

УДК 165.24:004.81

Андрій Синиця

ПРОБЛЕМА РОЗУМІННЯ В КОНТЕКСТІ КРИТИКИ КОМП'ЮТАЦІОНАЛІЗМУ

У статті досліджено основні труднощі комп'ютаціональної моделі свідомості, яку розробляли представники аналітичної філософії. Особливу увагу звернено на феномен розуміння, який не може бути відтвореним формально. Критично проаналізовано альтернативний до комп'ютаціоналізму варіант створення штучної свідомості – коннективізм, який ґрунтується на таких принципах функціонування мисленнєвих процесів, які наслідують діяльність кори головного мозку.

Ключові слова: розуміння, свідомість, комп'ютаціоналізм, коннективізм, аналітична філософія.

A. Synytsya. The problem of understanding in the context of computational theory of mind criticism

The phenomenon of understanding is analyzed in the article. Understanding is defined as the ability of setting the meaning. You can understand different things – text, a situation, a sentence, a word context. It is stated that the phenomenon of understanding has become to analytic philosophy in the middle of the twentieth century, when the intensive development of the theory of artificial intelligence had occurred. One of the initial theories explaining the nature of consciousness as a computer program, was the computational theory of mind. It was formulated by A. Turing, J. Fodor, S. Pinker, D. Dennett, H. Putnam. They considered thinking in general and the phenomenon of understanding in particular to be embodied on computers. However many thinkers in analytic philosophy criticized this theory. H. Dreyfus proposed to consider various types of mental activity. One of such types is an informal thinking, that can not be formalized. It is noted that the most famous critic of the computational theory of mind was John Searle, author of the well-known mental experiment "Chinese Room". John Searle argued that the computer calculation and understanding are different phenomena. Computer operates merely with the syntax, but not with the semantics. However, J. Searle reasons were also criticized (in particular by P. S. Churchland and P. M. Churchland). Although the author noted that their arguments are also too weak to refute J. Searle reasons. It is stated that the connectivism doesn't explain the nature of consciousness as well. Connectivism as an alternative to computational theory of mind based on the principles simulating our mental processes is critically analyzed. It is ascertained that the consciousness in general and the understanding in particular are still mysterious phenomena for contemporary science and philosophy.

Keywords: understanding, consciousness, computational theory of mind, connectivism, analytic philosophy.

A. Синица. Проблема понимания в контексте критики компьютеризации

В статье исследованы основные трудности компьютеризации модели сознания, созданной представителями аналитической философии. Особое внимание обращено на феномен понимания, который не может быть воспроизведенным формально. Критически проанализировано альтернативный компьютеризации вариант создания искусственного сознания – коннективизм, основанный на принципах функционирования мыслительных процессов, имитирующих деятельность коры головного мозга.

Ключевые слова: понимание, сознание, компьютеризация, коннективизм, аналитическая философия.

Стрімкий розвиток комп'ютерних наук у другій половині ХХ століття призвів до постановки питання про те, а чи можливо мисленнєві процеси, які притаманні людині, відтворити на штучних носіях. Прихильники комп'ютаціоналізму (А. Тюрінг, Дж. Фодор, С. Пінкер, Д. Деннет, Г. Патнем), у межах якого мисленнєві процеси розглядаються за аналогією із роботою комп'ютера, переконані, що мислення в цілому й феномен розуміння зокрема можуть бути реалізованими на обчислювальних машинах. Однак ті труднощі, які виникають при реалізації цієї ідеї, ставлять під сумнів можливість усього задуму. Метою даної статті є дослідження дискусії щодо природи розуміння, яка тривала в межах аналітичної філософії у другій половині ХХ ст. у зв'язку із розвитком комп'ютаціоналістських моделей свідомості.

