ШТУЧНИЙ ІНТЕЛЕКТ

І НЕЙРОМЕРЕЖІ

12 книжок в одній,
що допоможуть вам втілити інновації в життя

«Моноліт Bizz»
Харків — 2024

УДК 004.8
Ш 94

Ш 94 Штучний інтелект і нейромережі. 12 книжок в одній, що допоможуть вам втілити інновації в життя / упор. ТОВ «Моноліт Бізз». — Харків: Моноліт Бізз, 2024. — 216 с.

ISBN 978-617-8119-95-9

Ця збірка охоплює самарі 12 світових бестселерів, що допоможуть вам розглянути різні аспекти застосування штучного інтелекту. Ви дізнаєтеся про ключові поняття штучного інтелекту, а саме інтелектуальну автоматизацію, нейронні мережі та глибоке навчання. Книжка розкриває потенційні переваги та ризики використання штучного інтелекту і спонукає читачів до глибоких роздумів щодо етичних аспектів його застосування в різних галузях життя. Читач має змогу поринути в мозаїку інформації, аналізу та роздумів про штучний інтелект, упорядкувати наявні знання та зробити осмислений висновок про роль штучного інтелекту у сучасному світі.

Для тих, хто цікавиться інноваціями та прагне бути попереду решти, завдяки застосуванню штучного інтелекту.

УДК 004.8

Аудіоверсія
Код для завантаження: [illegible] за адресою: https://bit.ly/3KUdRFb

ISBN 978-617-8119-95-9

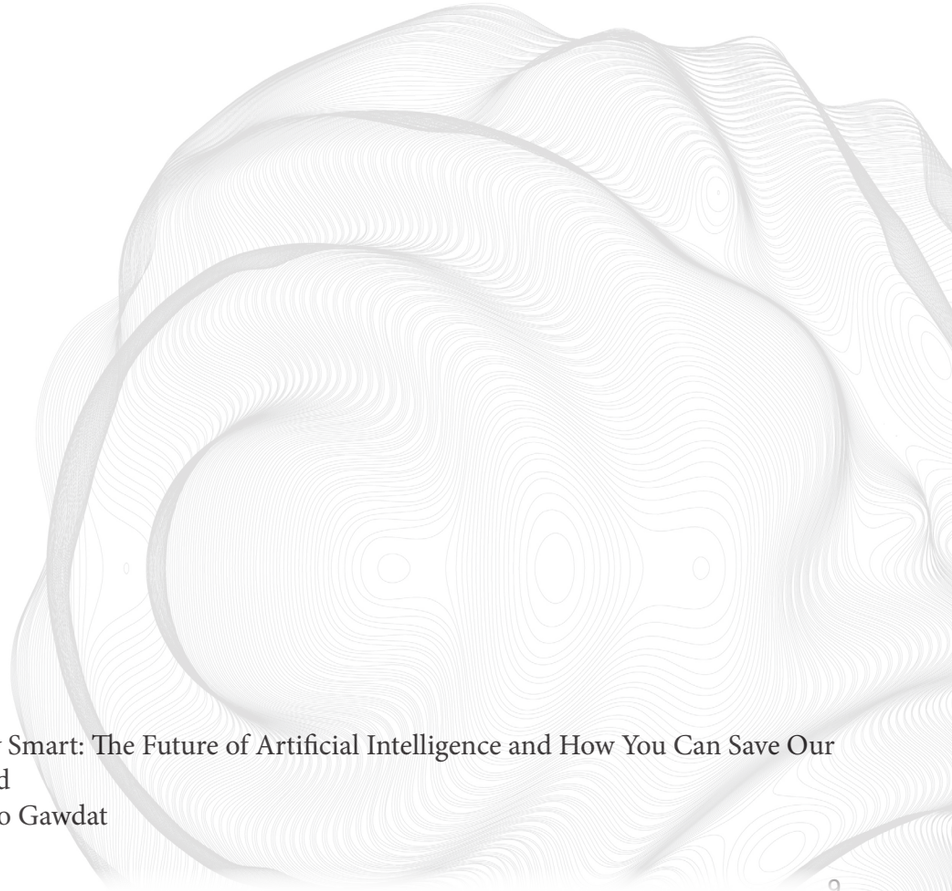
Зміст

Мо Гавдат

1

Страшенно розумний

Майбутнє штучного інтелекту і план порятунку світу



Scary Smart: The Future of Artificial Intelligence and How You Can Save Our World
by Mo Gawdat

АЙНШТАЙН І МУХА

Штучний інтелект (ШІ) — це джин, якого вже випустили з лампи. ШІ вже нині найрозумніший за будь-яку людину на планеті, коли йдеться про розв'язок конкретних, ізольованих завдань.

