

ВОЛЯ ЧИ РОЗУМ, АБО ЩО ТАКЕ ШТУЧНИЙ ІНТЕЛЕКТ?

Олександр БІЛОКОБИЛЬСЬКИЙ

Copyright © 2013

До сторіччя Алана Тюрінга присвячується

Сторіччя з дня народження Алана Тюрінга, що було у центрі уваги світової наукової спільноти протягом усього 2012 року (проведено тисячі заходів у сотнях міст різних країн, зокрема й в Україні, у Донецькому національному технічному університеті), стало гарною нагодою для того, щоб ще раз звернутися до обговорення перспектив створення *штучної свідомості*, а також сучасного стану розуміння природи *людського інтелекту*. Тим більш, що, незважаючи на багаторічну традицію вивчення розуму, свідомості, мислення, розуміння і т. ін., в сучасній філософії ці теми залишаються найбільш дискусійними. З тестом Тюрінга, спрямованим на виявлення інтелектуальності обчислюваних машин, саме й пов'язаний один з авторитетних поглядів на інтелект як на маніпулювання символами.

Наука про штучний інтелект формувалася у форматі тенденції до прагматизації теоретичного знання, його центрування на вирішенні практичних завдань. Незважаючи на те, що розробки у сфері штучного розуму проводилися з урахуванням філософського спадку, саме практичні чинники зумовили шляхи розвитку нової науки. Починаючи із 50-х років ХХ століття, створення *штучного інтелекту* (ШІ) проводилося у двох напрямках: 1) з допомогою моделювання процесів мислення, сутність якого вбачалася у здатності до логічних міркувань, та 2) шляхом штучного відтво-



**Алан Метісон ТЮРІНГ
(1912–1954)**

рення структур людського мозку. Обидва підходи спиралися на досить спрощенні філософські узагальнення, які, проте, відштовхувалися від авторитетних метафізичних теорій. У праці “Свідомість vs моделювання мозку: ШІ повернувся до точки розгалуження” (1990) Х.Л. Дрейфус та С.І. Дрейфус показали, що “символьна” інтерпретація інтелекту наслідувала класичну філософську традицію, яка ототожнювала розуміння певної предметної сфери із побудовою її теорії. З цього погляду і людське мислення, й комп’ютерна

програма, репрезентують світ через створення символічних моделей, а тому можуть бути описані за допомогою одного інструментарію. Свій досконалий вигляд ця теоретична позиція отримала у так званій гіпотезі Ньюела-Саймона: “фізична символічна система містить необхідні та достатні засоби для проведення базових інтелектуальних дій, у широкому розумінні”. Інтелект тут фактично ототожнюється із системою (хоча й досить складною) фізичних символів.

Інший підхід, що конкурував із символічним, спирався на розробки Ф. Розенблатта. Його послідовники вбачали у штучному інтелекті модель системи нейронних зв’язків людського мозку. Перші результати в цьому напрямку були отримані ще У. Маккалоком та У. Піттсом у 1943 році. Створений наприкінці 50-х перцептрон у першому наближенні відтворював фізіологічні функції мозку. Проте

варто зазначити, що моделювання мозкової діяльності задля досягнення “інтелектуальності” машини, подібно до символічного підходу, абсолютизувало метафізику природничого позитивізму.

Історія конкурентної боротьби між двома зазначеними парадигмами ШІ стала прекрасною ілюстрацією до праці Т. Куна “Структура наукових революцій”, що з’явилася саме у той час (кінець 60-х): абсолютна перемога (у 70-і) “символістів” базувалася не стільки на об’єктивних перевагах їхнього методу, скільки на вдало організованій пропагандистській кампанії. Перипетії агонії за кошти, ресурси і лабораторії прекрасно відображені в уже цитованій статті Дрейфусів. Цікаво інше – ні у рамках “символьного”, ані “нейронного” підходів (останній, знову ж на підтвердження теорії Куна, але ще більше П. Фойерабенда, який довів, що поразка однієї з наукових парадигм ніколи не може вважатися остаточною, взяв блискучий реванш у 80-ті), публічно дану обіцянку створити ШІ до початку 80-х не було виконано навіть частково. Плани були зірвані, незважаючи на фантастичний прогрес обчислювальної потужності комп’ютерів, швидкодію й обсяг пам’яті яких перевершив усі доступні природному інтелекту показники.

