

УДК 11/14.000.141./ 165.0+165.8.

DOI <https://doi.org/10.30970/PPS.2025.59.13>

ПРОБЛЕМА СУБ'ЄКТНОСТІ ШТУЧНОГО ІНТЕЛЕКТУ: ФІЛОСОФСЬКИЙ АНАЛІЗ

Олексій Крезе

*Державний торговельно-економічний університет,
кафедра філософії, соціології та політології
вул. Кіото, 19, 02156, м. Київ, Україна*

У статті здійснено спробу філософського аналізу фундаментальної проблеми можливості виникнення суб'єктності штучного інтелекту. На даний час у сучасному філософському дискурсі наявна певна термінологічна неузгодженість, оскільки штучний інтелект отожднюється із штучною свідомістю. Проте інтелект є лише однією з функцій свідомості. Водночас, питання суб'єктності безпосередньо пов'язане із таким аспектом свідомості як «самосвідомість», яка, власне, і передбачає усвідомлення себе як автономного суб'єкта, який наділений можливістю мати квалітативні переживання, виокремлювати себе зі світу, здійснювати пізнання та створювати нове знання. У свою чергу, попри значні технологічні досягнення, навіть найдосконаліші моделі штучного інтелекту обмежені алгоритмічно, не мають доступу до квалітативних переживань, не можуть усвідомлювати свої внутрішні стани та мати індивідуальний досвід. Отже, проблема відтворення самосвідомості, є не стільки технічною, скільки фундаментально-філософською, оскільки вона пов'язана із питанням самої природи суб'єктності та необхідності безпосереднього феноменального доступу до реальності для її виникнення.

Щодо цього, можна виділити декілька підходів до вирішення цієї проблеми. Так функціоналісти притримуються позиції каузальної обумовленості свідомості фізичними процесами, що відбуваються у складних системах. У свою чергу, їх опоненти, що притримуються концепції ідеалізму або дуалізму стверджують, що свідомість є проявом існування окремої нематеріальної субстанції.

Незважаючи на глибокі розбіжності в обох підходах, можна констатувати, що більшість сучасних філософів схильні розглядати самосвідомість та її центр – «Я» не в абстрактному метафізичному вимірі, а як складний динамічний процес, що має психофізіологічну обумовленість (але саме питання нейробіологічної редукції залишається відкритим). Таким чином, дослідження питання суб'єктності штучного інтелекту із необхідністю наближає нас до більш глибокого розуміння самого людського «Я» та підкреслює необхідність продовження фундаментальних філософських досліджень у нерозривному поєднанні із найновішими науковими досягненнями.

Ключові слова: суб'єктність, штучний інтелект, свідомість, самосвідомість, кваліа, функціоналізм, редукціонізм.

Постановка проблеми. Питання суб'єктності штучного інтелекту на сучасному технологічному етапі розвитку людської цивілізації набуло особливої актуальності. У першу чергу, це пов'язане із успіхом нейронних мереж, які стали доступними для широкого кола споживачів. Безпілотні автомобілі, голосові асистенти, мовні моделі вже є невід'ємною частиною нашого повсякденного життя. На даний час сучасні моделі штучного інтелекту здійснюють самонавчання на основі обробки гігантських обсягів даних, що містяться у різноманітних джерелах (наукових статтях, книгах, новинах тощо). Вони мають можливість формувати результат, оперуючи вірогідностями та статистичними зв'язками між даними, але поки що не здатні генерувати нове знання у тому сенсі, як це роблять вчені чи дослідники. Проте постійне ускладнення систем штучного інтелекту, фіксація проявів

емерджентності та непрозорість процесів отримання кінцевого результату ставлять перед дослідниками питання: чи є у сучасних філософів достатні підстави для здійснення однозначного негативного чи позитивного висновку про фундаментальну неможливість виникнення суб'єктності ШІ. Крім того, важливого значення набуває фундаментальність питання щодо епістемологічного потенціалу штучного інтелекту, а саме – чи може він бути джерелом і носієм знання. Водночас, слід зауважити, що технічний розвиток таких технологій відбувається, у певній мірі, автономно від філософських досліджень. Філософи сьогодення змушені, з одного боку, ретроспективно аналізувати існуючі феномени які проявилися у нашому суспільному бутті і вимагають осмислення, а з іншого – мають передбачити процеси, що безповоротно трансформують суспільну реальність вже у найближчі десятиліття.

