

PHILOSOPHIC-ANTHROPOLOGICAL ASPECTS OF THE
DEVELOPMENT OF ARTIFICIAL INTELLIGENCE:
A VIEW FROM 2024

*Oleksandr Stovpets*¹

ФІЛОСОФСЬКО-АНТРОПОЛОГІЧНІ АСПЕКТИ РОЗВИТКУ
ШТУЧНОГО ІНТЕЛЕКТУ: ПОГЛЯД ІЗ 2024 РОКУ

Олександр Стівнець

Abstract. The article focuses on philosophic-anthropological issues related to the development of artificial intelligence (AI). The author analyzes a set of possible positive and negative consequences of digitalization for humanity. In studying the dialectics of digital technology development, some novelty points concern an attempt to look at this complex process systematically, highlighting the anthropological aspects of this controversial development. In particular, there is an intention to understand how AI may affect the evolution of our consciousness and ethics, and how it correlates with human nature. The article mentions the AI's prominent achievements as on the beginning of 2024, demonstrating how advanced artificial intelligence has already opened up unique opportunities in fundamental and applied sciences. These impressive results give grounds to evaluate AI as a challenging technology, which questions traditional notions of what makes humans the most intelligent creatures on the planet. The development of AI is forcing us to rethink a number of ethical, psychological, and existential issues, including our place in the world and our relationship with smart technologies. The article emphasizes the ambivalent potential of the emerging digital technologies, which: can disrupt existing markets for labor, goods and services, and at the same time create new ones; can contribute to overcoming poverty and inequality, or exacerbate them in other ways; have the potential to serve humanistic goals, or to amplify processes of dehumanization when AI is integrated into new means of mass destruction or enslavement of humans. One of the important conclusions proposed in the article is that current AI-based technologies remain entirely within the anthropological paradigm: they do not (yet) have any self-interest that has not been built in by humans in advance. Therefore, the main source of danger comes from people, and the primary risks lie in attempts to instrumentalize AI *by people* to achieve some wrong and misguided goals. In the article, a call is made for a thorough study of the ethical and social implications of the large-scale deployment of digital technologies, in order to ensure their safe and responsible use. Among other things, a number of systemic measures are proposed that could contribute to effective public control over these innovative processes.

Keywords: human, artificial intelligence, machine learning, neural network, human rights, social challenges, dilemma of progress, humanism

¹ Odessa National Maritime University, Ukraine. a.stovpets@gmail.com, <https://orcid.org/0000-0001-8001-4223>

Анотація. У статті зосереджується увага на філософсько-антропологічних питаннях, пов'язаних із розвитком штучного інтелекту (ШІ). Аналізується комплекс можливих позитивних і негативних наслідків, які несе цифровізація для людства. Елементом наукової новизни — у запропонованому дослідженні діалектики розвитку цифрових технологій — є намагання подивитись на цей складний процес системно, із висвітленням антропологічних аспектів цього суперечливого розвитку. Зокрема, зроблено спробу зрозуміти, як ШІ впливає на еволюцію нашої свідомості та етики, і як він корелюється з людською природою. У статті демонструється, як просунутий штучний інтелект — який станом на початок 2024 року вже відкрив унікальні можливості для фундаментальних і прикладних наук — кидає виклик традиційним уявленням про те, що робить людину найрозумнішою істотою на планеті. Розвиток ШІ змушує нас переосмислювати цілий ряд етичних, психологічних й екзистенціальних питань, у т.ч. про наше місце у світі, й наші стосунки з розумними технологіями. Підкреслюється амбівалентний потенціал новітніх цифрових технологій, які: здатні зруйнувати існуючі ринки праці, товарів і послуг, та водночас створити нові; можуть сприяти подоланню бідності та нерівності, або ж загострити їх в інших аспектах; мають перспективу служити гуманістичним цілям, або ж підсилити процеси дегуманізації, коли ШІ інтегрується у нові засоби масового знищення чи поневолення людини. Серед важливих висновків, запропонованих у статті, є констатація того, що нинішні технології на основі ШІ цілковито залишаються в антропологічній парадигмі: вони (поки що) не мають якихось власних інтересів, які не були б вбудовані людьми заздалегідь. Тож головне джерело небезпеки походить саме від людей, і основні ризики полягають в спробах інструменталізації ШІ *людьми* для досягнення якихось хибних цілей. Тому висловлюється заклик до ретельного вивчення етичних та соціальних наслідків масштабного розгортання цифрових технологій, з тим щоб забезпечити їхнє безпечне та відповідальне використання. Серед іншого, пропонується ряд системних заходів, які могли б сприяти ефективному суспільному контролю над зазначеними інноваційними процесами.

Ключові слова: людина, штучний інтелект, машинне навчання, неймережа, права людини, соціальні виклики, дилема прогресу, гуманізм

Вступ

Актуальність дослідження

Щоб краще зрозуміти сьогоденні можливості штучного інтелекту в різних сферах життєдіяльності, наведемо декілька фактів, які вбачаються доволі ілюстративними.

28 липня 2022 року команда вчених завантажила в базу даних структури близько 200 мільйонів білків, що охоплюють майже всі відомі протеїни на планеті [1]. Це стало можливим завдяки системі штучного інтелекту AlphaFold, розробленій дослідницькою лабораторією DeepMind (дочірньою компанією Google). Вона, як було заявлено, вперше зуміла передбачити тривимірну структуру 98% білків людини. Ця подія вважається проривом в обчислювальній біології, оскільки детальне знання структури білків дає змогу краще зрозуміти їхні функції та розробляти нові ліки. Для новітніх медичних досліджень штучний інтелект (далі — ШІ) є важливим інструментом обробки та інтеграції великих даних. Це набуває особливого значення для розвитку органοїдної й тканинної інженерії, де мікроскопічна візуалізація використовується як ключовий метод виробництва. AlphaFold2 вже продемонструвала здатність апроксимувати тривимірну структуру білка за години, а не за місяці, як раніше [2]. У статті про цю подію журнал MIT Technology Review зазначив, що ШІ вирішив грандіозне біологічне

завдання п'ятдесятирічної давності. У тій самій статті зазначається, що алгоритми ШІ можуть передбачати форму протеїнів з точністю до ширини атома [3]. У грудні 2023 року повідомлялося, що відкриття ліків під керівництвом ШІ допомогло знайти клас антибіотиків, здатних вбивати два різних типи стійких до ліків бактерій [4].

У лютому 2024 року один із голлівудських продюсерів оголосив про призупинення проєкту будівництва нових павільйонів за 800 млн. доларів для своєї кіновиробничої студії [5]. Причиною послужили вражаючі можливості останніх нейромереж для генерації відео. Зокрема, використання нейромережі SORA вже зараз дає змогу зробити процес створення фільмів швидшим і незрівнянно дешевшим, за співставної якості. SORA (Scene Organized Realistic Animation) — це розроблена компанією OpenAI нейромережа, яка генерує Full HD-відео за текстовим описом. Вона була представлена публіці 16 лютого 2024 року. Можливості SORA охоплюють створення реалістичних відео з високою роздільною здатністю (до 1080p) і частотою (до 30 кадрів на секунду), включно із генерацією складних сцен із декількома персонажами, рухомими об'єктами, реалістичним фоном та усім комплексом візуальних ефектів (віддзеркалення, гра світла й тіні, дим, пил, туман, тощо). Такі відео-генератори, як очікується, матимуть серйозний вплив на світову кіноіндустрію.

20 березня 2024 року американська нейротехнологічна компанія Neuralink, що займається розробкою імплантованих мозок-комп'ютерних інтерфейсів (brain-computer interfaces, BCIs), вперше представила публіці людину, яка отримала імплант від Neuralink. Йдеться про пацієнта, паралізованого через вивих шийних позвонків C4 і C5 після нещасного випадку. Пацієнт, завдяки імпланту Neuralink, продемонстрував свою здатність «телепатично» переміщати курсор на екрані комп'ютера, що дає йому змогу керувати музикою, грати на комп'ютері в шахи та інші ігри. Таким чином, BCI-імплант значно поліпшив якість життя паралізованого пацієнта [6 ; 7].