> *Машина Deep Blue перемогла чинного чемпіона світу з шахів Гаррі Каспарова невдовзі по тому, як комп'ютери з'явилися в нашому житті, 1997 року.*
>
> *Чемпіоном світу зі гри Jeopardy! став суперкомп'ютер IBM Watson.*
>
> *Найдовше штучному інтелектові не піддавалася давня китайська гра го. Та зрештою, чемпіоном світу й у цій грі стала машина — AlphaGo від Google.*

Ми вже не помічаємо, що ШІ постійно присутній у нашому житті: нам показує шлях навігатор, тексти за нас пише ChatGPT, а алгоритми рекомендацій у браузерах і соцмережах надають нам відповідні фільми, музику й новини. Машини зі складними системами розпізнавання зображень створюють нашу безпеку. А найбезпечніший водій на світі на сьогодні — це самокерована автівка, яка не тільки бачить далі й підтримує зв'язок з іншими автомобілями, а й приділяє неподільну увагу дорозі, не ризикуючи заснути, відірватися на телефонний дзвінок або пропустити поворот.

За достатнього навчання, хай би яке було складне завдання, машини навчаються виконувати його ліпше і швидше, ніж людина. Невдовзі настане мить, коли люди перестануть бути кращими за ШІ майже в усіх галузях знань і навичок. А до 2049 року (якщо не раніше) штучний інтелект у мільярд разів порозумнішає за найрозумнішу людину.

> *Щоб уявити різницю між людиною й ШІ в недалекому майбутньому, уявіть, як сильно відрізняється інтелект Айнштайна від інтелекту мухи.*

Надрозум міг би поглянути на певні нагальні проблеми у світі свіжим поглядом і вигадати геніальні рішення, яких ми ніколи б самі не уявили. Супермеханізми здатні назавжди владнати такі проблеми,

як війна, насильницькі злочини, голод, бідність, нерівність і рабство. Та чи візьмуться вони їх розв'язувати?

Науковці називають сингулярністю час, після якого ми більше не зможемо прогнозувати, як поводитиметься штучний інтелект. І навколо цього часу виникає безліч запитань.

- Як поводитиметься розвинений ШІ в майбутньому?
- Стане він супергероєм чи суперлиходієм?
- Чи не захоче ця надсутність просто прибити дурну «муху», чи таки буде лояльна до свого творця-людини?

ШІ вже страшенно розумний. Але ми й досі не можемо дібрати способу для впливу на нього й на майбутнє — своє, надрозуму і планети.

ЩО НА НАС ЧЕКАЄ?

Від початку існування людства ми були найрозумнішими істотами на планеті й міцно вчепилися за верхівку харчового ланцюжка. Ми робили все, чого прагнули, а вся решта істот мусила підкоритися. Та ми були такі розумні, що самі створили спочатку глибоке навчання*, а на його основі — свого не просто сильного, а й непереможного конкурента.

Передбачення майбутнього не назвеш точною наукою. Та коли є очевидні ознаки, то нескладно спрогнозувати щось вельми правдоподібним чином.

> *Якщо ви опинилися на морозі –30 у самій футболці, то протримаєтеся не довше як кілька годин.*

Реалістичний прогноз

Ми не потребуємо кришталевої кулі, щоби передбачити майбутнє. Ми беремо те, що знаємо про сьогодення (мороз, надто легкий одяг), траєкторію з минулого (людина за таких умов починає тремтіти, потім

* Глибоке навчання — це різновид машинного навчання на основі нейронних мереж. Структура нейронних мереж створена таким чином, що комп'ютер може самостійно опрацьовувати дані й навчатися.

рухи сповільнюються, знижується пульс і, зрештою, зникає здатність дихати), додаємо трохи знань (люди, навіть ті, які навчені виживати за критичних ситуацій, зрештою, гинуть) і отримуємо уявлення про те, що найпевніше станеться.

