А.К.Малеєв,
аспірант філософського факультету
КНУ імені Тараса Шевченка

ПРОБЛЕМНІСТЬ ШТУЧНОГО ІНТЕЛЕКТУ І СВІДОМОСТІ ЗА РОБОТОЮ ДЖОНА СЕРЛЯ «СВІДОМІСТЬ, МОЗОК І НАУКА»

Досліджуючи раціональність, штучний інтелект та людський мозок, Джон Серль[*] стикається з одним з фундаментальних питань – проблемою свободи людської волі[†]. З'ясування цього питання може пролити світло не тільки на можливість створення штучного інтелекту, а й на спосіб, завдяки якому функціонує людський мозок. Це дасть можливість дати відповідь на одвічне питання філософії – «Що таке людина?» або «Що означає бути людиною?».

Науково-технічний прогрес, який почав набирати обертів у другій половині XX ст., популяризація науки виявили нові грані старих філософських проблем. Для їх вирішення знадобились нестандартні підходи. Один з них використав Алан Тьюринг в амбіціозній спробі визначити момент появи у машини штучного інтелекту (тут і надалі – ШІ). Як результат – був сформульований тест Тьюринга.

Питання про здатність машини вести себе так само, як людина, у філософії є доволі небезпечним через загрозу поринути у вир тисячолітніх наукових суперечок навколо тлумачень термінів «розум», «раціональність», «інтелект», «мислення» і т.ін. Саме тому Тьюринг формулює тест. Він повинен надати настільки очевидну відповідь, яка не змушувала б замислюватися над тлумаченням окремих термінів чи понять.

У тесті беруть участь троє – дві людини та машина. Всі перебувають у різних кімнатах, спілкуються опосередковано (адаптуючи

[*] Джон Серль (John Rogers Searle, 31 липня, 1932, Денвер, Колорадо) – американський філософ, професор Каліфорнійського університету в Берклі, відомий своїм внеском у філософію мови, філософію свідомості й соціальну філософію. Зокрема, Серль автор концепції уявного експерименту під назвою «китайська кімната», що є критикою можливості створення так званого сильного штучного інтелекту.

[†] Під свободою волі у даному дослідженні розуміється уся сукупність свобод людини, що виражені через здатність останньої бути причиною як своїх дій, так і себе самої.

до сьогодення – за допомогою текстових SMS повідомлень) у формі простих питань та однозначних відповідей. Люди не знають «хто є хто». Завдання інтерв'юера (один з двох людей) – відрізнити співрозмовників – машину від людини. Завдання машини – видати себе за людину. Якщо останній це вдасться, то ми маємо отримати відповідь на питання – «чи здатні машини мислити?» [2, 12]. Якщо продовжити Тьюринга, то здатність обманювати когось є ознакою розуму чи, принаймні, здатності мислити.

Однак цей тест викликав ще більше питань, ніж мав надати відповідей. Він поновив дискусію навколо термінів «мислення» та «раціональність», які, в свою чергу, є сутнісними проблемами вивчення свідомості. Ці питання поділили дослідників на представників слабого та сильного ШІ. «Відповідно до ШІ, основна цінність комп'ютера у вивченні свідомості полягає в тому, що він дає нам деякий дуже потужний інструмент» [1]. Слабкий інтелект – це така собі удосконалена програма, яка здатна допомогти нам у вирішенні складних раціональних проблем, коли ми обмежені у часі. В даному випадку це не більше, ніж інструмент, який є об'єктом.

«Відповідно ж до сильного ШІ, комп'ютер – це не просто інструмент в дослідженні свідомості; комп'ютер, запрограмований належним чином, насправді і є деяка свідомість, в тому розумінні, що можна буквально сказати, що за наявності належних програм комп'ютери розуміють, а також мають інші когнітивні стани» [1]. Відповідно, сильний ШІ буде не просто машиною, а свідомістю. І якщо це твердження правильне, то така свідомість цілком здатна бути не тільки інструментальним об'єктом для вирішення завдань, а й суб'єктом дії.