Прояснювати природу розуміння у випадку штучних пристроїв першим взявся А. Тюрінг. У статті "Обчислювальні машини й інтелект" (1950) [7] він вважав, що позитивну відповідь на запитання "Чи може машина мислити?", можна отримати в результаті проходження цією машиною деякого тесту. Суть цього тесту полягала в такому: суддя листується природною мовою з двома співрозмовниками, один з яких є людиною, а інший – комп'ютером. Якщо суддя не може визначити, хто є людиною, а хто – комп'ютером, значить комп'ютер пройшов тест. Кожен з відповідачів судді прагне, щоб людиною назвали саме його. Щоб полегшити цей тест, листування зводиться до обміну текстовими повідомленнями, які здійснюються через контрольовані проміжки часу (в ті часи це було необхідно, тому що

комп'ютери думали набагато повільніше, ніж люди, сьогодні – навпаки). Серед можливих варіантів критики цього тесту сам А. Тюрінг називав такі: теологічні зауваження, небезпека штучного інтелекту, про яку можемо й не здогадуватися, обмеженість формалізації, відсутність свідомості у машин, здатностей переживати дійсність, писати осмислені, ліричні вірші, проявляти ініціативу, любити, дружити, обманювати, робити помилки, а також те, що розум діє на основі інших, не цифрових, принципів й поведінка людини не є завжди запрограмованою, а є неформальною, крім того, людина володіє екстра-сенсорними здібностями. Хоча, з іншого боку, А. Тюрінг був переконаним, що, якщо нам не вдасться у процесі листування (спілкування) розпізнати, де людина, а де машина, то ми просто змушені будемо визнати, що комп'ютер наділений інтелектом.

Одним із перших розкритикував концепцію штучного інтелекту – Г. Дрейфус у своїй книзі “Чого не можуть обчислювальні машини: критика штучного розуму” (1972). Він пропонував розглядати різні види розумової діяльності. Адже мислення саме по собі не є феноменом, однаковим у своїх принципах роботи. Зокрема, Г. Дрейфус виділяє такі чотири типи розумової діяльності [2, с. 264–266]: асоціативний, простий і складний формальні – їх ще можна запрограмувати і неформальний, який не піддається програмуванню. При неформальному мисленні смисл мовних виразів залежить від контексту ситуації, який є не одразу зрозумілим. Неформального мислення можна навчити лише на основі конкретних прикладів. Його застосовують у процесі розгадування загадок, де описано ситуацію в алегоричній формі, в задачах, де потрібно застосувати інтуїцію, де має прийти певне осяяння, здогад, інсайт, де потрібно перекласти текст з однієї мови на іншу, врахувавши при тому контекст повідомлення, де потрібно упізнати видозмінені, спотворені образи. Таку розумову діяльність, вважає Г. Дрейфус, неможливо запрограмувати.

Не менш відомим критиком штучного інтелекту в цілому й тесту Тюрінга зокрема був Дж. Серл, який 1980 року запропонував широковідомий ментальний експеримент “Китайська кімната”. Він був переконаний, що не доцільно вважати, що машина може мислити, допоки вона не буде розуміти ті процеси, що з нею відбуваються, допоки вона не буде ідентифікувати себе як деяку самість, особистість тощо. Дж. Серл писав: “Питання полягає не в тому, чи можуть машини мислити, чи може артефакт мислити, а в тому, чи може мислити цифровий комп'ютер” [4]. На думку Дж. Серла, тест Тюрінга не може бути достатнім критерієм осмисленості процесів комп'ютером. Щоб розуміти чому, насамперед потрібно уточнити, що означає термін “розуміння”. Що можна розуміти? Текст, ситуацію, речення, слово, контекст. Розуміти – означає вміти надавати значення, вміти приписувати певному синтаксису – послідовності символів (букв, картинок, звуків) відповідну семантику. Для Дж. Серла фраза “розуміти текст” означає вміти віднаходити не лише очевидну інформацію в тексті, але й таку, що є прихованою. Тобто коли ми щось розуміємо, то можемо його проаналізувати, вивести нове знання, дуже часто спираючись на знання нашого особистого досвіду про ті чи інші правила поведінки, культурні норми, юридичні закони тощо.

