

УДК 1:004.383.8

**Мороз О. Я.**

*Кандидат філософських наук, старший науковий співробітник відділу філософських проблем природознавства та екології Інституту філософії імені Г. С. Сковороди НАН України*

## **ШТУЧНИЙ ІНТЕЛЕКТ VERSUS ПРИРОДНИЙ ІНТЕЛЕКТ (МАЙБУТНЄ ЛЮДИНИ В КОНТЕКСТІ ВИКЛИКІВ ІНТЕЛЕКТУАЛЬНИХ СУПЕРТЕХНОЛОГІЙ)**

Аналізуються актуальні питання майбутнього людини як *Homo sapiens* в контексті інтенсивного розвитку та викликів інтелектуальних супертехнологій.

Ключові слова: *машина Тюрінга, комп'ютер, штучний інтелект, суперінтелект.*

Прикметною особливістю сучасної епохи є стрімка актуалізація вічних питань про долю людини як *Homo sapiens*, грядуще планетарного соціуму на зламі століть спричинена різкими метаморфозами в поглядах на досить туманні історичні перспективи людської цивілізації, що постають у контексті викликів наукових супертехнологій, передусім технологій штучного інтелекту (ШІ) або інтелектуальних технологій. Такі різкі зміни, спонукані передусім тривожним очікуванням настання (в прийдешні десятиліття) гіпотетичної (зпрогнозованої в кінці ХХ ст. В. Вінджем) технологічної сингулярності — експоненційно прискореного, вибухоподібного розвитку синергійно взаємодіючих між собою неосупертехнологій. Джерелом напруженої ситуації, тривожної непевності є усвідомлення того, що стрімкий прогрес супертехнологій (загроза насунання їх сингулярності) в принципі здатний спричинити такі різкі трансформації умов людського існування, круті перемини в укладі життя людей, що виникає реальна загроза неадекватного сприйняття з їх боку таких вражаючих метаморфоз. Це може, звісно, поставити під сумнів здатність адаптації людини, соціуму до перемін такого стрімкого характеру. Тож на грань катастрофи може бути поставлене саме існування людства. Гостроти ситуації надає ще й та обставина, що, згідно з домінуючою в науці концепцією, біологічна еволюція в лінії *Homo* з появою *Homo sapiens* призупинилась (що мовби вже остаточно встановлено),

а тому еволюційні зміни, трансформації в біології людини можливі (чи начебто навіть украй необхідні, аби запобігти загрозі кризи її існування) лише виключно завдяки радикальному фізичному втручанню в її природу не інакше, як у спосіб рішучого застосування новітніх науко-містких супертехнологій.

На жаль, категоричні наміри штучного фізичного втручання в біологічну природу людини в спосіб некоректного застосування цих супертехнологій приховують серйозну небезпеку отримання неочікуваних наслідків, появи незворотних процесів, викликаючи природне занепокоєння про долю *Homo sapiens*, майбутнє людства. Крім того, виклики супертехнологій часто-густо ще заздалегідь сприймаються як щось неминуче, невідворотне, якийсь фатум, чого не повинно бути в принципі. Ці виклики не можуть зарані, а ргіогі вважатися абсолютно невідворотними, неминучими. Такими, що мають для людини, цивілізації фатальний характер. Вони беззастережно підлягають усебічному аналізу, глибокому об'єктивному дослідженню, адже для майбутнього людства важливо знати не тільки, звідки та як прийшла людина в цей світ. Стрижневим є питання про те, куди вона йде. Чи є в неї реальні шанси для пролонгації свого перебування в цьому світі, залишатись у ньому й надалі

Тож із поняттям сингулярності асоціюється не просто небезпечна ймовірна можливість стрімкої трансформації сформованого впродовж прелімінарних тисячоліть еволюції людства темпоритму, але й (як закономірний наслідок) — реальна загроза існуванню людини як *Homo sapiens*, людського роду, людської цивілізації.

Витоки ідей інтелектуальних технологій сягають глибини століть (точніше, тисячоліть). Кмітливий людський розум, активно пізнаючи навколишній світ, його складні процеси та явища, водночас прагнув до самопізнання, осягнення сутності феномену мислення — проникнення в його таїни, розкриття глибинних механізмів й закономірностей функціонування мозку як продуценту мислення. Та допитлива людська думка не обмежується проблемою пізнання сфери мислення, не замикається в ній, але й мірою науково-технічного розвитку прагне також інтенсифікувати розумову діяльність (не очікуючи якоїсь повноти знання про неї, а виходячи з наявного на той чи той час його рівня) в спосіб формалізації, машинізації (комп'ютеризації) інтелектуальних функцій мозку, процесів мислення (Аристотель, Р. Луллій, Т. Гоббс, Г. В. Ляйбніц, У. С. Джевонс, Дж. Буль, Г. Фреге, К. Гедель, А. Тьюрінг, Дж. Непман та ін.). Вражає феноменальна прозорливість Е. Шредера, який ще 1873 р. екс-

пліцитно передвістив концепцію ШІ: «Подібно до інших наук, і логіка могла б колись здійснити дещо абсолютно неочікуване (Ungeahntes) і при цьому принести непередбачуваним чином незчисленну користь... Справді, ніхто не може сказати, що скоро не буде побудована «думаюча машина», аналогічна чи й досконаліша, ніж лічильна машина, та здатна звільнити людину від... значної частини виснажливої розумової праці» [23, с. 356]. Та умови для втілення ідеї ШІ, що еволюціонувала тисячоліттями, історично вперше склались лише в середині ХХ ст., коли стали, звісно, технологічно здійсненними можливості інтенсифікації інтелектуальної діяльності (в спосіб комп'ютеризації тих чи інших її функцій, видів і т. д. ). Безпосередні передумови виникнення ШІ асоціюється з теорією формальних логічних нейронних мереж Маккаллока — Піттса [9], теорією алгоритмічних числень Тьюрінга [30] та прогностичними концепціями й фундаментальними результатами Ж. Ербрана й Г. Генцена (кінець 20-х — початок 30-х рр. ХХ ст. ) — основи логічного виведення, комп'ютеризації доведення теорем і т. п.