Огляд останніх досліджень. Проблематика співвідношення штучного інтелекту та властивостей людської свідомості є однією з центральних тем сучасного філософського дискурсу. Слід зазначити, що філософські наукові розвідки у цій сфері здійснюються у тісному поєднанні із дослідженням, власне, людської свідомості. Аналіз сучасних філософських досліджень у предметній галузі свідомості та штучного інтелекту обумовлює необхідність звернення до ключових робіт, які обумовили основи сучасного наукового дискурсу у даній сфері. Передусім, це стаття Т. Нагеля «Що означає бути кажаном», дослідження Г. Дрейфуса «Що комп'ютери не можуть робити», роботи Дж. Сьорла, зокрема, стаття «Свідомість, мозок і програми» у якій представлений ментальний експеримент «Китайська кімната». Істотний внесок в обговорення проблеми свідомості здійснив Ф. Джексон, який запропонував ментальний експеримент «Кімната Мері» у своїй статті «Епіфеноменальні квалії».

У подальшому сформувалися декілька підходів до розуміння свідомості у т.ч. у контексті можливості її відтворення. Д. Деннет у низці своїх робіт, зокрема у книзі «Пояснена свідомість» послідовно розвиває фізикалістський підхід, заперечуючи роль «квалії» як основи свідомості. У свою чергу, Д. Чалмерс (у роботі «Усвідомлюючий розум» та ін.) формулює основи «панпсихізму, визначаючи свідомість як фундаментальну властивість самої реальності.

Також серед сучасних досліджень останніх років не можна не виділити роботу Ю. Харарі «21 урок для XXI століття», де проблематика суб'єктності штучного інтелекту розглядається у контексті впливу сучасних технологій на глобальні соціальні процеси. Т. Метцингер у книзі «Тунель его» приділяє увагу можливості виникнення у машин внутрішнього образу «Я». Г. Кьоніг у книзі «Кінець індивідуума» аналізує вплив штучного інтелекту на традиційні уявлення про знання, свободу волі та творчість, а також підіймає етичні аспекти взаємодії людини і машини.

У своїй праці «Розум і дух» Р. Негарестані досліджує концепцію «раціонального я» на відміну від нейробіологічно обумовленого «феноменального самосприйняття». Г. Каппелен та Дж. Денвер у роботі «Зробити ШІ зрозумілим. Філософські засади» зосереджують фокус свого дослідження на метасемантичному контексті функціонування систем штучного інтелекту і некоректності застосування антропоцентричного підходу до дослідження таких систем. Р. Беркемер, М. Гроттке у статті «На що насправді здатний ШІ» досліджують питання підвищення ефективності моделей ШІ на основі підходу, що не заснований на людських практиках, а самостійно створює оптимальний алгоритм.

Українські дослідники також здійснюють активну наукову розробку даного напрямку, зокрема, вважаємо за необхідне виділити монографію Д. Сепетія «Свідомість як суб'єктивність: таємниця Я», дисертаційне дослідження Швиркова О. І. «Проблема штучного інтелекту і людиновимірність штучних інтелектуальних систем», статті А. Кадикало, О. Комар, В. Воронкової та В. Никитенко, А. Шевченка, та ін.

Мета статті. Метою статті виступає аналіз сучасних філософських досліджень штучного інтелекту у контексті актуальних підходів до вирішення проблеми можливості створення «синтетичної свідомості» із притаманними для неї властивостями самоусвідомлення, отримання і створення нового знання.

Виклад основного матеріалу. Вважаємо, що основною методологічною проблемою вивчення питання суб'єктності штучного інтелекту виступає складність чіткого виокремлення самого предмету дослідження, оскільки такий предмет не є цілісним і розпадається на декілька складових: визначення поняття суб'єктності як такої; встановлення співвідношення суб'єктності та свідомості; формулювання поняття штучного інтелекту і, нарешті, вирішення питання чи може взагалі існувати штучна свідомість.

