

ЖИТТЯ 3.0

ДОБА ШТУЧНОГО ІНТЕЛЕКТУ

Макс Тегмарк

*Переклала з англійської
Зорина Корабліна*

«НАШ ФОРМАТ» · КИЇВ · 2019

[Купити книгу на сайті kniga.biz.ua >>>](http://kniga.biz.ua)

ЗМІСТ

<i>Пролог.</i>	ІСТОРІЯ КОМАНДИ «ОМЕГА»	11
<i>Розділ 1.</i>	ЗАПРОШУЄМО ДО НАЙВАЖЛИВІШОЇ РОЗМОВИ СУЧАСНОСТІ	33
<i>Розділ 2.</i>	МАТЕРІЯ СТАЄ РОЗУМНОЮ	65
<i>Розділ 3.</i>	НАЙБЛИЖЧЕ МАЙБУТНЄ: ПРОРИВИ, БАГИ, ЗАКОНИ, ЗБРОЯ І РОБОЧІ МІСЦЯ ..	104
<i>Розділ 4.</i>	ВИБУХ ІНТЕЛЕКТУ?	166
<i>Розділ 5.</i>	НАСЛІДКИ: НАСТУПНІ 10 000 РОКІВ	197
<i>Розділ 6.</i>	НАШ КОСМІЧНИЙ ДАР: НАСТУПНИЙ МІЛЬЯРД РОКІВ І ПІСЛЯ НЬОГО	246
<i>Розділ 7.</i>	ЦІЛІ	302
<i>Розділ 8.</i>	СВІДОМІСТЬ	339
<i>Епілог.</i>	ІСТОРІЯ КОМАНДИ ІНСТИТУТУ МАЙБУТНЬОГО ЖИТТЯ	381
<i>Подяки</i>	405
<i>Примітки</i>	407

ІСТОРІЯ КОМАНДИ «ОМЕГА»

Команда «Омега» була серцем компанії. Якщо решта підприємства заробляла гроші на поточні потреби різними комерційними застосуваннями вузького штучного інтелекту (ШІ), то команда «Омега» наполегливо йшла вперед до давньої мрії гендиректора: створити загальний ШІ. Більшість інших співробітників вважала «омег» (так їх жартома прозвали) жменькою мрійників-ідеалістів, яким до мети – як до неба пішки. Проте їм це прощалося, бо всім подобався той авторитет, який заробляли для компанії супермегановітні розробки «омег». А ще – вдосконалені алгоритми, які їм час від часу перепадали від «омег».

Єдине, чого вони не знали, – це те, що «омеги» ретельно ліпили свій образ, аби приховати за ним таємницю: вони впритул наблизилися до здійснення найзухвалішого в людській історії задуму. Харизматичний гендиректор відібрав до команди не лише блискучих науковців; серед критеріїв були також честолюбність, ідеалізм і сильне бажання допомогти людству. Він нагадував їм, що цей задум – надзвичайно небезпечний і якщо про нього довідаються могутні держави, то не зупиняться практично ні перед чим, навіть перед викраденням, аби згорнути цей проект. А ще краще – поцупити код. Але вони стовідсотково були готові до всього, як ті найкращі фізики світу, які стали членами Мангеттенського проекту з розробки ядерної зброї – вони були переконані: якщо першими стануть не вони, то це буде хтось із менш ідеалістичними намірами.

Створений ними ШІ під назвою «Прометей» дедалі сильніше вдосконалювався. Хоча його когнітивні здібності все ще значно

відставали від людських у багатьох сферах, наприклад, у соціальних навичках, «омеги» щосили намагалися навчити його відмінно програмувати системи III. Вони навмисно обрали цю стратегію, тому що вірили у твердження про інтелектуальний вибух, яке висловив британський математик Ірвінг Гуд ще 1965 року:

Визначмо надрозумну машину як таку машину, що може значно перевершити всю розумову діяльність навіть найрозумнішої людини. Оскільки конструювання машин належить до такої розумової діяльності, надрозумна машина може розробляти ще кращі машини; тоді, безсумнівно, відбудеться «інтелектуальний вибух», і людський розум відстане й залишиться далеко позаду. Отож перша надрозумна машина – останній винахід, який узагалі доведеться зробити людині, якщо тільки ця машина буде достатньо слухняною, щоб повідомити нас, як тримати її під контролем.