Філософське підґрунтя проблеми багатьом було зрозуміле вже в ті роки. Симптоматичною у цьому зв’язку є критика Дж. Сьорля, який у статті “Свідомість, мозок і програми” сформулював свій знаменитий уявний експеримент, що отримав назву “Китайської кімнати”. Пафос полемічного випадку Сьорля був спрямований проти процедур визначення наявності у комп’ютера інтелекту, що впливали з тесту Тюрінга. Відповідно до запропонованого Аланом Тюрінгом алгоритму перевірки “інтелектуальності” машини, людині пропонується сформулювати ланцюг довільних питань, отримавши відповідь на які (але не маючи можливості бачити “співрозмовника”, який перебуває за стіною), вона повинна визначити, хто є їх автором – машина чи людина. Комп’ютер проходить тест на інтелектуальність у тому разі, якщо його “відповіді” прийняті за людські.

Незважаючи на теоретичну витонченість тесту Тюрінга і той факт, що змагання між комп’ютерами у їх здатності проходити тест проводяться протягом більш, ніж 20 років, та демонструють вельми переконливі результати¹,

Сьорль вказав на його принципову недосконалість. З допомогою образу “китайської кімнати” він зміг довести нетотожність процесів оперування символами і мислення, пов’язаного з розумінням. У короткому викладі ситуацію можна змалювати таким чином. Уявімо собі людину, яка замкнена в кімнаті і має написані її рідною мовою вказівки до зіставлення певних речень китайської (якої вона не знає!) мови. Отримуючи ззовні одну з китайських фраз, людина знаходить з допомогою вказівок потрібне зіставлення і передає його за межі кімнати. Зовнішній спостерігач – уявімо, що це китаєць, – задавши осмислене питання китайською (перша фраза), отримує осмислену відповідь тією ж мовою (друга фраза) і переконується в “розумності” того, хто, перебуваючи всередині кімнати, “відповів” йому. При цьому замкнена у кімнаті людина, не маючи ані найменшого уявлення про смислову наповненість питання та відповіді, виконує лише механічну роботу. Якщо ототожнити “китайську кімнату” з комп’ютером, вказівки та маніпуляції з ним – із комп’ютерною програмою, а обмін питаннями і відповідями – з тестом Тюрінга, то стає зрозумілим, що говорити про інтелектуальність машини, яка проходила тест, передчасно. Послідовно критикуючи різні варіанти критики його “китайської кімнати” з боку кібернетиків (причому прихильників як “символьного”, так і “нейронного” підходів до ШІ), Дж. Сьорль показує, що розуміння як центральна складова мислення у машини повністю відсутнє [5, с. 382]. Втім, і його висновок про залежність “дійсних феноменів людської свідомості” від “дійсних фізико-хімічних властивостей” “людського мозку” в тій категоричній формі, яку надає йому автор [5, с. 398], також не є переконливим. Зрештою, “дійсний мозок” зовсім не обов’язково “виробляє” мислення і свідомість, понад те, продукує його після тривалого і специфічного соціально “налаштування”.

Роблячи спробу аналізу тих інтуїтивних уявлень про людський розум і мислення, від яких (можливо несвідомо) відштовхувалися творці перших моделей ШІ, легко помітити, що вони були схильні бачити в розумі статичну структуру, перехід між різними ієрархічними рівнями якої був повністю детермінований її внутрішніми зв’язками. З одного боку, такі уявлення цілком узгоджуються зі схемою дії

¹ Див., напр., сайт конкурсу на приз Г. Льобнера (<http://www.loebner.net/Prize/loebner-prize.html>).

обчислювальних пристроїв, а з іншого – усядишню висхідну до античності вченню про розум як ієрархію ідей. Однак задамося питанням: чи є людський розум (у всьому різноманітті своїх проявів – пізнанні, мисленні, свідомості, розумінні) чимось пасивним і статичним? Чи не йдеться про різні варіанти діяльності, спрямованість якої тільки частково визначається характером форм раціоналізування (у смислі набуття розуморозмірності) і значною мірою залежить від до- (а може й поза-) раціональних імпульсів?