Слід зауважити, що саме поняття «Штучний інтелект» у сучасному філософському дискурсі майже завжди отожднюється із «Штучною свідомістю», хоча «інтелект» є лише одним із функціональних проявів свідомості (тобто похідним від неї). На це звертає увагу А. Шевченко, керівник інституту проблем штучного інтелекту МОН і НАН України, зауважуючи також, що «штучний інтелект, з одного боку, розглядається як «чорна скринька», яка представляє собою складну обчислювальну структуру, що має: а) спроможність до навчання й самонавчання, б) здатність розпізнавати й синтезувати образи (мову, зображення тощо), в) сенсорну (рецепторну) систему, тобто штучні органи відчуття та ін». Водночас штучний інтелект виступає функцією штучної свідомості, яка «представлена створеною та контрольованою системою алгоритмів, забезпечує самонавчання згідно з наявною інформацією, набутими знаннями, правилами, законами суспільства та своїм досвідом» [18, с. 8–9].

Джон Мак Карті, якого вважають «батьком» терміну «Штучний інтелект» послідовно ухилявся від його чіткого і універсального визначення, зазначаючи що, передусім, інтелект є обчислювальною складовою, яка спрямована на досягнення цілей у світі. При цьому ми не можемо чітко охарактеризувати, які саме обчислювальні процедури слід називати інтелектуальними [23].

Таким чином, для побудови проблемного поля, ми маємо чітко визначитись, що суб'єктність як потенційна властивість штучного інтелекту може розглядатися лише за умови припущення фундаментальної можливості створення штучної (синтетичної) свідомості, якій, у свою чергу, буде притаманна самосвідомість. Слід зауважити, що дана проблема є фундаментальною і не має безпосередньо залежати від рівня технологій сьогодення у даній сфері. Як слушно зазначає О. Швирков «вирішення проблеми штучного інтелекту повинне розглядатися як істотно орієнтоване на невизначене майбутнє, так само, як вирішення фундаментальних проблем нашого світогляду» [19, с. 7].

Проте вирішення цього фундаментального питання необхідно стикається із плюралізмом підходів до визначення свідомості та самосвідомості. Традиційно свідомість прийнято визначати як: «1) особливий спосіб діяльності психіки, пов'язаний з переробкою та інтерпретацією інформації, що надходить у мозок від зовнішнього світу та від організму; 2) відображення дійсності суб'єктом у формах, пов'язаних (прямо чи опосередковано) з практичною діяльністю, суб'єктивне переживання зовнішнього світу та власного життя; протиставлене несвідомому [14, с. 201]. У свою чергу, під «самосвідомістю» традиційно розуміється: 1) усвідомлення людиною себе як особистості й суб'єкта життєтворчості; власних фізичних, розумових, духовних якостей, потреб, мотивів, цілей, емоцій, дій, світоглядних цінностей; як носія соціальних статусів і ролей; ставлення до інших людей та світу в цілому; 2) інтенція свідомості на саму себе, а не на зовнішній світ [14, с. 199].

При цьому інтелект визначається як: 1) у широкому сенсі – розум, здатність до мислення; 2) у гносеології – здатність до абстрактного пізнання, що включає функції

порівняння, узагальнення, утворення понять, суджень, умовиводів, аналізування тощо [2, с. 97]. У даному випадку для нас представлятиме цінність саме гносеологічне розуміння, оскільки у широкому сенсі інтелект ототожнюється із свідомістю взагалі. У свою чергу, гносеологічне розуміння інтелекту як здатності до пізнання із необхідністю обумовлює існування суб'єкта пізнання. Тобто суб'єкт стає «активним компонентом пізнавального відношення, протилежним до пізнавальної діяльності – об'єкта» [15, с. 613].

Деякі українські вчені звертають увагу на те, що у розмінні суб'єктності також використовується такий термін як агентність (у більш вузькому розумінні) [11].

Щодо цього, на наш погляд, методологічно вірним буде розглядати питання суб'єктності штучного інтелекту двох вимірах: фундаментально-філософському та агентно-соціальному. Перший вимір стосуватиметься безпосередньої можливості наявності у штучного інтелекту саморефлексії, здатності до самонавчання та створення нового знання. У свою чергу, другий вимір буде пов'язаний із соціальною роллю штучного інтелекту як свого роду «актора» у розумінні Б. Латура, тобто його самостійному впливі на людей і суспільство у цілому, незалежно від наявності самосвідомості та епістемічної суб'єктності.