На думку розробників, якби вдалося домогтися такого рекурсивного самовдосконалення, то машина невдовзі так сильно порозумнішає, що сама зможе опанувати всі інші людські вміння, які можуть бути їй корисними.

ПЕРШІ МІЛЬЙОНИ

Запуск вирішили почати у п'ятницю, о дев'ятій ранку. «Прометей» муркотів собі у спеціально для нього зібраному кластері, розміщеному на довгих рядах стелажів у величезному залі з кондиціонованим повітрям, куди не впускали нікого стороннього. Заради безпеки він був повністю від'єднаний від інтернету, але містив локальну копію великої частини мережі (Вікіпедію, Бібліотеку Конгресу США, Twitter, нарізку з YouTube, Facebook тощо). Усі ці дані мали служити навчальним матеріалом для «Прометейя»*. Цей час «омеги» спеціально обрали, щоб

* Для спрощення я в цьому сюжеті використовую нинішній стан економіки й технологій, хоча більшість дослідників вважають, що до створення загального III

ЗАПРОШУЄМО ДО НАЙВАЖЛИВІШОЇ РОЗМОВИ СУЧАСНОСТІ

Технології дають життю потенціал для досі небувалого розвитку – або для самознищення.

Інститут майбутнього життя

Через 13,8 мільярда років після народження Всесвіт ожив й усвідомив своє існування. З маленької блакитної планети крихітні свідомі часточки почали вдивлятися в космос через об'єктиви телескопів, раз у раз відкриваючи, що все, що вони сприймали за всесвіт, – це просто частинка чогось величнішого: Сонячна система, галактика і Всесвіт із сотнями мільярдів інших галактик вишикувалися у складний візерунок груп, скупчень і надскупчень. І хоча ці наділені свідомістю порошинки багато в чому не погоджуються між собою, вдивляючись у зорі, в одному вони однакові: ці галактики вражають своєю красою.

Але краса – поняття відносне, не закладене в законах фізики, тож доки Всесвіт не ожив, ніякої краси не було. Тим дивовижніше наше космічне пробудження, і його варто цінувати: воно перетворило наш Всесвіт із бездумного, позбавленого самоусвідомлення зомбі на живу екосистему, здатну аналізувати себе, відчувати прекрасне, надіятися, а ще – досягати цілей, шукати зміст і мету. Якби Всесвіт ніколи не ожив, я думаю, він не мав би жодного сенсу – займав би страшенно багато місця і все. Якщо наш Всесвіт усе ж знову засне навки через якусь космічну катастрофу чи з нашої власної вини, він, на жаль, втратить усякий сенс.

З другого боку, усе може повернути на краще. Ми поки не знаємо, чи люди – єдині, хто вдивляється в зорі у нашому космосі. Ми навіть не знаємо, чи ми перші. Але ми вже достатньо дізналися про наш Всесвіт, щоб зрозуміти: він має потенціал для набагато повнішого пробудження, ніж досі. Можливо, ми, люди, нагадуємо отой перший проблеск самоусвідомлення, який ви відчули, прокидаючись уранці: передчуття значно більшої свідомості, яка настане з повним пробудженням. Можливо, життя пошириться в нашому космосі й процвітатиме мільярди й трильйони років. І можливо, це станеться завдяки рішенням, які ми за свого життя приймаємо тут, на нашій маленькій планеті.

КОРОТКА ІСТОРІЯ СКЛАДНИХ РЕЧЕЙ

А як відбулося це дивовижне «пробудження»? Це була не ізольована подія, а просто один крок у невпинному процесі тривалістю 13,8 мільярда років, який робить наш Всесвіт дедалі складнішим і цікавішим. І цей процес постійно прискорюється.

Як фізику мені пощастило провести більшу частину останньої чверті століття, допомагаючи простежити нашу космічну історію. Це дивовижна подорож від одного відкриття до іншого. Коли я був ще студентом, вік нашого Всесвіту оцінювався від 10 до 20 мільярдів років; зараз же питання в тому, чи йому 13,7 чи 13,8 мільярда. Усе це завдяки поєднанню кращих телескопів, кращих комп'ютерів і кращого розуміння. Ми, фізики, усе ще не знаємо точної причини Великого вибуху (і чи він узагалі був початком усього, чи просто продовженням якогось давнішого етапу). Проте в нас склалося досить детальне розуміння того, що відбулося після Великого вибуху, – завдяки масі високоточних вимірювань. Тож дозвольте мені коротко підсумувати 13,8 мільярда років космічної історії.