Дж. Серл розрізняв поняття сильного і слабого штучного інтелекту. Будь-які машини можуть мати лише слабкий штучний інтелект, який ґрунтується на обчисленні певних математичних операцій. Проте він виключає можливість, що запрограмований комп'ютер може володіти когнітивними станами, що він може з допомогою програм пояснити людську свідомість (сильний штучний інтелект). Для того, щоб довести власний погляд, Дж. Серл у статті “Розуми, мізки і програми” (1980) [6] пропонує ментальний експеримент “Китайська кімната”. Припустімо, що Дж. Серла закрили в кімнаті й дали рукопис китайською мовою, якої він взагалі не знає. Проте у нього є ще й другий текст, в якому можна знайти правила оперування символами. Цей текст він розуміє, оскільки ці правила написані англійською. Коли ж йому дають третій рукопис китайською мовою, то він знову нічого не розуміє. У третьому рукописі міститься питання до тексту, який є в першому рукописі. У другій книзі Дж. Серл для того, щоб відповісти на ці питання, може скористатися правилами, що ставлять відповідно до кожного питання правильну відповідь. (У звичному значенні слова ця множина правил є комп'ютерною програмою). Цю відповідь він і видає за межі кімнати. Таким чином, китайським науковцям, що знаходяться за межами цієї кімнати, видаватиметься, що Дж. Серл прекрасно розуміє китайську мову й веде осмислений діалог. Хоча сам Дж. Серл і далі нічого не знає китайською. Він просто вдало оперує із символами (синтаксисом), залишаючись інстанцією комп'ютерної програми.

Дж. Серл переконує, що у такий спосіб він описав діяльність комп'ютера, який здатний маніпулювати синтаксисом і не більше. Комп'ютер не розуміє, що відбувається. Хоча людям як користувачам комп'ютера і може видаватися, що він щось розуміє, адже відповідає на наші питання. Однак таке приписування розуміння комп'ютерам є не більше, ніж метафоричним. Таке розуміння позбавлене тієї інтенціональності, яка притаманна людському мисленню.

Для того, щоб попередити скепсис критиків, уже в тій самій статті “Розуми, мізки і програми” Дж. Серл детально відповідає, зокрема, на такі можливі варіанти заперечень:

– по-перше, можна зауважити, що якщо й людина не розуміє смислу текстових повідомлень, то їх може розуміти система як єдине ціле. У такому випадку розуміння притаманне не індивідові, а цілій системі, одним із елементів якої є індивід. Це зауваження Дж. Серл спростовує так: можна припустити,

що індивід вивчив усі правила з другого рукопису і вміє швидко віднаходити правильні варіанти відповідей на поставлені питання, проте це зовсім не означає, що він розумітиме про що в них йдеться. У випадку з англійською мовою він розумітиме діалог, у випадку з китайською – ні, хоча з боку слухача все виглядатиме абсолютно однаково в обох випадках;

– по-друге, навіть якщо ми приписуватимемо інтелект не лише комп'ютеру, а роботу, що здатний пересуватися в просторі й реагувати на появу предметів у полі зору своїх візуальних датчиків, це все одно нічого не змінить, оскільки його принципи роботи все одно ґрунтуватимуться на основі маніпуляцій із синтаксисом. Схоже цей варіант відповіді на зауваження критиків просто не помітив (чи не хотів помітити) В. Васильєв, який у статті "Кока-кола і секрет Китайської кімнати" пише таке: "Припустимо, що перед експериментом я зайшов у кімнату і поставив банку кока-коли на стіл, за яким буде сидіти Серл. І ось ми починаємо наш діалог. Я запитую Серла китайською: "Скажіть, що знаходиться прямо перед Вами?". Очевидно, що він має звернутися до програми, яка повинна містити відповідь: "Банка кока-коли". Але як людина, що склала цю програму, могла знати, що я принесу в китайську кімнату саме цей предмет?" [1]. На нашу думку, у випадку з роботом можна сказати, що в другому рукописі має знаходитися програма для розпізнавання предметів. Зрозуміло, що у випадку з експериментом "Китайська кімната" Дж. Серл переслідував мету – показати, що сильний штучний інтелект є неможливим, оскільки маніпуляції із синтаксисом не можуть призвести до появи семантики, до виникнення розуміння. Для нього було важливо пояснити сам принцип, а не спростувати одним прикладом всі можливі варіанти критики;