Отже, розглянемо перший вимір. У цьому контексті вважаємо слушним звернутися до відомого філософа-футуролога Ю. Харарі, який сформулював три можливих варіанти: 1. Свідомість якось пов'язана з органічною біохімією у такий спосіб, який ніколи не дасть змоги створити її в неорганічних системах. 2. Свідомість не пов'язана з органічною біохімією, однак пов'язана з інтелектом у такий спосіб, що комп'ютери можуть розвинути свідомість, і комп'ютери муситимуть розвинути в собі свідомість, якщо хочуть перейти певний поріг інтелекту. 3. Не існує істотного зв'язку між свідомістю й органічною біохімією чи високим інтелектом. Тож комп'ютери можуть розвинути свідомість, однак необов'язково. Вони можуть стати суперінтелектуальними, водночас маючи нульову свідомість [17, с. 101–102].

Фактично, лише другий варіант містить у собі потенційне поле для подальших наукових досліджень, пов'язаних із суб'єктністю штучного інтелекту як носія свідомості і суб'єкта пізнавального процесу, оскільки перший констатує неможливість неорганічної свідомості, а третій варіант, вочевидь, містить у собі протиріччя (інтелект виступає функцією свідомості, отже, некоректно говорити про можливість створення надінтелекту, якщо створення надсвідомості принципово неможливо). При цьому найважливішим елементом другого варіанту виступає те, що розвинення у себе свідомості є необхідним для формування інтелекту. У свою чергу, як ми вже визначили, інтелект як функція свідомості неможливий без наявності усвідомлюючого себе суб'єкта пізнання. Отже, питання вирішення проблеми можливості створення пізнаючого штучного інтелекту невід'ємно пов'язане із більш фундаментальним питанням – онтологічної допустимості існування синтетичної свідомості. У контексті даної проблематики О. Комар слушно зауважує, що питання про створення розуму, повністю тотожного людському, вже є неправильним, оскільки однозначно вирішується негативно. Натомість цілком доречним виступає питання про можливість створення розумної альтернативи людській свідомості, оскільки «проект штучного інтелекту, спрямований на імітацію першочергово мислення, увиразнює відмінності між людською свідомістю як mind (з включенням несвідомого), а не лише consciousness» [9, с. 60].

Також, як слушно зазначає О.В. Стовець, «не маючи абсолютно достовірного та повного розуміння, чим є людська свідомість, як вона функціонує та яким чином виникла, сумнівно, що ми найближчим часом будемо здатні повністю замінити свідомість людини синтетичним розумом нейромереж (навіть якщо в цьому і був би якийсь сенс)» [13, с. 81].

Щодо цього слід зазначити, що питання сутності свідомості та наявності у ній центру самосприйняття завжди було одним з центральних у філософських роздумах, починаючи з найдавніших часів, і залишається остаточно невирішеним. Так, Платон у своїх діалогах *Тімей*, *Федон*, *Федр* послідовно говорить про існування «одвічної душі», яка сама є причиною свого існування, при цьому ум виступає кормчим душі, а сама душа має триєдину структуру, причому розум відокремлюється від вольової та чуттєвої складової.

У свою чергу, Арістотель відходить від платонівського віднесення душі до ідеального світу, зауважуючи, що душа є властивістю природного тіла (ентелехією). Оскільки душа є сутність буття і форма тіла, так само як, якщо б око було живою істотою, то його душею був би зір. У «*Нікомаховій етиці*» філософ послідовно розвиває ідею того, що індивідуальне «Я» формується крізь діяльність (практику і моральність), а індивідуальні властивості людської особистості формуються через практику і звичку і не є даними від природи [1]. Таким чином, «Я» у Арістотеля – це не вроджена сутність, а результат життєвої практики.

Новий час охарактеризувався формуванням принципово нової парадигми свідомості, ознаменованої, передусім, дослідженнями Р. Декарта. У своїй праці «*Медитації про першу філософію*» він стверджує, що основною ознакою «Я» виступає здатність до мислительної діяльності (*res cogitans*) [6]. Саме Декартом було сформульовано основну тезу дуалізму у розумінні свідомості: існують дві незалежні субстанції – мислення і протяжність. Саме з такого дуалізму і випливає у подальшому проблема можливості машини отримати «справжню» свідомість. Також у своїй праці «*Міркування про метод*» Декарт першим ставить важливе питання, яке потім відлунням відізветься у філософських спорах сучасних дослідників, а саме – критерії вирізнення машини від людини (до машин Декарт відносить і тварин). Філософ формулює наступні критерії такого розмежування: наявність мови, оскільки тварини і машини можуть робити різні звуки і навіть вимовляти слова, але не можуть робити це свідомо; обмеженість у функціях, оскільки машини за певних умов можуть здійснювати щось краще за людини, але їх поведінка є детермінованою. У свою чергу, людина характеризується наявністю універсальності, оскільки, на відміну від машини, «розум є знаряддям універсальним, яке може слугувати за всіляких обставин» [7, с. 74–75].

