Шевцов Андрій. Розділ XVII. Критичний розгляд поняття «свідомість» у сучасній методології штучного інтелекту. Штучний інтелект у науці : монографія / [авт. колектив]; за ред. Яцишина Андрія та Яцишин Анни. – Київ: ФОП Ямчинський О.В., 2025. – С. 223-231. ISBN 978-617-8830-09-0

РОЗДІЛ XVII. КРИТИЧНИЙ РОЗГЛЯД ПОНЯТТЯ «СВІДОМІСТЬ» У СУЧАСНІЙ МЕТОДОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ

DOI: 10.33407/lib.NAES.id/748284

Шевцов Андрій[1] [0000-0002-7307-7768]

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

dr_shevtsov@ukr.net

Анотація. Розглянуто сучасні підходи до певних методологічних психолого-філософських питань у сфері 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 з точки зору фізикалістського редукціонізму. Наведений аналіз сучасного рівня розробок ШІ свідчить, що алармістські побоювання з приводу «війни мислячих машин з людством» наразі не підтверджуються фундаментальними працями у сфері філософії та психології.

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

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.

Понятійно-термінологічне поле кіберпсихології

На особливому місці знаходяться проблеми понятійно-термінологічного поля науки про СШІ. 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, психології та філософії, а відповідно до сфери своєї основної професійної діяльності. Отже 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.

При цьому доречним у сучасних дослідженнях ШІ наведення певних ремінісценції із робіт наших вітчизняних психологів-класиків, зокрема праць Костюка Г.С. і його теорії особистості.

«Сильний» та «слабкий» штучний інтелект

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 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 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. (Поки що…!).

«Штучна особистість»

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 без свідомості не має особистості!

«Штучна свідомість»

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 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, як машина 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.

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 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.

Кваліа та «штучна особистість»

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, кваліа, 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 таким 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 і 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 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?

«Машинне розуміння»

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

Всі ці аспекти можна успішно 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. 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 [8].

Висновки. В даному матеріалі ми проаналізували деякі праці філософів з точки критиків існування «штучної свідомості». Звичайно в літературі можна знайти також і безліч прихильників позитивного прогнозу щодо перебігу подій у кіберпсихології таким чином, що майбутні штучні інтелектуальні системи стануть настільки складними, що питання свідомості може отримати нову інтерпретацію. Проте наведений вище аналіз ситуації у сфері ШІ свідчить, що алармістські побоювання з приводу «війни мислячих машин з людством» наразі не підтверджуються на сучасному рівні розвитку кіберпсихології.

У подальшому як теоретичний, так і прикладний аспект можуть мати дослідження зазначених питань щодо перспектив розвитку «штучної особистості» з точки зору методологічного доробку прихильників футуристичної теорії «сильного штучного інтелекту».

Список джерел

1. Костюк Г. С. (1988). Про психологію розуміння. Вибрані психологічні праці, 304 с.

2. Костюк Г.С. (1988). Проблема особистості у філософському та психологічному аспектах. Вибрані психологічні праці. 304 с., С.81.

3. Рыбалка В.В. (2015). Теории личности в отечественной философии, психологии и педагогике: Пособие. – Житомир : 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.

* * *