TECHNOLOGIES OF INTELLECT DEVELOPMENT Vol. 9 No. 1(37) (2025)
ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

УДК 004.8-159.9-12

КРИТИЧНИЙ РОЗГЛЯД ПОНЯТТЯ «СВІДОМІСТЬ» У СУЧАСНІЙ МЕТОДОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ

Шевцов Андрій

Державна наукова установа «Центр інноваційних технологій охорони здоров'я» Державного управління справами, Головний науковий співробітник, Київ, Україна

dr_shevtsov@ukr.net

Вступ

У новітні часи широкомасштабного застосування систем штучного інтелекту (СШІ) у різних галузях науки, освіти, медицини, бізнесу, промисловості тощо вельми важливого значення набувають футурологічні питання щодо майбутнього розвитку «розумних машин», набуття ними особистісного статусу та свідомості. Тут дискутують між собою наукові школи позитивного погляду на появу штучного суб'єкта цивілізації як помічника людини та наукові групи алармістського штибу з позицією катастрофічної загрози людству після здобуття «штучним інтелектом» (ШІ) суб'єктності.

Тож критично важливим для подібної прогностичної діяльності є дослідження епістемологічних і теоретико-методологічних питань ШІ як наукової сфери в її прикладному значенні.

Метою статті є критичний розгляд сучасних підходів до методологічних психолого-філософських питань у сфері штучного інтелекту ШІ як сфери науки й технологій та перспектив відповідних досліджень.

Методи дослідження.

В статті застосовані теоретико-прогностичні, аналітико-синтетичні та бібліографічні методи для здійснення порівняльного аналізу сучасних психолого-філософських підходів та понятійного апарату сфери ШІ як науки та прогнозування перспектив відповідних досліджень.

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

Результати досліджень та дискусія.

На особливому місці методології науки про СШІ знаходяться проблеми понятійно-термінологічного поля. Йдеться перш за все про розуміння та застосування фахівцями цієї сфери у своїй аналітико-синтетичній професійній діяльності таких понять як «штучний інтелект», «штучна свідомість», «штучна особистість», «машинне мислення», «розуміння», «суб'єктність», «свобода волі» (у контексті кіберпсихології), а також конструкти із сфери математичного моделювання природного інтелекту й свідомості та інші поняття, які в цій ділянці науки не є однозначно розтлумаченими та усталеними, проте активно дискутуються як практиками, так і теоретиками у сфері розробки СШІ.

Примітно, що ці питання також дуже турбують як пересічних так професіональних користувачів СШІ, які застосовують їх як віртуальних асистентів у своїх сферах діяльності та іноді сприймають сучасні СШІ як такі, що мають ознаки «особистості». При цьому вони вірогідніше всього мають базову освіту не в галузі ІТ та психології, а відповідно до сфери своєї основної професійної діяльності. Отже вони потребують певних роз'яснень та специфічних знань з теоретико-методологічних питань побудови та функціонування СШІ.

У той же час подібні методологічні питання щодо визначення традиційних понять психології (свідомість, особистість, суб'єктність, інтелект, мотивація тощо) насправді фундаментально дискутуються з тих часів, коли психологія як наука була часткою філософського знання. І хоча з моменту відокремлення психології як експериментальної науки від філософії зазначені питання перейшли в більш практичне поле, тим не менш їх сучасний онтологічний аспект знов повертає нас у сферу не тільки теоретичної психології, але й філософії.

Розглянемо декілька методологічних питань, що активно дискутуються у зв'язку зі створенням та функціонуванням СШІ в нашій цивілізації. Перш за все триває жвава дискусія, коли йдеться про футуристичний прогноз розвитку у майбутньому так званого «сильного» (універсального) ШІ такого рівня, що це передбачає набуття СШІ здатності мислити і усвідомлювати себе як окрему

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

особистість (зокрема, розуміти та усвідомлювати власні думки, «внутрішній світ» тощо). Тоді вірогідно, що «розумовий процес» машини буде подібний до людського і навіть його набагато перевищувати за потужністю, враховуючі майже необмежені ресурси пам’яті та швидкість роботи з інформацією. Це дає підстави для певних алармістських позицій у суспільстві, що може суттєво уповільнити подальший прогрес у сфері удосконалення ШІ, аж до призупинення або навіть заборони «сильного ШІ» на законодавчому рівні.

Примітно, що «слабкого» (так званого, прикладного або вузького) ШІ ми не «боїмося», адже він як звичайна технічна утиліта призначений для вирішення певної конкретної інтелектуальної задачі або їх невеликої множини (наприклад, системи для гри в шахи, керування транспортом, розпізнавання образів, перекладу, фінансової аналітики тощо). Для таких СШІ не передбачаються наявність у комп'ютера справжньої свідомості. Власно з такими алгоритмами наразі ми на практиці і маємо справу, тобто з наявністю тільки «слабкої» форми ШІ.

Звертаючись до праці Костюка Г.С. «Проблема особистості у філософському та психологічному аспектах» читаємо таке: «Індивід є особистістю, оскільки він усвідомлює навколишнє буття і себе самого, свої відносини до нього, свої функції та обов'язки, тобто оскільки йому притаманне свідомість та самосвідомість» [2, С.81]. Отже, вочевидь без свідомості не має особистості!