Іншу традицію у розумінні свідомості започаткував Д. Г'юм, як передтеча фізикалістів. У своїй праці «*Трактат про людську природу*» ним було сформульовано ідею, що «Я» являє собою лише «пучок перцепцій» і не можна говорити про наявність сталого носія свідомості. На думку філософа, «тотожність, яку ми приписуємо свідомості, є фіктивною і належить до того ж виду, що й та, яку ми приписуємо рослинам і тваринам. Тому вона не може бути іншою за походженням, а повинна походити від схожих дій уяви стосовно схожих об'єктів» [5, с. 232]. Тобто «істинна ідея людського розуму полягає в тому, що його слід розглядати як систему різних сприйнять або різних існувань, які пов'язані відношенням причини і наслідку, і які взаємно створюють, знищують, впливають і змінюють одне одного». [5, с. 234]. Певною мірою людська свідомість постає перед нами не як цілісна система, а як «театр», на сцену якого одне за одним виходять різні сприйняття. [5, с. 228]. На протигагу такому підходу І. Кант у своїй праці «*Критика чистого розуму*» вводить концепцію «трансцендентальної єдності аперцепції». Тобто «Я» не дається нам як річ, але виступає центром сприйняття і поєднання отриманого досвіду. Для того, щоб ми могли сприймати та розуміти світ, наші відчуття підпорядковуються двом основним умовам: «По-перше, все, що ми спостерігаємо, організоване в просторі та часі. По-друге, усі наші уявлення мають бути об'єднані за допомогою "Я мислю". Отже, саме простір, час і єдність свідомості є необхідними умовами для нашого досвіду [8, с. 107].

На відміну від І. Канта, Й. Фіхте надавав «Я» статусу окремого творчого суб'єкта, яким сам себе конститує. При цьому навіть усе існуюче є продуктом діяльності «Я». «Отже, поняття мислення, що повертається до самого себе, та поняття Я взаємно вичерпують одне одного. Я покладає самого себе, а покладання самого себе становить Я. Лише те, що постає завдяки поверненню твого мислення до самого себе, є Я» [16, с. 98–99]. Також Фіхте наголошував на наявності нерозривного зв'язку між свідомістю та усвідомленням: «Ти усвідомлюєш себе як свідомця (*des Bewußten*) лише тому, що ти усвідомлюєш себе як усвідомлювача (*Bewußtseienden*). Але тоді усвідомлювач знову стає свідомцем і ти наново маєш усвідомлювати усвідомлювача цього свідомця, і так нескінченно. Таким побитом ти можеш побачити, як ти досягаєш першої свідомості» [16, с. 101].

Г. Гегель намагався вирішити цю проблему на принципово інших засадах, пропонуючи діалектичний підхід. На думку філософа, «Я» і об'єкт взаємозалежні. «Я» за Гегелем є центром усвідомлення змін які відбуваються, але є суперечливим, оскільки постійно рухається між тим, що мало місце «зараз» та тим, що минуло, намагаючись осмислити ці зміни. «Я», фактично, створює реальність і само набуває процесуальності, виступаючи і суб'єктом і частиною діалектичного руху свідомості [3, с. 96–97].

З іншого боку до проблеми суб'єктності підійшли представники феноменології, зокрема її засновник – Е. Гусерль, шукаючи сутність самоусвідомлення у самоусвідомлення феноменального досвіду. Тобто суб'єкт віднаходить себе як когось, хто сприймає, уявляє, думає, відчуває, жадає тощо; і при цьому актуально пов'язаний із дійсністю, що його оточує [4, с. 63].

Друга половина двадцятого століття ознаменувала собою «когнітивний поворот» в у філософії свідомості, що був, передусім, спрямований на подолання розуміння свідомості як метафізичної субстанції і переведення фокусу досліджень на пошук чітких наукових орієнтирів. Так Г. Райл у роботі «Поняття свідомості» аргументує помилковість картезіанського концепту «примари у машині», стверджуючи, що самосприйняття формується не внаслідок наявності безпосереднього доступу суб'єкта до ментальних станів, а у результаті діяльності [26]. У свою чергу, Т. Нагель у своїй праці «Що означає бути кажаном» наголосив, що не можна не брати до уваги того, що індивідуальний досвід завжди є суб'єктивним. Отже, відтворення суб'єктивного квалітативного досвіду на підставі спостереження є неможливим [25].