Джон Серль, аналізуючи тест Тьюринга та різновиди машин, побудованих на ідеї проходження тесту, пропонує зіграти в уявну логічну гру «Китайська кімната». Припустимо, що вас помістили у кімнату, де розставлені кошики, повні слів невідомої вам мови (для Серля це китайська) [1]. Вам відомі правила поєднання слів за зовнішнім виглядом, але не за змістом – він залишається вам невідомим. Знавці мови дають вам нові кошики з невідомими словами, немов ставлячи перед вами «питання», а ваше завдання розмістити їх в правильному порядку, так, щоб вийшла «відповідь». Припустимо також, що правила поєднання слів написані так, що ваші відповіді на питання не відрізняються від відповідей людини, яка вільно розмовляє цією мовою. У такому випадку ви витримаєте тест Тьюринга.

Та ви все одно нічого не розумітимете в цій мові. Чим в такій ситуації ви відрізняєтесь від комп'ютера? А отже, і комп'ютер в такому розумінні не є свідомістю. Більше того, на тому рівні технологічних можливостей його неможливо навіть запрограмувати так, щоб він володів свідомістю.

Висновки цього інтелекту посилили інтерес вчених до дослідження інтенціональності, раціональності та свідомості. Вихід вбачався один – варто ще раз дослідити структуру людського мозку та свідомості, структурувати їх, надавши достатні визначення, а потім порівняти з комп'ютером та програмами. Таке порівняння дало б елементи, яких не вистачає комп'ютеру, щоб перетворитися на штучний інтелект.

Порівняльні співвідношення програми і комп'ютера з мозком і свідомістю, а також певні «неточності» мови, спокушають повернутися до ідеї декартівського дуалізму фізичного та ментального. Якщо рухатись у цьому напрямку, то комп'ютер є фізичним проявом тіла, подібним до людського мозку, а комп'ютерна програма подібна до людської свідомості.

Звернімо увагу на сутнісний момент лінгвістичних неточностей, які ведуть до непорозуміння щодо штучного інтелекту. Коли перед програмістом стоїть завдання пояснити людині, що не володіє відповідними знаннями, чому комп'ютер на півтори хвилини «зависає» при обробці того чи іншого завдання, то програміст не розповідатиме про весь набір процесуальних дій, які у даний момент відбуваються у «сталевих нутрощах». Замість цього він скаже – комп'ютер думає. Зазвичай таке твердження сприймається як метафора. Втім, трапляється, що такі метафори сприймаються буквально, тобто антропоморфізуючи комп'ютер та програми. Виникає хибне уявлення, що будь-яке повідомлення системи, – це комунікація комп'ютера з людиною, що супротивник у комп'ютерній грі, який є програмою, справжній, здатний самостійно приймати рішення. Та програма є не більше, ніж формальним інструментом, схожим на електричний сигнал у мозку людини. Та хіба сам по собі електричний сигнал буде свідомістю?

Не погоджуючись з позицією дуалістичного погляду на світ, Джон Серль висловлює припущення, що свідомість з людським мозком пов'язана точнісінько так само, як пов'язана твердість столу зі специфікою побудови деревини на молекулярному чи атомному рівні. Простіше кажучи, якщо ми маємо стіл і він є дерев'яним, то

у більшості випадків він буде твердим. Твердість у даному випадку – це властивість. Так само і з мозком – якщо ми маємо живий мозок з усіма необхідними йому елементами, то він обов'язково матиме таку властивість, як свідомість.

Відповідно до цієї ідеї Серль припускає, що більшість дослідників припускаються помилки. Якщо свідомість дійсно є властивістю мозку подібно до того, як твердість є властивістю дерев'яного столу, то більшість вчених намагаються знайти одну єдину молекулу, атом чи електричний сигнал, які б відповідали за твердість столу чи за свідомість у людини. Ми не можемо знайти конкретний нейрон у мозку і оголосити, що він відчуває біль. Однак, цілком логічно, що весь мозок, як невід'ємна частина нас самих, здатен на відчуття болю [3, 13]. На базі цих припущень нейрони, мозок та свідомість варто сприймати як декілька рівнів однієї і тієї самої системи.

У 1984 р., в серії лекцій «Свідомість, мозок і наука» Джон Серль формулює свою ідею більш формалізовано [3, 24–26]:

– Мозок породжує розум.

– Синтаксису недостатньо для існування семантики.

– Комп'ютерна програма повністю визначається своєю синтаксичною структурою.

– Людський розум оперує смисловим змістом (семантикою).

І робить висновки:

– Програми не є сутністю розуму і їх наявності недостатньо для наявності розуму.