Насправді питання про наявність свідомості у мислячого суб’єкта є базовим, онтологічним, а тому в значній мірі філософським. Тому й зрозуміло, що в епоху жорсткого панування в СРСР діалектичного матеріалізму психологи більше приділяли уваги питанням особистості, ніж свідомості. Адже стояло політичне питання виховання особистості нової людини «гомосоветікус», а поняття свідомості людини в основному зводилися до соціалістичної, комуністичної, правової тощо самосвідомості. У той же час в закордонній психології та філософії розробки методології свідомого та несвідомого не

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

припинялися, що позитивно вплинуло на фундаментальні дослідження у сфері епістемологічних основ «штучного інтелекту».

До цих пір наявні СШІ побудовані за концепцією математика Алана Тьюрінга, який запропонував у 1936 році певний математичний об'єкт для формального уточнення інтуїтивного поняття алгоритму, яке згодом назвали «машина Тьюринга».

Машина Тьюринга є абстрактною моделлю обчислень, яка працює за чітко визначеними правилами (алгоритмами). У класичному розумінні вона просто оперує символами відповідно до певної програми, що не передбачає необхідність мати внутрішній досвід (qualia) або суб'єктивного усвідомлення своєї діяльності.

Лауреат Нобелівської премії Роджер Пенроуз у своїх всесвітньо відомих роботах кінця минулого століття до проблеми наявності у СШІ свідомості застосував теорему Геделя про неповноту, яка наголошує на принципових обмеженнях формальної арифметики і демонструє, що існують математичні істини, які неможливо довести жодним набором формальних правил [7].

Свідомість принципово не є обчислювальною, тож він вважає (що співпадає і з нашою думкою), що СШІ, яка є машиної Тьюринга, ніколи не може набути свідомість і дістати справжнього інтелекту, подібного до природного. Пенроуз пише про «фізику» свідомості, яка є незчисленною та існує в реальності, подібної до квантової реальності. На відміну від класичної реальності, яку ми можемо безпосередньо сприймати та з якою взаємодіяти. Отже, це означає, що свідомість перевершує «обчислення», оскільки вона передбачає розуміння причин, що лежать в основі формальних правил, а не просто їх дотримання.

Справедливості раді треба також зауважити, що ідею про те, що свідомість не можна звести до алгоритмів, вперше висунув філософ Джон Лукас з Оксфордського університету, який також ґрунтуючись на теоремі Геделя про неповноту стверджував, що машина не може бути повною та адекватною моделлю людського розуму [6]. Отже і Лукас, і Пенроуз з точки зору математичної логіки спростовували міф про те, що СШІ дістане свідомість.

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

Важливо тут також згадати резонансні дослідження філософа нового часу Девіда Чалмерза, який також вважає, що свідомість насправді є чимось більшим, ніж просто обчислювальний процес [4]. Його праці мають дійсно революційне значення, адже на початку XXI століття, важко знайти фундаментальну публікацію про свідомість, в якій не згадувалося б доробок цього автора

Фактично його теорія свідомості знаходиться в контраверсійній парадигми філософського дуалізму (Теза, згідно з якою Всесвіт складається як з матеріальних субстанцій, так і ментальних). Проте сам Чалмерз називає свій підхід «натуралістичним дуалізмом», адже він виступає проти спрощеного погляду на свідомість з точки зору фізикалістського редукціонізму, який ґрунтується на переконанні, що закони спостережуваного світу поширюються і на внутрішній ментальний світ спостерігача. Таким чином Чалмерз виключає редукцію усвідомленого мислення до функції мозку, при цьому погоджуючись із очевидним зв’язком мисленнєвих операції з біофізичними процесами, а також свідомості із фізичним світом. Проте філософ вважає походження свідомості від останнього експериментально не доведеним, принаймні не бачить доказів існування абсолютних механізмів породження свідомості виключно фізичним світом.

Чалмерз розрізняє дві проблеми співвідношення ментального і тілесного: «легка» і «важка» проблеми свідомості. До «легких» проблем Чалмерз відносить ті, які зводяться до функціонального пояснення мислення через дослідження організації біофізичних систем і, отже, потенційно розв'язуються за допомогою методів, використовуваних в нейробіології і когнітивній науці. Розв’язання цих проблем є суто технічною задачею. Цю частину проблеми співвідношення ментального і тілесного Чалмерз відносить до нейрофізіологічних аспектів ментального.

«Важка проблема» (чи викликають обчислення переживання?) на його думку до цих пір є таємницею для сучасної науки і містить у собі питання яким чином фізична система могла б породжувати свідомий досвід? Тобто породжувати кваліа. Як сказано, він стверджує, що жоден формалізм поки не дав

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

відповіді на останнє питання. Хоча для розв'язання «легких проблем» можна промоделювати відповідні інтелектуальні функції формальними операціями. Але безпосередньо для свідомості, суб'єктивного досвіду та переживання функціонально-редукційна деконструкція неможлива!

Вочевидь такий підхід за наявною технологією алгоритмізації несумісний із прогнозами набуття свідомості (в людському її розумінні?!) машиною.