Дж. Сьорл у статті «Розуми, мізки та програми» описав відомий ментальний експеримент «китайської кімнати» з якого випливає, що алгоритмічна обробка інформації немає нічого спільного із розумінням [27]. Оже і машини можуть отримувати правильні результати, але цей процес жодним чином не свідчить про наявність самосвідомості. Ф. Джексон у своїй статті «Епіфеноменальні кваліа» описує відомий експеримент «Кімната Мері» і формулює висновок про те, що знання про феномен не може замінювати саме феноменальне сприйняття [22].

Д. Деннет послідовно просуває фізикалістський підхід, який полягає у тому, що свідомість – це лише сукупність когнітивних обчислювальних процесів, які не пов'язані з необхідністю наявності центрального суб'єкта, а «кваліа» є лише ілюзією, яка вводить нас в оману на шляху знання про свідомість та закономірності її функціонування. Також філософ послідовно стверджує, що не існує єдиного «Я», натомість всі види сприйняття — а також усі форми мислення чи розумової діяльності — здійснюються в мозку через паралельні, багатоканальні процеси інтерпретації та опрацювання сенсорних даних [21, с. 111].

У свою чергу, Д. Чалмерс у книзі «Свідомий розум» стверджує, що самосвідомість, тобто «Я» не може бути редуковане до фізичних процесів, оскільки саме суб'єктивне переживання виходить за межі можливостей функціонального пояснення. Для цього він пропонує концепт «феноменального зомбі», який фізично функціонально є ідентичним людині,

але не має суб'єктивного досвіду [20, с. 95] Тобто свідомість є чимось більшим ніж просто фізичні процеси. У даному описі неважко побачити аналогію зі «штучним інтелектом» як синтетично створеним аналогом людини. Проте цей концепт був аргументовано спростованим А. Синицею у його статті "Філософські зомбі: логіко-когнітивна інтерпретація", де він зазначив, що Чалмерс припустився логічної помилки, оскільки приймаючи умову, що структура мозку «зомбі» тотожна мозку людини, ми опиняємося у ситуації одночасного визнання і заперечення існування свідомості у самої людини [12, с.62–63].

Вагомий інтерес представляє робота Т. Метцингера «Тунель Его» у якій він стверджує, що сутність свідомості полягає, передусім, у створенні моделі світу для суб'єкта. На думку дослідника, «для того, щоб світ міг з'явитися для машини, їй мають бути властиві дві особливості: перша – це організація внутрішнього потоку інформації таким чином, щоб він створював психологічний ефект переживання «зараз», тобто виділяти окремі події з фізичного світу і уявляти їх як одночасні. Друга властивість – це забезпечення того, щоб ці внутрішні структури сприймалися не як внутрішньо сконструйовані, а «прозорі» тобто такі, що дозволяють синтетичній системі сприймати себе, причому такою, що знаходиться у середині світу» [24, с. 192].

Таким чином, можна побачити, що проблема відтворення «Я» як центру свідомості у штучному інтелекті не просто не вирішена у сучасних дослідженнях, ба більше – лише поверхово окреслена, без чіткого проблемного поля. Для нас поки що залишається невідомою сама природа суб'єктивного переживання та автономного самовизначення. Безумовно, достатньо переконливими вбачаються тези фізикалістів про те, що наше «Я» є не сталим уявним метафізичним центром, а лише одним з багатьох процесів, що відбуваються у свідомості, але це жодним чином не наближує нас до розуміння самої можливостей відтворення таких процесів у системах штучного інтелекту. Тобто ми лише здатні створити алгоритм, який буде імітувати процеси, що відбуваються у свідомості. Найдосконаліші нейромережі, що позиціонуються як такі, що набули здатність до самонавчання, не мають нічого спільного із свідомим отриманням знання людиною та його розумінням, не виступають суб'єктом пізнавального процесу.