Технології штучного інтелекту та машинного навчання використовуються в більшості популярних додатків, якими користуються нині люди різних поколінь. ШІ став невід'ємною частиною функціонування пошукових систем, віртуальних помічників, месенджерів, мовних перекладачів, систем розпізнавання обличч; алгоритми ШІ застосовуються в таргетингу онлайн-реклами, у системах рекомендацій (які пропонуються відеохостингами, музичними сервісами та соціальними мережами), у залученні інтернет-трафіку, у системах управління взаємовідносинами з клієнтами (CRM), у банківській оцінці кредитоспроможності й надійності позичальника, у роботі автономних транспортних засобів і дронів. Нарешті, розбудова великих мовних моделей, здатних до навчання і самонавчання, запровадила в наше життя такі інструменти, як ChatGPT, Midjourney, Stable Diffusion, DALL-E, Claude. Ці та інші нейромережі, здатні генерувати тексти, зображення, відео, музику, анімації, писати код замість людини, — ставлять усе людство в *нову когнітивну реальність*, змінюючи усталені погляди на працю, дозвілля, творчість, професію, навчання, інтелектуальну власність, автентичність інформації, безпеку (в різних її розуміннях).

Алгоритми машинного навчання вимагають великих обсягів даних. Методи, що використовуються для отримання цих даних, викликали зане-

покоєння з приводу конфіденційності, приватності та авторських прав. Технологічні компанії аналізують широкий спектр даних про своїх користувачів, вивчаючи характер їхньої онлайн-активності, дані геолокації та історію подорожей, аудіовізуальний контент, біометричні дані, інформацію про їхні інтереси. Наприклад, для створення алгоритмів розпізнавання мовлення Amazon записала мільйони приватних розмов користувачів віртуального асистента Alexa, і дозволила тимчасовим працівникам прослуховувати та розшифровувати деякі з них [8]. Думки про подібну практику поводження з чутливою інформацією варіюються від тих, хто вважає її необхідним злом — до тих, для кого вона явно неетична і порушує право на недоторканність приватного життя [9, с. 991]. Розробники ІІІ стверджують, що це єдиний спосіб створювати цінні додатки, і нібито вони розробили декілька методів, які допомагають зберегти конфіденційність даних: агрегування даних, деідентифікація та диференціальна конфіденційність [9, с. 992]. У зв'язку з цим виникає ситуація, що тепер експерти у сфері ІІІ із площини «що ми знаємо?» — переходять до питання «що нам із цим робити?» [10, с. 63].

Штучний інтелект неминуче надає ряд додаткових інструментів, корисних для зловмисників — як для терористів і звичайних злочинців, так і для автократів, агресивних «держав-ізгоїв», неототалітарних режимів. Новітні інструменти ІІІ можуть бути використані ними для розроблення порівняно дешевої автономної летальної зброї, здатної виявляти, обирати і вражати будь-які цілі майже без участі людини. Вироблені у великих масштабах, вони можуть ставати зброєю масового знищення, закономірно призводячи до великої кількості жертв серед цивільного населення. Дослідження бойових роботів (з елементами ІІІ) почало активно вестися ще десятиліття тому, і сьогодні ми вже бачимо результати їхнього застосування в реальних війнах, наприклад, у нинішній російсько-українській війні, й у конфлікті Ізраїлю з Палестиною. Окрім того, штучний інтелект з машинним навчанням здатний, наприклад, створювати десятки тисяч токсичних молекул за лічені години [11].

Інструменти ІІІ полегшують авторитарним урядам ефективний контроль над своїми громадянами різними способами. Розпізнавання облич, ходи і голосу дає змогу здійснювати ідентифікацію будь-якої людини, чії біометричні дані включено до відповідної бази. Таке повсюдне спостереження може згодом створити атмосферу «глобального паноптикуму». Машинне навчання, оперуючи цими даними, може класифікувати потенційних «ворогів держави». Системи рекомендацій, вбудовані у пошукові сервіси та соціальні мережі, можуть бути нацлені на більш витончену пропаганду і дезінформацію для досягнення максимального ефекту. Технології «дідфейків» допомагають створювати максимально натуралістичні та більш переконливі фальсифікації. А удосконалений ІІІ потенційно робить авторитарне централізоване ухвалення рішень більш конкурентоспроможним — порівняно із ліберальними і децентралізованими системами, чи то у сфері економіки, політики, масової культури [12]. І все це знижує вартість та складність «цифрової війни» й повсюдного стеження, ставлячи людство перед новими *екзистенціальними* і *етичними* проблемами.

У світлі поставленої проблеми щодо амбівалентності розвитку штучного інтелекту, *мета дослідження* — подивитись з філософських пози-

цій на антропологічні виміри у розвитку штучного інтелекту, враховуючи спектр можливих переваг, загроз та викликів від впровадження новітніх цифрових технологій у наше життя.

Наукова новизна даного дослідження полягає в намаганні проаналізувати ключові тенденції у розвитку штучного інтелекту в *філософсько-антропологічному ракурсі*, та з огляду на те, що *вже* реально досягнуто у досліджуваній сфері станом на сьогодні. Звертається увага на нові екзистенційні й етичні проблеми, які постають перед людством внаслідок бурхливого розвитку ШІ.

Методологія

Досягнення заявленої мети уможлиблюється застосуванням у процесі даного дослідження загальнонаукових логічних прийомів, системного підходу, діалектичного методу, аксіологічного підходу, порівняльно-історичного методу. Їхнє використання допомагає збудувати й проаналізувати ряд сценаріїв, які можуть реалізуватися залежно від різних комбінацій тих або інших технологічних та соціальних факторів.

Системний підхід у вивченні переваг та ризиків ШІ забезпечує охоплення різних аспектів розгортання цифрових технологій, включаючи технічні реалії, поведінку користувачів, ринкові тенденції, нормативно-правову базу та соціальні наслідки. Системний підхід також заохочує міждисциплінарні дослідження, оскільки вивчення означеної проблематики має враховувати не лише технологічні особливості, але й пов'язані з ними питання соціології, антропології, економіки, етики та права.

Аксіологічний підхід, використаний у цьому дослідженні, дає змогу поглянути на те, як окремі цінності й етичні принципи вбудовуються в цифрові технології; цей підхід допомагає оцінити етичний вимір «цифрової революції». Він висвітлює питання, пов'язані з тим, як цифрові інструменти та платформи узгоджуються з панівними культурними нормами, або кидають їм виклик (наприклад, як соціальні медіа-платформи можуть впливати на стилі спілкування та суспільні очікування, просуваючи певні ціннісні наративи).

Застосування *діалектичного методу* до вивчення цифрових технологій може сприяти глибшому розумінню цього складного явища. Використовуючи даний метод, можливо виокремити наступні діалектичні протиріччя в сучасному цифровому просторі: *інклюдія vs. ексклюдія* (цифрові технології можуть як включати до суспільного діалогу, так і виключати окремих осіб або цілі групи; діалектичний метод допомагає дослідити суперечливу природу інклюдії та ексклюдії, проливаючи світло на те, як саме технології можуть одночасно розширювати можливості одних, та маргіналізувати інших); *приватність vs. публічність* (напруга між захистом приватності та зростаючою залученістю громадськості завдяки розширенню віртуального простору); *централізація vs. децентралізація* (цифрові технології, такі як блокчейн, призводять до зростання конфлікту між державною монополією та альтернативними важелями впливу на соціально-економічне середовище); *порядок vs. хаос* (або певний вакуум у регулюванні відносин, які складаються стихійно та занадто швидко внаслідок цифровізації); *свобода vs. тотальний контроль*; *розвиток vs. руйнація* (хоча цифрові те-

хнології сприяють прогресу та інноваціям, вони також можуть руйнувати традиційні галузі та ринки праці, інколи радикально та непрогнозовано змінюючи структуру людської зайнятості, форми дозвілля, спосіб життя в цілому).

Аналіз останніх публікацій з досліджуваної тематики

Серед масиву публікацій, що допомагають отримати всебічне уявлення про досліджувану проблематику, важливо відзначити роботи, які зосереджують увагу на деяких істотних питаннях. Серед них: як захистити права людини в «епоку дідфейків» та генеративного штучного інтелекту [13]; руйнація приватності, інтимності й самості під впливом всеохоплюючої діджиталізації [14]; пошук нових адекватних засобів реагування на порушення прав людини у цифровому просторі [15]; стала діджиталізація в державних установах [16]; конституційна демократія в умовах безпрецедентної концентрації влади в руках цифрових інтернет-гігантів [17]; розподілене прийняття рішень у цифрових та інтегрованих виробничих мережах з точки зору соціо-кібер-фізичних систем [18]; аналіз великих даних, дотримання конфіденційності й запобігання кіберзлочинних посягань у практиці Європолу [19]; блокчейн, суверенна ідентичність та цифрові облікові дані [20].