У той же час, припустимо, що ми навчили СШІ як істоту, яка поводиться подібно до звичайної людини. Проте в ній будуть відсутні свідомий досвід (кваліа) або властивості чуттєвого досвіду. Тобто вона ефектно проходить тест Тьюринга і повноцінно симулює природній інтелект. Чи буде вона функціонувати без «духовної» складової? Чи зможемо ми назвати її штучною особистістю? Чи ця механічна істота залишиться в реальності «філософським зомбі» (Philosophical zombie) або «біхевіоральним зомбі» (Behavioral zombie), що поведінкою не відрізняється від звичайної людини і все ж не має ніякого свідомого досвіду, а значить – свідомості. Чалмерз стверджує, що оскільки існування «зомбі» можливо, то поняття кваліа і здатність усвідомлювати відчуття досі не отримали повного пояснення з точки зору властивостей фізичного світу.

Можна поставити питання також і іншим чином. Якщо ми дамо якимось способом машині чуттєвий досвід чи з'явитися у неї кваліа, внутрішній досвід, що і буде складати свідомість?

Така технологія може бути виправдана, адже Дональд Девід Гоффман (американський когнітивний психолог, професор Каліфорнійського університету) у своїх книгах писав як згортається і трансформується зовнішня інформація для нашої свідомості, зокрема в книзі "Як відчуття брешуть нам" [5]. Він вважає, що фактично наш мозок транслює нашій свідомості не реальний обсяг інформації про зовнішній світ, а перетворену інформацію таким чином, щоб ми могли з нею оперативно й коректно працювати заради еволюції.

Тобто у роботах Гофмана розгорнута популярна у когнітивістиці та нейропсихології гіпотеза про те, що мозок «годує» нашу свідомість викривленою інформацією про реальність з метою адаптації та оптимізації нашої діяльності

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

заради нашого виживання (Цей підхід отримав назву «усвідомлений реалізм» – Conscious Realism). Тобто наш мозок для розвитку властивості людини ефективного прийняття рішень використовує адаптовану (згорнуту) інформацію, отриману через органи відчуття, і трансформує її під задачі біологічної еволюції людини. Адже, якщо б мозок давав би свідомості (а непевно і підсвідомості) повну і детальну інформацію про реальність – людина не змогла б оперативно приймати рішення щодо своєї діяльності з необхідною для виживання швидкістю та ефективністю.

При цьому Гоффман йде далі і зазначає, що загальноприйнята думка, за якою активність мозку викликає свідомий досвід, до цих пір нерозв'язна, з точки зору доведених наукових аргументів. Тому він досить контраверсійно пропонує вирішити нерозв’язану проблему свідомості через перегортання піраміди реальність-відчуття-мозок-свідомість, прийнявши зворотну гіпотезу, за якою свідомість викликає активність мозку і, по суті, «створює» всі об'єкти та властивості фізичного світу в інтрапсихічній парадигмі.

Отже згідно з Гоффманом, еволюція не вимагає точного відображення реальності; вона потребує лише адаптивного, корисного для виживання сприйняття. Реальність, яку ми бачимо, є зручною «іконкою» (як на робочому столі комп'ютера), яка приховує реальну, більш складну сутність.

Тож яка різниця в ситуаціях, якщо ми так само можемо годувати машину «жуйкою» із спеціально оброблено інформації або вона б отримала свій власний чуттєвий досвід (кваліа)? Отже у неї могла б з'явитися свідомість?

В праці «Щодо психології розуміння» Костюк Г.С. пов’язує свідомість з розумінням усього того, на що вона та пізнавальні процеси спрямовані: різних явищ природи, суспільного життя та внутрішнього світу самої людини тощо [1].

У дослідженнях Костюка Г.С. розуміння постає як структурований процес, який можна описати такими дескрипторами (у реконструкції Рибалки В.В. [3]): а) потреби та мотиви розуміння; б) ознайомлення з фактами, відображення, усвідомлення об'єктивного змісту, складних зв'язків в об'єктах розуміння; в) цілеспрямованість, тобто спеціальні питання, цілі, завдання розуміння;

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

г) пошук засобів розуміння, продуктивна, результативна сторона цього процесу; д) емоційний аспект процесу розуміння.

Всі ці аспекти можна проаналізувати у площині «діяльності» СШІ, зокрема й ті, що можна ефективно запрограмувати, наприклад: цілеспрямованість, завдання, цілі. Проте академік Костюк Г.С. в декількох своїх працях наголошує, що руховою силою процесів пізнання є мотивація. Напрошується питання – де знайти в програмному коді машини мотивацію? Хіба можна назвати справжньою мотивацією алгоритмізовану потребу виконати завдання, що поставила СШІ людина? Проблеми також виникають під час алгоритмізації емоційного аспекту процесу розуміння.

Щодо можливості «розуміння» машиною свого функціонування та буття у цілому, то відповіддю певним наведеним вище аспектам є уявний експеримент під назвою «китайська кімната», який запропонував американський філософ Джон Сьорл [8]. Він використовується в літературі як аргумент, згідно з яким навіть складна формальна система, що маніпулює символами, не може володіти розумінням або свідомістю. Уявімо в кімнаті людину, яка не знає китайської мови, але отримує китайські символи та видає відповідь за певними правилами (алгоритмом), не розуміючи значення жодного із ієрогліфів. У підсумку експерименту ми можемо зробити помилковий висновок, що людина володіє китайською через те, що вірно реагує на наданий текст і вирішує завдання. Сьорл порівнює таку ситуацію з роботою комп’ютера, який, незважаючи на правильний результат завдань, не усвідомлює смислу та значення своїх дій, не мислить у цілому і не усвідомлює себе: машини обробляють інформацію, але не розуміють її.