Певною мірою спробою подолання такої фундаментальної прогалини виступає теорія інтегрованої інформації Дж. Тононі, згідно якої свідомий досвід має унікальну, інтегровану та визначену структуру, що не зводиться до окремих частин, а його кількість вимірюється показником Φ (Фі), який відображає рівень інтегрованої інформації [28, с. 452], тобто система, що має відповідні характеристики рівня інтеграції та складності зможе безпосередньо переживати своє існування. При цьому вчений зауважує, що певні мережеподібні системи, схожі за своєю топологією на задню кору головного мозку, можуть мати високий рівень свідомості без проявів інтелектуальної поведінки, а цифрові комп'ютери, що виконують складні програми на основі архітектури фон Неймана, не будуть мати свідомості [28, с. 460].

Висновки. Питання суб'єктності штучного інтелекта безпосередньо пов'язане з фундаментальною проблемою існування «Я» як властивості свідомості, що визначає виокремлення суб'єкта зі світу. Це «Я» вочевидь обумовлене наявністю суб'єктивного квалітативного досвіду, можливість відтворення якого залишається одним із самих складних завдань як для філософії так і для нейронауки. Проте сучасні дослідження дозволили здійснити значний крок уперед у розумінні «Я» не як абстрактного трансцендентального центру, а як процесу, що має нейрофізіологічну природу. Це відкриває нові перспективи для подолання прогалини між фізичними процесами та виникненням самосвідомості. У контексті можливості виникнення суб'єктного штучного інтелекту ці дослідження мають принципове значення. Якщо «Я» і самосвідомість дійсно є результатом складних нейрофізіологічних процесів, то теоретично

можливе відтворення аналогічних процесів у штучних системах. Доки ці питання залишаються без відповіді, створення штучного інтелекту, що має суб'єктність, виступає скоріше предметом філософських дискусій ніж практичним завданням. Проте, прогрес нейронауки та когнітивних досліджень надають можливість розширити наше сприйняття і створюють основу для фундаментальних філософських наукових розвідок.

Список використаної літератури

1. Арістотель. Нікомахова етика. Київ: Аквілон-Плюс, 2002. 480 с.
2. Велика українська енциклопедія. Тематичний реєстр гасел з напрямку «Філософські науки» (філософія, логіка, етика, естетика) / Укладачі: Арістова А. В., Шліхта І. В.; за заг. ред. д. і. н., проф. Киридон А. М. Київ: Державна наукова установа «Енциклопедичне видавництво», 2019. 256 с.
3. Гегель І. В. Ф. Феноменологія духу. Київ: Вид-во Соломії Павличко «Основи», 2004. 548 с.
4. Гусерль Е. Ідеї чистої феноменології і феноменологічної філософії. Книга перша. Загальний вступ до чистої феноменології. Харків: Фоліо, 2020. 348 с.
5. Г'юм Д. Трактат про природу: Спроба запровадження експериментального методу міркувань про об'єкти моралі. Київ: Видавничий дім «Всесвіт», 552 с.
6. Декарт Р. Медитації про першу філософію. Метафізичні медитації. *Sententiae*. 2009. № 1. С. 158–229.
7. Декарт Рене Міркування про метод, щоб правильно спрямовувати свій розум і відшукувати істину в науках. Київ: ТанDEM, 2001. 104 с.
8. Кант І. Критика чистого розуму. Київ: Юніверс, 2000. 504 с.
9. Комар О. Тесеєві мандри свідомості. *Філософська думка*. 2016. № 2. С. 57–70.
10. Платон Діалоги. Київ: Основи, 1999. 395 с.
11. Смільсон М. Л., Мещеряков Д. С. Суб'єкт і суб'єктність у зарубіжній психології: нюанси термінології та перекладу. *Технології розвитку інтелекту*. 2021. № 1 (29). Vol 5. URL: https://psytir.org.ua/index.php/technology_intellect_develop/article/view/552/137 (дата звернення: 23.02.2025).
12. Синиця А. С. Філософські зомбі: логіко-когнітивна інтерпретація. *The Journal of VN Karazin Kharkiv National University. Series «Theory of Culture and Philosophy of Science»*. 2013. № 2(1029). С. 59–69.
13. Стовпець О. В. Проблема людської свідомості в «епоху нейромереж». Концепти соціокультурної трансформації сучасного суспільства. *Матеріали V Міжнародної конференції здобувачів вищої освіти і молодих учених* (Одеса, 25–26 травня 2023 р.). Південноукр. нац. пед. ун-т ім. К. Д. Ушинського, кафедра філософії, соціології та менеджменту соціокультурної діяльності; [О. М. Єременко, М. В. Пальчинська (рец.)]. Одеса: ПНПУ імені К. Д. Ушинського. С. 77–81.
14. Філософія: словник термінів та персоналій / В. С. Бліхар, М. А. Козловець, Л. В. Горохова, В. В. Федоренко, В. О. Федоренко. Київ: КВІЦ, 2020. 274 с.
15. Філософський енциклопедичний словник / НАН України, Ін-т філософії імені Г. С. Сковороди; [редкол.: В. І. Шинкарук (голова) та ін.]. Київ: Абрис, 2002. 742 с.
16. Фіхте Й. Г. Спроба нового викладу вчення про науку (1797) / переклад Іващенко І. *Sententia*. 2013. № 2 (XXIX). С. 97–107.
17. Харарі Ю. Двадцять один урок для двадцять першого століття. Київ: Book Chef, 2022. 416 с.
18. Шевченко А. Штучний інтелект: український вимір. *Світгляд*. 2022. № 3 (101). С. 8–9.
19. Швирков О. І. Проблема штучного інтелекту і людиновимірність штучних інтелектуальних систем: автореф. дис. ... канд. філос. наук: 09.00.09. Київ, Київський національний ун-т ім. Тараса Шевченка. 2006. 14 с.