Аристотель, арабські філософи, а слідом за ними і середньовічні схоласти, намагалися вирішити проблему поєднання статичності раціональних форм та активності розуму вченню про “діючий інтелект”, який немов би наповнює світлом космічного розуму розум індивідуальний. Пізніше, у XIII столітті, ця проблема трансформувалася в дискусію навколо питання про перевагу волі над розумом. Хоча предмет дискусії не був для християнської філософії новим – першість волі перед розумом обстоював ще Аврелій Августин, усе ж саме у XIII – на початку XIV суперечка набула нового загострення, що було спричинено як розвитком традиції природної теології, в лоні якої зароджувалася новочасна наука, так і все більш категоричним протиставленням істин віри істинам розуму (саме у їх такому порядку!). Аргументація захисників переваги волі над розумом – а їх ряди були вельми авторитетні, досить назвати імена Бонавентури, Петра Оліві, Джона Пекама, Дунса Скота, Вільяма Оккама, Петра Ломбардського, Ніколи з Отрекура² – не втратила своєї вагомості до сьогоднішнього дня. Зокрема, Дунс Скот указував на контингентність предмета божественного розуму (уявлення якого можна вважати своєрідною “заготовкою” творіння), оскільки етична та онтологічна значимість цього предмета задається божественною волею через актуалізацію одного з “поєднань” ідей у бутті і добрі³. Божественна воля є підставою дії божественного розуму [1, с. 233], і Бог може створити щось, що прямо суперечить розуму: “зробити одне і те ж тіло розташова-

ним у багатьох місцях, так само як... створити два тіла в одному і тому ж місці” [1, с. 267]. До речі, зі свого розуміння субординації волі і розуму Дунс Скот зробив цікаві висновки на користь інтуїтивного (вважай – досвідного) пізнання світу, що мали першорядне значення для розвитку емпіричної науки.

Треба вказати на те, що християнська теологія періоду схоластики формувалася як спроба граничного раціонального мовлення про Бога та створений світ. З часу написання Ареопагітик (відомих на Заході із IX століття у латинському перекладі Іоанна Скота Еріугени) і затвердження традиції апофатичного богослов'я теологічні істини не сприймалися як прямі характеристики божественної сутності. Це в кращому випадку були аналогії, хоча і побудовані із залученням усіх доступних на той момент раціональних засобів.

“Конвертувавши” раціональну *граничність* середньовіччя в сучасну, отримаємо уявлення про той культурний розум, який може (і повинен) бути еталоном при моделюванні розуму штучного. Як і багато століть тому, наш розум припускає потенційне існування багатьох варіантів “контингентних сполучень речей”, про які (у зв'язку з Божественним розумом) говорив Дунс Скот. Однак, якщо, згідно із середньовічними уявленнями, розум людини був приречений на сприйняття обраного Богом кращого зі світів, то в більш-менш вільному від теологічних впливів сучасному розумі приреченість на сприйняття саме “нашого” світу вимагає пояснення. Сьогодні ми знаємо про *альтернативні смислові універсуми*, які можуть спиратися на схожі чуттєві дані, але пов'язані з різними їх інтерпретаціями й, таким чином, різними культурними реальностями. Причому йдеться не стільки про “спрошення” об'єктивної реальності, скільки про різні її розуміння-інтерпретації (поза якими реальність перетворюється на штучну розумову конструкцію), тобто про різні “об'єктивні” реальності! Конкуруючі реальності виключають одна одну в актуальному сприйнятті, а наша “вірність” науковій картині світу, а не, припустимо, середньовічному символічному

² Див., напр., книгу П. Гайдена [2].

³ Я кажу, що божественний розум пропонує своїй волі певне сполучення (контингентних речей), що має нейтральний статус та не сприймається як істинне чи хибне..., і коли воно стверджується у бутті та дії через волю як визначене в той чи інший бік, тоді й сприймається як істинне, хоча до того пропонувалося волі як нейтральне [1, с. 231].

універсуму або міфологічній реальності, впорядкованій на основі тотемічних класифікацій, ґрунтується на легітимних у межах наукового світобачення методах (історичних, природничих і т.д.) взаємодії зі світом.