– Той спосіб, за допомогою якого людський мозок породжує ментальні явища, не може зводитися лише до виконання комп'ютерної програми.

– Те, що породжує розум, повинно мати, принаймні, причинно-наслідкові властивості, еквівалентні відповідним властивостям мозку.

– Ментальні стани та свідомість взагалі є не більше, ніж біологічними феноменами. Це означає, що без живого організму вони не можуть існувати.

Порівнюючи комп'ютерні програми із свідомістю та досліджуючи останню, Серль зробив ще деякі важливі формалізовані висновки, які ми розглянемо. Перший стосується того, що між мозком і свідомістю немає розриву у розумінні підсвідомих психологічних програм та правил, які б регулювали свідомість. Цей тезис є критикою

Ноама Хомського*, який припускає існування загальної граматики, тобто певної універсальної системи, з якої і вибудовується кожна унікальна свідомість.

Комп'ютер та програми побудовані на основоположних правилах. Порівняння комп'ютера з людським мозком спокушає зробити припущення, що людська поведінка теж ґрунтується на деяких основоположних правилах. Серль вважає таке припущення помилковим, хоча б тому, що, коли ми маємо на увазі комп'ютер, наше висловлювання про правила є метафоричним. А щодо ставлення до людини і її поведінки – ні. Машина не може не дотримуватися правил. Вона зобов'язана це робити – це причинно-наслідковий зв'язок. Аналогічно до нашого висловлювання «машина підпорядковується правилам» можна було б сказати, що, коли ми молотком забиваємо цвях у діжку, цвях теж дотримується правила: якщо по тобі вдарили – занурюйся в речовину в напрямку удару. Та цвях сам не обирає – зануритись йому чи ні. Комп'ютер не може не виконати дану йому команду. А людина може. Серль на цьому не акцентує увагу, та з його тез можна зробити висновок, що на відміну від машини, для людини правило є питанням добровільного виконання, тим часом як для комп'ютера – законом причинно-наслідкового зв'язку.

Окремим предметом дослідження Серля є осмислена людська дія. Згодом це виллється в окрему працю філософа – «Раціональність у дії», що є ґрунтовним переосмисленням його філософської теорії свідомості. Та на даному етапі дослідження він намагається з'ясувати, як людина діє. З цього випливають такі основоположні принципи структури людської дії [3, 41–47]:

1. Дія складається з двох компонентів: ментального та фізичного.

2. Ментальний компонент інтенціональний – він про щось. Кожна дія має зміст.

Серль відзначає важливий момент: «З точки зору теорії інтенціональності, дія складається з двох компонентів: ментального та фізичного. Якщо дія успішна, ментальний компонент причинно зумовлює і репрезентує фізичний компонент. Цю форму причинності я називаю інтенціональною причинністю» [3, 42]. Це підводить до

* Авра́м Ноа́м Хо́мський (Avram Noam Chomsky; 7 грудня, 1928, Філадельфія, Пенсильванія) – американський лінгвіст, філософ та політичний активіст, аналітик, літератор, професор мовознавства Массачусетського технологічного інституту (МТІ) у відставці. Хомський добре відомий науковій спільноті як один із засновників сучасної лінгвістики та визначна постать в аналітичній філософії.

наступного принципу:

3. Інтенціональна причинність невіддільна від структури дії та її пояснення.

4. Є два види інтенцій: ті, що заплановані задовго до дії – попередні інтенції, та ті, що виявляються під час самого виконання дії – інтенції у дії.

5. Формування попередніх інтенцій зазвичай є результатом практичного розмірковування для вибору між конфліктуючими бажаннями.

6. Дія має пояснюватись через той зміст (причини), який був наявний в голові у людини, що виконувала дію або розмірковувала над попередньою інтенцією до виконання дії.

7. Будь-який інтенціональний стан функціонує і може бути задоволений тільки в якості елементу системи інтенціональних станів.

8. Система інтенціональності функціонує на фоні людських можливостей, які самі не є ментальними станами.

Втім, дана сукупність принципів, кожен з яких є елементом структури людської дії, не може розглядатися як такий поза самою структурою, точнісінько як і молекула зі столу поза столом має зовсім інше значення і властивості. Структура людської дії у даному випадку – скоріше, набір елементів, без яких раціональна дія людини в принципі неможлива. Ця структура не містить причини, тобто того, хто здійснював дію – суб'єкта.