– по-третє, що буде, якщо нам вдасться повністю змоделювати діяльність нейронів у корі головного мозку, коли наш робот функціонуватиме на основі тих принципів, які є в нашому мозку, паралельно опрацьовуючи дані від різних систем. Але й у такому випадку Дж. Серл вважає, що це нічого не змінює: "Оскільки моделюється тільки формальна структура послідовності нейронних збуджень у синапсах, то не моделюється та сторона функціонування мозку, яка якраз і має значення, а саме: каузальні властивості мозку, його здатність продукувати інтенціональні стани" [6, с. 421]. Таким чином, на думку Дж. Серла, без інтенціональності не може бути розуміння;

– по-четверте, розуміння не виникне навіть коли ми спробуємо одразу поєднати усі попередні варіанти зауважень в єдине ціле. Робот, поведінка якого не буде відрізнятися від нашої, все одно, вважає Дж. Серл, не матиме інтенціональності. В цьому варіанті зауважень Дж. Серл наче випереджає майбутні дискусії щодо філософських зомбі – істот, поведінка яких аналогічна людській, проте вони позбавлені свідомості і т. д.

Таким чином Дж. Серл здійснює критику комп'ютаціоналізму у зв'язку з тим, що жодні обчислювальні маніпуляції не можуть призвести до виникнення інтенціональних процесів, які є виключно результатом нейробіологічної діяльності кори головного мозку. Подібно до того, як ми не можемо ствердити, що комп'ютерна імітація пожежі спричинить пожежу чи, що комп'ютерна імітація дощу, змусить нас промокнути, так само ми не можемо сказати, що імітація комп'ютером мисленнєвих процесів людини змусить комп'ютери мислити аналогічно до того, як мислить людина. Дж. Серл відстоює позицію біологічного натуралізму: "Чим би ще не була інтенціональність, вона біологічний феномен, і її буття настільки ж імовірно, наскільки воно каузально залежне від таких конкретних біохімічних особливостей її походження як лактація, фотосинтез і інші біологічні феномени" [6, с. 424].

Пол М. Черчленд і Патриція С. Черчленд у статті "Чи може машина мислити?" (1990) [5], загалом позитивно оцінюючи заслуги Дж. Серла в питанні дослідження філософських основ проблематики штучного інтелекту, все ж таки не погодилися із тими висновками, які запропонував Дж. Серл. Усі його міркування в експерименті "Китайська кімната", на їх думку, зводяться до чотирьох аксіом:

1. Комп'ютерні програми – це формальні (синтаксичні) об'єкти.
2. Розум людини має смислову компоненту (семантику).
3. Синтаксис сам по собі не є сутністю семантики і його недостатньо для існування семантики.
4. Програми не є сутністю розуму, їх наявності не достатньо для існування розуму.

Незадовільною, на їх думку, є саме 3-тя аксіома. Для того, щоб спростувати це твердження, вони вирішили застосувати метод аналогії й запропонувати такий набір аксіом:

1. Електрика і магнетизм – це фізичні сили.
2. Суттєвою властивістю світла є світіння.
3. Сили самі по собі не є сутністю ефекту світіння і не достатні для його наявності.
4. Електрика і магнетизм не є сутністю світла і не достатні для його наявності.