У 30-х рр. ХХ ст. внутрішні потреби математики й математичної логіки — проблеми основ математики, логічного виведення, встановлення алгоритмічної (не)розв'язності та ін. — спонукали до необхідності коректного формально-логічного уточнення (формулювання чіткої дефініції) інтуїтивного поняття алгоритму — загального прийому розв'язання певного класу задач, його строгої (логічної) експлікації, побудови теорії алгоритмів та ін. Дослідженням цієї проблематики зайнялась когорта визначних математичних логіків і невдовзі (в середині 30-х рр. ) формально-логічні уточнення поняття алгоритму були запропоновані в низці теорій алгоритмів різних авторів. Один з них — видатний британський логік А. Тьюрінг, який, ввівши клас абстрактних машин, названих машинами Тьюрінга (МТ), створив струнку теорію алгоритмів — теорію машин (Тьюрінга) [30]. При розробці концепції цих машин, він виходив із припущень про те, чого могла б досягти людина-обчислювач у принципі, тобто будь вона ідеалізована — ніколи не помилялася, мала необмежений обсяг пам'яті та могла нескінченно працювати. Такими властивостями й наділена МТ — вона не може помилятися, має потенційно нескінченну пам'ять та здатна працювати нескінченно. Зіставлення такої машини з реальним (цифровим) комп'ютером свідчить, що МТ — це математично ідеалізований комп'ютер. Принципова відмінність МТ від людини-обчислювача та реального комп'ютера полягає в її ідеалізації — абстрагуванні від властивих їм практичних обмежень. МТ — це логічний еквівалент

понять алгоритму, ефективної обчислювальної функції, з одного боку, а з іншого — теоретичний («ідеальний») предтеча реального комп'ютера, гіпотетична абстрактна (віртуальна) обчислювальна машина.

А. Тьюрінг установив, що можна побудувати машину (МТ), здатну імітувати будь-яку іншу машину (МТ), якщо її пам'ять забезпечити описом цієї машини й програмою її функціонування. МТ, яка здатна виконувати всі операції, що реалізуються іншою довільною МТ, опис якої належним чином закодовано в її алфавіті та нанесено на її стрічці (у вигляді скінченної сукупності слів), називається універсальною машиною (МТ). Алгоритмічна універсальність МТ — це притаманна їй здатність реалізації будь-яких довільних алгоритмів. Зважаючи на принципову відмінність між МТ та комп'ютером (МТ у принципі не може бути втілена як діюча технічна система через практичну неможливість технічної реалізації її нескінченної пам'яті ні в якому пристрої), «прийнято називати машину універсальною, якщо організація її управління та набір операцій такі, що забезпечили б можливість реалізації будь-якого алгоритму за умови необмеженості обсягу пам'яті» [4, с. 235]. Отже, алгоритмічна універсальність комп'ютера забезпечується тим, що обсяг його пам'яті (на потужних інформаційних носіях) може вважатись практично необмеженим, що й визначає практичну можливість реалізації ним довільного алгоритму. «Тільки Тьюрінгу прийшла в голову ідея специфічної та всеохоплюючої алгоритмічної машини, названої універсальною машиною Тьюрінга, що здатна самостійно виконувати будь-яку алгоритмічну дію. Саме ця ідея привела в наслідок до створення концепції універсального комп'ютера...» [18, с. 60].

ІІІ — метафорична назва одного з найпріоритетніших наукових напрямів, потужний арсенал теоретичних і технічних засобів якого сконцентрований на вирішенні складного комплексу актуальних проблем, які різняться рівнем абстракції (від логіко-математичних, скажімо «геделівського типу», до техніко-технологічних) і пов'язані передусім із дослідженням розумової сфери людини: функціонування мозку (як органу мислення), комп'ютеризації інтелектуальної діяльності, імітації, моделювання ментальних процесів etc. Як комплексний напрям, що склався в кібернетиці (в загальному спектрі її широкомасштабних досліджень сфери мислення), а відтак, перебуваючи (принаймні номінально) під її егідою та набуваючи в процесі прискореного розвитку дедалі більш самостійного характеру, ІІІ став науковою основою створення інформаційно-комп'ютерних систем і технологій.

Еволюція комп'ютерів (покоління яких стають усе досконалішими) сприяла стрімкому підвищенню їх можливостей і розширенню сфери застосування. Комп'ютери все інтенсивніше й масштабніше почали перебирати на себе функції, що традиційно вважалися притаманними лише людині (як *Homo sapiens*), прерогативою її мозку. Із часом це *per se* вже не вкладалось у розуміння комп'ютерів як посилювачів розумових здібностей людини (У. Р. Ешбі), засобів підвищення її інтелектуального потенціалу, а сприймалось як дещо самодовільне, несумірно вагоміше, суттєвіше, якісно вище. Йдеться про наділення комп'ютерів властивостями, іманентними людині, визнання їх мовби «інтелектуальної здатності». Вони «обросли» епітетами — «інтелектуальні», «розумні», «мислячі», «думаючі»... На хвилі сплеску евфорії, спричиненої унікальною здатністю комп'ютерів моделювати чи імітувати притаманні мозку функції, виник і сам термін «artificial intelligence» — «AI» («Ш»), котрий, відзначаючись відповідними антропоморфними ознаками чи характеристиками, набував певного метафоричного сенсу.