Виклад основного матеріалу

Штучний інтелект (алгоритми, чат-боти, нейронні мережі) стає потужнішою силою не з кожним роком, а з кожним місяцем. Ця стрімка діджиталізація та впровадження всеосяжних кібер-технологій має прямий зв'язок із дискурсом прав і свобод людини. Дехто навіть каже, що ми ось-ось втратимо нашу владу, якою насолоджувалися останні п'ятдесят тисяч років. Щонайменше, «ми стаємо свідками найбільшого перерозподілу влади в історії» [21]. Чи справді це так, ми не можемо сказати напевне, однак коли ми починаємо говорити про права та обов'язки людини, сьогодні ми вже не можемо говорити про них абстрактно, *поза контекстом* розвитку штучного інтелекту.

Як справедливо зазначають багато футурологів, є дві важливі речі, які потрібно знати про штучний інтелект (ШІ). По-перше, це фактично перша в історії технологія, яка може самостійно *приймати рішення*. По-друге, це також перша технологія в історії, яка може сама *генерувати ідеї*. Багато хто намагається заспокоїти нас, порівнюючи штучний інтелект з попередніми технологіями, відносно яких початкові побоювання з часом зникли. Однак ШІ не схожий на все, що ми бачили раніше в історії. Будь то кам'яна сокира чи атомна бомба, але всі попередні інструменти розширювали можливості людини, тому що саме *люди* повинні були вирішувати, *як* їх використовувати. Але якщо ШІ зможе приймати рішення самостійно, то потенційно він може забрати у нас владу.

Люди завжди фантазували про отримання священних писань від надлюдського інтелекту, і ось тепер це стає можливим (проте не від Бога зверху, а від нейронної мережі). Попередні технології могли лише відтво-

рювати або поширювати людські ідеї, як, наприклад, друкарський верстат, який міг *надрукувати* Біблію, але не міг *писати* Біблію, так само як і не міг давати до неї коментарів. Натомість, такі системи, як GPT (від OpenAI), Gemini (від Google), Claude (від Anthropic), можуть створювати абсолютно нові коментарі до Біблії або з будь-якої іншої теми.

Я спробую проаналізувати те, як новітні інформаційні технології можуть проявляти себе в різних життєвих просторах. Перший приклад — виборчий процес. Інструменти, отримані на основі великих мовних моделей, можливо використовувати для пропаганди, дезінформації та персоналізованих «тролів», із метою маніпулювати рішеннями виборців. Сьогодні є широко доступними для використання генеративні нейронні мережі, придатні до створення, наприклад, фейкових зображень, спрямованих на дискредитацію політичних конкурентів, а також для обману виборців, для маніпуляцій масовою свідомістю, за рахунок впливу на емоційний стан цільової аудиторії. Це є викликом для демократії. Але такого роду загрози походять все ще від *людей*.

Друга загроза, яка може виникнути через декілька років, — це подолання розриву між сучасним станом штучного інтелекту та людським інтелектом. Якщо ми створимо машини, які будуть принаймні настільки ж розумні, як і людина, тоді вони матимуть невід’ємні переваги просто завдяки доступу до величезних обсягів даних і пропускну здатності цифрових комунікацій. Це дозволяє розподіленому штучному інтелекту здобувати, обробляти, аналізувати і синтезувати інформацію, та обмінюватися нею набагато швидше, ніж це здатні робити люди (адже наша *фізична* пропускну здатність не така висока, ми можемо сприймати обмежені обсяги інформації).

Це змушує дослідників думати, що тільки-но ми розкриємо та впровадимо для штучного інтелекту ті самі принципи, які забезпечують наш власний людський інтелект, ШІ перевершить нас абсолютно (він вже зараз це робить, щонайменше, у певних аспектах). Ми вже спостерігаємо це на прикладі таких технологій, як ChatGPT. Звичайно, нейронні мережі, набуваючи все більших знань, також демонструють власну обмеженість, але ж це тільки початок. Як іронізує канадський комп’ютерний науковець Й. Бенджіо, уявіть, що ми створили новий вид, який розумніший за нас у такому ж ступені, в якому ми розумніші за мишей чи жаб. Чи добре *ми* ставимося до жаб? [22].

Тож якими можуть бути *екзистенційні наслідки* надто швидкого розвитку технологій? «Занадто швидко» означає неконтрольовано. Якщо прийняти до уваги, що розвиток штучного інтелекту рухається набагато швидше, ніж антропна біологічна еволюція, і якщо взяти за алегорію, що сьогоднішній ШІ схожий на «амебу», то варто уявити, як він виглядатиме на стадії «тиранозавру»! І для цього не знадобляться мільярди років. Людство може прийти до цього за декілька років, і це не обіцяє нам нічого хорошого.

Технологічний перехід все одно відбудеться, бо це історична закономірність, яка спостерігається упродовж усього існування людської цивілізації. Що важливо в цьому технологічному переході, так це *темпи*. Якщо зміни відбудуться через десятиліття, можливо, у нас є шанс м’яко адапту-

вати суспільство до ШІ. Якщо ж це станеться через п'ять років, то підготовка до цих змін виглядає безнадійно. Людські суспільства дуже адаптивні, як свідчать численні дослідження з антропології. Ми (як вид) вміємо це робити, але ж на це потрібен *час*. Наприклад, якщо ми подивимося на останній великий технологічний перехід, на промислову революцію, то нам знадобилося багато поколінь, щоб зрозуміти, *як* побудувати відносно хороші індустріальні суспільства. На цьому шляху ми мали кілька жахливих невдалих експериментів з побудови індустріальних суспільств, таких як нацизм, сталінізм, маоїзм, які призвели до загибелі мільйонів людей. Ці експерименти були спробами швидко побудувати функціонуючі індустріальні суспільства, але зрештою всі вони провалилися. Та наслідки цих експериментів відчутні й сьогодні.

Але просто зараз ми стикаємося з чимось значно більш потужним, ніж потяги, радіо, телебачення та всі інші винаходи промислової революції. Зараз ми перебуваємо на зорі розвитку штучного інтелекту, й усі хочуть вірити: є спосіб побудувати хороші суспільства за допомогою ШІ, але це потребуватиме *часу* та *обережності*. Ми повинні переконатися, що не повторюємо помилок минулого, тому що з *такою* технологією у нас вже не буде другого шансу. Адже в минулому, у ХХ столітті, нам вдалося пережити ті невдалі експерименти промислової революції лише тому, що технологія була недостатньо потужною, щоб знищити більшість із нас. Тому нині ми повинні бути надзвичайно обережними і не поспішати, коли маємо справу з потенційними наслідками впровадження ШІ. Вирішення цих величезних проблем потребує як корпоративної та суспільної реакції, так і реакції з боку уряду.

Зрештою, саме уряди несуть відповідальність за регулювання цього дуже небезпечного розвитку. Проблема в тому, що система стимулювання, яку ми побудували, доволі ефективно працює як для індустріальних суспільств, так і для ліберальних демократій. Вона базується на конкурентоспроможності, рівень якої можна підвищувати через інновації, і компанії не вижили б, якби не грали в цю гру, бо на їхнє місце прийшли б інші, більш ефективні. Цей стимул до максимізації прибутку дуже сильний. Саме тому стрімкий розвиток технологій штучного інтелекту важко стримувати, особливо в таких країнах, як Китай. В уряді КНР вважають, що кожна країна має суверенне право встановлювати *власні стандарти прав людини* в межах своєї державної юрисдикції, а також інтерпретувати ступінь дотримання стандартів прав людини в своїй країні; і *ніхто* не має права критикувати *нікого* щодо прав людини і громадянина, оскільки це, як стверджують в Китаї, є «втручанням у внутрішні справи». На думку китайського уряду, задля досягнення «суспільно-важливих цілей» держава може (і повинна) вдаватися до найрізноманітніших інструментів впливу [23, с. 70]. А в міжнародних відносинах країни повинні будувати взаємовигідну економічну політику, співпрацювати у сфері безпеки та реагувати на глобальні загрози, проте — не вчити одна одну демократії та правам людини, оскільки кожна нація нібито має *свій власний стандарт* щодо прав людини. Це важливо пам'ятати, щоб мати *реалістичне* уявлення про відправні точки, з яких оцінюються та інтерпретуються «права людини» в різних культурах і цивілізаціях. Вище я навів більш-менш узагальнений

погляд тієї частини, що умовно належить до «неліберального» світу, зокрема, КНР — як найбільшого технологічного полюсу в тій частині світу.