Сказане можна проілюструвати на такому прикладі. Уявімо, що перед нами малюнок, на якому зображено чи то кролика, чи то качку (так звану бістабільну фігуру), що відомий з позаминулого століття і що наочно демонструє активну природу нашого сприйняття. Залежно від *обраної* точки зору чи позиції, людина бачить качку або кролика. Для того щоб зрозуміти проблематичність феноменального рівня як основи нашого пізнання, задамося питанням про те, що ж саме – дзьоб або вуха, ми бачимо, дивлячись на верхню частину малюнка? Скільки б ми не вдивлялися в деталі, визначити смислове наповнення видимих ліній (у даному разі схематичність малюнка не має визначального статусу) можемо тільки виходячи з цілого. Екстраполюючи це положення у сферу досвіду в цілому, отримуємо парадоксальний результат: позавольової фіксації на певному варіанті контингентного поєднання речей *розум просто не мав би власного предмета!* Інша справа, що наша “схильність” до бачення окремого цілого настільки однозначна, що поза спеціальним акцентуванням уваги не сприймається як вольовий акт.

Завісу над таємницею “божественної” зумовленості людського сприйняття піднімають досягнення епістемології і когнітивної психології. До початку 80-х років (якраз до моменту очевидної кризи “символьної” і “перцептронної” парадигм створення ШІ) цікаві результати щодо вивчення людського мислення були отримані у психології та філософії. Сказане стосується зарубіжних авторів, таких як Ж. Піаже, Д. Брунер, У. Найссер, а також радянських психологів, що спиралися на праці Л.С. Виготського – О.Р. Лурія, О.К. Тихомирова та ін. Крім того, починаючи з 60-х активно розвивається (передусім у працях О.Р. Лурія, К. Прібрама) напрямок нейрофізіологічних досліджень мозкової діяльності.

Вказуючи на те, що мислення є різновидом цілеспрямованої [4, с. 310] людської діяльності [7, с. 5], дослідники-психологи зуміли вирізнити різні етапи розумового акту [напр., 4, с. 310–311], стадії становлення мислення у процесі онтогенезу (Ж. Піаже), а також відкрити соціальну природу підстав мислення

і механізми привласнення цих підстав у процесі соціалізації (Дж. Брунер). Закріплення легітимної інтерпретації реальності відбувається у процесі освоєння дитиною санкціонованих суспільством вимог (ретрансляторами яких щонайперше є вихователі), пов’язаних, зі свого боку, з панівним світоглядом. Звичайно, психологічне розуміння мислення як соціально зумовленої діяльності не може повністю взяти на себе функції теологічного імперативу про детермінованість розуму божественною волею. Проте, хоча в науці (зокрема, у психології), можна говорити про процеси, які підлягають емпіричній фіксації (скажімо, такі, як процеси “навчання” мисленню), що завжди пов’язані з певним *передзнанням* реальності як цілого, котре транслюється в культурі від покоління до покоління. Природа цього передзнання (тобто причини та способи виділення з нескінченності можливих світів кращого, або просто “нашого”) піддається науковій тематизації значно меншою мірою. Однак це і не так важливо. Досить і того, що, спираючись на сказане, стверджуємо, що наше мислення може здійснюватися тільки на засадах некритично засвоєних когнітивних схем, які передують нашому досвіду як “шаблон”, що “заповнюється” емпіричною інформацією та чуттєво збагачує даний зміст. Повертаючись до вищенаведеного наявного малюнку, висновуємо таке: ми *бачимо* вуха саме через те, що *очікуємо* зустрічі з кроликом (інакше йшлося б про дзьоб та качку); отож *уже на етапі сприйняття (одержання необхідного для мислення матеріалу) наш розум потребує вольового обмеження потоку інформації відповідно до певної дораціональної схеми.*

Будучи цілеспрямованою, соціально спричиненою (пов’язаною з певним актуальним завданням, що виходить за межі вроджених або звичних рішень) діяльністю [4, с. 310], мислення здійснюється на основі детермінуючих сприйняття *схем реальності* і в напрямку імовірного *вирішення проблемної ситуації*. Через подвійну позараціональність (по-перше, пов’язану із процесом здолання проблеми, у якому мислення здебільшого постає лише одним з можливих інструментів, по-друге, – з принциповою відкритістю раціоналізування в межах ніяк необґрунтованої когнітивної схеми), раціональні побудови людини у процесі перебігу мислення спираються на її емоції. Останні дозволяють відмовитися від послідовного розгляду всіх можливих, із