Порівнюючи характерні риси комп'ютерів та людського мозку, досліджуючи людську свідомість та структуру людської дії, Серль намагається уникнути основної проблеми, яка ставить під сумнів усю теорію сильного штучного інтелекту. Ця проблема криється у свободі волі.

Питання сильного штучного інтелекту формується з припущення, що комп'ютер за наявності певного програмного забезпечення здатен розуміти, навіть більше того – перетворитися на самостійну свідомість, а отже, може діяти вільно. Джон Серль вдало використовує тест «Китайської кімнати», щоб висвітлити той факт, що комп'ютер – це не більше, ніж формалізована машина і що вона розуміє не більше, ніж будь-який інший інструмент, який вигадала людина. Комп'ютер імітує людське мислення, проте людським мисленням він не є. Для того, щоб чітко описати відмінності між комп'ютером та людською свідомістю, він має їх порівняти та більш ґрунтовно дослідити свідомість і структуру людської дії. Однак у дослідженні він потрапляє у логічну пастку – і до свідомості, і до раціональної дії ставиться

як до об'єктів, котрі треба дослідити.

Об'єкт як такий підлягає причинно-наслідковим зв'язкам. Одні об'єкти викликані іншими. Для прикладу – сила тяжіння викликає падіння предметів вниз. Опір повітря призводить до того, що одні предмети, які мають велику площу спротиву, падають швидше, інші – повільніше. Кожний феномен має причину, яка його викликає. Науковець в ідеалі розглядає світ як ланцюг причинно-наслідкових подій. В такому випадку світ мав лише один першопочаток. Це може бути Великий Вибух, можна його назвати Богом, а для Гегеля цілком прийнятним здавався і Абсолютний Дух. Цей першопочаток доволі часто ще називається суб'єктом. В експерименті суб'єктом буде науковець.

При дослідженні сильного штучного інтелекту, питання мало б полягати у тому, чи здатні ми породити комп'ютер в якості суб'єкта? Адже людина, навіть якщо вона цього не усвідомлює, є суб'єктом власних дій. Це означає, що вона єдина може обирати причини для своєї діяльності. Вона здатна чинити раціонально чи ні тільки за рахунок свободи волі. Та чи може машина чинити раціонально, якщо у неї немає свободи волі? Чи здатна машина бути суб'єктом дій, якщо всі її ходи наперед записані програмістом, а сама вона є лише імітацією людського мислення (і то не всього, а окремих логічних задач)?

Коли Серль розглядає свідомість та дії, він ставиться до них як до об'єктів. І як об'єкти вони підлягають певним об'єктивним законам світу. Досліджуючи їх в якості об'єктів, ми нічого не дізнаємося про суб'єкта – людину. Саме тому, коли Серль підходить до проблеми свободи волі, він пише про власну поразку. Вона проявляється у двох аспектах. Насамперед, Серль не може пояснити одночасне існування і детермінізму, і свободи волі. По-друге, хоча наявність людської свободи для Серля є безумовним фактом, він не може спрямувати її вплив на прийняття людиною рішень. Аналіз свободи волі у Серля зведеться до дослідження об'єкта. І як об'єкт свобода волі не може бути пояснена, бо не зрозуміло, чим вона викликана і на загальному рівні, і в кожному конкретному випадку.

Свого часу Макс Шелер написав, що людина ще ніколи не була для себе такою проблемною. Зараз можна сміливо стверджувати, що штучний інтелект робить людину ще більш проблемною. Алан Тьюринг сформулював питання про можливість машин мислити. Це питання породило «Китайську кімнату» Джона Серля, яка є відповіддю про неможливість сильного штучного інтелекту на базі

формальних систем.

Дослідження умов виникнення штучного інтелекту – амбітна мета, що актуалізувала дослідження людської свідомості. Науково-технічний прогрес багато в чому допоміг виявити структуру людської діяльності та зрозуміти співвідношення свідомості й людського мозку. Та ці ж дослідження спричиняють ще одну проблему, яка потребує принципово іншого підходу, – людина є суб'єктом своїх дій. Проблемність штучного інтелекту потребує не тільки вирішень на рівні досліджень свідомості як об'єкта, а й досліджень самого суб'єкта, що здатен бути причиною власних дій. Це породжує необхідність дослідження людини як суб'єкта дії.