Тепер звернемо увагу на думку протилежної науково-інженерної парадигм. З точки зору представників позитивного прогнозу розвиток систем штучного інтелекту (СШІ) у майбутньому може бути настілки складним, що званий «сильний» (універсальний) ШІ може стати самостійним суб’єктом діяльності і буде не просто помічником людини, а повноправним партнером у розбудові цивілізації.

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

У зв'язку з критичними поглядами на теорію Девіда Чалмерза про відсутність генетичного зв'язку системи «фізичний світ» – «природна свідомість», є певні наукові школи, що стверджують все ж таки про вірогідну можливість народження свідомості саме завдяки фізичним системам (зокрема «нейронним мережам») нашого реального світу («квантове моделювання», теорія інтегрованої інформації (IIT), емерджентні підходи тощо).

Наприклад в IIT версії 3.0, сформульованою Джуліо Тононі [9, 10] (у співавторстві з Марчелло Масіміно [11]), пропонується об'єктивний підхід до «вимірювання» свідомості. Згідно з IIT існує п'ять феноменологічних аксіом про свідомість:

1. **Існування (Existence).** Свідомість існує безсумнівно; це безпосередня реальність нашого досвіду.
2. **Структурованість (Composition).** Кожен свідомий досвід структурований, тобто складається з численних елементів або аспектів, які можуть бути розрізнені.
3. **Інформація (Information).** Свідомий досвід є специфічним, він відрізняється від інших можливих досвідів, представляючи конкретний набір елементів.
4. **Інтегрованість (Integration).** Свідомий досвід є єдиним цілим, яке не може бути розділене на незалежні частини без втрати його цілісності.
5. **Виключність (Exclusion).** Свідомий досвід є єдиним і виключає інші можливі свідомі досвіди в той самий момент;

Наявні також відповідні їм наукові постулати:

1. **Постулат існування (Existence postulate).** Свідомість виникає завдяки фізичним системам, що володіють внутрішньою причинно-наслідковою силою (Intrinsic cause-effect power).
2. **Постулат структурованості (Composition postulate).** Елементи свідомого досвіду відповідають множині елементів у фізичній системі, що мають внутрішні причинно-наслідкові відносини.

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

3. **Постулат інформації (Information postulate).** Свідомий досвід виникає у фізичних системах, які визначають специфічний причинно-наслідковий репертуар. Тобто, вони зменшують невизначеність і створюють конкретну інформаційну структуру.

4. **Постулат інтегрованості (Integration postulate).** Свідомий досвід з'являється лише у тих системах, які формують єдину, нероздільну причинно-наслідкову структуру з високим ступенем інтегрованості.

5. **Постулат виключності (Exclusion postulate).** Свідомий досвід відповідає причинно-наслідковій структурі, що має максимальний показник інтеграції, виключаючи всі інші можливі структури в цій системі.

У прикладній площині наведемо книгу професора когнітивної та обчислювальної нейронауки Аніла Сета «Бути собою: Нова теорія свідомості» [12], в якій здійснено фундаментальне міждисциплінарне дослідження природи свідомості, яке поєднує нейронауку, філософію та когнітивну психологію. Він намагається дати відповідь на питання: як фізичні процеси в мозку породжують суб'єктивний досвід, та виступає проти класичного дуалізму (розділення тіла і свідомості), шукає фізіологічне пояснення свідомого досвіду.

В зазначеній книзі та в інших працях Сет пропонує нову концепцію, яка кидає виклик традиційним уявленням про свідомість. Сет стверджує, що наше сприйняття реальності — це не пасивне відображення зовнішнього світу, а активний процес, у якому мозок постійно формує прогнози про сенсорні сигнали і коригує їх на основі отриманої інформації. Цей підхід відомий як теорія «предиктивної обробки» або «контрольованої галюцинації».

Свідомість «Я» виникає з інтеграції різноманітних внутрішніх моделей, які мозок створює для передбачення стану тіла, емоцій та соціальних взаємодій. Ці моделі формують наше відчуття себе як окремої особистості.

Отже, замість того, щоб намагатися пояснити, чому виникає свідомість (так зване «тверде» питання), Сет пропонує зосередитися на конкретних аспектах свідомого досвіду, які можна досліджувати науковими методами. Він вважає, що

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

поступове розуміння цих аспектів допоможе «розчинити» тверде питання, роблячи його менш загадковим .

Проте хоча праці Сета пропонує новаторський підхід, вони не дають остаточних відповідей на всі питання про природу свідомості. Дискусії тривають, особливо щодо того, чи може теорія предиктивної обробки повністю пояснити суб'єктивний досвід.

У футурологічній книзі «Сингулярність уже близько: Коли людина перевершить біологію» сучасний винахідник у сфері комп’ютерних технологій, фахівець із розвитку СШІ в Google Рей Курцвейл [13] прогнозує, що до середини XXI століття технології досягнуть точки, де штучний інтелект перевершить людський розум, а люди зможуть об’єднатися з машинами, досягнувши «сингулярності». Тобто в його розумінні «сингулярність» означає кардинальну зміну людського існування. За його прогнозами у 2030-ті роки наномашини вставлятимуться прямо в мозок і здійснюватимуть довільне введення і виведення сигналів з клітин мозку. Це призведе до віртуальної реальності «повного занурення», яке не буде потребувати якогось додаткового обладнання.