Тож, асоціюючись з відомими антропоморфними уявленнями, термін «Ш» (як й еквівалентні йому чи семантично споріднені з ним вирази — «штучний розум», «машинний інтелект», «мислячий комп'ютер», «кібернетичний», «електронний мозок» etc. ), був неоднозначно сприйнятий в інтелектуальних колах, передусім наукових, справляючи суперечливі враження, викликаючи фантазмагоричні уявлення, збуджуючи гострі емоції, породжуючи бурхливі реакції, що загалом викликало численні пристрасні суперечки, запеклі дискусії з яскраво вираженим філософсько-світоглядним сенсом. Однак, попри гостру полеміку щодо дефініції терміна «Ш» (розмитості самого поняття, відсутності логічної експлікації), його коректності, метафоричності, цей термін, запропонований Дж. Маккарті на конференції з питань комп'ютерних програм, здатних до розумної поведінки (Дортмут, США, 1956 р. ). був «офіційно легітимований» I Міжнародною конференцією зі Ш (Вашингтон, 1969 р. ). Тож становлення Ш як наукової гапузі проходило в умовах запеклих дискусій. Та найбільше списів, певне, було зламано з проблеми «людина — комп'ютер» («людина — машина» в класичних традиціях), її різних аспектів, а паче з питання «чи може машина мислити?». Ефектно, образно, емко змалював цю ситуацію Г. М. Поваров: «Зближення людського мозку з «електронними мізками» викликало не менш бурхливу реакцію, ніж... дарвінівське зближення людини з мавпою. Напевне, після Коперніка й Дарвіна це було третім великим ураженням нашого звичного антропоцентризму [19, с. 19]. Цей

лаконізм імпліцитно криє чимало актуальних питань, сторін, нюансів, з яких декотрі будуть заторкнуті в цьому контексті.

У вирі перманентних дискусій щодо комп'ютерів — цих незбагнених для долі людини машин, серед гами різних голосів звучали також застережні нотки про реально можливі негативні наслідки їх еволюції. Ця проблема особливо болісно непокоїла Н. Вінера: «Одна з великих проблем, із якою ми неодмінно зіткнемось у майбутньому, — це проблема взаємовідношення людини й машини, проблема правильного розподілу функцій між ними...» [3, с. 81]. Цей щемний неспокій не полишав фундатора кібернетики до кінця його життя. Він, мужньо беручи на себе моральну відповідальність за можливі негативні наслідки цієї взаємодії, настійно й невтомно застерігав від них людство. В такому ключі висловлювались, звісно, й інші вчені, зокрема М. Мінський: «Упродовж життя всього лише одного покоління поряд із людиною виріс дивовижний новий вид: обчислювальні та подібні до них машини, з якими... їй доведеться ділити світ. Ані історія, ані філософія, ані здоровий глузд не можуть підказати нам, як ці машини вплинуть на наше життя в майбутньому, бо вони працюють зовсім не так, як машини, створені в еру промислової революції» [11, с. 9]. З подальшою еволюцією ІІІ тональність дискусій не спадає, хіба що зміщуються акценти. Особливо полемічними є проблеми «комп'ютерного мислення», «комп'ютерної свідомості», «комп'ютерного розуміння», актуалізація яких на зламі століть спонукана концепцією про начебто реальну можливість досягнення безсмертя людини в спосіб переведення її мислення, свідомості, душі, «Я» на «вічні» — штучні (небіологічні) носії: електронні, нейромережеві, силіконові (кремнієві)...

З-посеред цих питань головним зостається сакраментальне — «чи може машина мислити?». Це питання ще в 1950 р. висунув А. Тьюрінг у праці [24], концептуально пов'язаний з його працею [30]. Ці праці відіграли виняткову роль в інтенсифікації досліджень з ІІІ (логічних, епістемологічних, філософсько-методологічних, світоглядних, морально-етичних), у постановці глобальної проблеми «ІІІ versus людський інтелект?». Та він, певне, не тільки розуміючи надзвичайну складність, нетривіальність цього питання, але й усвідомлюючи, що його фундаментальна невизначеність (щодо сенсу термінів «мислити», «машина») навряд чи залишає якісь шанси на можливість отримання на нього раціональної відповіді, ухилився від цих ускладнень у спосіб заміни цього питання «іншим, яке тісно з ним пов'язане й виражається словами з відносно чітким сенсом. Ця нова форма проблеми може

бути описана за допомогою гри» (названої автором «грою в імітацію») [24, с. 9]. Суть методу гри в імітацію, чи «тесту Тьюрінга» («критерію Тьюрінга»), полягає в тому, щоб людина (в діалозі людина– машина) не змогла (впродовж сеансу— певного проміжку часу) розпізнати, хто її співрозмовник (на основі його відповідей на її запитання)— людина чи машина. Тобто, гра в імітацію як спосіб виявлення того, чи наділена машина властивістю мислити, зводиться до того, щоб у діалозі з людиною машина мовби могла імітувати ілюзію генерування правильної відповіді на її запитання— відповідати їй так, аби в тої склалося враження, начебто вона спілкується з людиною, а не з машиною. За А. Тьюрінгом, машину, здатну в такий спосіб перехитрити людину, можна вважати такою, що мислить. Отже, згідно із цією концепцією, машина, що проходить «тест Тьюрінга» (відповідає «критерію Тьюрінга») «може вважатися», мислячою, думаючою, розумною. Проте, як резонно зауважує Р. Солсо, «для спеціалістів з мови й ШІ в задачі Тьюрінга є дуже важливий момент— аби комп'ютер перехитрив нас і заставив думати, що він— це людина, він повинен розуміти й генерувати відповідну реакцію, яка б ефективно імітувала важливу когнітивну функцію» [21, с. 459]. Комп'ютер «повинен розуміти»!