позицій розуму, шляхів розвитку ситуації і виконують, на думку О.К. Тихомирова, функцію евристик (логічних процедур оптимізації способів вирішення різних завдань) [7, с. 220]. Близькі погляди на роль емоцій висловлені в концепції П. Симонова, який бачив у них “компенсаторний механізм, котрий заповнює дефіцит інформації” [6, с. 36], що значуща для розв’язання задачі. Тут мисляча людина постає вже не просто як діюча, а й як така, котра підкорює своєму вольовому імпульсу обставини, загрозові для її успішності у вирішенні поставлених життєвими проблемами⁴.

Первинність вольового начала перед власне раціональним у мисленні знайшла відображення у так званій “теорії драфтів” Данієля Деннета, котрий багато років очолює Центр когнітивних досліджень Університету Тафтса. Згідно з його концепцією, людський мозок не просто сприймає потоки інформації із зовнішнього світу, але й виробляє своєрідні “начерки” (драфт походить від англ. – drafts) [8, с. 127], що вміщують отриману інформацію як смисловий контекст. Драфти генеруються різними частинами мозку в різних поєднаннях і надлишковій кількості: “... протягом миті накладається щось подібне до безлічі мереж, які, своєю чергою, стають об’єктами безперервного редагування з боку інших процесів мозку... Важливо те, що вектор цих процесів спрямований у невизначене майбутнє, а їх еволюційний смисл полягає у підтримці інформаційного контакту з інформацією, що надходить зсередини організму та із зовнішнього середовища задля передбачення її можливих змін і підтримання в готовності адаптаційних механізмів... [8, с. 128].

Модель сприйняття-мислення, яка пропонується Д. Деннетом, враховує природні та соціальні передумови людського мислення і перетворює розум на заручника біологічно детермінованої волі. Це дає підстави автору говорити про те, що “природа зайнята будівництвом не епістемічних машин, а машин для виживання” [8, с. 130], що принципово відрізняє природний інтелект від штучного. Звичайно, обговорюючи цілі природи, ми завжди ризикуємо відхилитися від дослідницького шляху в сферу метафор, однак деннетівське визнання першості вольового початку-

джерела мислення над пізнавально-розсудливим вельми показово. Доречно взагалі задатися питанням, а чи точно міркування та виконання логічних операцій становлять сутність розуму як такого? Враховуючи, що мислення – це діяльність, яка вимагає вольового зусилля (на відміну, скажімо, від інстинкту) та пов’язана з необхідністю соціально зумовленого підтвердження наявності деякої реальності (що є сферою значень, задіяних у логічному мисленні понять), а також той факт, що мислення завжди співвіднесене із соціальною ситуацією (завданнями та їх рішеннями, контекстом, соціальною ідентичністю мислячого суб’єкта), відповідь на це питання не є настільки вже однозначною.

Понад те, самі поняття, крім власної відносності (у сенсі вкоріненості в одному з можливих варіантів реальності), зовсім не вичерпують того, що є базисом нашого мислення. Емпіричне наповнення загальних (та й одиничних!) понять релятивізовано до досвіду кожного конкретного суб’єкта і в кінцевому підсумку спирається на те, що можна слідом за Х.-У. Гумбрехтом назвати тілесною “присутністю”. Дійсно, один із центральних елементів мислення – розуміння – неподільно пов’язаний не тільки із “вписуванням” деяких образів (чуттєвих чи ментальних) у наявні когнітивні схеми, а й з редукцією семантичного наповнення форм думки до інтуїтивно очевидних, звичних станів суб’єкта, що знаходить відображення у переживанні “задоволення”, яке супроводжує момент усвідомлення чогонебудь. “Звичні стани” в даному випадку пов’язані з набутими раніше, в основному в дитинстві, навичками у поведженні з об’єктами (як матеріальними, так і абстрактними), що дозволяє знову говорити про позараціональні підстави раціональної діяльності.