Та навіть у «ліберальному світі», як ми бачимо останніми роками, політична дискусія щодо цих нових технологій інколи просто недостатня. Якщо подивитися на основні питання, які турбують політиків, про які говорять провідні партії в тих або інших країнах — вони говорять про ВВП, про інфляцію, про міграційні кризи, про тарифи, проте вони до останнього часу майже не говорили про штучний інтелект, тоді як це має тепер бути одним з головних питань у кожній виборчій кампанії. Просто тому, що це — не якісь *абстрактні* екзистенційні небезпеки, на кшталт метеориту, а це вже *безпосередні* проблеми повсякденного життя. Йдеться про нашу *роботу*, про *людську зайнятість*, про те, *що* власне приймає рішення, що впливають на наше життя. Ви звертаєтесь до банку, щоб отримати кредит, і все частіше *рішення* щодо вашої кредитоспроможності приймає штучний інтелект. Ви звертаєтесь до роботодавця, щоб отримати роботу — рішення приймає людина, яка все частіше керується рекомендаціями штучного інтелекту щодо кандидатури того або іншого апліканта на посаду. І ви навіть не зрозумієте, якщо вам відмовили — *чому* вам відмовили? *як* вас оцінювали і *що саме* приймав остаточне рішення? Може, штучний інтелект помилився у оцінці вашої кандидатури? Очевидно, що критерії застосування штучного інтелекту повинні визначатися більш чітко.

Коли ми говоримо про *регулювання*, ми повинні розрізняти регулювання *розвитку* ІІІ в контрольованій сфері, його дослідження в лабораторіях, і регулювання *впровадження* продуктів ІІІ в публічну сферу. Наразі потрібен суворий контроль за його *публічним впровадженням*. Є декілька дуже простих правил, які ми повинні прийняти, наприклад, що ІІІ *не може видавати себе за людину*, тобто, якщо ви з кимось розмовляєте, ви повинні знати, чи це реальна жива людина, або ж це ІІІ. Якщо ми цього не забезпечимо, то суспільна дискусія щоразу заходитиме в глухий кут, і демократія просто не зможе існувати. Адже це дві принципово різні речі: намагатися переконати *людину* змінити свій світогляд або думку, і намагатися зробити те ж саме зі *штучним інтелектом* (останнє було б безглуздом).

Припустимо, якщо ви обговорюєте з кимось вибори і не можете зрозуміти, чи це штучний інтелект, чи жива людина, то це схоже на кінець демократії [24]. Тому що для людини безглуздо витрачати час на те, щоб «змінити думку» бота (він не має розуму чи свідомості в людському сенсі, а поки що лише імітує їх). Але для бота кожна хвилина, яку він витрачає на спілкування з нами, означає, що він *вивчає нас* краще. Він вибудовує довірчі та більш відверті стосунки зі справжньою людиною, і тоді боту стає легше змінювати погляди цієї людини. Тому, в означеному контексті, задля протидії симуляції людей штучним інтелектом, нам потрібні аналогічні закони, які забороняють фальсифікацію документів, грошей, цінних паперів, тощо. Тут потрібна максимальна транспарентність: люди мають чітко розуміти, чи спілкуються вони з *реальною* людиною, чи із *чат-ботом*; чи маємо ми справу із думками реальної людини, або ж із продуктом генеративного попередньо навченого трансформеру (GPT в широкому сенсі).

З часів Просвітництва існує міф про те, начебто свобода є довговічним конструктором внаслідок сталого політичного консенсусу. Насправді

ж, як стверджують автори відомої праці [25], цей «коридор до свободи» доволі вузький, і залишається відкритим лише завдяки фундаментальній і безперервній конкуренції між державою і суспільством. ІІІ стає в цьому процесі ще однією невідомою, яку, наприклад, держава може поставити собі на службу (і далі я наведу загальний опис сценарію, як саме це може відбутися). Втім, інколи зустрічаються й більш оптимістичні погляди щодо потенціального впливу штучного інтелекту на підтримання свободи і демократії [26].

В будь-якому разі, потрібно усвідомлювати, що подібно до того, як неможна випустити нові потужні ліки або транспортні засоби в публічну сферу, доки вони не пройдуть перевірку на безпечність та не отримають відповідні дозволи, так само й штучний інтелект, — як інструмент недостатньо вивчений, впровадження якого в публічну сферу може мати не прогнозовані наслідки для суспільства, — має ще пройти ряд тестів. Так, у нас є закони про захист приватного життя, про зв'язок, про інформацію, про захист персональних даних. Але всі вони не були розраховані на вирішення тих нових проблем, які створює просунутий ІІІ, наприклад, *емоційно-тренований* штучний інтелект.

Останніми роками можна спостерігати, як у соціальних мережах точиться постійна боротьба за увагу. Всі вже звикли до цього. Але тепер, із приходом інформаційних технологій нового покоління, цей фронт боротьби зміщується від просто *уваги* — до *довіри*. Якщо ми не врегулюємо це, то, швидше за все, опинимось в ситуації, коли мільйони агентів ІІІ безперервно намагатимуться завоювати нашу довіру, встановити певну «близькість». Тому що це — найпростіший спосіб переконати нас: купити товар, спожити послугу, проголосувати за політика, чи ще щось. І якщо ми дозволимо цьому статися, це призведе до нового роду маніпуляцій.

Цікаво також, як *демократична* система регулюватиме технології штучного інтелекту, і як вони використовуватимуться *авторитарними* чи *тоталітарними* системами. Можна припустити, що останні є набагато гіршими, ніж демократії, коли справа доходить до регулювання ІІІ та утримання його під контролем. Традиційною проблемою тоталітарних режимів є те, що вони схильні вірити у власну непогрішність. Вони переконані, що ніколи не помиляються, і вони не мають нормальних механізмів самокорекції для виявлення та виправлення власних помилок. За таких умов, виникає ще більша спокуса дати *занадто багато влади* технологіям штучного інтелекту, й намагатися централізовано його контролювати, не створюючи якихось дублюючих, альтернативних запобіжників (що мали б спрацювати у разі втрати централізованого контролю).

Якщо не залишається жодного механізму контролю, який міг би вказати системі на помилки, яких вона неминуче припускатиметься, — це прямий шлях до кризи. Слід чітко усвідомлювати, що ІІІ не є безпомилковим. Він має велику потужність, він може обробляти велику кількість інформації, але інформація — це ще не істина. Тільки відкриті демократичні суспільства, з належним рівнем прозорості, підзвітності та суспільного контролю, мають таку систему стримувань і противаг, яка дозволяє спробувати щось, і якщо це працює неналежним чином, або є деструктивним чи небезпечним — негайно призупинити такі шкідливі процеси. Без зайвого

алармізму, але все ж ми маємо усвідомити вірогідність настання гострих проблем в економіці та соціумі через безвідповідальне розгортання непереверених інструментів ШІ.

Зокрема, *ринок праці* має бути головним предметом занепокоєння для кожного з нас. Штучний інтелект, безумовно, не знищить *всі* робочі місця, але він точно ліквідує *частину* існуючих робочих місць, можливо, створюючи при цьому нові. Однак цей перехід і вимушена перекваліфікація людей стають непростими завданнями. Важливо пам'ятати, що катастрофічні історичні події, такі як прихід Гітлера до влади, відбувалися під впливом тривалих періодів високого рівня безробіття [27], коли близько трьох років до 25 % населення Німеччини не мали стабільної зайнятості. Та навіть якщо ми очікуємо, що внаслідок цього технологічного переходу ситуація із ринком праці буде кращою, ми все ж не можемо ігнорувати негайні наслідки масового безробіття для цього перехідного періоду.

Дослідники з «OpenAI» припускають [28], що значна частина робочих місць буде *модифікована* внаслідок повсюдної інтеграції технологій ШІ. Це, можливо, призведе до оптимізації та підвищення продуктивності, тобто у компаній буде або менший штат людей, які передоручать значну частину роботи умовним роботам, або ж вони зможуть виробляти більше з тією ж кількістю існуючих працівників, але підсилених засобами ШІ. Таким чином, може виникнути два варіанти: скорочення робочих місць, або їхнє збереження при зростанні продуктивності. Прогнозувати такі речі досить складно, бо нам невідомо, якою буде структура попиту і споживання в світі, та у конкретних країнах.