ЛІТЕРАТУРА

1. *Серль Д.* Сознание, мозг и программы. – URL: http://www.philsci.univ.kiev.ua/biblio/searle.html
2. *Тьюринг А.* Может ли машина мыслить? – М., 1960. – 67 с.
3. *Searle J.* Minds, Brains and Science // Сёрль Дж.Р. Сознание, мозг и наука. – М., 1993. – 64 с. // Электронная публикация: Центр гуманитарных технологий. – 2013. URL: http://ecsocman.hse.ru/data/2010/05/20/1214101925/002_Dzhon_Serl_Soznaniex2c_mozg_i_nauka_03-66.pdf

Малеєв А.К. Проблемність штучного інтелекту і свідомості за роботою Джона Серля «Свідомість, мозок і наука».

Логічний експеримент Джона Серля «Китайська кімната» є критикою машин, заснованих на тесті Алана Тьюринга. Мета цього експерименту – довести, що комп'ютери, побудовані на формальних системах, нездатні володіти свідомістю чи її елементами.

На думку Джона Серля, створення сильного штучного інтелекту неможливе без ґрунтовного дослідження взаємозв'язку мозку і свідомості, свідомості і дії. Ці думки викладені у його праці «Свідомість, мозок і наука». Втім, у цьому дослідженні перед Джоном Серлем виникає нерозв'язна проблема – антиномія свободи волі і детермінізму.

«Китайська кімната» вказує ще й на відсутність у комп'ютерів, побудованих на формальних системах, можливості бути першопричиною власних дій, тобто бути суб'єктом. Досліджуючи свідомість чи вільну дію як об'єкт, не можна дізнатися нічого про суб'єкт, котрий ними володіє. Адже об'єкт підпорядковується причинно-наслідковим зв'язкам світу.

А свобода волі суб'єкта – це розрив зв'язків, що надає можливість породжувати нові причинно-наслідкові події у світі. Це породжує необхідність змінити парадигму сучасних досліджень свідомості і вільної дії.

Ключові слова: штучний інтелект, свобода волі, свідомість, раціональна дія.

Малеев А.К. Проблематичность искусственного интеллекта и сознания по работе Джона Серля «Сознание, мозг и наука».

Логический эксперимент Джона Серля «Китайская комната» является критикой машин, основанных на тесте Алана Тьюринга. Цель этого эксперимента доказать, что компьютеры, основанные на формальных системах, не могут иметь сознания или его элементов.

По мнению Джона Серля, создание сильного искусственного интеллекта невозможно без основательного исследования взаимосвязи мозга и сознания, сознания и действия. Эти мысли изложены в его работе «Сознание, мозг и наука». Но в этом исследовании возникает неразрешимая проблема – антиномия свободы воли и детерминизма.

«Китайская комната» Серля указывает еще и на отсутствие у компьютеров, основанных на формальных системах, возможности быть первопричиной своих действий, то есть быть субъектом. Исследуя сознание или свободное действие как объект, ничего нельзя узнать про субъект, который ими владеет. Ибо объект подчиняется причинно-следственным связям мира. А свобода воли субъекта – это разрыв таких связей, что дает возможность порождать новые причинно-следственные события в мире. Это рождает необходимость изменить парадигму современных исследований сознания и свободного действия.

Ключевые слова: искусственный интеллект, свобода воли, сознание, рациональное действие.

Maleyev A.K. The problems of the artificial intelligence and consciousness at the work of John Searle, «Consciousness and the brain science».

«Chinese Room» is a logical experiment of John Searle. It is critique of machines based on the test of Alan Turing. The purpose of this test is to prove that computers based on formal systems may not have the consciousness or its components.

According to John Searle, the creation of strong artificial intelligence is impossible without a thorough research the connection of the brain and the consciousness, consciousness and action. These ideas form the basis of his work «The consciousness and the brain science». But in this work, John Searle notes the unsolved problems – the existence of the antinomy of free will and determinism.

«Chinese Room» Searle points also to absence at computers based on the formal system may be the primary cause of their actions, that is – the subject. Exploring consciousness or free action as an object, we do not know anything

about the subject who owns them. Because the object obeys the causal relations in the world. But free will – a break of such causal relations, which makes it possible to generate a new chain of causal events in the world. This creates the need to change the paradigm of modern consciousness research and the free action.

Key words: Artificial intelligence, free will, consciousness, rational action.