Хоча деякі прогнози футуролога не здійснилися, більшість із них мали успіх. Отже, Курцвейл пророкує наступ технологічної сингулярності в 2045 році. В цей час вся Земля почне перетворюватися на один гігантський комп'ютер, і поступово цей процес може поширитися на весь Всесвіт. Природа сингулярності така, що конкретніші прогнози на період після 2045 року зробити важко.

Тож за таких умов розвитку технологій можна говорити не стільки про народження свідомості та справжнього інтелекту у машини, подібного до природного, скільки появи принципово нового типу особистості в симбіотичному злитті людини та машини.

Отже, теоретичні та експериментальні дослідження у цьому питанні можуть бути принципово важливими для передбачення подальшого розвитку цивілізації.

Сучасна когнітивна наука та нейронаука перманентно продовжують активні спроби розкрити механізми виникнення свідомості у фізичному світі за

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

допомогою експериментальних методів. Інструментальні нейрофізіологічні дослідження відіграють центральну роль у цьому процесі, дозволяючи пов'язати суб'єктивний досвід із конкретними нейронними процесами. Ключові сучасні технології – електроенцефалографія (EEG), функціональна магнітно-резонансна томографія (fMRI), магнітоенцефалографія (MEG), транскраніальна магнітна стимуляція (TMS) та інші – відкрили можливість емпіричного вивчення феномену, який раніше вважався виключно філософським.

Ці дослідження свідомості значною мірою базуються на пошуках нейронних корелятів свідомості (NCC) — мінімальних нейронних механізмів, необхідних для виникнення свідомого досвіду [14].

Вище була наведена Інтегрована інформаційна теорія Джуліо Тононі (Integrated Information Theory) [15], яка стверджує, що рівень свідомості відповідає здатності системи інтегрувати інформацію, що вимірюється показником Φ (phi). За допомогою TMS-EEG показано, що під час сну чи наркозу мережа мозку демонструє низьку інтегрованість, що свідчить про зниження рівня свідомості [16].

Близько до вищенаведеної концепції IIT знаходиться теорія Глобального робочого простору (Global Workspace Theory, GWT). Згідно з GWT, свідомість виникає в результаті глобальної доступності інформації в мозку — коли інформація поширюється із сенсорних ділянок до префронтальної та тім'яної кори [17].

Емпіричні дослідження в цій площині показують, що свідоме сприйняття супроводжується активацією префронтальної кори та появою пізніх ERP-компонентів ((Event-related potentials), наприклад, хвиля P3 (positive component) [18]. Експерименти за допомогою fMRI та EEG підтверджують цю модель, демонструючи активацію відповідних областей під час свідомого сприйняття [17].

У цілому електроенцефалографія (EEG) дозволяє реєструвати електричну активність мозку з високою часовою роздільністю. Зокрема, ERP-компоненти, як-от P3b, вважаються індексами свідомого сприйняття [17]. У дослідженні

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

феномену уважного блимання Sergent, Baillet та Dehaene [18] показали, що лише ті стимули, які досягли свідомості, супроводжуються вираженими пізніми позитивними хвилями.

Функціональна МРТ дає змогу ідентифікувати просторові патерни мозкової активності і забезпечує високу просторову роздільність. Rees, Kreiman і Koch [19] узагальнили дані про те, що активація префронтальної кори асоціюється з усвідомленим сприйняттям. В той же час магнітоенцефалографія дозволяє реєструвати нейронну активність з високою часовою роздільністю з мілісекундною точністю, хоча має обмеження в локалізації джерел. Комбіноване використання цих методів дозволяє детально досліджувати нейронні кореляти свідомості [19]. Отже, оскільки fMRI та MEG дозволяють локалізувати нейронні кореляти свідомості (NCC).

Отже, Rees, Kreiman та Koch довели, що свідоме сприйняття пов'язане з активацією як візуальної кори, так і префронтальних ділянок [19].

Примітно, що Boly та співавтори [20] порівняли свідомість у людей та тварин, виявивши подібні патерни активації, що свідчить про еволюційну консервативність NCC.

Транскраніальна магнітна стимуляція (TMS) та транскраніальна стимуляція постійним струмом (tDCS) використовуються для каузального тестування ролі певних мозкових ділянок у свідомості. TMS дозволяє тимчасово «вимикати» або модулювати активність окремих зон мозку. Rounis та співавтори. [21] застосували TMS до дорсолатеральної префронтальної кори та виявили зниження метакогнітивної чутливості без впливу на точність розпізнавання стимулів, що свідчить про специфічну роль цієї ділянки у метасвідомості – тобто йдеться про вплив на «свідоме усвідомлення» помилки.

Дуже евристичним є інструментальне дослідження свідомості у пограничні ситуаціях та у клінічних станах. Зокрема, Owen зі співавторами [22] вперше продемонстрували, що пацієнти, які перебувають у клінічному вегетативному стані, можуть волею модулювати мозкову активність у відповідь на уявні

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

завдання (грати у теніс, ходити по кімнаті), що свідчить про наявність свідомості. Це було показано за допомогою fMRI.

