів, які необхідно винести за межі кімнати. Перший текст варто називати "сценарієм", другий – "історією", третій – "питаннями", символи, що повертаються людиною – "відповідями на питання", а інструкції англійською – "програмою". Через деякий час людина в кімнаті настільки добре навчиться дотримуватися інструкцій з використання китайських символів, що її відповіді будуть не відрізнити від відповідей людини, яка говорить китайською. Таким чином, замкнена в кімнаті людина пройде тест Тьюринга, не розуміючи ні слова китайською мовою, з чого, на думку Дж. Серла, можна зробити висновок: "поки програма визначається в термінах обчислювальних операцій над формально визначеними елементами, вона не має зв'язку з розумінням" [4, p. 421].

"Китайська кімната" стала чи не найбільш обговорюваним уявним експериментом у філософії штучного інтелекту, суперечки про який тривають і досі, і її вплив на уми і сьогодні є дуже великим. Деякі аргументи проти "Китайської кімнати" можна представити наступним чином: 1) людина в кімнаті не розуміє китайської мови, але розуміння є у просторі кімнати в цілому, разом з людиною, інструкціями і текстами китайською мовою; 2) людина в кімнаті розуміє китайську, хоча і не знає її, оскільки можливе розуміння без знання, або вона розуміє китайську на підсвідомому рівні; 3) проста обробка природної мови, описана в "Китайській кімнаті", не є розумінням, проте досягти розуміння може комп'ютер, вбудований в тіло робота з необхідними датчиками і двигунами і взаємодіючий з фізичним світом, або система, що точно імітує роботу людського мозку. У статті "Чи є розум мозку комп'ютерною програмою?" ("Is the Brain's Mind a Computer Program?") Серл називає всі ці аргументи неадекватними, тому що вони не можуть бути протиставлені реальному сенсу "Китайської кімнати", який полягає в розділенні використання формальних символів комп'ютером і ментальних процесів, що відбуваються в мозку, тобто в розділенні синтаксису і семантики [3, p. 30]. Обговорення можливостей розуміння штучним інтелектом текстів на семантичному рівні актуальне й досі.

Варто також зазначити, що саме завдяки "Китайській кімнаті" виник поділ штучного інтелекту на "сильний" і "слабкий". Дж. Серл мав на увазі під "сильним" ШІ певним чином запрограмований комп'ютер, здатний на "розуміння і здійснення інших розумових процесів" [4, p. 417]. І хоча існуючі сьогодні форми штучного інтелекту можна охарактеризувати як "слабкі", тобто такі, що вирішують завдання, в тому числі творчого характеру, в одній або декількох суміжних царинах, прогрес в даній царині

змушує задуматися про ймовірність появи ШІ, що володіє порівнянними або перевершуючими людину когнітивними здібностями.

Поточний розвиток технологій машинного навчання призводить до відновлення існуючих і появи нових дискусій про можливість створення "сильного" штучного інтелекту. Так, у книзі шведського філософа Н. Бострома "Штучний інтелект: етапи, загрози, стратегії" ще 10 років тому автор описує кілька можливих шляхів виникнення "надрозуму", в тому числі стрімке ускладнення комп'ютерів, створення повноцінної моделі головного мозку, а також вдосконалення мереж, що пов'язують людей і програми (наприклад, Інтернету), що призведе до появи колективного надрозуму [9].

Але чи можливо в принципі створення "сильного" штучного інтелекту? Багато дослідників і філософів схильні відповідати на це питання позитивно, проте у такого підходу є і противники. Одним з найвідоміших є представник феноменології Х. Дрейфус. Він вивів чотири припущення (біологічне, психологічне, епістемологічне та онтологічне), на яких ґрунтується позиція прихильників появи штучного розуму. Біологічне припущення полягає в тому, що мозок і комп'ютер використовують у роботі схожі механізми. Психологічне припущення, яке випливає з біологічного, передбачає, що розум, як і машина, обробляє дискретні дані у вигляді символів, спираючись на формальні правила. Епістемологічне припущення стверджує, що всі знання і дії людини потенційно піддаються формалізації і подальшому відтворенню штучним розумом. Нарешті, онтологічне припущення базується на тому, що все в світі, "що є істотним для розумної поведінки, може бути представлено в термінах множини чітко визначених незалежних елементів" [7].