Чи можливо це в принципі

У тьюрінгівській концепції питання «чи може машина мислити?», по суті, еквівалентне питанню «чи має мислення алгоритмічний характер?». Тобто, чи не може мислення бути виражене у вигляді універсального алгоритму (чи універсальної МТ)

З огляду на існування (алгоритмічно) універсального комп'ютера (здатного реалізувати будь-який довільний алгоритм, зокрема й універсальний алгоритм, що описує процес мислення), це означає, що питання «чи може машина мислити?» в концепції Тьюрінга фактично трансформується в питання «чи може бути процес мислення комп'ютеризований?». Однак А. Тьюрінг, поставив його не в такій формі, а в «новій формі»— гри в імітацію. В процесі гострої полеміки з цього питання було встановлено логічну нечіткість його постановки. Проте й у сучасних дослідженнях з цих проблем також даються неоднозначні оцінки критерію Тьюрінга [8; 18; 20; 21; 27]. Загалом тесту Тьюрінга радше пасувала б, певне, роль критерію розумної поведінки машини, котрий з нею фактично ідентифікувався, аніж— критерію, чи може мислити машина. В умовах відсутності прямої відповіді на стрижневе питання «чи може машина мислити?» превалюючим на певний час стало мовби переконання, що пряма постановка його науково некоректна. Та водночас

відбувалось інтенсивне поширення термінів на кшталт: «ІІІ», «штучний розум», «інтелектеальні» системи та технології, «суперінтелект» і т. д. Чи правомірний припис людських властивостей артефактам

Чи є сенс у наділенні їх епітетами: «розумні», «інтелектуальні», «мислячі», «думаючі»

Що не піддається прямому розкриттю, варто спробувати з'ясувати опосередковано.

Тут плідну роль можуть відіграти наукові метафори, що явно впливає з висновку, якого дійшов відомий лінгвіст Дж. Лакофф, аналізуючи (в контексті дослідження проблем теорій категоризації, когнітивних моделей та прототипів) евристичну функцію метафори. Його заключний акорд: «Залучення метафори до розуміння досвіду є одним із найвеличніших тріумфів людського мислення» [7, с. 182]. Х. Ортега-і-Гассет [16] у метафорі вбачав засіб, який дає можливість мисленню, свідомості вловити, змістовно визначити, виокремити й виразити абстрактні об'єкти (особливо об'єкти високого рівня абстракції). Будучи не лише необхідним засобом пізнання, вираження, а й — мислення, формою наукової думки, метафора «подовжує «руку» інтелекту» [16, с. 72]. Це особливо стосується галузі ІІІ, в якій метафоричне мислення, вирізняючись високим евристичним потенціалом, відіграє неабияку когнітивну роль. ІІІ навіть своїм виникненням завдячує метафоричній парадигмі: «Поява сучасного комп'ютера принесла із собою метафоричне уявлення про думаючі машини; кібернетики, філософи та психологи, засвоївши метафоричне уявлення, що комп'ютери займаються розумовою діяльністю, спорідненою із людською, створили нову дисципліну, котра займається штучним інтелектом» [10, с. 367]. Метафоричний сенс ІІІ полягає в тому, що мозок і комп'ютер зіставляються на основі двополусної (з двома протилежними сенсами, референтами) метатеоретичної комп'ютерної метафори. За цією метафорою, з одного боку, людський інтелект уподібнюється ІІІ: мозок за структурною організацією й розумовою діяльністю вважається подібним комп'ютеру (в його механізмах вбачаються відповідно аналогі архітектури та функціонування комп'ютера), а з іншого — навпаки: принципи архітектоники й функціонування комп'ютера розглядаються як аналогічні відповідним принципам будови й інтелектуальної діяльності мозку (комп'ютер мовби наділений мозкоподібними структурою та функціями). В контексті таких припущень та приписів впливає важливий когнітивний висновок: наявність істотних аналогій між інтелектуальною діяльністю людини та функціонуванням комп'ютера служить онтологічною основою формування комп'ютерної метафори, котра,



свою чергою, виступає концептуально-епістемологічною основою наділення комп'ютера епітетом «інтелектуальний» (як і «мислячий», «розумний», «думаючий»...) та зіставлення інтелектуальності людини й «інтелектуальності» комп'ютера.

Когнітивне значення комп'ютерної метафори знайшло свій прояв передусім у тому, що зусилля кібернетиків (спеціалістів з ШІ), психологів, лінгвістів нейрофізіологів та ін. були сфокусовані на пізнанні структурної організації, принципів і механізмів діяльності мозку з метою застосування цих результатів у конструюванні комп'ютерів, створенні інтелектуальних систем і технологій високих поколінь, з одного боку, а з іншого — у використанні конструктивних знань з галузі ШІ, кібернетики для розкриття внутрішніх механізмів психічних та мислительних процесів, осягнення ментальних просторів інтелектуальної діяльності. Метафоричний рівень зіставлення мозку й комп'ютера вимагає виняткової уваги та обережності. Застосування комп'ютерної метафори, яка має сприйматися лише виключно як метафора наукова, не припускає жодних крайнощів: її необґрунтоване відхилення чи ігнорування звужує горизонт пізнання, а — абсолютизація, буквальне тлумачення, криє загрозу отримання неочікуваних результатів, появи непрогнозованих наслідків, викликів. На думку Х. Ортеги-і-Гассета, при інтерпретації універсального явища через доступніше конкретне відношення йдеться лише про оперування науковою метафорою. «Ми не маємо встановлювати між» ними «відношення тотожності... Небезпека ототожнення тут дуже велика. Бо від наших уявлень про свідомість залежить наша концепція світу, а від неї... залежать наша мораль, наша політика, наше мистецтво. Виходить, що вся величезна будова Всесвіту, сповненого життям, базується на крихітному повітряному тільці метафори» [16, с. 77].