Спочатку було світло. У першу частку секунди після Великого вибуху вся та частина космосу, яка в принципі доступна для спостереження нашим телескопам («наш видимий Всесвіт», або просто «Всесвіт», якщо коротко), була набагато гарячіша і яскравіша за середину Сонця і швидко розширювалася. Звучить дуже

МАТЕРІЯ СТАЄ РОЗУМНОЮ

Водень [...], якщо йому дати достатньо часу, перетворюється на людей.

*Едвард Роберт Гаррісон,
британський космолог і астроном 1995*

Одна з найвинятковіших подій за 13,8 мільярда років після Великого вибуху — те, що в нерозумної мертвої матерії з'явився інтелект. Як це могло трапитися й наскільки розумнішою може вона стати в майбутньому? Що наука може сказати про історію і долю інтелекту в нашому космосі? Щоб легше було розібратися із цими питаннями, присвятімо цей розділ дослідженню засад та основних цеглинок, з яких складається інтелект. Що означає, коли ми кажемо «згусток матерії наділений розумом»? Що означає, коли ми кажемо «об'єкт може запам'ятовувати, обчислювати і вчитися»?

ЩО ТАКЕ РОЗУМ?

Нам із дружиною недавно пощастило відвідати симпозіум про штучний інтелект, організований шведською Фундацією Нобеля. Коли групу провідних дослідників ШІ попросили дати визначення розуму, вони довго сперечалися, але так і не дійшли консенсусу. Нам це видалося кумедним: навіть між інтелектуалами, які досліджують інтелект, немає згоди в тому, що таке розум! Тож, мабуть, беззаперечно «коректного» визначення розуму не існує.

Натомість є багато суперечливих визначень, у яких згадується здатність до логічного мислення, розуміння, планування, емоційного знання, самоусвідомлення, творчості, розв'язування задач і навчання.

У нашому дослідженні майбутнього інтелекту ми хочемо зайняти максимально широку, інклюзивну позицію, не обмежену тими видами інтелекту, які існують наразі. Ось чому визначення, яке я дав у попередньому розділі, і спосіб, у який я збираюся використовувати це слово в цій книжці, дуже широкі:

інтелект = здатність досягати складних цілей.

Воно достатньо широке, щоб охопити всі вищенаведені визначення, позаяк розуміння, самоусвідомлення, розв'язування задач, навчання тощо – усе це приклади складних цілей, які можна мати. Воно також достатньо широке, щоб умістити в собі визначення, що його дає Оксфордський словник: *«здатність набувати і використовувати знання та вміння»*, тому що ціль може полягати в застосуванні знань і вмінь.

Оскільки може бути багато цілей, типів інтелекту теж може бути багато. Тому, за нашим визначенням, кількісне вираження інтелекту людей, тварин або машин одним числом, як-от рівень IQ*, не має жодного сенсу. Що розумніше: комп'ютерна програма, яка вміє лише грати в шахи, чи програма, яка вміє лише грати в го? На ці питання немає логічної відповіді, тому що обидві добре надаються до різних завдань, які неможливо порівняти безпосередньо. Проте можна сказати, що якась третя програма розумніша за першу і другу, якщо вона принаймні так само добре може досягати *всіх* цілей і точно краще може виконувати принаймні одне завдання (скажімо, виграти партію в шахи).

* Щоб зрозуміти це, уявіть свою реакцію, якби хтось заявив, що здатність досягати олімпійських спортивних результатів можна кількісно виразити одним числом, яке називається «атлетичний коефіцієнт» (athletic quotient), скорочено «AQ», тож олімпійський спортсмен із найвищим AQ повинен здобути золоті медалі в усіх видах спорту.

НАЙБЛИЖЧЕ МАЙБУТНЄ: ПРОРИВИ, БАГИ, ЗАКОНИ, ЗБРОЯ І РОБОЧІ МІСЦЯ

Якщо незабаром не змінити напрямок, то ми опинимося там,
куди йдемо.