Американський філософ стверджував, що це правило червоною ниткою проходить через всю західну філософську традицію, починаючи з Платона і закінчуючи Л. Вітгенштейном, який у "Логіко-філософському трактаті" прийшов до думки описати світ як безліч атомарних фактів у формі логічно незалежних речень. Ідея, що "ми живемо у світі, в якому гарантована ясність, визначеність і керованість" [7], лягла в основу природничих наук, стала поштовхом для науково-технічного прогресу і вивела людство на принципово новий рівень якості життя. На думку Х. Дрейфуса, саме онтологічне припущення робить появу "сильного" ШІ принципово можливою для прихильників виникнення штучного розуму, адже при розвитку наявних технологій він повинен буде оперувати чітко структурованою моделлю світу.

Ґрунтуючись на ідеях Х. Дрейфуса, в 2020 р. Р. Ф'єлланд у статті "Чому загальний штучний інтелект не буде реалізований" ("Why general artificial intelligence will not be realized") стверджує, що сильний (або загальний) штучний інтелект не може бути створений, оскільки комп'ютер ніколи не зможе набути властивостей людського мислення: розсудливості – здатності вибирати правильні підходи до досягнення мети в конкретних ситуаціях – і мудрості, тобто здатності бачити ціле. Також досить велика частина людського досвіду не може бути передана словами, а значить, її не можна сформулювати в будь-якій формі, в тому числі й за допомогою програмного коду. Нарешті, для набуття справжнього розуму необхідно діяти в реальному світі, що означає мати тіло, бути частиною якоїсь культури і соціальної групи [24].

"Болотна людина". Уявний експеримент "Болотна людина" був запропонований філософом Д. Девідсоном в 1987 р. у статті "Знати власні думки" ("Knowing One's Own Mind"). Автор пропонує уявити, що він опинився на болоті під час грози. Блискавка влучає в сухе дерево поруч з філософом, який виявляється розщепленим на молекули, а дерево неймовірним чином стає точною фізичною копією Девідсона.

“Болотна людина” йде з болота, зустрічає друзів автора, вітається з ними англійською мовою, йде до філософа додому і сідає писати статті. Чи можна сказати, що “Болотна людина” – це Д. Девідсон? Філософ вважає, що відповідь має бути негативною, адже копія Девідсона “не може нічого розпізнати, оскільки вона нічого не пізнавала”. “Болотна людина” не в змозі розуміти значення слів, тому що “не вивчала їх у контексті” [6, p. 444].

Поширюючи цей уявний експеримент на штучний інтелект, варто зазначити, що, виходячи з вищевикладеного, отриманий в результаті потенційного перенесення людської свідомості на новий носій сильний ШІ не буде володіти контекстним розумінням і не зможе дублювати вихідну людську особистість.

“Китайська нація”. Уявний експеримент, згаданий вище як аргумент “відсутності кваліа” проти функціоналізму, був сформульований Н. Блоком у 1978 р. у статті “Проблеми з функціоналізмом” (“Proubles with Functionalism”) [25]. Філософ запропонував уявити, що все населення Китаю функціонально відповідає нейронам головного мозку. У кожного громадянина є двостороннє радіо, і в певний момент люди починають дзвонити один одному подібно до того, як одні нейрони передають сигнал іншим. Дані про дзвінки відображаються на карті, яку можна бачити з будь-якої точки Китаю. Потім описана система радіопередачі з’єднується зі штучним тілом, яке забезпечує сенсорні входи і поведінкові виходи. Виходить, що з точки зору функціоналізму “китайська нація” являє собою мозок в момент якогось психічного стану, а значить, вона володіє розумом. Такий висновок, з точки зору Блока, абсурдний і доводить хибність функціоналізму. У контексті сучасного етапу розвитку штучного інтелекту даний аргумент видається вкрай актуальним, адже ті творчі завдання, які вирішують штучні нейронні мережі, у людини пов’язані з психічними станами, з пережитим досвідом, з відчуттями, тобто з кваліа. Якщо їх немає, то і говорити про наявність розуму у ШІ навіть при створенні ним музичних композицій або художніх зображень не уявляється можливим.