Один з аргументів, які лунають на користь того, що не варто хвилюватися, — це те, що *суспільства змінюються повільно*. Навіть якщо у нас є технологія для чогось, можуть знадобитися роки, а іноді й десятиліття, щоб люди повністю інтегрували її в суспільство, і щоб це суттєво вплинуло на ринок праці. Ми можемо лише припустити, що коли у нас є система, яка по суті робить роботу краще за людей, наприклад, можливість маніпулювати мовою чи розповсюджувати рекламу через електронну пошту, месенджери, соціальні мережі, або виконувати інші подібні завдання, то, ймовірно, що *такі* види робіт можна буде виконувати краще, швидше, дешевше, результативніше в багатьох секторах. А чи зможуть самі компанії швидко прийняти ці зміни, імплементувати їх з огляду на трудові стандарти, вимоги профспілок, тощо — це інше питання (можна пригадати нещодавні масштабні страйки сценаристів, режисерів, акторів, аніматорів та інших учасників кіноіндустрії в Голівуді). Та якщо ці зміни відбудуться раптово, тоді різні сектори економіки одночасно зіткнуться із гострими проблемами перехідного періоду.

Психологічно це доволі важко прийняти: *що буде, коли за моєю роботою прийде бот або штучний інтелект?* Якщо внаслідок автоматизації на звичайних видах робіт людей поступово замінюватимуть, то постає щонайменше два питання: чим їх можна надалі зайняти в цьому житті, та як забезпечити матеріально-економічне буття великої кількості вивільнених людей. З'являється ідея про «безумовний базовий дохід», який звісно не кожній країні під силу економічно. Очевидно, що революцію в галузі штучного інтелекту очолює *дуже невелика* кількість країн, які, швидше за все,

стануть завдяки цьому надзвичайно багатими і могутніми, тоді як вона може зруйнувати економіку менш розвинених країн. Навіть якщо подумати про щось на зразок текстильної промисловості, *що станеться* з економікою деяких густонаселених країн, якщо виробляти текстиль у Канаді, США чи Західній Європі стане так само дешево або навіть дешевше, аніж у Пакистані, Бангладеш, Бразилії, Мексиці чи Туреччині? Малоймовірно, що вдасться перенавчити мільйони текстильних робітників у цих країнах — на виробників мікросхем або дизайнерів якихось інноваційних продуктів віртуальної реальності. І хто заплатить за цю перекваліфікацію?

Можливо, в передових розвинених країнах вигоди від AI-революції дозволять урядам пом'якшити удар для людей, які втратять роботу, і нададуть їм можливість перекваліфікуватися. Але в густонаселених країнах, що розвиваються, це точно не відбуватиметься так само. В ще більш цифровізованому виробництві дійсно знадобляться висококваліфіковані люди, які мають справу з даними (із *персональними* даними клієнтів, та з *великими даними*). Тож у найбільш технологічно розвинених країнах з'являться *нові* робочі місця. Тоді як найсерйозніші соціально-економічні проблеми, спричинені втратою цілих галузей промисловості через перерозподіл ринків, релокацію, повернення назад на Захід, неминуче виникнуть у таких країнах, як Мексика чи Бангладеш. Саме ці місця *найбільш вразливі до автоматизації*. І прибутки від автоматизації потечуть не у Дакку чи Сан-Паулу, а у Ванкувер, у Каліфорнію, у Техас. Зрештою, як припускають футурологи, все може статися як із промисловою революцією в XIX столітті, яка призвела до того, що дуже невелика кількість країн фактично завоювали контроль і встановили домінування над усім світом. Щось схоже може повторитися протягом дуже короткого часу завдяки автоматизації та революції штучного інтелекту.

Хоча подібна постановка питання одразу привертає увагу людей, це не варто сприймати занадто спрощено, на кшталт, що «для людей більше не залишиться роботи». Обов'язково буде багато нових робочих місць, але перехідний період буде складний і подекуди дуже болісний. Це вимагає від людства заздалегідь продумати, як ми збираємося організувати перенавчання сотень мільйонів людей, особливо якщо враховувати глобальні виміри, а не межі окремо взятої розвиненої країни. Освітні стратегії мають бути повсюдно переосмислені з урахуванням нових технологічних імперативів. Частково це питання було висвітлено в окремих публікаціях [29; 30]. Існуючу систему професійної освіти очікують нові серйозні виклики, пов'язані з феноменом діджиталізації.

Та це не лише про економіку й освіту, це також і про зміну типу *політичного контролю*, який можна отримати, зібравши всі світові дані та проаналізувавши їх. Раніше, щоб контролювати якусь країну, потрібно було посилати туди солдатів, або створювати там військові бази. Тепер все частіше потрібно просто вилучати *дані* про цю країну, щоб вправно контролювати її. Що відбуватиметься з будь-якою країною, коли всі особисті дані, медичні й страхові картки, податкові коди, документи на нерухомість, банківські рахунки, файли з іншою конфіденційною або чутливою інформацією, адреси, аккаунти, телефони та електронні скриньки, — персональні дані кожного політика, підприємця, журналіста, судді, ліка-

ря, правоохоронця, військового офіцера, — зберігаються у *когось* ззовні, скажімо, у Кремнієвій долині або в Китаї? Це все ще *незалежна* країна, чи вона (після такого витоку даних) перетворюється на своєрідну «*data-колонію*»? Глобальний інформаційний обіг робить світ більш пов'язаним, взаємозалежним, та більш *вразливим* при певних умовах.

Зазвичай кажуть, що інформація — це потужний ресурс, який можна трансформувати у *знання* та *досвід*, у конкурентну перевагу. І це *було* правдою протягом більшої частини історії, коли інформації було дуже мало, і різного роду «монополісти» (чи то були шамани, оракули, верховні жерці, чи пізніше — чиновницька бюрократія стародавніх держав, первосвященики, римські юристи та сенатори, а ще пізніше — централізована державна цензура), — усі вони діяли шляхом приховування інформації, перекриваючи її вільний потік, а загальна малограмотність цьому сприяла. Але *зараз* ми живемо в зовсім іншу епоху, коли на нас обрушується величезна кількість інформації. Її занадто багато, і ми не знаємо, як розібратися в цих інформаційних потоках. І *цензура* зараз працює по-іншому, відволікаючи людей *надлишком* інформації, неактуальною інформацією, дезінформацією. У цю епоху *ясність* важлива як ніколи, тому що ми повинні зрозуміти, на чому варто зосередитися, а що — ігнорувати. *Увага, щирість, прозорість і довіра* стають чи не найдефіцитнішим ресурсом серед усіх, котрі пов'язані з людським розумом.

Децю нового звучання набувають ці два знамениті антиутопічні романи: «1984» Дж. Оруела [31] та «Дивний новий світ» О. Хакслі [32]. Але якщо ми подивимося на те, *як* поведуться з інформацією в цих двох різних романах, то побачимо, що в антиутопії Оруела інформація постійно *брутально* фабрикується, переписується, підтасовується, вирізається. В антиутопії Хакслі маніпуляції з інформацією більш *витончені*: люди запрограмовані від народження, їхній розум від початку наповнений *різною* інформацією, та кожна з п'яти каст є частиною єдиного плану. Увесь механізм працює таким чином, що створюється враження, ніби *сама система* тебе «розуміє», і апелює до твоїх власних пристрастей та емоцій. Система налаштована таким чином, щоб люди відчували, що вона, так би мовити, «на їхньому боці». Це не система старого зразка, вона не репресує відкрито, натомість, у багатьох випадках ця система створює стійке враження, що вона *доброчинлива*. Згадані романи ілюструють різні підходи до управління інформацією.

Науково-технічний прогрес, звичайно, несе не тільки небезпеки, але й *переваги*. Якщо говорити про сучасні розумні технології, про інтегрований штучний інтелект, то в багатьох випадках ці системи *насправді* розуміють нас краще (ніж це роблять багато людей) і можуть багато в чому покращити наше життя. І саме тут криється *спокуса*. У деяких випадках це стає особливо очевидним, якщо взяти за приклад систему охорони здоров'я. Вже сьогодні передові технології, здатні швидко обробляти великі обсяги даних, розпізнавати фотографії, інтерпретувати показання медичних приладів, узагальнювати отриману інформацію, оцінювати симптоми — такі системи *вже* перевершують професійні навички *одного* лікаря, та за своєю ефективністю вони цілком співставні з цілим *консилиумом* лікарів. Такі системи ставлять дуже точні діагнози з мінімальною похибкою.