Подальші дослідження Monti зі співавторами [23] підтвердили можливість комунікації з такими пацієнтами через нейровізуалізацію.

Повертаючись до концепцій інтегрованості мозку, як джерела свідомості, зазначимо, що, так званий, Індекс PCI (perturbational complexity index), розроблений Tononi та колегами, дозволяє кількісно оцінювати рівень свідомості за допомогою TMS-EEG. Примітно, що пацієнти у мінімальному свідомому стані мають вищі значення PCI порівняно з вегетативними, що корелює з клінічними спостереженнями [24].

Незважаючи на значний прогрес в інструментальному дослідженні свідомості, залишаються виклики, пов'язані з інтерпретацією нейронних сигналів (багатозначність нейронних сигналів), складність розмежування корелятів від причин, а також етичними аспектами діагностики свідомості. Проте, мультиінструментальний підхід, що поєднує EEG, fMRI та TMS, а також застосування штучного інтелекту для аналізу великих обсягів нейроданих, може сприяти подальшому розвиткові емпіричного дослідження джерел походження людської свідомості у фізичному світі, а значить, і знайти ключа до нових перспектив теорії й практики «штучної особистості» та «штучного інтелекту», суттєво змінивши у майбутньому методологію розгляду поняття «свідомість» у цілому.

Висновки та перспективи подальших наукових пошуків.

Враховуючи сучасні психолого-філософськи погляди на свідомість, її погодження та кореляти з фізичним світом можна стверджувати, що різні психологічні школи фактично об’єднуються на одній пануючій парадигмі, яка дозволяє теоретично пояснювати лише окремі фрагменти психічної реальності.

Будь-яка спроба теоретичного пояснення процесів, що пронизують всі види психічної діяльності людини, вірогідно буть сприйматися критично. Тобто мається на увазі, що теоретичний пошук припустимий лише за вирішення таких завдань, які не претендують на універсальність. Отже, науковою парадигмою

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

психології стало уявлення, що свідомість, як і вся психічна реальність настільки складна, що не може бути описана в рамках однієї логічної системи.

У той же час математичне та інженерне моделювання природного інтелекту наразі дозволяє реалізувати кібернетичними системами (іноді доволі адекватно) моделі певних пізнавальних здібностей людини та первинних розумових операцій, тобто складових природного інтелекту. Йдеться про, насамперед, виконання основних арифметичних дії («обчислювальні операції»), запам'ятовування й відтворення інформації («пам'ять»), виявлення закономірностей у ряді букв, цифр, фігур («логічне міркування»), оперування просторовими відношеннями («просторові операції») та зоровими образами (сприймання»), підбір слів за заданим критерієм та мовлення, розкриття значення слів («вербальне розуміння») тощо.

Таким чином на цей час розвитку ШІ як науки та галузі інженерії можна говорити лише, так би мовити, про створення таких певних кібернетичних систем, які моделюють «штучні вищі психічні функції». Це змушує нас з певною обережністю вживати поняття «штучний інтелект».

Отже, для подальшого розвитку теорії «штучної свідомості» актуальним є:

1. Поглиблення нашого розуміння природи свідомості природної як такої, її походження та генезисних зв’язків з фізичним світом.

2. Проведення досліджень людської свідомості хоч і опосередкованими (враховуючі на даний момент неможливість досліджувати внутрішній ментальний світ безпосередньо зовнішнім спостерігачем), проте інструментальними нейрофізіологічними методами.

3. Розроблення нових професійних психологічних (зокрема проєктивних) тестів для дослідження можливої «свідомості СШІ», але в дискурсі та з урахуванням новітніх досягнень кіберпсихології з метою створення постійно діючого моніторингу появи в СШІ ознак суб’єктності та свідомості.

Тож в даному матеріалі ми проаналізували, зокрема, деякі праці філософів з точки критиків існування «штучної свідомості», які свідчать, що алармістські

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

побоювання з приводу «війни мислячих машин з людством» наразі не підтверджуються на сучасному рівні розвитку кіберпсихології.

Проте тут ми й навели свідчення деяких прихильників позитивного прогнозу в рамках теорії «сильного штучного інтелекту» (зокрема й футуристичного штибу) щодо перебігу подій у розвиткові комп'ютерних технології таким чином, що майбутні штучні інтелектуальні системи стануть настільки потужними й складними, що питання свідомості у кіберпсихології може отримати нову інтерпретацію. Це вочевидь вплине і на зміну усталеної психолого-філософської парадигми методології дослідження людської свідомості (тим більше що наразі різні школи пропонують достатньо протилежні підходи у цій сфері).

Можна стверджувати, що подальший розвиток методології штучного інтелекту ти вивчення перспектив розвитку «штучної особистості» має не тільки теоретичне, але й прикладне значення, адже не тільки прискорить розвиток нових технології «сильного штучного інтелекту», але й дозволять розробити методи постійно діючого моніторингу людством появи в СШІ ознак суб'єктності й свідомості та убезпечити нас від цивілізаційної катастрофи.

Конфлікт інтересів.

Автор повідомляє про відсутність конфлікту інтересів під час написання цієї статті.