Цей логічний умовивід Черчлендів до появи робіт Максвелла міг би виявитися вагомим аргументом на користь того, що світло й електромагнітні хвилі не є ідентичними. Однак насправді коливання магнітних хвиль і породжують світло. Черчленди доходять висновку: "Аналогічно маніпулювання символами у відповідності з визначеними правилами може являти собою розум, хоча у системи, що знаходиться в "китайській кімнаті" Дж. Серла й ґрунтується на застосуванні правил, справжнє розуміння начебто й відсутнє" [5]. Іншими словами, на їх думку, є цілком допустимим той факт, що маніпуляції із

синтаксисом на основі певних правил можуть зумовити появу семантичних властивостей. Наразі відсутність видимих ознак семантики може пояснюватися тим, що функціональна архітектура обчислювальних машин є кардинально іншою, ніж нейрофізіологічна структура мозку. Комп'ютери працюють з допомогою процесора, який послідовно обробляє сигнали один за одним. Робота мозку відрізняється від роботи комп'ютера, оскільки в мозку паралельно (одночасно) опрацьовуються сигнали з допомогою мільйонів різних нейронних мереж. Нейрони виступають елементарними "процесорами", характер дії яких, якщо можна так зазначити, є не цифровим, а аналоговим, адже може безперервно змінюватися. Крім того, взаємодія між нейронами відбувається й у зворотному напрямку. Мозок є динамічною системою, на відміну від статичних комп'ютерів. Таким чином, Черчленди набагато кращою моделлю штучного інтелекту, що пояснює процес роботи мозку, вважають не комп'ютаціоналізм, а коннективізм, що прагне унаслідувати основні принципи функціонування нейронних мереж мозку.

У відповідь Дж. Серл у статті "Розум мозку – комп'ютерна програма?" (1990) зауважує, що "... П. М. і П. С. Черчленди у своїй статті ("*Чи може машина мислити?*" – А. С.), як мені видається, не зовсім правильно зрозуміли суть моїх аргументів. На їхню думку, сильний штучний інтелект стверджує, що комп'ютери в кінцевому підсумку можуть набути здатності до мислення і що я заперечую таку можливість, розмірковуючи лише на рівні здорового глузду. Однак сильний штучний інтелект стверджує інше, і мої доведення проти цього не мають нічого спільного зі здоровим глуздом" [3]. Його висновки, що стосуються неможливості породження семантики в результаті маніпуляцій із синтаксисом, як він зазначає, стосуються не лише послідовних машин, але й паралельних, хоча їх функціонування є поки-що маловивченим. Але це нічого не змінює, оскільки будь-які функції, що обчислюються на паралельних машинах, можна обчислити й на послідовних. В основі паралельних машин лежать такі ж функціональні процеси, як і в основі послідовних машин. Ці процеси прекрасно моделюють роботу нейронів, хоча їх фізико-хімічна природа суттєво відрізняється. Проте мозок породжує й специфічні нейробіологічні процеси, які не можна так просто змодельовати на інших носіях з допомогою програм. Тому приклад, який пропонують Черчленди, є незадовільним, оскільки порівнюються речі зовсім різної природи: "Аналогія з формальними символами недоречна, оскільки формальні символи не мають фізичних причинних властивостей" [3]. Єдині, хто розумів, що відбувається в китайській кімнаті, були програмісти, які і запрограмували весь процес її роботи. Але їм нічого не заважало тим же символам надати інших значень. Наприклад, усі дії Дж. Серла в середині кімнати можна було б проінтерпретувати як гру в шахи, як передбачення курсу акцій на валютній біржі, як якусь картярську гру і т. д. Кількість цих семантик є безкінечною, тому що синтаксис сам по собі не містить внутрішньої семантики.

Хоча Дж. Серл і був скептично налаштованим щодо перспектив коннективістських моделей свідомості, починаючи з 1980-х рр., увага до них різко зростає. Так, уже Черчленди вказували на низку переваг коннективістських моделей. По-перше, вони можуть швидше обробити інформацію, оскільки цей процес відбувається паралельно в різних нейронних мережах. По-друге, навіть якщо в одній із паралельних систем сталася помилка, то в цілому система може й надалі продовжувати свою діяльність. Вона без проблем може запам'ятовувати велику кількість інформації та за потребою розподіляти її у своїх структурних елементах. Звісно, коннективістська система програє у швидкості комп'ютаціоналістській у тому випадку, коли потрібно провести якісь обчислення, що вимагають мільйонів тих самих повторень. Проте коннективістська система швидше може розпізнавати контури предметів, ідентифікувати розмову, що супроводжується сторонніми шумами, і т. д.