20. Chalmers D. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996. 414 p.
21. Dennett D. *Consciousness Explained*. New York: Back Bay Books, 1991. 511 p.
22. Jackson F. Epiphenomenal qualia. *Consciousness and Emotion in Cognitive Science*. Routledge. 1998. Vol. 3. P. 197–206.
23. McCarthy J. What is Artificial Intelligence? URL: <https://www-formal.stanford.edu/jmc/whatisai.pdf> (дата звернення: 23.02.2025).
24. Metzinger T. *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. New York: Basic Books, 276 p.
25. Nagel T. What is it like to be a bat? *The Language and Thought Series*. Harvard University Press, 1980. P. 159–168.
26. Ryle G. *The Concept of Mind*. London and New York: Routledge, 2009. 314 p.
27. Searle J. R. Minds, Brains, and Programs. *Behavioral and Brain Sciences*. 1980. Vol. 3. № 3. P. 417–424.
28. Tononi G. et al. Integrated Information Theory: From Consciousness to Its Physical Substrate. *Nature Reviews Neuroscience*. 2016. Vol. 17. № 7. P. 450–461.

THE PROBLEM OF SUBJECTIVITY OF ARTIFICIAL INTELLIGENCE: A PHILOSOPHICAL ANALYSIS

Oleksii Kreze

*State University of Trade and Economics,
Department of Philosophy, Sociology and Political Science,
Kyoto St., 19, 02156, Kyiv, Ukraine*

The article attempts to provide a philosophical analysis of the fundamental problem of the possibility of the artificial intelligence subjectivity. At present, there is a certain terminological inconsistency in modern philosophical discourse, as artificial intelligence is often equated with artificial consciousness. However, intelligence is the only one of the functions of consciousness. At the same time, the issue of subjectivity is directly related to such an aspect of consciousness as "self-awareness," which implies the ability to perceive oneself as an autonomous subject, capable of having qualitative experiences, distinguishing oneself from the world, gaining knowledge, and creating new knowledge.

Despite significant technological advancements, even the most advanced artificial intelligence models remain algorithmically limited. They do not have access to qualitative experiences, cannot be aware of their internal states, and do not possess individual experience. Therefore, the problem of reproducing self-awareness is not so much a technical issue as a fundamental philosophical one, as it is connected to the very nature of subjectivity and the necessity of direct phenomenal access to reality for its emergence.

In this regard, several approaches to solving this problem can be distinguished. Functionalists adhere to the position that consciousness is causally determined by physical processes occurring in complex systems. Their opponents, who follow the concepts of idealism or dualism, argue that consciousness is an expression of the existence of a separate immaterial substance.

Despite profound differences between these approaches, it can be stated that most modern philosophers tend to consider self-awareness and its core—the "Self"—not as an abstract metaphysical entity but as a complex dynamic process with psychophysiological conditioning (although the question of neurobiological reduction remains open). Thus, the study of artificial intelligence subjectivity inevitably brings us closer to a deeper understanding of the human "Self" and highlights the necessity of continuing fundamental philosophical research in close connection with the latest scientific researches.

Key words: subjectivity, artificial intelligence, consciousness, self-consciousness, qualia, functionalism, reductionism,