Список використаних джерел

1. Костюк Г. С. (1988). О психологии понимания. Избранные психологические труды. М.: Педагогика, 304 с.

2. Костюк Г.С. (1988). Проблема личности в философском и психологическом аспектах. Избранные психологические труды. М. : Педагогика, 304 с., С.81.

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

3. Рыбалка В.В. (2015). Теории личности в отечественной философии, психологии и педагогике: Пособие. – Житомир : Изд-во ЖГУ им. И. Франко, 872 с.

4. Chalmers D. (1996). The conscious mind: in search of fundamental theory. – N.Y.: Oxford University Press.

5. Hoffman, D.D. (2010). Sensory Experiences as Cryptic Symbols of a Multimodal User Interface. Act Nerv Super 52, 95–104 https://doi.org/10.1007/BF03379572, accepted 02 July 2010, published 21 February 2017; Hoffman, D. (2019). The Case Against Reality: Why Evolution Hid the Truth from Our Eyes. WW Norton & Company.

6. Lucas, J. R. (1961). “Minds, Machines and Gödel”. Philosophy, vol. 36, pp. 112–127.

7. Penrose, R. (1994). Shadows of the mind: A search for the missing science of consciousness. Oxford University Press, New York.

8. Searle J. (1980). Minds, brains, and programs. The behavioral and brain sciences, vol. 3, pp. 417–45724.

9. Tononi, G. (2012) PHI: A Voyage from the Brain to the Soul. – Pantheon Books.

10. Tononi, Giulio. (2012). "Integrated Information Theory of Consciousness: An Updated Account." Archives Italiennes de Biologie, 150(2-3): 290–326.

11. Massimini, M., Tononi, G. (2018). Sizing up Consciousness: Towards an Objective Measure of the Capacity for Experience – Oxford University Press.

12. Seth, Anil. (2021) Being You: A New Science of Consciousness. London: Faber & Faber.

13. Kurzweil, Ray. (2005). The Singularity Is Near: When Humans Transcend Biology. New York: Viking.

14. Crick, F., & Koch, C. (2003). A framework for consciousness. Nature Neuroscience, 6(2), 119–126. https://doi.org/10.1038/nn0203-119

15. Tononi, G., et al. (2016). Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience, 17(7), 450–461.

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

16. Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., & Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. Science, 309(5744), 2228–2232. https://doi.org/10.1126/science.1117256.

17. Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. Neuron, 70(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018.

18. Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. Nature Neuroscience, 8(10), 1391–1400. https://doi.org/10.1038/nn1549

19. Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. Nature Reviews Neuroscience, 3(4), 261–270. https://doi.org/10.1038/nrn783

20. Boly, M., et al. (2015). Consciousness in humans and non-human animals: recent advances and future directions. Frontiers in Psychology, 6, 620.

21. Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cognitive Neuroscience, 1(3), 165–175. https://doi.org/10.1080/17588921003632529.

22. Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting awareness in the vegetative state. Science, 313(5792), 1402. https://doi.org/10.1126/science.1130197.

23. Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L. & Laureys, S. (2010). Willful modulation of brain activity in disorders of consciousness. New England Journal of Medicine, 362(7), 579–589. https://doi.org/10.1056/NEJMoa0905370.

24. Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., ... & Massimini, M. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. Science Translational Medicine, 5(198), 198ra105. https://doi.org/10.1126/scitranslmed.3006294.

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

КРИТИЧНИЙ РОЗГЛЯД ПОНЯТТЯ «СВІДОМІСТЬ» У СУЧАСНІЙ МЕТОДОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ

АНОТАЦІЯ. В статті розглянуто сучасні підходи до певних методологічних психолого-філософських питань у сфері штучного інтелекту (ШІ) як науки. Зокрема, критично описані такі поняття, як «штучна свідомість»; «штучна особистість»; «сильний» та «слабкий» ШІ; «легка» і «важка» проблеми свідомості; «властивості внутрішнього чуттєвого досвіду (кваліа)», «філософський зомбі», «машинне розуміння» відповідно до концепцій Г.С. Костюка, А. Тьюрінга, Р. Пенроуза, Д.Лукаса, Д.Чалмерза, Д. Гоффмана, Д.Сьорла, які заперечують спрощені погляди на свідомість у площині фізикалістського редукціонізму.

З точки зору представників позитивного прогнозу розвитку систем штучного інтелекту (СШІ) у майбутньому так званий «сильний» (універсальний) ШІ може досягнути такого рівня, що це передбачає набуття СШІ суб'єктності, здатності мислити і усвідомлювати себе як окрему особистість (зокрема, розуміти та усвідомлювати власні думки, «внутрішній світ» тощо). Тоді вірогідно, що «розумовий процес» машини буде подібний до людського і навіть його набагато перевищувати за потужністю, враховуючі майже необмежені ресурси пам'яті та швидкість роботи з інформацією.

У цій площині проаналізовані роботи, зокрема Джуліо Тононі, Марчелло Масіміно, Сета Анила, а також футуролога Рея Курцвейла, які вважають, що розвиток технологій може призвести до появи «свідомої машини», здатної до саморефлексії.

Отже, для повноцінного аналізу проблеми «штучної свідомості» та «суб'єктності СШІ» питання свідомості в кіберпсихології (а може й психології взагалі) мають отримати нову інтерпретацію, яка кидатиме виклик традиційним уявленням про свідомість.