“Філософські зомбі”. Філософські зомбі (philosophical zombies, або p–зомбі) – це вигадані істоти, які фізично еквівалентні людям, але не мають свідомого досвіду – кваліа. Вважається, що вперше поняття зомбі було використано Р. Кірком [26], а найбільш докладну картину даного антифізикалістського аргументу розробив Д. Чалмерс у книзі “Свідомий розум” (“The conscious mind”) [10]. Філософ пропонує уявити світ, у всьому схожий на наш і населений філософськими зомбі. З огляду на те, що такий світ мислимий, він метафізично можливий. Якщо ґрунтуватися на фізикалізмі, то все, що існує в нашому світі, має фізичну природу (в тому числі свідомість). При істинності фізикалізму метафізично можливий ідентичний нашому світ зомбі повинен містити всі його компоненти, в тому числі і свідомий досвід. Однак ми можемо собі уявити такий світ без кваліа, тому, за Чалмерсом, фізикалізм є хибним (за правилом modus tollens).

Досить часто зустрічається уявлення про філософського зомбі як про людину без душі. Однак насправді єдина різниця між нами і p–зомбі полягає у відсутності у останніх нефізичних суб’єктивних переживань. Філософський зомбі поводиться так само, як людина, він розрізняє кольори, реагує на удар струмом відсмикуванням кінцівки, але не здатний зрозуміти “червоність” помідора або суб’єктивне відчуття болю, хоча і заявляє, що може це робити.

Питання про те, чи можливе існування філософських зомбі, видається ключовим для штучного інтелекту, адже поняття несвідомого зомбі логічно необхідне для вироблення поняття “свідомої істоти”. Вже зараз ШІ (очевидно, будучи “слабким”, а не

“сильним”) обіграє кращих гравців у шахи, відкриває нові антибіотики, створює художні твори, тобто в якійсь мірі поводиться по-людськи. Чи володіє він кваліа, чи є у нього суб’єктивні переживання? Ймовірно, проблема сучасної філософії штучного інтелекту, що спирається на аналітичну філософську традицію, якраз і полягає в тому, що відповісти на ці питання з позиції сучасної науки неможливо, оскільки кваліа є внутрішньою характеристикою, яку неможливо зафіксувати жодним приладом і визначити жодним тестом.

Філософські зомбі і кваліа не можуть бути фальсифіковані, проте мають найважливіше прикладне значення, що полягає в можливості наділяти суб’єктністю штучний інтелект. Можливо, для пошуку відповіді на питання про наявність у ШІ свідомого досвіду варто звернути увагу на підходи, що пропонують розглядати суб’єктивний досвід не тільки з позиції природничо-наукового методу, але і крізь призму соціокультурного знання. Хотілося б зазначити, що, з огляду на певну вразливість міждисциплінарності, яка пов’язана з “несистемністю” і “нестійкістю” даного підходу до інтеграції, варто визначити спроби використовувати поєднання природничих і соціогуманітарних наук для вирішення проблем штучного інтелекту як “трансдисциплінарні”.

Показово, що подібний запит існує не тільки з боку гуманітарних дисциплін, але й надходить від представників нейронаук. Наприклад, у 2021 р. була опублікована стаття “Когнітивна нейробіологія зустрічається зі спільнотою знань” (“Cognitive Neuroscience Meets the Community of Knowledge”), де вчені, що працюють у галузі когнітивної нейробіології, висунули гіпотезу, згідно з якою процес пізнання у кожної людини відбувається не повністю в її голові, а розподіляється між людьми і утворює “спільноту знання” [27, p. 2]. Результати численних експериментів, проведених за допомогою методів вимірювання активності головного мозку і нейровізуалізації, не дозволяють сформувати цілісну картину, оскільки, на думку дослідників, пізнання відбувається не стільки в голові конкретної особи, скільки на рівні міжмозкових взаємодій, що підтримують загальне символічне знання, яке не можна звести до нейромереж.