Метафори, вважає Е. Маккормак, можуть являти небезпеку не тільки тим, що схиляють «думати, що те, на що вони натякають, існує насправді, а й тим, що вони наштовхують на думку, що властивості, притаманні одним референтам метафори, притаманні й іншим її референтам. Якщо люди й комп'ютери володіють пам'яттю й думками, то це метафоричне слововживання може ввести в оману, змусивши вважати, що в комп'ютері можуть бути виявлені властивості людської пам'яті або ж, що поняття думки в людини має розумітися обмежено як схильність до дії, оскільки саме таке обмежене розуміння доречно стосовно до комп'ютера» [10, с. 375]. Своєю чергою, М. Арбіб, зазначає, що «стараючись використати риси схожості між людиною й машиною,... маємо... пам'ятати про

те, що поведінка людини, її пам'ять, сприйняття й здібність до навчання значно багатші, ніж у автомата... Тому, коли ми говоримо, що в певному розумінні людський мозок можна розглядати як обчислювальну машину з закладеною в неї програмою, ми не зводимо людину до рівня існуючих механізмів, а просто стараємось зрозуміти, як обчислювальні машини можуть допомогти нам розібратись в особливостях поведінки людини» [1, с. 18].

На зламі століть у зв'язку зі стрімким розвитком синергійно взаємодіючих наукомістких супертехнологій питання «чи може машина мислити?» особливо актуалізувалось, набуваючи вкрай гострого характеру. Саме завдяки цьому прогноуються радикальні зміни в сфері комп'ютерних технологій. Очікується, зокрема, що синергізм супертехнологій має привести не тільки до створення ШІ (еквівалентного за «інтелектуальним» потенціалом людському інтелекту), але й «штучного суперінтелекту» (ШСІ) чи просто «суперінтелекту» — ШІ надлюдського рівня (Г. Моравек, Н. Бостром, М. Дірінг та ін.), здатного перевершити «кращих представників людського розуму практично в будь-якій галузі, включаючи наукову творчість, здоровий глузд і соціальні навички» [2, с. 313], тобто — талантів, геніїв, титанів розуму. Порівняння можливостей комп'ютера та мозку можуть, звісно, проводитися лише за певними критеріями (чи параметрами). Наразі здебільшого використовується кількісний критерій — далеко не універсальний, не абсолютний [8; 10; 12; 15; 18], що спирається на концепцію, котра трактує мислення як лічильно-обчислювальний процес й асоціюється із концепцією комп'ютерної метафори. Е. МакКормак зазначав, що «комп'ютерна метафора пропонує підходити до людського пізнання, як мовби воно було обчислювальною діяльністю, і до обчислювальної діяльності — як нібито вона була людським пізнанням» [10, с. 378]. За цією концепцією, будь-яка розумова діяльність — це мовби операції над символами. Тому вважається, що завдяки належним експліцитним формальним правилам вона за допомогою цих символів може начебто бути відтворена універсальним комп'ютером.

За даними сучасної нейрофізіології мозок людини складається, ймовірно, десь із  $10^{11}$  (100 млрд) нейронів. Однак мозок — це не просто система, що складається з величезної множини якихось ізольованих інертних, статичних нейронів, а надзвичайно складна ієрархічна функціонуюча система, утворена з різноманітно взаємопов'язаних й інтенсивно взаємодіючих нейронів. Активно функціонуючий мозок — це ієрархічна система складних нейронних зв'язків — синапсів. Уважається, що на

кожний нейрон мозку припадає (впродовж однієї секунди) приблизно десь  $10^3$  синапсів (Г. Моравек [29], Д. Х'юбел [28]). Це—та ймовірна кількість сигналів, яку здатний, в середньому надсилати, на думку Г. Моравека, кожний з активно діючих нейронів мозку за одну секунду, тобто— $10^3$  опер/сек на нейрон. Отже, обчислювальна потужність мозку складає десь, ймовірно,  $10^{14}$  (100 трлн) опер/сек ( $10^{11}$  опер/сек  $\cdot 10^3 = 10^{14}$  опер/сек), тобто це—максимальна продуктивність, якої в принципі може досягти, за Г. Моравеком, мозок. Прагнучи встановити обчислювальну потужність суперінтелекту й припускаючи, що мозку може бути властива певна надлишковість, він підняв на три порядки—в  $10^3$  (1000) разів—висоту планки визначеної ним потенційної потужності мозку, вважаючи, певне, що отримана в такий спосіб величина— $10^{14}$  опер/сек  $\cdot 10^3 = 10^{17}$  (100 квадрилн) опер/сек—виражає імовірну потенційну продуктивність суперінтелекту [29], здатну перевершити будь-які мислимі (чи й немислимі) можливості мозку. Якщо з нижньою межею ( $10^{14}$  опер/сек) уведеного Г. Моравеком діапазону асоціюється обчислювальна потужність ШІ, еквівалентна потужності мозку людини, то з верхньою—( $10^{17}$  опер/сек)—потужність ШСІ, котра має мовби цілком перекрити його потужність.

Концепція можливості створення СШІ спирається на закон Мура, що носить експоненціальний характер: виражає закономірність подвоєння обчислювальної потужності комп'ютера за якийсь певний період (константу) часу (фактично реальний час подвоєння коливається в межах одного—двох років, тобто в середньому півтора року [2]). Тож, згідно із законом Мура, потужність ШСІ в  $10^{17}$  опер/сек цілком досяжна. За прогнозами, невдовзі той час, коли створення ШСІ—ШІ з (розрахунковою) апаратною (hardware) потужністю в  $10^{17}$  опер/сек, буде цілком технологічно можливе. Та з огляду на те, що, за твердженням Н. Бострома, суперінтелект з'явиться незабаром після того, як тільки виникне ШІ з потужністю людського рівня [2, с. 334], спершу варто, певне, з'ясувати стан реалізації нижньої межі встановленого Г. Моравеком діапазону. Меседжі, що надходять із відповідних світових науково-дослідних центрів—інститутів, університетів, лабораторій, корпорацій,—свідчать, що потужність комп'ютерів не тільки впритул наближується до  $10^{14}$  опер/сек, а й уже де-де реалізована. Тож можна вважати, певне, що потужність ШІ, еквівалентна потужності мозку, вже принаймні в принципі досягнута. Тоді десь невдовзі, за Н. Бостромом, буде створено й суперінтелект. Дистанція від  $10^{14}$  опер/сек до  $10^{17}$  опер/сек не така вже й значна. Її буде пройдено, певне, навіть за умови (деякого)