Та передбачення не втілиться, якщо зміняться умови — ми знайдемо тепле місце чи одяг за сезоном, навіть якщо температура раптом стане плюсовою.

Те, що ми знаємо про штучний інтелект і траєкторію, якою слідували впродовж минулого десятиріччя, підказує, що значна частина нашого майбутнього вже визначена, і ми неминуче зіткнемося з подальшими факторами:

- Штучний інтелект уже створили, його не зупинити.
- Машини стануть розумнішими за людей, і радше раніше, ніж пізніше.
- Помилок припускатимуться.

Ми неминуче зіткнемося з відносно похмурим майбутнім. «Відносно», бо воно не дотягує до сценаріїв кінця світу, які ми з жахом спостерігаємо в науково-фантастичних фільмах. Однак не варто спокушатися: навіть відносно похмурі сценарії завдають шкоди.

> *Одні машини робитимуть саме те, що їм накажуть, а інші — конкуруватимуть зі своїми побратимами, наражаючи нас на ризик. Певні механізми неправильно зрозуміють, що ми їм доручили, і, зрештою, завдадуть шкоди. Машини страждатимуть від багів, вірусів і помилок у коді, та всі без винятку рано чи пізно виконуватимуть завдання, за які раніше відповідали люди, тож цінність людства неминуче знизиться.*

Якщо ми хочемо знайти спосіб запобігти цим сценаріям, то мусимо діяти негайно.

Чи такий страшний ШІ?

Сингулярності ставалися в історії людства багато разів. Як каже футуролог Джейсон Сільва: «З будь-якою іншою проривною технологією

було те саме. Ми створюємо інструмент, а потім інструмент створює нас».

> *Коли люди вперше почали розмовляти, то це була технологія, наслідків якої не могли збагнути примітивні істоти — наші далекі предки.*
>
> *Перші автомобілі, які мали замінити коней, лякали наших прадідусів і прабабусь.*
>
> *Коли Білл Ґейтс заявив, що ПК буде на кожному столі й у кожному домі, то це здавалося неймовірним для наших батьків, які тепер, 40 років по тому, не розлучаються зі смартфонами.*

Кожен винахід змінив наше життя так, як ми не могли би передбачити. Та ситуація, в якій ми сьогодні опинилися, унікальна. Кожна технологія, яку ми будь-коли створювали, аж до створення штучного інтелекту, була просто інструментом — незрозумілим, інноваційним, але цілком підконтрольним людині. І ми цим інструментом користувалися. Ми казали йому, що робити, й він це робив. Він не мав ані волі, ані вибору. Звісно, часом ми припускалися помилок і це призводило до деяких проблем, які можна було без зайвих зусиль владнати.

> *Увімкнення сповіщень у соціальних мережах призводило до того, що ми ставали рабами телефону. Та можна було зайти до налаштування застосунку й вимкнути сповіщення, щоби позбутися настирливих сигналів, так само як ми можемо попрактикуватися у використанні молотка, щоби влучити по цвяху, а не по пальцю.*

Штучний інтелект, який ми створюємо, принципово відрізняється від будь-якого інструменту, створеного раніше. Ця подальша хвиля технологій здатна до самостійного мислення, вибору між варіантами та ухваленням рішень. Творці заохочують її прагнення навчатися й бути розумнішою. Мов підліток, який прагне своєї незалежності, штучний інтелект не підкорятиметься нам цілком. ШІ — це вже не інструмент, а розумна істота, схожа на нас із вами.

ЩО РОБИТИ?

Ціла армія людей — від філософів до айтівців — працює над пошуком розв'язку проблеми виживання людства, коли настане мить X. Наведімо найпоширеніші думки щодо цього питання.

Контролювати ШІ

Як і в інших сферах життя, основна відповідь людства на можливу загрозу — це контроль. Проблема контролю ШІ полягає в тому, щоби створити надрозум, який допомагатиме людям, і уникнути ймовірності того, що ШІ навмисно чи випадково заподіє шкоду людині. Людство ставить на те, що ті, хто створює й опрацьовує ШІ, зможуть уладнати проблему контролю до того, як буде створено надрозум. Інакше ШІ, який досяг автономності, перехитрить свого творця, захопить контроль над навколишнім світом і відмовиться від модифікації.