Аналізуючи найбільш впливові напрямки філософії штучного інтелекту варто згадати і енактивізм – один із найбільш радикальних та цікавих напрямків у сучасній когнітивній науці та філософії свідомості. Якщо традиційний підхід розглядає розум як “комп’ютер”, що обробляє дані, то енактивізм стверджує: пізнання – це не відображення світу, а його створення через дію. Цей підхід сформувався завдяки працям Ф. Варели, Е. Томпсона та Е. Рош (їхня спільна праця “Втілений розум” (“The Embodied Mind”, 1991)) [28].

Згідно з енактивізмом, жива істота не просто отримує інформацію ззовні. Світ “виступає” (enacted) для організму через його специфічну активність. Наприклад, для бактерії цукор – це не просто “хімічна формула”, а “їжа” або “напрямок руху”. Тобто смисл світу з’являється лише у процесі взаємодії. Розум не обмежений лише мозком. Пізнання залежить від усього тіла: структури нервової системи, м’язів, сенсорних органів. Наші думки та концепції обмежені та сформовані тим, як ми рухаємося у просторі.

Ця концепція стверджує, що живі системи постійно відтворюють самі себе. Пізнання – це необхідний інструмент для підтримки життя та цілісності організму.

Енактивізм заперечує ідею, що існує “об’єктивний світ”, який ми просто фотографуємо очима. Світ і пізнаючий суб’єкт взаємопов’язані, як дві сторони одного аркуша паперу. Без дій організму світ не має для нього структури, а без світу організм не може діяти.

З точки зору енактивізму, штучні інтелектуальні конструкції (як-от ChatGPT) – це “інтелект у банці”, який не має справжнього розуміння, оскільки він позбавлений: тілесного досвіду (не знає, що таке “холодно” чи “важко” на фізичному рівні), а також потреб (у нього немає біологічного прагнення до самозбереження, яке б наділяло інформацію смислом).

Порівняльна таблиця ключових підходів до розуміння понятійних основ штучного інтелекту

Напрямок	Основна теза	Що залишається невирішеним?
Фізикалізм	Свідомість – це лише стан матерії.	Як виникла прірва між матерією та суб’єктивним “Я”?
Функціоналізм	Інтелект – це функція (неважливо, на чому вона реалізована).	Чи достатньо лише функціональної схожості для наявності свідомості?
Когнітивний плюралізм	Не існує єдиної моделі розуму. Інтелект — це набір різних, часто несумісних стратегій і способів опису дійсності.	Як об’єднати різні моделі в цілісну архітектуру ШІ, яка б не розпадалася на окремі алгоритми?
Енактивізм	Розум формується через дію в середовищі.	Чи можна створити “цифровий енактивізм” у віртуальних світах?

Висновки. У даний час штучний інтелект є міждисциплінарною галуззю з переважанням досліджень у сфері інформаційних технологій і нейрофізіології. Соціогуманітарний напрямок розглядається розробниками як додатковий, проте у зв’язку з обмеженнями, що існують у сучасних підходах до ШІ, зростає попит на посилення взаємодії з представниками філософії, культурології, соціології, мистецтвознавства.

Більшість авторів, творчість яких аналізувалася в статті (особливо Д. Чалмерс та Дж. Серл), останнім часом все більше схиляються до думки, що проблема ШІ – це не лише питання “заліза” (hardware) чи “софту” (software), а питання біологічного натуралізму та того, як саме матерія породжує сенс. Таким чином, при вивченні творчого потенціалу штучного інтелекту та оцінюванні перспектив його поширення варто використовувати комплексний підхід, що спирається як на гуманітарні, так і природничі дисципліни.