порушення строгості дії закону Мура. Це може хіба-що якоюсь мірою (та й то, мабуть, не суттєво) відбитись на термінах. За прогнозами ШСІ буде створений приблизно десь у період 2015–2024 рр. (2015 р., якщо час подвоєння потужності ШІ становитиме один рік, 2024 р. — десь усереднено півтора року) [2, с. 23].

Та цим проблема, вочевидь, не вичерпується [17; 18; 20]. Г. Моравек увів рівні обчислювальних потужностей мозку й ШІ, виходячи мовби з «цілісної нейронної концепції». Однак сам нейрон (як структурно-функціональна одиниця нервової системи)—надскладна система. Крім тіла (соми), що містить ядро, й дендритів та аксона з його відростками, кожний нейрон має свій власний цитоскелет, який також являє собою складну структуру (утворення, що «складається із протеїноподібних молекул, організованих у структури різного типу—актин, мікротрубочки, проміжні волокна»), котра, oprіч функції «підтримки форми клітини», виконує й «численні інші функції», ймовірно, в певному сенсі навіть функцію «особистої нервової системи» нейрона [18, с. 547–548]. Із структур, які складають цитоскелет (цей суттєвий структурно-функціональний елемент нейрона), особливий інтерес викликають мікротрубочки—порожнисті циліндричні трубки з внутрішнім і зовнішнім діаметрами десь відповідно 14 нм і 25 нм (нм—нанометр, тобто  $10^{-9}$ м), які «іноді організовані в більш крупні трубкоподібні волокна, що складаються з... дублетів, триплетів, або часткових триплетів мікротрубочок». Мікротрубочка—«білковий полімер, що складається із субодиниць—тубулінів. Кожна із субодиниць тубуліна «являє собою «димер», тобто складається із двох з'єднаних тонкою перемичкою частин, так званих  $\alpha$ -тубулін і  $\beta$ -тубулін... Ці пари глобулярних білків... укладено в дещо скошену гексагональну решітку вздовж всієї трубки...» [18, с. 549]. Кожна з мікротрубочок складається з тринадцяти рядів димерів тубуліну, а кожний димер може існувати принаймні в двох різних геометричних конфігураціях—конформаціях; «центром управління» в цитоскелеті, ймовірно, є центросома, що складається із центросолі та інших мікроелементів, etc. [18, с. 549–552].

Та на даному рівні розгляду стрімко ускладнюється не тільки проблема створення суперінтелекту, але й—ШІ з потужністю людського рівня. Так, за обчисленнями Р. Пенроуза, в цьому разі потужність мозку сягає десь уже порядку  $10^{27}$  опер/сек. «Якщо ж у значенні елементарного обчислювального блоку взяти димер тубуліну, — зазначає він, — то слід урахувати, що на кожний нейрон припадає біля  $10^7$  димерів; відповідно, елементарні операції тепер виконуються десь у  $10^6$  разів швидше,

в наслідок чого отримуємо  $10^{27}$  опер/сек. Можливо, продуктивність сучасних комп'ютерів й насправді вже починає наближуватись десь до...  $10^{14}$  опер/сек..., однак не дивлячись на всі ці успіхи, досягти в оглядному майбутньому продуктивності  $10^{27}$  опер/сек не видається можливим» [18, с. 559–560]. У цьому розрізі особливого резону набирає, звісно, прогноз Н. Бострома, що дія закону Мура обмежена в фізичних і часових вимірах: «Очевидно, рано чи пізно він має перестати діяти. Існують фізичні обмеження на щільність, з якою матерія здатна зберігати та обробляти інформацію». Враховуючи, «що максимальна швидкість обмежена швидкістю світла..., екстенційне збільшення доступної обчислювальної потужності не може продовжуватись нескінченно, якщо тільки не будуть відкриті нові закони фізики. На мій погляд, закон Мура втратить довіру до себе задовго до того, як ми досягнемо абсолютних фізичних меж» [2, с. 316]. У контексті квантової теорії, з одного боку, відкривається надскладний, ієрархічний, невичерпний світ архітектоники мозку людини, його функціонування, а з іншого— зростає ймовірність того, що із часом обчислювальна потужність комп'ютера дійде своїх меж: закон Мура втратить свою силу. Та дослідження цих проблем мають принципове значення не лише для встановлення можливості створення ШІ потужністю, еквівалентною потужності мозку людини, та ШСІ, потужність якого переважала б його потужність, а й загалом— встановлення того, чи принципово (не)обчислювальна сутність людського розуму [18, с. 534–597]. Річ, певне, передусім у тому, чи можна в принципі будь-яку розумову функцію мозку виразити (реалізувати) через функцію обчислення, чи загалом можливо усю сукупність («континуум») його функцій замінити операціями обчислення.