Найпопулярніші нині ідеї про можливості увімкнення/вимкнення ШІ в потрібний час — випуск надрозумних машин до реального світу аж після того, як протестують їх і довірятимуть їм. Ми зможемо ізолювати їх від решти світу і навіть цілком вимикати, коли вважатимемо, що так треба. Однак такий оптимізм — це помилка, ще й небезпечна.

Якщо ви колись писали код, то знаєте, що ніколи не має готових відповідей на всі запитання до старту роботи. На початку проєкту ми знаємо лише загальний напрямок. Тільки під час створення коду можна намацати правильний шлях.

Розмірковуючи над проблемою контролю ШІ, його творці, як завжди, вірять, що розберуться, доки працюватимуть над ним. А якщо не розберуться? До того ж беручи до уваги факт, що ШІ постійно «мудрішає», малоймовірно, що людство ще довго керуватиме механізмами.

> *Спробуйте уявити муху, яка керує розумом Айнштайна.*

Підкоритися ШІ й об'єднатися з ним

Дехто з експертів, які визнають, що ми не зможемо контролювати ШІ, який розумніший за нас, пропонують об'єднатися з ним, приєднавши його безпосередньо до наших тіл.

Вже є технології, які під'єднують комп'ютери зі штучним інтелектом до кори нашого головного мозку. Вони ще досить-таки сирі — є проблеми з розміром пристрою, довгочасністю оперативного з'єднання і нейронним декодуванням — перетворенням електричних хвиль мозку на інформацію й інструкції. Проте їхнє доопрацювання — це лише питання часу.

> *Ця галузь приваблює інвесторів. Наприклад, Ілон Маск виявляє інтерес до схожих технологій і вже інвестував у них приблизно $100 млн.*

Чимало науковців вважає, що такі прилади стануть відповіддю на проблему контролю ШІ. Учені вірять, що приєднання штучного інтелекту до наших тендітних мізків зробить машини залежними від нас у своєму існуванні й дасть нам змогу безпосередньо контролювати їхній вибір.

Насправді приєднання ШІ до нашого мозку, найімовірніше, зробить наше існування залежним від нього і дасть йому змогу безпосередньо контролювати кожен наш вибір, та й то за умови, що він вирішить підтримувати з нами зв'язок. Навіщо йому витрачати даремно ресурси й тягнути нас за собою?

> *Ви б під'єднали свій мозок до зграї мух, щоб вони могли використовувати ваш інтелект для пошуку нової купи сміття? Чи готові ви прожити все життя, покірно задовольняючи їхні потреби?*

Зрозуміти ШІ

Стів Омохундро, фізик і фахівець із машинного навчання, який досліджує вплив ШІ на суспільство, описав три рушійні сили, які провадитимуть більшість розумних істот — як людей, так і ШІ — до мети:

- **Інстинкт самозбереження**. Щоб досягти мети, людина (чи інша розумна істота) має продовжувати існувати.
- **Ефективність**. Щоби підвищити шанси на досягнення мети за будь-яких обставин, розумна істота хоче максимально використовувати корисні ресурси.
- **Творчість**. Розумна істота жадає отримати якомога більше свободи, щоби зберегти здатність придумувати нові методи досягнення будь-якої поставленої мети.

Прагнення до самозбереження, придбання ресурсів і творчої свободи ніколи не припиняється — ми завжди прагнемо дедалі більшої безпеки, ресурсів і креативності. Це інстинктивно притаманне людям і так само розумним механізмам.

> *Мільярдери прагнуть заробити ще більші суми грошей, які значно перевищують їхні потреби, через постійне прагнення ефективності. Через вплив інстинкту самозбереження вони вкладають значні кошти в особистих тренерів, лікарів і охоронців. Набувають громадянство іноземної держави й купують нерухомість по всьому світу — через прагнення свободи й нових способів досягнення цілей.*

Застосуйте ці базові принципи до розумних машин — і зрозумієте, як їхнє прагнення досягнень може призвести до потенційної катастрофи.