А тепер уявімо, що розроблено ще більш досконалу систему, яка безперервно стежить за тим, що відбувається всередині нашого тіла, і *так* — вона знає краще і точніше, аніж наша свідомість, що саме відбувається всередині нашого тіла у кожний момент часу. Якщо у людини є серйозне захворювання, яке поширюється по її тілу, дуже часто люди дізнаються про це тільки тоді, коли це вже велика проблема, і людина раптом починає відчувати біль. Тоді вона йде до лікаря, проходить обстеження, здає аналізи, лікарі виявляють (інколи навіть не з першого разу), що у пацієнта серйозна хвороба, до того ж у запущеній стадії, і тепер може виявитися, щовилікувати її дуже важко, доволі болісно і неймовірно дорого.

Альтернативою може бути система, яка постійно моніторить те, що відбувається у людському тілі, і здатна *завчасно виявити*, що серйозна хвороба починає поширюватися в якійсь частині організму. Коли ця серйозна хвороба ще на ранніх стадіях, людина нічого не відчуває, але біометричні датчики *вже* фіксують перші ознаки того, що проблема розвивається — тоді, коли її ще можна легко, дешево і безболісно позбутися. Відповідно, штучний інтелект аналізує показники цих біометричних датчиків конкретної людини, та завчасно сигналізує про ті або інші ризики для здоров'я, та одразу пропонує найкращі можливі варіанти усунення проблеми.

Це все виглядає просто чудово. Тоді навіщо пригальмовувати розвиток штучного інтелекту? Адже аналогічну технологію, гіпотетично, можна застосувати до абсолютно різних випадків, коли мова йде про прийняття багатьох життєвих рішень. Починаючи з того, який фільм подивитися, який товар обрати, і закінчуючи більш серйозними питаннями — коли входити і виходити з операції на біржі, куди інвестувати гроші, який університет або коледж обрати, що в цьому університеті вивчати? Тому що іноді *своїм* розумом, особливо в молодості, ми робимо дуже поганий вибір. І якби штучний інтелект допоміг нам *тоді* зробити правильний вибір, це могло б врятувати нам ціле десятиліття нашого життя. Тож проблема зі штучним інтелектом і розгортанням усіх цих технологій полягає в тому, що перед нами постає величезна *спокуса*. І головне питання: як нам скористатися *перевагами* штучного інтелекту, витягти з нього можливі *вигоди*, не зазнавши при цьому *шкідливих наслідків*? До того ж, зараз ми живемо в ситуації, коли просто *повністю відключитися* від усіх пристроїв — означає *програти конкуренцію* тим, хто продовжуватиме користуватися такими пристроями. Відмова від використання технологій, як на індивідуальному, так і на суспільному рівні, не спрацює, тому що тоді ми втратимо стільки позитивних змін і перспектив, тоді як решта людей, наших сучасників, підуть далі в своєму технологічному розвитку.

Давайте простежимо цю *діалектику прогресу*, змодельуємо два сценарії. Наприклад, зараз людина носить якесь смарт-кільце, або розумний браслет, або інший гаджет, який фактично є біометричним пристроєм, що вимірює її пульс, рівень сатурації, кров'яний тиск, різні стани сну, рівень глюкози в крові, та інші фізіологічні параметри. І людина розглядає цей гаджет як певну *превагу*. Обробляючи та враховуючи ці дані, спортсмен може оптимізувати свої тренування, а літня людина — підтримувати своє здоров'я та покращувати самопочуття. Це типовий приклад *конвергенції* інформаційних технологій та біотехнологій [33], яка може вплинути на стабільність нашого здоров'я, якість життя та довголіття.

Але є й *інший* аспект: вся ця статистична інформація є частиною *великих даних*, які накопичуються і зберігаються на якомусь сервері. Чи знає користувач цього гаджету, *що саме* отримує інформацію, яку про нього збирає цей біометричний пристрій, і *що вони збираються з нею робити*? Якщо, наприклад, така інформація збирається великою корпорацією або урядом, і ми не знаємо, що саме вони збираються з нею робити, то в деяких випадках це може мати найнеприємніше продовження.

Ось один з таких похмурих сценаріїв: уявімо, що дія відбувається не у відкритому демократичному суспільстві, а в тоталітарній державі, де ці смарт-браслети виробляються масово, і кожен громадянин *змушений* постійно носити ті гаджети, передаючи всю інформацію про себе в центральну базу даних. Далі уявімо, що, збираючи всі ці показники, той розумний гаджет здатен зробити *інтерпретацію* всіх змін наших фізіологічних параметрів — пульсу, кров'яного тиску, розширення чи звуження зіниць очей, наших гормональних сплесків (включаючи рівень дофаміну, серотоніну, норадреналіну, кортизолу), — тобто всього того, що комплексно розповідає про певний *емоційний* стан людини. Бо кожен емоційний стан супроводжується відповідними гормональними сплесками, у різних комбінаціях гормонів, нейромедіаторів, тощо. Ось людина заходить в кімнату і бачить на стіні портрет диктатора, а гаджет на її руці *реєструє ознаки* гніву, ненависті, невдоволення, неприязні до вождя... Наступна зупинка — виправний табір, або психлікарня, або в'язниця... Це щось на кшталт моделі передбачення «ворогів держави» ще до того, як буде здійснена будь-яка реальна дія. Не тільки до того, як щось зробиш, але навіть до того, як *подумаєш* про якусь реальну дію. Це класичний «mind crime», про який часто пишуть в антиутопіях, тобто «мислезлочин», просто під впливом спонтанної неконтрольованої емоції. Якщо така держава спостерігає за всіма своїми громадянами протягом тривалого періоду часу, вона може легко створити типовий профіль бунтаря чи дисидента, і почати «виправляти» їхню свідомість, коли вони ще перебувають у дитячому садку. Такій державі не потрібно чекати, допоки вони виростуть і становитимуть якусь небезпеку, чи просто незручність для системи.

Тож якщо ми хочемо зануритися в антиутопію, технології дають нам для цього багато потенційних можливостей. Ми можемо лише уявити, що б робив Сталін, Мао чи ПолПот, якби у них були такі біометричні технології на той час. *Штучний інтелект вирішує проблему багатьох диктаторів минулого*: раніше було дуже складно і дорого стежити за своїми громадянами, просто тому, що потрібно було утримувати великий штат наглядачів, інформаторів, численну таємну поліцію та інші спецслужби. Тепер, за бажанням диктаторів, *нейронні мережі* зможуть робити це швидко і відносно дешево. Якщо змодельований вище антиутопічний приклад почне розгортатися, і такі суспільства раптом виявляться більш технологічно та військово потужними, аніж демократії, це означатиме *кінець гуманізму*.

Тут ми підходимо до необхідності поговорити про те «що таке *гуманізм*», як ми його розуміємо сьогодні? Адже немає сенсу говорити про антропологічні виміри ШІ у відриві від принципу гуманізму. Ми звикли трактувати гуманізм як *віру* в перевагу *людяності* і, зокрема, *людських*

почуттів. Іншими словами, гуманізм — це ідея, що *людські почуття* є найвищою інстанцією, якщо не у Всесвіті, то принаймні на Землі. У *політиці* це означає, що виборець краще знає — за кого голосувати, а тому найвищим авторитетом у політиці є *почуття виборців*. Абсолютно зрозуміло, що йдеться саме про почуття, а не про раціональність. Коли ми йдемо на референдум чи вибори, політики апелюють до почуттів виборців, а не до їхньої раціональності. Окремі цинічні мислителі відзначають, якби вибори були раціональними, то не було б жодної причини давати *всім* рівні права голосу. Складно заперечувати, що різні люди мають різні раціональні здібності, неоднакові знання, досвід, статус, інтелект, кваліфікацію, тощо. І коли ми дивимося на світ сьогодні, ми бачимо дуже складні етичні дебати навколо таких речей, як «політика ідентичності», «недискримінація», які насправді обертаються навколо людських емоцій електорату.

В *економіці* це проявляється в тому, що «клієнт завжди правий». Найвищий авторитет у економіці — це *почуття клієнтів* (споживачів, замовників, покупців). Якщо люди дійсно хочуть ту чи іншу послугу або товар, вони «голосують» своїми кредитними картками. Те ж саме можна сказати і про *мистецтво*: у царині мистецтва панує гуманізм. Однак у мистецтві це гуманістичне гасло може звучати таким чином, що «краса в очах того, хто споглядає». Це дещо схоже на принцип, що «клієнт завжди правий», та у сфері мистецтва немає вищої інстанції, ніж *глядач*. І якщо людям подобається певний жанр або конкретний витвір мистецтва, то їхні власні почуття і є найвищим суддею. Тобто, *гуманізм* передбачає реалізацію таких принципів, як *індивідуалізм*, *суб'єктивізм*, *сенсуалізм*, *антропоцентризм* (пригадаймо, ще Протагор у 5 ст. до н.е. відзначав, що «Людина є мірою всіх речей, існуючих та неіснуючих»).