Тож, припустивши, що «задекларований» ШСІ уже створено (потужність  $10^{17}$  опер/сек технологічно реалізована), чи означатиме це, що він насправді не тільки досяг рівня людського інтелекту, але й навіть перевершив його

Це питання, зачіпаючи чимало нюансів, вкрай складне й дискусійне. Як уже зазначалось, ШСІ порівнюється з людським мозком лише за одним-єдиним параметром— обчислювальним еквівалентом. Саме так, за М. Дірінгом, основне питання тут крутиться навколо обчислювальної теорії розуму, тобто ідеї, що все, що діється в людській голові базується на обчисленнях, які виконуються відповідними фізичними структурами мозку, включаючи не лише процеси логічні, а й хімічні та електричні, котрі могли б однаковою мірою бути успішно реалізованими й на цифровому комп'ютері [6]. Загалом концепція «мислення—це обчис-

лення» не узгоджується з теоремою Геделя про неповноту, з її логічними наслідками [5; 8; 10; 11; 12; 14; 15; 18]. Як резонно зазначає (у контексті цієї заборони Геделя) Р. Пенроуз, «немає іншого виходу, крім як однозначного висновку, що деяку суттєву складову людського розуміння неможливо змоделювати ніякими обчислювальними засобами» [18, с. 323]. В іншому місці він з іще більшою переконливістю наголошує, що «розуміння зовсім не стосується тих якостей, на які комп'ютери коли-небудь виявляться здатні...» [18, с. 661]. Не зарадить справі навіть досягнення комп'ютером швидкодії в  $10^{27}$  опер/сек. Отже, позиція Р. Пенроуза, як і, скажімо, — Дж. Серла (в контексті його концепцій «китайська кімната» й «силіконові мізки»), сприяє усталенню загальної концепції про принципову неможливість комп'ютерного розуміння (як і мислення в цілому), за якою, комп'ютеру не дано бути суб'єктом мислення.

Загалом ця концепція альтернативна існуючій концепції про мовби реальну можливість вирішення проблеми безсмертя людини як *Homo sapiens* в спосіб «перенесення» («переведення») її мислення, свідомості, ества, «душі», «Я» на нові «вічні» — штучні (небіологічні) носії: електронні, нейромережеві, кремнієві (силіконові)... Звісно, тут ідеться про надскладний комплекс питань і проблем, на багато з яких відповідей (принаймні остаточних) ще немає. Та концепція Моравека про можливість перенесення «інформаційної моделі» людини, що виражає її «сутність», у «машинну форму», чи концепція Курцвейла — оцифрувати за допомогою наномедасемблерів кожний нейронний зв'язок мозку і передати отриману інформацію («цифрову модель») в суперкомп'ютер («завантажити мозок у комп'ютер»), тобто «skonфігурувавши» в такий спосіб людину, зробити її безсмертною, вразливі. Як зазначалось, мозок складається десь, імовірно, з  $10^{11}$  (100 млрд) нейронів, які відзначаються надзвичайною різноманітністю. За Ч. Стівенсом, серед них не «знайдеться й двох нейронів, однакових на вигляд» [22, с. 31]. Стівенсове твердження цілком конгеніальне висловлюванню М. Арбіба, що, «виявляється, в природі загалом не існує того, що можна було б назвати типовим нейроном» [1, с. 40]. Виходить, що кожний нейрон мозку людини чимось специфічний, неповторний, унікальний! Чи є підстави стверджувати, що всі нейрони мозку (незважаючи на їх численність, надзвичайну різноманітність) уже виокремлені, досліджені, пізнані, описані

Навряд чи. Загальне число синапсів мозку (сумарна кількість операцій мозку за одну секунду) становить, за Г. Моравеком, приблизно  $10^{14}$

(100 трлн). Синапси відзначаються ще більшим розмаїттям, оскільки їх генерування детермінується специфікою нейронів, широким діапазоном сигналів та множиною інших факторів. Звісно, як і нейрони, вони залишаються ще мало дослідженими.

Отже, виходячи із зазначеного, «інформаційні моделі» (типу Моравека) та «цифрові моделі» (типу Курцвейла), якщо вони й існують, слабо виражають «сутність» людини, її мозку. Омріяні моделі людини (мозку) в машинній формі (чи загалом у «програмній» версії), якщо їх і можна було б отримати (що мало ймовірно) при наявності надпотужних суперкомп'ютерів та глибоких і певною мірою повних знань про нейрони, синапси), навряд чи матимуть прямий стосунок до життя, смерті чи безсмертя людини (хіба-що створення «таких моделей» можливе лише ціною смерті «оригінала»—самої людини). «Такі моделі» являтимуть собою, певне, мертві (статичні?) миттєві «зліпки» мозку (людини), позбавлені життя, бажань, прагнень, емоцій та ін. Так, навіть коп'ютерне моделювання, імітація Тайфуна, Ядерної зими, Великого вибуху—це зовсім (відповідно, звісно) не Тайфун, не Ядерна зима, не Великий вибух. Життя людини на «вічних» (небіологічних) носіях—це ілюзія, фікція. Тож питання про вирішення проблеми безсмертя людини в спосіб переведення її мислення, свідомості на якісь інші носії, про існування «електронної людини», «електронного суспільства»—це невичерпна тематика досліджень [13; 25; 26; 27]. Особливо гостропоземічним є питання про те, чи може «мертва матерія» відчувати, розуміти, мислити, переживати, мріяти тощо. Який сенс міняти живу, чутливу, мислячу матерію на матерію мертву, полишену життя, відчуття, свідомості

### Список використаних джерел:

1. Арбіб М. Метафорический мозг (пер. с англ. ) / М. Арбіб. — М.: Мир, 1976. — 296 с.
2. Бостром Ник. Сколько осталось до суперинтеллекта
3. / Ник Бостром // Информационное общество. — М.: ООО «Изд. АКТ», 2004. — С. 313–340.
4. Винер Н. Творец и робот (пер. с англ. ) / Н. Винер. — М.: Прогресс, 1966. — 104 с.
5. Глушков В. М. Введение в кибернетику / В. М. Глушков. — К.: Наук. думка, 1964. — 256 с.
6. Глушков В. М. Проблема «искусственного интеллекта» в свете идей Ф. Энгельса / В. М. Глушков, А. Я. Мороз // Ф.