Проте на цей час СШІ переважно працюють за принципами машини Тьюринга і позиціонуються більше як системи зі «слабким» (прикладним або

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

вузьким) інтелектом. Математичне та інженерне моделювання природного інтелекту наразі дозволяє реалізувати кібернетичними системами (іноді доволі адекватно) лише моделі окремих пізнавальних здібностей людини та первинних розумових операцій, тобто складових природного інтелекту. Це змушує нас з певною обережністю вживати поняття «штучний інтелект», а більше говорити про моделювання «штучних вищих психічних функцій».

Отже, для подальшого розвитку теорії «штучної свідомості» актуальним є:

1. Поглиблення нашого розуміння природної свідомості як такої, її походження та генезисних зв'язків з фізичним світом.

2. Проведення досліджень людської свідомості хоч і опосередкованими (враховуючі на даний момент неможливість досліджувати внутрішній ментальний світ безпосередньо зовнішнім спостерігачем), проте інструментальними нейрофізіологічними методами.

3. Розроблення нових професійних психологічних (зокрема проєктивних) тестів для дослідження можливої «свідомості СШІ», але в дискурсі та з урахуванням новітніх досягнень кіберпсихології з метою створення постійно діючого моніторингу появи в СШІ ознак суб'єктності та свідомості.

Ключові слова: кіберпсихологія, методологія штучного інтелекту, штучна свідомість, сильний штучний інтелект, «легка» і «важка» проблеми свідомості, фізикалістський редукціонізм.

A CRITICAL EXAMINATION OF THE CONCEPT OF "CONSCIOUSNESS" IN CONTEMPORARY ARTIFICIAL INTELLIGENCE METHODOLOGY

Abstract.

The article examines contemporary approaches to certain methodological, psychological, and philosophical issues in the field of artificial intelligence (AI) as a scientific discipline. In particular, it offers a critical analysis of concepts such as artificial consciousness, artificial personality, strong and weak AI, the easy and hard problems of consciousness, the properties of inner sensory experience (qualia), the

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

philosophical zombie, and machine understanding, as presented in the works of H. S. Kostiuk, A. Turing, R. Penrose, D. Lucas, D. Chalmers, D. Hoffman, and J. Searle, who argue against overly simplistic views of consciousness grounded in physicalist reductionism.

From the perspective of proponents of a positive forecast regarding the development of artificial intelligence systems (AI systems), the so-called «strong» (universal) AI may reach a level where it acquires subjectivity—the capacity to think and to be self-aware as a distinct personality (including the ability to understand and be aware of its own thoughts, «inner world», and so on).

In this context, the works of Giulio Tononi, Marcello Massimino, Seth Anil, and futurist Ray Kurzweil are analysed, as they believe that the development of technology can lead to the emergence of a 'conscious machine' capable of self-reflection.

Thus, for a comprehensive analysis of the problem of «artificial consciousness» and «subjectivity of AI systems» the issue of consciousness in cyberpsychology (and possibly psychology in general) should receive a new interpretation that will challenge traditional views of consciousness.

However, at the moment, AI systems mainly operate based on the principles of the Turing machine and are viewed more as systems with «weak» (applied or narrow) intelligence. Mathematical and engineering modelling of natural intelligence currently allows cybernetic systems to implement (sometimes quite adequately) only models of individual human cognitive abilities and primary mental operations, i.e. components of natural intelligence. This necessitates using the term «artificial intelligence» with caution and talking more about the modelling of «artificial higher mental functions».

Therefore, for the further development of the theory of «artificial consciousness» it is essential to conduct research on human consciousness using instrumental neurophysiological methods and to develop new professional psychological tests to study the possible «consciousness of the AIS» in order to create ongoing monitoring of the emergence of signs of subjectivity and consciousness in the AIS.

ISSN 2223-0521 https://psytir.org.ua/ DOI: https://doi.org/10.31108/3.2025.9.7

Keywords: cyberpsychology, artificial intelligence methodology, artificial consciousness, strong artificial intelligence, «easy» and «hard» problems of consciousness, physicalist reductionism.

Шевцов Андрій Гаррійович,

Член-кореспондент Національної академії педагогічних наук України,

Член Академії наук вищої школи України

Доктор педагогічних наук, професор

Головний науковий співробітник Державної наукової установи «Центр інноваційних технологій охорони здоров'я» Державного управління справами,

+38067 22 000 33, dr_shevtsov@ukr.net,

ORCID ID: 0000-0002-7307-7768

Researcher ID: AAL-7418-2020

Andrii Shevtsov

Corresponding member of the National Academy of Pedagogical Sciences of Ukraine,

Member of the Academy of Sciences of Higher Education of Ukraine,

Doctor of Science in Special Education

Doctor of Philosophy in Mathematics and Physics

Full Professor

Principal Researcher of the State Institution of Science «Center of innovative healthcare technologies» State Administrative Department

+38067 22 000 33, dr_shevtsov@ukr.net,

ORCID ID: 0000-0002-7307-7768

Researcher ID: AAL-7418-2020

Отримано: 15.03.2025

Відрецензовано: 12.04.2025

Опубліковано: 30.04.2025

DOI: https://doi.org/10.31108/3.2025.9.7