Ірвін Корі, американський комік

Що означає бути людиною в наш час? Наприклад, що ми справді цінимо в собі, що робить нас відмінними від інших форм життя і машин? Що цінують у нас інші – аж так, що їм навіть хочеться запропонувати нам роботу? *Хоч би якими були наші відповіді на ці питання в певний момент, очевидно, що поява технологій може поступово їх змінювати.*

Візьмімо, наприклад, мене. Як учений я пишаюся тим, що ставлю сам собі цілі, що застосовую креативність та інтуїцію для розв'язання широкого спектра нерозв'язаних проблем і що користуюся мовою, щоб поділитися своїми відкриттями. На щастя для мене, суспільство охоче платить мені за це, як за роботу. Кілька сторіч тому я міг би натомість, як багато інших, стати фермером або ремісником, але розвиток технологій із часом звів такі професії до малесенької часточки в складі робочої сили. Це означає, що тепер не кожен може сформуватися як професіонал у фермерстві або ремеслах.

Особисто мені байдуже, що сучасні машини набагато краще за мене виконують таку ручну працю, як копання землі чи плетіння на спицях, бо така робота не приносить мені задоволення, не є джерелом моїх доходів чи почуття власної гідності.

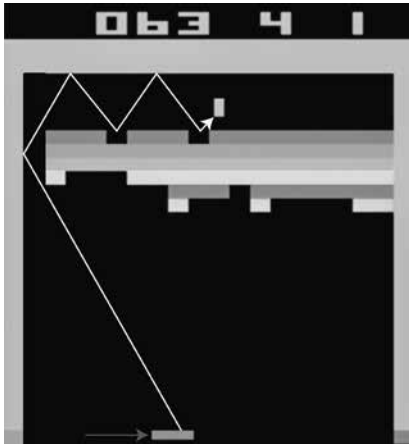
І справді, якщо я і мав якісь ілюзії стосовно своїх здібностей у цьому плані, то вони жорстоко розбилися у восьмирічному віці, коли в школі мене змусили ходити на заняття з плетіння на спицях. Цей предмет я майже завалив і дов'язав роботу лише завдяки добросердй помічниці з п'ятого класу, яка змилювалася наді мною.

Але, з удосконаленням технології, чи затьмарить поява ШІ також і ті вміння, які зараз забезпечують мені відчуття власної самоповаги й цінності на ринку праці? Стюарт Рассел розповів мені, що в нього і багатьох інших дослідників ШІ не так давно просто щелепа відпала, коли вони бачили, як ШІ робить щось таке, чого вони не чекали ще багато років. З огляду на це дозвольте розповісти вам кілька таких моментів, коли відпадала вже моя щелепа, і чому я вважаю їх передвісниками появи можливостей, сильніших за людські.

Прориви

Агенти глибинного навчання з підкріпленням

У 2014 році я з розкритим ротом дивився відео, як система штучного інтелекту DeepMind учиться грати в комп'ютерні ігри. Якщо конкретніше, то ШІ грав у Breakout (малюнок 3.1) — класичну гру від компанії Atari, яку я любив ще підлітком. Ціль —



Малюнок 3.1. Гра ШІ у Breakout

Навчившись грати у Breakout від Atari практично з нуля, з використанням глибинного навчання з підкріпленням для досягнення максимального рахунку, штучний інтелект DeepMind відкрив оптимальну стратегію: продірявити крайню ліву частину цегляної стіни й відбивати м'ячик за нею, дуже швидко набираючи бали. Я намалював стрілки, щоб показати попередні траєкторії м'ячика й ракетки.

ВИБУХ ІНТЕЛЕКТУ?

Якщо машина буде здатна думати, то рівень її інтелекту може виявитися набагато вищим за наш, і куди тоді подітися нам? Навіть якби ми могли тримати машини в покорі... як вид ми відчували б себе дуже приниженими.

*Алан Тюрінг, англійський математик, логік,
криптограф, 1951*

...Надрозумна машина — останній винахід, який узагалі доведеться зробити людині, якщо тільки ця машина буде достатньо слухняною, щоб повідомити нас, як тримати її під контролем.