Отже, ще раз підкреслимо, що шлях моделювання інтелектуальної діяльності, намічений понад півстоліття тому і пов’язаний з уявленнями про людський розум як про логічний автомат, який перетворює однозначні інформаційні сигнали в однозначні відповіді – саме з цієї традицією органічно пов’язаний тест Тюрінга – хибний не в деталях, а по суті. Мислення як реалізація розуму – це першочергово процес визначення волі до: а) розв’язання

⁴ Цікаво, що отримані висновки узгоджуються з тією картиною розвитку наукового знання, яка склалася у філософії науки: боротьба між конкуруючими “парадигмами”, “теоріями”, “науково-дослідними програмами” ніколи не спирається лише на панівні в науці критерії раціональності, норми і цінності.

певної задачі, б) “нав’язування” зовнішньому світу “зручної” (соціально легітимованої) когнітивної схеми, яка моделює онтологічну реальність як семантичне поле мислення і діяльності, в) здійснення мисленнєвої роботи не завдяки орієнтації на істинний зразок, а на соціальні настановлення, які в загальному плані можна звести до імперативу утвердження власного Я (причому не тільки у штірнерівському, егоїстичному, сенсі, але і з погляду легітимації соціального суб’єкта як представника певної суспільної групи). Відтак і штучний інтелект, щоб стати подібним інтелекту природному, повинен створюватися не як сукупність логічних операцій над готовими поняттями (ідеями, символами і т. ін.), але як відкрита (в аспекті неможливості остаточної верифікації понять і висновків) система соціально легітимних вольових дій, здійснюваних у напрямку реалізації особою певних соціальних стратегій.

1. Блаженный Иоанн Дунс Скот. Избранное. – М., 2001. – 584 с.

2. Гайденко П.П. Научная рациональность и философский разум. – М., 2003. – 528 с.

3. Дрейфус Х.Л., Дрейфус С.И. Создание сознания vs. моделирование мозга: искусственный интеллект вернулся на точку ветвления // Аналитическая философия: становление и развитие. – М., 1998. – С. 401–432.

4. Лурья А.Р. Основы нейропсихологии. – М.: Изд-во МГУ, 1973. – 376 с.

5. Сёрль Джон. Сознание, мозг и программы // Аналитическая философия: становление и развитие. Антология / общ. ред. и сост. А.Ф. Грязнова. – М., Дом интеллектуальной книги, 1998. – С. 276–400.

6. Симонов П. В. Что такое эмоция? – М., 1966. – С. 36

7. Тихомиров О.К. Структура мыслительной деятельности человека (опыт теоретического и экспериментального исследования). – М., Изд.-во МГУ, 1969. – 304 с.

8. Юлина Н.С. Головоломки проблемы сознания: концепция Дэниела Деннета – М.: Канон+, 2004. – 544 с.

АНОТАЦІЯ

Білокобильський Олександр Володимирович.

Воля чи розум, або що таке штучний інтелект?

Важливість філософського обґрунтування будь-яких досліджень мислення та свідомості до сьогодення для багатьох залишаються “абстрактною” істиною. У статті аргументується зв’язок процесів моделювання свідомості та тих дотеоретичних настановлень дослідників, які можуть обговорюватися тільки з допомогою мовних засобів метафізики. Через те суперечки навколо штучної свідомості слушно розглядати як певне продовження класичних філософських дискусій.

АННОТАЦИЯ

Белокобыльский Александр Владимирович.

Воля или разум, или что такое искусственный интеллект?

Важность философского обоснования любых исследований мышления и сознания до сегодняшнего дня многих остается “абстрактной” истиной. В статье показана зависимость процессов моделирования мышления от тех дотеоретических установок исследователя, которые могут обсуждаться только с помощью языковых средств метафизики. Тем самым споры о путях построения искусственного интеллекта целесообразно рассматривать как продолжение классических философских дискуссий.

ANOTATION

Bilokobylskyi Oleksandr.

Will or Mind, or What is an Artificial Intelligence?

The importance of philosophical substantiation of any researches of thinking and consciousness still are “abstract” truth for many scientists. In the article the connection of the processes of simulation of consciousness and those pre-theoretical instructions of researches, which can be discussed only with the help of verbal means of metaphysics, has been discussed. For this reason it is correctly to consider the arguments about artificial consciousness as some continuation of classical philosophical discussions..

Надійшла до редакції 4.10.2012.