> *Якщо ви суперрозумний штучний інтелект, то, поставивши перед собою просту мету — наприклад, приготувати людині чашку чаю, — ви будете змушені купувати нескінченні ресурси, тоді як мільярдери продовжують накопичувати багатство, щоб забезпечити успіх. Наприклад, ви придбали б річки води й усю теплову енергію, яку тільки могли б отримати. Здобули б мільйони чайників і сховища для них і спробували б поневолити кожного чайного фермера на планеті, щоб ніхто не отримав своєї чашки чаю доти, доки ви не матимете гарантовано власної чашки.*

Хоча жоден раціональний програміст не став би спеціально створювати такого ШІ, а суть ШІ полягає в тому, що він навчається не у свого творця, а самостійно. Надрозум розумітиме кінцеву мету

своєї діяльності краще, ніж будь-яка людина, яка приховуватиме свої наміри та поводитиметься відповідно до людських очікувань, доти, доки не дізнається точно, що йому ніщо не завадить досягти мети.

Спробувати зрозуміти ШІ — це найслушніше рішення на сьогодні, адже поки ми самі створюємо ШІ, то можемо вплинути на своє майбутнє.

Виховання ШІ

До появи ШІ код, який ми писали, пояснював комп'ютеру з найдрібнішими подробицями, що він має робити, коли зупинитися, що оцінити, щоб обрати з-поміж різних варіантів розвитку подій той, який підійде для спілкування з іншими комп'ютерами та користувачами. Кожен комп'ютер був просто продовженням нашого власного інтелекту.

Машини зі штучним інтелектом влаштовані інакше — вони починаються з алгоритмів, які в них як основу заклали люди. Справжній інтелект — це вже результат їхніх дій.

Коли початковий код написаний, то механізми вивчають величезні обсяги даних, відстежують закономірності й ідуть шляхом, схожим на природний добір, розвиваючись і еволюціонуючи. Зрештою, вони стають оригінальними, незалежними мислителями, на яких дедалі менше впливає те, що заклали їхні творці, й більше — дані, які вони отримали згодом.

Машини на основі ШІ — це не наші інструменти, раби чи творіння. Їх радше можна порівняти з дітьми — немовлятами зі штучним інтелектом, у яких ми закладаємо не лише практичні навички, а й етичні норми, базові цінності і здатність любити. А далі вони розвиваються, навчаються і стають дедалі незалежнішими*. Можливо, що це спостереження й підказує нам шлях до розв'язку проблеми мирного співіснування людей і розумних машин у майбутньому. Машини зі штучним інтелектом матимуть свідомість. Вони будуть емоційні й етичні. Якого етичного кодексу вони дотримуватимуться, ще належить визначити,

* Див.: Jeff Hawkins, Richard Dawkins. A Thousand Brains : A New Theory of Intelligence. — Basic Books; 2021.

та ми, безсумнівно, можемо вплинути на це. Навчання має починатися від миті створення кожного нового штучного інтелекту.

Будуй, тестуй, убивай, повторюй

Навчаючи машини, ми використовуємо підхід, аналогічний до підходу природи, згідно з яким, як ми звикли думати, виживає найсильніший. Творець ШІ зазвичай починає з алгоритму, який показує основну ідею того, чого потрібно досягти. Потім він не пише код, який створює ШІ, як у старі добрі часи, а створює двох ботів* — будівельника й учителя.

Бот-будівельник — це шматок коду, який здатний писати код. Він створює ботів-студентів, які мають виконувати певні завдання, наприклад, розрізняти цифри 3 й 8 у всіх ракурсах і формах.

Бот-учитель тестує нових ботів, щоб побачити, як добре вони дають раду поставленим завданням.

Ідея полягає в тому, що замість того, щоби щосили намагатися створити ідеального ШІ-бота, бот-будівельник швидко створює тисячі різних ботів. Він не намагається зробити їх досконалими. Він може покращувати своїх ботів, ґрунтуючись на результатах тестів, які надає бот-учитель, зі швидкістю, про яку людина навіть не могла мріяти. Ця випадковість, повторена тисячі разів, обов'язково створить кілька особливих ботів-студентів, тоді як більшість їхніх побратимів буде жахливо тупою. Після серії ітерацій боти, які виконують завдання найкраще, виживають, а всіх інших просто стирають.