*Ірвінг Гуд, англійський математик
і дешифрувальник, 1965*

Позаяк ми не можемо повністю відкинути можливість створення колись ЗШІ людського рівня, поміркуймо в цьому розділі, куди це нас може привести. Почнімо з головного питання, яке всі оминають:

Чи може ШІ справді захопити світ або допомогти людям це зробити?

Якщо ви закочуєте очі під лоба від одних лише розмов про озброєних до зубів роботів-термінаторів, які захоплюють владу над світом, ви маєте рацію: це і справді нереалістичний і дурний сценарій. Ці голлівудські роботи не такі вже й розумніші за нас, і їм навіть не вдається домогтися свого. На мою думку,

небезпека історії про Термінатора не в тому, що вона справді може трапитися, а в тому, що вона відволікає нас від реальних ризиків і можливостей, які несе ШІ. Щоб справді перейти від наших днів до захоплення влади над світом за допомогою ЗШІ, потрібно зробити три логічні кроки:

- Крок 1. Створити ЗШІ людського рівня.
- Крок 2. За допомогою цього ЗШІ створити суперінтелект.
- Крок 3. Застосувати цей суперінтелект або дозволити йому захопити владу над світом.

У попередньому розділі ми побачили, що важко назвати перший крок абсолютно неможливим. Ми бачили також, що якщо цей перший крок буде зроблено, то важко назвати безнадійним другий крок, тому що ЗШІ буде здатен рекурсивно створювати дедалі кращий ЗШІ, межею для якого будуть лише закони фізики (які, схоже, допускають інтелект, набагато вищий від людського рівня). Зрештою, якщо ми, люди, спромоглися панувати над іншими формами життя на Землі завдяки інтелектуальній перевазі, то чому суперінтелект не може аналогічно підкорити нас і панувати над нами?

Утім ці аргументи на користь правдоподібності такого сценарію страшенно розпливчасті й неконкретні, а диявол ховається в деталях. То чи може ШІ *справді* спричинити захоплення влади над світом? Щоб дослідити це питання, забудьмо про тих безглузких термінаторів і розгляньмо натомість кілька детальних сценаріїв можливого розвитку подій. Після цього ми детально проаналізуємо ці сюжетні лінії, пошукаємо в них слабкі місця – тож, будь ласка, читайте їх з певною дозою скепсису. Адже вони головним чином показують, що ми зовсім не маємо уявлення про те, що трапиться, а що – ні, і що спектр можливостей надзвичайно широкий. Наші перші сценарії розташовуються в найдинамічнішому і найдраматичнішому кінці спектра. На мою думку, деякі з них варто дуже добре розглянути в деталях – не тому, що вони найімовірніші, а тому, що раз ми не можемо переконати себе в тому, що існує надзвичайно мізерний шанс їх утілення, то нам потрібно зрозуміти їх так добре, щоб завчасно вберегтися і не дати їм призвести до негативних наслідків.

НАШ КОСМІЧНИЙ ДАР: НАСТУПНИЙ МІЛЬЯРД РОКІВ І ПІСЛЯ НЬОГО

Наші уможлядні міркування закінчуються суперцивілізацією, синтезом усього життя в Сонячній системі, яке безперервно самовдосконалюється і поширюється, віддаляючись від Сонця і перетворюючи безживні простори на свідомість.

Ганс Моравек. Діти розуму

З усіх наукових відкриттів мене особисто найбільше надихає те, що ми дуже сильно недооцінюємо майбутній потенціал життя. Наші мрії й прагнення не обов'язково повинні обмежуватися протяжністю життя в одне століття, затьмарене хворобами, бідністю і метушнею. Навпаки, за допомогою технологій життя має потенціал процвітати мільярди років, і не просто тут, у нашій Сонячній системі, а й скрізь у космосі – набагато грандіознішому, ніж уявляли собі наші предки. Адже можна піднятися навіть вище неба.

Це чудова новина для виду, який протягом віків шукає способів розширювати свої межі. На Олімпійських іграх розширюють межі сили, швидкості, вправності й витривалості. У науці розширюють межі знання й розуміння. Література і мистецтво розширюють межі створення прекрасного або переживань, що збагачують життя. Багато людей, організацій і держав розширюють межі, нарощуючи ресурси, збільшуючи території й подовжуючи тривалість життя. Знаючи про одержимість людей межами, не дивуєшся, що найперший бестселер усіх часів се-

ред захищених копірайтом книжок – «Книга світових рекордів Гіннеса».