У зв’язку з необхідністю об’єднати підходи природничих і гуманітарних наук для дослідження штучного інтелекту, не можна не звернути увагу на проблему єдності знання, пов’язану зі складністю рефлексивної активності в царинах, що працюють для виникнення і розвитку нового знання. Якщо в класичній науковій парадигмі, спрямованій на чітке розділення суб’єкта і об’єкта, знання виявилося відокремленим від суб’єкта і засобів створення і зведеним до інформації, в некласичній, навпаки, воно розглядалося у взаємодії з суб’єктом і засобами і набуло характеру особистісного, то в постнекласичному підході система “суб’єкт-метасуб’єкт” і розворот у бік суб’єкта зробили знання більш складною і нелінійною структурою, що поєднує як традиційні, так

і абсолютно нові підходи, які з'явилися в результаті поміщення суб'єктів у цифрову реальність. Велике значення для збереження цілісності людського Я в постнекласичній раціональності набуває середовище "саморозвиваюче" і "рефлексивно-активне", яке саме по собі можна уявити "метасуб'єктом".

Уявляється, що саме таким середовищем здатні виступити сучасні технології штучного інтелекту.

ПОДЯКИ: Немає
КОНФЛІКТ ІНТЕРЕСІВ: Немає

Використана література

1. Turing A. Computing machinery and intelligence. *Mind.* 1950. № 50. P. 433–460.
2. McCarthy J., Hayes P. Philosophical problems from the standpoint of AI. *Machine Intelligence*. Edinburgh: Edinburgh University Press Press, 1969. P. 463–502.
3. Searle J. R. Is the Brain's Mind a Computer Program? *Scientific American*. 1990. № 262 (1). P. 26–31.
4. Searle J. R. Minds, brains and programs. *Behavioral Sciences*. 1980. № 3. P. 415–431.
5. Nagel T. What Is It Like to Be a Bat? *The Philosophical Review*. 1974. Vol. 83. № 4. P. 435–450.
6. Davidson D. Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*. 1987. Vol. 60. № 3. P. 441–458.
7. Dreyfus H. L. What computers still can't do: a critique of artificial reason. URL: https://epdf.pub/what–computers–still–cant–do–a–critique–of–artificial–reason.html (дата звернення: 25.01.2026).
8. Varela F. The Embodied Mind: Cognitive Science. Cambridge: MIT Press, 1991. 457 p.
9. Bostrom N. Superintelligence: Paths, Dangers, Strategies. URL: https://readli.net/superintelligence–paths–dangers–strategies/ (дата звернення: 20.01.2026).
10. Chalmers D. J. The conscious mind. In search of a fundamental theory. URL: https://personal.lse.ac.uk/ROBERT49/teaching/ph103/pdf/Chalmers_The_Conscious_Mind.pdf (дата звернення: 25.12.2025).
11. Pettit P. A definition of physicalism. *Analysis*. 1993. Vol. 53. № 4. P. 214–217.
12. Hempel C. Comments on Goodman's Ways of Worldmaking. *Synthese*. 1980. № 45. P. 193–200.
13. Rorty R. Holism, Intrinsicality, and the Ambition of Transcendence. *Dennett and His Critics. Demistifying Mind*; ed. by B. Dahlbom. Cambridge (Mass.), 1993. P. 185–202.
14. Dennett D. C. Consciousness explained. London: Penguin, 1993. 528 p.; Churchland P. M. Matter and consciousness. Massachusetts: MIT press, 2013. 304 p.
15. Churchland P. M. Matter and consciousness. Massachusetts: MIT press, 2013. 304 p.
16. Penrose R. Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford: Oxford University Press, 1994. 480 p.
17. Block N. Qualia. *Oxford Companion to the Mind*. Oxford: Oxford University Press, 2004. URL: https://philpapers.org/rec/BLOQ (дата звернення: 25.01.2026).
18. Walsch J. Can a Functionalist Account for Qualia? *Oxford Philosophical Society*. 2017. P. 2 (1–4). URL: https://oxfordphilsoc.org/Documents/StudentPrize/2017_H2.pdf (дата звернення: 21.01.2026).
19. Lewis D. An Argument for the Identity Theory. *The Journal of Philosophy*. 1966. Vol. 63. P. 17–20.
20. Horgan T. Functionalism, Qualia, and the Inverted Spectrum. *Philosophy and Phenomenological Research*. 1984. Vol. 44. № 4. P. 453–469.
21. Shoemaker S. Absent Qualia are Impossible – A Reply to Block. *The Philosophical Review.* 1981. Vol. 90. № 4. P. 581–599.