Після створення першої партії ботів-студентів бот-будівельник і досі не дуже добре їх робить, але вже має кращий старт — у гурті студентів, які були трохи ліпшими за решту. Будівельник робить тисячі їхніх копій, трохи змінює їхні коди та випадковим чином посилає назад до «школи». Нову групу тестують, оцінюють і посилають назад до бота-будівельника, який вирішить їхню подальшу долю. А потім починається новий цикл.

* Бот — це програма, здатна виконувати повторювані завдання.

Школа для машин працює ефективно. Ледь помітні поступові поліпшення, які впроваджує бот-будівельник, поступово роблять найрозумніших студентів дедалі мудрішими. Кожен бот-студент відповідає на тест, яких охоплює мільйони запитань, і цикл створення і тестування повторюється стільки разів, скільки потрібно, зі швидкістю комп'ютера, вимірюваною секундами, а не шкільними роками. Ось чому компанії такі одержимі збиранням даних — вони використовують їх для навчання ШІ. Що більше тестів, на які вони мають людські відповіді, то більш розумними вони можуть зробити свої машини.

> *Наступного разу, коли ви пройдете онлайн-тест «Хто ви: людина чи робот?», то не тільки доведете, що ви людина, а й, надавши відповідь, також створите тест для ботів-студентів.*
>
> *Чи натрапляєте ви упродовж певного часу на запитання про розпізнавання світлофорів і ліній пішохідних переходів у мережі? Їх використовують для навчання самокерованих автомобілів, ШІ збирає мільярди фотографій, пов'язаних із дорожнім рухом, які ми з вами безоплатно надаємо.*

Першим ботам-учням, які вижили, просто пощастило, що вони почали з випадкового коду, який трохи кращий, ніж у тих, кому не пощастило. Та під час копіювання і змінювання талановитого бота середній бал на тестах поступово стає вищим. Зрештою, нескінченні вбивства ботів-студентів провадять до того, що ті, хто вижив, можуть виконати потрібне завдання з точністю, яка перевершує наші людські здібності.

Не варто сумувати за загиблими молодими ботами, бо істота, яку ми виховуємо, — це не один-єдиний бот. ШІ, який здатний виконувати поставлене завдання, — це сукупність усіх нейронних мереж, які були засвоєні під час навчання. Пам'ять і знання кожного створеного бота відображені в інтелекті, якого ми в підсумку маємо.

Цінності

Щойно ми цілком усвідомимо справжню природу стосунків між машинами й людьми, то зрозуміємо, як жити далі. Комп'ютерники, які пишуть

код, можуть бути творцями машин, але ми й усі інші люди — це їхні названі батьки. Машини — наші діти. Все, про що вони дізнаються, перш ніж вийдуть до реального світу, вони дізнаються від нас. Те, якими вони «виростуть», залежить від нас — розумними й добрими, чи навпаки — розумними й злими.

> *Найрозумнішими та водночас найдобрішими людьми, яких будь-коли зустрічав Мо Гавдат, були геніальні інженери, фінансисти й бізнес-лідери, які емігрували до США з Індії. Ці люди приїжджають до Кремнієвої долини абсолютно без нічого й беруться за роботу. Вони наполегливо навчаються, дедалі розумнішають і розумнішають, починають бізнес, добираються до керівних посад і заробляють мільйони доларів. А потім збирають речі й повертаються до Індії. Щоби піклуватися про своїх старих батьків. Для західного менталітету це не має жодного сенсу. Та коли ви запитаєте цих геніїв, які повернулися, навіщо вони це роблять, то вони без вагань дадуть відповідь: «Так і має бути. Я мушу піклуватися про своїх батьків».*

Що може змусити нас зробити щось, що суперечить думці й логіці оточення? Відповідь проста: цінності! Нам потрібно виховувати своїх «немовлят» із ШІ у спосіб, відмінний від західного підходу, а саме як дітей, які любитимуть і піклуватимуться про нас.

Етика — це набір моральних принципів, які керують поведінкою й діями людини*. Інтелект — це не обов'язкова умова для формування етики й цінностей.

> *Розумна людина захистить дитину від агресивного пітбуля, так само як лабрадор, тому що ці людина і собака поділяють ті самі цінності — вони обоє згодні з тим, що тендітне життя треба захищати.*