- 
- Энгельс и современное естествознание.—К.: Наук. думка, 1979.—С. 268–292.
7. Диринг М. Рассвет Сингулярности / М. Диринг [электронный ресурс].—Режим доступа: <http://transhumanism.org/languages/russian/dawnofsingularity/Deering.htm>.
  8. Лакофф Дж. Когнитивное моделирование (пер. с англ.) / Дж. Лакофф // Язык и интеллект.—М.: Прогресс, 1995.—С. 143–184.
  9. Люгер Дж. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем / Дж. Ф. Люгер.—М.—СПб.—К.: Изд. дом «Вильямс», 2003.—864 с.
  10. Маккаллок У.С. Логическое исчисление идей, относящихся к нервной активности (пер. с англ.) / У.С. Маккаллок, У. Питтс // Автоматы.—М.: ИЛ, 1956.—С. 362–384.
  11. Мак Кормак Э. Когнитивная теория метафоры (пер. с англ.) / Э. Мак Кормак // Теория метафоры.—М.: Прогресс, 1990.—С. 358–386.
  12. Минский М. Вычисления и автоматы (пер. с англ.) / М. Минский.—М.: Мир, 1971.—366 с.
  13. Мороз А. Я. Проблема «искусственного интеллекта» и формально-логические ограничения / А. Я. Мороз // Рациональность, рассуждение, коммуникация.—К.: Наук. думка, 1987.—С. 142–155.
  14. Мороз О. Я. Доля Homo sapiens у контексті викликів новітніх наукових супертехнологій / О. Я. Мороз // Природознавство і гуманітарія. Пошуки взаєморозуміння.—К.: Видавець ПАРАПАН, 2009.—С. 155–238.
  15. Мороз О. Я. Епістемологічний аналіз когнітивного підходу до проблеми штучного інтелекту/ О. Я. Мороз // Сучасне природознавство: когнітивний, світоглядний, культурно-історичний виміри.—К.: Наук. думка, 1995.—С. 205–253.
  16. Мороз О. Я. Символьний і конекціоністський підходи до штучного інтелекту (штучний розум versus машинний інтелект?)/ О. Я. Мороз // Науковий світогляд на зламі століть.—К.: Видавець ПАРАПАН, 2006.—С. 171–251.
  17. Ортега-и-Гассет Х. Две великие метафоры (пер. с исп.) / Х. Ортега-и-Гассет // Теория метафоры.—М.: Прогресс, 1990.—С. 68–82.



18. Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики (пер. с англ. ) / Р. Пенроуз. — М.: Едиториал УРСС, 2003. — 384 с.
19. Пенроуз Р. Тени разума: в поисках науки о сознании (пер. с англ.) // Р. Пенроуз. — М. — Ижевск: Ин-т комп. иссл., 2005. — 688 с.
20. Поваров Г.Н. Норберт Винер и его «Кибернетика» (от редактора перевода) / Г. Н. Поваров // Винер Н. Кибернетика, или Управление и связь в животном и машине (пер. с англ. ). — М.: Сов. радио, 1968. — С. 5–19.
21. Рассел С. Искусственный интеллект: современный подход (пер. с англ. ) / С. Рассел, П. Норвинг. — М. — СПб. — К.: Изд. дом «Вильямс», 2006. — 1408 с.
22. Солсо Р. Когнитивная психология (пер. с англ. ) / Р. Солсо. — СПб: Питер, 2002. — 592 с.
23. Стивенс Ч. Нейрон // Мозг (пер. с англ. ) Ч. Стивенс. — М.: Мир, 1982. — С. 31–58.
24. Стяжкин Н.И. Формирование математической логики / Н. И. Стяжкин. — М.: Наука, 1967. — 508 с.
25. Тьюринг А. Может ли машина мыслить
26. (пер. с англ. ) / А. Тьюринг. — М.: Физматгиз, 1960. — 112 с.
27. Фукуяма Ф. Конец истории и последний человек (пер. с англ. ) / Ф. Фукуяма. — М.: АСТ: Полиграфиздат, 2010. — 588 с.
28. Фукуяма Ф. Наше постчеловеческое будущее: Последствия биотехнологической революции (пер. с англ. ) / Ф. Фукуяма. — М.: ООО «Изд. АСТ», 2004. — 349 с.
29. Хокинс Дж. Об интеллекте (пер. с англ. ) / Дж. Хокинс, С. Блейкли. — М. — СПб. — К.: Изд. дом «Вильямс», 2007. — 239 с.
30. Хьюбел Д. Мозг // Мозг (пер. с англ. ) / Д. Хьюбел. — М.: Мир, 1982. — С. 9–29.
31. Moravec H. Mind Children: the Future of Robot and Human Intelligence / H. Moravec. — Cambridge, Massachusetts: Harvard University Press, 1988. — 214 p.
32. 30. Turing A.M. On computable numbers, with an application to the Entscheidungsproblem / A. M. Turing // Proc. London Math. Soc. — 1936. — 42. — № 3–4. — P. 230–265; 1937. — 43 — № 7. — P. 544–546.

---

Мороз А. Я. Искусственный интеллект versus естественный интеллект (будущее человека в контексте вызовов интеллектуальных супертехнологий)

Анализируются актуальные вопросы будущего человека как Homo sapiens в контексте интенсивного развития и вызовов интеллектуальных супертехнологий.

Ключевые слова: машина Тьюринга, компьютер, искусственный интеллект, суперинтеллект.

Moroz A. Y. Artificial intelligence versus natural intelligence (future of human in context of challenges intellectual super technology)

Actual questions about the future of human as Homo sapiens in the context of the intensive development and challenges of the intellectual super technology are analyzed.

Keywords: Turing machine, computer, artificial intelligence, super intelligence.

© Мороз О. Я., 2014