Загалом відмінність коннективістського і комп'ютаціоналістського підходів полягає у тому, що перший намагається цілковито імітувати діяльність людського мозку, в той час як другий у своїй діяльності використовує інші принципи. Коннективісти прагнуть навчати систему шляхом утворення нових зв'язків між штучними нейронними мережами (системами, що складаються з певної кількості пристроїв – штучних нейронів, з'єднаних між собою), в той час як комп'ютаціоналісти мають справу виключно із маніпуляцією символами. Коннективісти особливу увагу приділяють процесу самонавчання машин, пошуку можливих шляхів розв'язання проблеми (утворення нових нейронних зв'язків), за умови, що відомо дані входу й правильні відповіді. В комп'ютаціоналістських моделях, навпаки, прагнуть до дрібниць з допомогою заданих алгоритмів продумати всі можливі варіанти відповіді комп'ютера. Таким чином, бачимо значну перевагу саме коннективістських моделей штучного інтелекту, які набагато краще наслідують діяльність людського мозку. Хоча основні принципи цієї діяльності й до сьогодні не до кінця зрозумілі. Але принаймні, як і у випадку комп'ютаціоналізму, ці системи можуть використовувати основні принципи оперування поняттями, логічно виводити одне знання з іншого тощо.

Саме в межах коннективістського підходу до побудови систем штучного інтелекту, на нашу думку, і може бути вирішеною проблема розуміння природної мови. Таке розуміння проявляється в ситуації реагування на фрази людини, в розпізнаванні семантичних і прагматичних особливостей мовлення, в здатності здійснювати переклад з однієї мови на іншу. Крім того, такі системи повинні уміти навчатися.

Щоправда можливості і перспективи теорії штучного інтелекту і в цьому випадку не переконують скептиків у питанні того, що машини коли-небудь зможуть розуміти так, як розуміють люди. Однак

питання, чи є розуміння виключно прерогативою організмів, фізико-хімічна природа яких є такою як у людини, а чи розуміння як функція людського мозку, може бути реалізоване на інших, штучних, носіях, наразі залишається відкритим. Проте сама ідея створення штучного інтелекту є дуже цікавою й досі надзвичайно перспективною в науці і філософії.

Література:

1. Васильев В. В. Кока-кола и секрет Китайской комнаты [Электронный ресурс]. – Режим доступа : http://ecsosman.hse.ru/data/2010/04/23/1213594898/Vasilev_Koka-kola.pdf.
2. Дрейфус Х. Чего не могут вычислительные машины: Критика искусственного разума / Хьюберт Дрейфус. – М. : Прогресс, 1978. – 334 с.
3. Сирл Дж. Разум мозга – компьютерная программа? Джон Сирл [Электронный ресурс]. – Режим доступа : <http://www.raai.org/library/books/sirl/ai.htm>.
4. Сёрль Дж. Сознание, мозг и наука / Джон Сёрль [Электронный ресурс]. – Режим доступа : http://ecsosman.hse.ru/data/2010/05/20/1214101925/002_Dzhon_Serl_Soznaniex2c_mozg_i_nauka_03-66.pdf.
5. Черчленд П. М. Может ли машина мыслить? / П. М. Черчленд, П. С. Черчленд [Электронный ресурс]. – Режим доступа : <http://www.raai.org/library/books/sirl/ai.htm>.
6. Searle J. R. Minds, brains and programs / J. R. Searle // Behavioral and Brain Sciences. – 1980. – № 3. – P. 417–424.
7. Turing A. M. Computing machinery and intelligence / A. M. Turing // Mind. – 1950. – № 59. – P. 433–460.

Рекомендовано до друку рішенням кафедри гуманітарних дисциплін Львівського державного університету фізичної культури від 9 липня 2013 р., протокол № 11