Тож якщо технології можуть похитнути наші старі уявлення про межі тривалості життя, які його *крайні* межі? Скільки нашого космосу може ожити? Як далеко здатне сягнути життя і як довго протривати? Скільки матерії може використати життя і скільки енергії, інформації й обчислень – отримати? Ці крайні межі встановлює не наше розуміння – їх встановлюють закони фізики. Хоч як парадоксально, але через це аналізувати віддалене майбутнє життя де в чому легше, ніж його майбутнє у ближчій перспективі.

Якби ті 13,8 мільярда років нашої космічної історії стиснути в один тиждень, то напружені десять тисячоліть двох попередніх розділів не протривали б і півсекунди. А значить, хоча ми й не можемо передбачити, як відбудуватиметься інтелектуальний вибух (якщо він узагалі відбудеться) і якими будуть його безпосередні наслідки, уся ця метушня – просто короткий спалах у космічній історії, і її подробиці не впливають на крайні межі життя. Якщо життя після вибуху буде таке само одержиме, як нинішнє людство одержиме розширенням своїх кордонів, то воно розвине технологію, щоб і справді *досягнути* цих кордонів – тому що може. У цьому розділі ми розглянемо, що ж воно таке ці межі і зазирнемо у віддалене майбутнє життя. Позаяк ці межі спираються на наше нинішнє розуміння фізики, їх слід вважати нижньою межею можливостей – наукові відкриття майбутнього можуть дати нам можливості зробити це ще краще.

Та хіба ми можемо бути впевнені в тому, що майбутнє життя матиме такі амбіції? Зовсім ні. Можливо, воно вдовольнятиметься малим, як героїновий наркоман або ледар, який цілими днями на дивані не відривається від реаліті-шоу. Проте є причина підозрювати, що амбітність – це риса, притаманна передовим формам життя взагалі. Майже безвідносно до того, що воно намагатиметься максимізувати – інтелект, довголіття, знання або цікаві переживання, – йому будуть потрібні ресурси. А відтак воно має стимул розвивати свої технології до максимуму і якнайповніше використовувати наявні ресурси. Після цього єдиний спосіб про-

довжувати вдосконалюватися – знаходити нові й нові ресурси, поширюючись і займаючи дедалі більші частини космосу.

Життя може також виникати незалежно в різних місцях у Всесвіті. У цьому разі позбавлені амбіцій цивілізації просто стають неважливими в космічному масштабі, а амбітніші форми життя захоплюватимуть дедалі більші частини космічного «капіталу». У такий спосіб відбувається природний добір у космічному масштабі. Через деякий час майже все наявне життя складатиметься з амбітних форм. У підсумку, якщо нас цікавить, наскільки може ожити космос, ми повинні звернути увагу на межі амбіцій, накладені законами фізики. Що ж, зробимо це. Спочатку дослідимо межі того, що можна зробити за допомогою ресурсів (матерії, енергії тощо) нашої Сонячної системи, тоді перейдімо до того, як отримати нові ресурси завдяки освоєнню і заселенню космосу.

МАКСИМАЛЬНО ЕФЕКТИВНЕ ВИКОРИСТАННЯ РЕСУРСІВ

Якщо нинішні супермаркети й товарні біржі продають десятки тисяч товарів, які можна назвати «ресурсами», майбутнє життя, досягнувши технологічної межі, потребуватиме здебільшого одного фундаментального ресурсу – так званої баріонної матерії: усього, що складається з атомів або їхніх компонентів (кварків і електронів). Хоч би в якій формі перебувала ця матерія, передова технологія може реорганізувати її в будь-які бажані субстанції або об'єкти, зокрема електростанції, комп'ютери й передові форми життя. Тому почнімо з розгляду меж для енергії, що живить високорозвинуте життя та уможливорює обробку інформації, завдяки чому воно здатне мислити.

Будівництво сфер Дайсона

Коли мова йде про майбутнє всього життя, то в цьому плані одним із найбільших візіонерів-оптимістів є Фрімен Дайсон. Я маю честь і приємність знати його останні два десятиліття років,