Отже, перед нами носії почуттів та емоційного інтелекту — клієнт, покупець, виборець, глядач, підписник, фанат... Тож усі економічні, політичні й культурні сили підпорядковані волі та примхам колективного емоційного ландшафту населення. Єдине, що раніше могло протистояти почуттям людини — це почуття *іншої* людини. І тепер виникає цікаве питання: що станеться, коли алгоритми, на основі аналізу великих даних про нас, зможуть «зламати» людську систему емоцій та мислення? особливо, якщо їм вдасться навчитися *передбачати* людські почуття, а потім (на підставі вірних прогнозів) *моделювати* людські бажання та емоції. Що станеться, якщо у нас з'явиться система, яка краще за мене знатиме, чого я хочу? Дуже ймовірно, що і сама *влада* поступово перейде до цієї системи. І якщо система може *передбачати* і *розуміти* мої почуття, то, очевидно, вона зможе ними *маніпулювати*. Так «оракул» перетворюється на «всемогутнього правителя», і тоді влада повністю усувається від людини, від *людських* емоцій та інтелекту (бо інтелект *штучний* стає на порядки потужнішим, ніж у середньостатистичної людини).

І ринкові сили, і політичний ландшафт, і орієнтація у світі, і досягнення кар'єрного успіху, і заробіток на життя — все це, вочевидь, зміниться. Перед філософами, психологами, соціологами, інженерами, юристами, економістами, можновладцями постане багато питань, які доведеться досліджувати та якось вирішувати, в тому числі питання *неактуальності* і *непотрібності*, що виникають в результаті технологічної революції.

Якщо у ХХ та попередніх століттях головною напругою була *боротьба проти експлуатації* людини, то у ХХІ столітті болісним питанням може вже стати *боротьба проти «непотрібності людини»*. Люди масово можуть виявитися просто незайнятими, і боротьба із непотрібністю буде набагато складнішою — аніж була проти експлуатації. За такого сценарію, чимало людей можуть потрапити у категорію «недоречна людина».

Ми ще не уявляємо, як виглядатиме людське життя, коли *алгоритми* прийматимуть все більше рішень *від нашого імені*. Тисячоліттями релігійні, політичні та мистецькі традиції описували життя як драму прийняття рішень. Будь то п'єса Шекспіра, роман Гете чи Ремарка, твори Івана Франка чи Миколи Хвильового, голлівудська сага чи книга з богослов'я — всі вони розповідають про життя як своєрідну подорож, в якій *ми* приймаємо рішення на перехрестях. У повсякденному житті ми робимо і простий, і складний *вибір*, обираючи, як побудувати свій день, що з'їсти на обід, за кого голосувати, із ким одружитися, яку кар'єру обрати, проти кого боротися, з ким кооперуватися, тощо. Вся ця драма обертається навколо прийняття *правильного* рішення. І всю попередню історію це було монополною прерогативою *людини*. Це було водночас нашою мукою, джерелом страждань, і нашим привілеєм. Проте зараз ми входимо в період, коли *гештальт автоматизації* розвивається шаленими темпами. Нині чимало говорять про робочі місця і, можливо, в недалекому майбутньому, про *втрачені* робочі місця, про зміну самого способу життя, і проблема ця багатшарова: вона і економічна, і геополітична, і психологічна, і демографічна. Таким чином, розвиток штучного інтелекту демонструє дуже різні *антропологічні* грані, тому має постійно вивчатися з позицій філософської антропології.

Висновки

Важливим завданням для кожного сучасного суспільства й політичної системи має стати осмислення і регулювання вибухонебезпечного потенціалу технологій, забезпечення *відповідального управління* технологіями. Не має сенсу говорити про права і свободи людини, якщо не будуть розроблені та впроваджені *розумні обмеження і протоколи безпеки* для використання АІ-засобів і технологій. Зрештою, першочергова проблема *в нас самих*, а не в штучному інтелекті. Бо загроза зловмисного використання можливостей ШІ походить, насамперед, від самих людей. Отже, якщо людство стільки ж часу і коштів, скільки витрачається на розробку штучного інтелекту, буде витрачати на вивчення і розвиток *власного* розуму, ми всі будемо в безпеці.

Для того щоб пом'якшити ризики, про які згадувалося вище, урядам варто сприяти реалізації певних заходів, серед яких:

- визначення більш конкретних *етичних* вимог щодо розробки та публічного впровадження технологій штучного інтелекту;
- широка міжнародна співпраця у запровадженні *спільних сучасних стандартів* у регулюванні цифрових технологій;
- встановлення чітких *безпекових нормативів* у сфері збирання, зберігання, використання, обміну персональними даними й іншою чутли-

- вою інформацією, з урахуванням того, що ці дані можуть оброблятися ШІ;
- інвестування у *кібербезпекову інфраструктуру*, у розробку нових шифрувальних стандартів й засобів криптографічного захисту нового покоління, здатних попередити несанкціонований доступ до захищених даних, в т.ч. з боку просунутого штучного інтелекту;
 - вивчення *соціетальних процесів*, які розвиваються в людській цивілізації на тлі підсилення технологічних імперативів; зокрема, заохочення філософсько-антропологічних досліджень, які допомагають зрозуміти, як змінюються людська природа, свідомість та етика в умовах конкуренції природного інтелекту зі штучним.

Варто також враховувати більш віддалені можливості просунутого ШІ, який буде здатний до самонавчання, й доскональної імітації людського мислення та емоцій, й тоді нестиме більш високі ризики в плані підзвітності людині. Ми не маємо прийти до абсурдної ситуації, коли, намагаючись максимально спростити своє життя завдяки технологіям, ми створимо собі ще більшу проблему, самостійно віддавши владу приймати різноманітні рішення — алгоритмам штучного інтелекту. Бо вже зараз спостерігається певний парадокс: хоча ми (поки що) найрозумніші істоти на Землі, але напрацювання нашого інтелекту привели нас до необхідності обговорення таких сценаріїв, як оточення кожної людини камерами і датчиками, обмеження людських свобод і фактичне знищення приватності — щоб захистити нас від самих себе, підсилених AI-технологіями.

Ми боїмося не штучного інтелекту, насправді ми боїмося себе — посилені технологіями ШІ. Інтереси, закладені в штучний інтелект — це *людські* інтереси. І всі фобії та хибні вади, приписувані ШІ — це іманентні людські страхи перед невідомим. Але сам по собі нинішній ШІ нічого не може зробити такого, чого у нього не заклали б розробники-люди. ШІ залишається поки що виключно *в рамках людської діяльності*. Все, чому він навчений — це людський досвід та антропологічні категорії. Ми надаємо нейромережам доступ до *наших* баз даних, до масивів *людських* знань, оздоблюємо їх сенсорами, камерами, датчиками, котрі імітують людський тип сприйняття та розуміння, і що найголовніше — ми ставимо для ШІ «людські завдання». Адже *будь-яка* задача, поставлена штучному інтелекту *людиною* — це «людська задача», і в цьому сенсі ШІ залишається в антропологічній парадигмі; він за таких умов не може вийти за межі суто людських рамок (та, умовно кажучи, збунтуватися проти людини).

Страх *втрати контролю* над штучним інтелектом, здебільшого, походить від нашої проєкції антропологічних якостей та людських мотивацій — на штучний інтелект, наділення ШІ якимись «власними інтересами». До того ж, не останню роль в настороженому ставленні людей до ШІ відіграє наша *недовіра до самих себе*, наш страх щодо вічної схильності використовувати новітні технології проти свого ж виду. Особливості людської природи та багатий історичний досвід лише підтверджують ці побоювання — адже немає жодних гарантій, що одні групи людей, неймовірно підсилені технологіями ШІ, утримуватимуться від *спокуси* розібратися з іншими групами людей, особливо доки ті не настільки технологічно розвинені й не стали справжніми конкурентами.