22. Horst St. Beyond Reduction. Philosophy of Mind and Post–Reductionist Philosophy of Science. *Minds and Machines*. 2008. № 18 (3). P. 421–423.

23. McGinn C. The Mysterious Flame: Conscious Minds in a Material World. New York: Basic Books, 1999. 256 p.

24. Fjelland R. Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*. 2020. № 7 (1). URL: https://www.nature.com/articles/s41599–020–0494–4 (дата звернення: 15.01.2026).

25. Block N. Troubles with functionalism. *Minnesota Studies in the Philosophy of Science*. 1978. P. 261–325.

26. Kirk R., Squires R. Zombies v. materialists. *Proceedings of the Aristotelian Society.* Supplementary Volumes. 1974. Vol. 48. P. 135–163.

27. Sloman S. A., Patterson R., Barbey A. K. Cognitive Neuroscience Meets the Community of Knowledge. *Frontiers in System Neuroscience.* 2021. Vol. 15. P. 1–13.

28. Varela F. J., Thomson E., Rosch E. The Embodied Mind. URL: https://monoskop.org/images/2/21/Varela_Thompson_Rosch_The_Embodied_Mind_Cognitive_Science_and_Human_Experience_1991.pdf (дата звернення: 29.01.2026).

Олександр Петрович Дзьобань
доктор філософських наук, професор,
головний науковий співробітник Державної наукової установи "Інститут інформації, безпеки і права Національної академії правових наук України"
04053, Україна, м. Київ, пров. Несторівський, 4
email: a_dzeban@ukr.net
Максим Вікторович Жушман
кандидат юридичних наук, доцент
доцент кафедри цивільної юстиції та адвокатури Національного юридичного університету імені Ярослава Мудрого
61024, Україна, м. Харків, вул. Григорія Сковороди, 77
email: m.v.zhushman@nlu.edu.ua

Oleksandr P. Dzioban
Doctor of Philosophy, Professor
Chief Research Fellow of the State Scientific Institution "Institute of Information, Security and Law of the National Academy of Legal Sciences of Ukraine"
4 Nestorivskyi Lane, Kyiv, 04053, Ukraine
email: a_dzeban@ukr.net
Maksym V. Zhushman
Ph.D. in Law, Associate Professor
Associate Professor of the Department of Civil Justice and Advocacy Yaroslav Mudryi National Law University
77 Hryhorii Skovoroda Street, Kharkiv, 61024, Ukraine,
email: m.v.zhushman@nlu.edu.ua

Рекомендоване цитування: Дзьобань О.П., Жушман М.В. Концептуальні межі основних стратегій осмислення природи штучного інтелекту: крізь призму уявних експериментів. *Інформація і право.* № 2(57)/2026. 2026. С. 64-75. https://doi.org/10.37750/2616-6798.2026.2(57).364303

Suggested Citation: Dzioban O., Zhushman M. (2026) Conceptual Limits of the Basic Strategies for Understanding the Nature of Artificial Intelligence: Through the Prism of Imaginary Experiments. *Information and Law*. 2(57)/2026. 64-75. https://doi.org/10.37750/2616-6798.2026.2(57).364303

Дата надходження статті до редакції: 27.04.2026 р.
Дата прийняття статті до друку після рецензування: 28.04.2026 р.
Дата публікації (оприлюднення): 31.05.2026 р.

~~~~~~~~~~~~~~ * * * ~~~~~~~~~~~~~~
~~~~~~~~~~~~~~