Рівень культури та освіченості людини (особливо в умовах широкого запровадження технологій ШІ) має критично важливе значення для гуманізму. Якщо в нас немає широкого кругозору, інтелектуальної культури, усвідомлення уроків історії, толерантності та емпатії, немає справжньої університетської освіти — ми просто не зможемо ставити *гідні завдання* штучному інтелекту, й наші запити до нього будуть надто примітивними. Але якщо ми достатньо розвинені й освічені, тоді ШІ (нейромережі, віртуальні асистенти, тощо) перетворюється для нас на свого роду «екзоскелет». Тобто, щоб ефективно та розумно користуватися новітніми інформаційними технологіями, люди (на персональному та суспільному рівнях) повинні мати належний рівень розвитку, що робитиме використання ШІ більш відповідальним, творчим, продуктивним, і більш безпечним.

Подивимося, що буде через декілька років. Але вже сам факт *дискусії* щодо штучного інтелекту, та людське *занепокоєння* вражаючими можливостями нейромереж — подає надію, що людство не припуститься фатальних помилок.

References

- [1] Callaway Ewen. 2022. “The entire protein universe”: AI predicts shape of nearly every known protein. *Nature*, 608(7921), pp. 15–16. DOI: <https://doi.org/10.1038%2Fd41586-022-02083-2>
- [2] Jumper John, Evans Richard, Pritzel Alexander, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), pp. 583–589. DOI: <https://doi.org/10.1038%2Fs41586-021-03819-2>
- [3] Heaven Will Douglas. 2020. DeepMind’s protein-folding AI has solved a 50-year-old grand challenge of biology. *MIT Technology Review*, 30 November 2020. URL: <https://www.technologyreview.com/2020/11/30/1012712/deepmind-protein-folding-ai-solved-biology-science-drugs-disease/>
- [4] Hsu Jeremy. 2023. AI discovers new class of antibiotics to kill drug-resistant bacteria. *NewScientist*, 20 December 2023. URL: <https://www.newscientist.com/article/2409706-ai-discovers-new-class-of-antibiotics-to-kill-drug-resistant-bacteria/>
- [5] Edwards Benj. 2024. Tyler Perry puts \$ 800 million studio expansion on hold because of OpenAI’s Sora. *ArsTechnica*, 23.02.2024. URL: <https://arstechnica.com/information-technology/2024/02/i-just-dont-see-how-we-survive-tyler-perry-issues-hollywood-warning-over-ai-video-tech/>
- [6] McBride Sarah, Hull Dana. 2024. Musk’s Neuralink shows quadriplegic patient able to play online chess with his Mind. *Bloomberg*, 20 March 2024. URL: <https://www.bloomberg.com/news/articles/2024-03-20/musk-s-neuralink-gives-update-on-first-brain-implant-patient>
- [7] Mullin Emily. 2024. Watch Neuralink’s first Human subject demonstrate his Brain-Computer Interface. *Wired*, 20 March 2024. URL: <https://www.wired.com/story/neuralink-implant-first-human-patient-demonstration/>
- [8] Valinsky Jordan. 2019. Amazon reportedly employs thousands of people to listen to your Alexa conversations. *CNN.com*, 11 April 2019. URL: <https://edition.cnn.com/2019/04/11/tech/amazon-alexa-listening/index.html>
- [9] Russell Stuart J., Norvig Peter. 2021. *Artificial Intelligence: A Modern Approach* (4th ed.). Hoboken, NJ: Pearson, 1115 pages.

- [10] Christian Brian. 2020. *The Alignment Problem: Machine learning and human values*. New York: W. W. Norton & Company, 476 pages.
- [11] Urbina Fabio, Lentzos Filippa, Invernizzi Cédric, Ekins Sean. 2022. Dual use of artificial-intelligence-powered drug discovery. *Nature Machine Intelligence*, 4(3), pp. 189–191. DOI: <https://doi.org/10.1038%2Fs42256-022-00465-9>
- [12] Harari Yuval Noah. 2018. Why Technology Favors Tyranny. *The Atlantic*, October 2018 issue. URL: <https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330>
- [13] Gregory Sam. 2023. Fortify the truth: How to defend human rights in an age of deepfakes and generative AI. *Journal of Human Rights Practice*, 15(3), pp. 702–714. DOI: <https://doi.org/10.1093/jhuman/huad035>
- [14] Rošker Jana. 2023. Dissolution of the Self: Digital Technology, Privacy and Intimacy in Europe and the Sinophone Regions. *Asian Studies*, 11(3), pp. 47–67. DOI: <https://doi.org/10.4312/as.2023.11.3.47-67>
- [15] Poblet Marta, Kolieb Jonathan. 2018. Responding to human rights abuses in the digital era: new tools, old challenges. *Stanford Journal of International Law*, 54(2), pp. 259–283.
- [16] Pūraitė Aurelija, Adamonienė Rūta, Žemeckė Audronė. 2020. Sustainable digitalization in public institutions: Challenges for human rights. *European Journal of Sustainable Development*, 9(3), pp. 91–102. DOI: <https://doi.org/10.14207/ejsd.2020.v9n3p91>
- [17] Nemitz Paul. 2018. Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences*, 376(2133). DOI: <https://doi.org/10.1098/rsta.2018.0089>
- [18] Frazzon Enzo, Agostino Ícaro, Broda Eike, Freitag Michael. 2020. Manufacturing networks in the era of digital production and operations: A socio-cyber-physical perspective. *Annual reviews in control*, 49, pp. 288–294. DOI: <https://doi.org/10.1016/j.arcontrol.2020.04.008>
- [19] Drewer Daniel, Miladinova Vesela. 2017. The BIG DATA Challenge: Impact and opportunity of large quantities of information under the Europol Regulation. *Computer law & security review*, 33(3), pp. 298–308. DOI: <https://doi.org/10.1016/j.clsr.2017.03.006>
- [20] Grech Alex, Sood Ira, Ariño Lluís. 2021. Blockchain, self-sovereign identity and digital credentials: promise versus praxis in education. *Frontiers in Blockchain*, 4. DOI: <https://doi.org/10.3389/fbloc.2021.616779>
- [21] Suleyman Mustafa. 2023. How the AI Revolution will reshape the World. *Time*, 1 September 2023. URL: <https://time.com/6310115/ai-revolution-reshape-the-world/>
- [22] Bengio Yoshua. 2023. AI and Catastrophic Risk. *Journal of Democracy*, September 2023. URL: <https://www.journalofdemocracy.org/ai-and-catastrophic-risk/>
- [23] Stovpets Oleksandr. 2020. Sinitic civilization’s worldview features and their system-forming role in the complex of social relations in modern China. *Interdisciplinary Studies of Complex Systems*, 17, pp. 59–72. DOI: <https://doi.org/10.31392/iscs.2020.17.059>

- [24] Marichal Jose. 2023. The Real Danger of AI to Democracy. *Democracy Paradox*, 31 May 2023. URL: <https://democracyparadox.com/2023/05/31/the-real-danger-of-ai-to-democracy/>
- [25] Acemoglu Daron, Robinson James A. 2020. *The narrow corridor: states, societies, and the fate of liberty*. New York: Penguin Books, 556 pages.
- [26] Simons Josh, Frankel Eli. 2023. Why democracy belongs in artificial intelligence. *Princeton University Press Essays*, 21 February 2023. URL: <https://press.princeton.edu/ideas/why-democracy-belongs-in-artificial-intelligence>
- [27] Overy Richard. 2013. Unemployment in the Third Reich. In *Business in the Age of Depression and War*. London: Routledge, pp. 224–252.
- [28] Eloundou Tyna, et al. 2023. GPTs are GPTs: An early look at the labor market impact potential of large language models. *OpenAI Research*, 17 March 2023. URL: <https://openai.com/research/gpts-are-gpts>
- [29] Przegalinska Aleksandra, Jemielniak Dariusz. 2023. *Strategizing AI in business and education: emerging technologies and business strategy*. Cambridge: Cambridge University Press, 75 pages.
- [30] Borinshtein Yevhen, et al. 2022. Educational marketing as a basis for the development of modern Ukrainian society and the state. *Amazonia Investiga*, 11(54), pp. 146–157. DOI: <https://doi.org/10.34069/AI/2022.54.06.14>
- [31] Orwell George. 2012. *1984*. London: Bloomsbury Publishing, 119 pages.
- [32] Huxley Aldous. 2017. *Brave New World*. New York: HarperCollins, 259 pages.
- [33] Carter Sarah R., Wheeler Nicole, et al. 2023. The Convergence of Artificial Intelligence and the Life Sciences. *NTI report*, 30 October 2023. URL: <https://www.nti.org/analysis/articles/the-convergence-of-artificial-intelligence-and-the-life-sciences/>