

<https://doi.org/10.15407/fd2021.03.180>  
УДК 117:004.8

**Михайло БОГАЧОВ**, магістр філософії,  
аспірант 3-го року навчання, відділ соціальної філософії,  
Інститут філософії ім. Г.С. Сковороди НАН України,  
01601, Київ, вул. Трьохсвятительська, 4  
m.bogachov@gmail.com  
<https://orcid.org/0000-0002-3223-7227>

## ШТУЧНИЙ ІНТЕЛЕКТ ЯК АНТРОПОТЕХНОЛОГІЯ

---

*Штучний інтелект — це комп'ютерна система, яка мислить або діє як людина. Особливості ШІ втілюють імпліцитні уявлення його розробників про природу людини. «Сильний» ШІ, який має загальні когнітивні здатності дорослої людини, досі не створено, натомість «слабкі» ШІ вже є частиною планетарної обчислювальної інфраструктури. Нейронно-мережеві ШІ імітують специфічні типи людської поведінки, узагальнюючи дані про повсякденне життя користувачів. Такий підхід у створенні ШІ відповідає філософському мейнстріму ХХ сторіччя, коли повсякденне життя розглядали як джерело мовних та соціальних наперед-даностей, що забезпечують порозуміння. Також цей підхід спирається на класичну дихотомію «людина—машина» й відповідне уявлення, що людська природа є стабільною та незалежною від технологічного стану. Проте у постметафізичну добу, коли взаємодія людини з технікою є комунікативною, а не інструментальною, дані про людський побут не можуть бути незалежним зразком людськості. ШІ-застосунки не лише кодифікують дескриптивні аспекти людської природи, а й дисциплінують своїх користувачів, адже цифрове середовище, в якому можуть бути зібрані дані про повсякдення, уже впорядковане за допомогою ШІ. Відповідно, у цифровому середовищі люди змушені відтворювати нові норми поведінки, кодифіковані за допомогою ШІ, який є одним із способів видозміни людської природи самою людиною. Вплив ШІ непомітний, позаяк цифрове середовище, в якому люди взаємодіють з ШІ, організовано недостатньо зрозумілим для користувачів чином. Антропотехнологічний характер ШІ є побічним наслідком розвитку платформ, тому творці ШІ не перебирають на себе відповідальність за втілені в їхніх розробках норми.*

**Ключові слова:** штучний інтелект, антропотехнологія, повсякденне життя, людська природа, інструментальність, цифрове середовище.

### 1. Вступ

Класичну етику ґрунтовано на передсудах, що суб'єкт етики (людина) має незмінну природу, а техніка<sup>1</sup> є етично нейтральною. Попри те, що ні перше,

---

<sup>1</sup> Тут і далі *техніка* означитиме «сукупність усіх технологій», а у значеннях «спеціальний метод досягнення певної мети [technique]» чи «технічні прилади, інструменти та агре-

ні друге уявлення вже не здаються прийнятними, перше з них продовжує виконувати регуляторні функції у практичній філософії. Г. Йонас відзначав, що класична етика — з її окресленими вище недоліками — не здатна відповідати на виклики сучасного технологічного стану [Йонас, 2001, с. 20], але згодом все ж апелював до статичного квазісуб'єкта — «ідеї людини, яка вимагає присутності її втілень у світі» [Йонас, 2001: с. 72], перед яким ми й будемо відповідальними, якщо пристанемо на його категоричний імператив етики технологічної доби, згідно з яким людська природа, утілюючись у людях, не повинна виходити за межі автентичності. Інакше кажучи, замість традиційного уявлення, що людині властива незмінна природа, отримуємо твердження, що людина *мусить* мати незмінну природу.

Крім матеріальних чинників, які можуть впливати на ступінь присутності ідеї людини у світі (модифікація геному, кліматичні зміни, ядерна війна), існують соціальні чинники, які теж можна розглядати крізь «евристичку страху». Проте чи впізнаємо ми в них небезпеку для ідеї людини, це залежить від наших уявлень про взаємозв'язок людини та техніки.

Другий передсуд класичної етики («Техніка є етично нейтральною») перетворився у ХХ сторіччі на тезу «Техніка є небезпекою за визначенням». Проте обидві позиції спираються на уявлення, що дійсність узагалі поділяється на технічну та нетехнічну царини, де автентичне, відповідне ідеї людини життя є чимось нетехнічним. У цій статті я розгляну деякі аспекти впливу технології штучного інтелекту на природу людини, які стали побічним продуктом розгляду людської природи як чогось, що за визначенням є автономним від техніки.

Штучний інтелект — технологія, яка має прямий стосунок до ідеї людини, оскільки зразком для створення штучного інтелекту є «природний» людський інтелект. Проте наразі ШІ навряд чи є «дзеркалом людської природи», а уявлення про нього як про виключно технологічне явище є некоректним. Я доводитиму, що сучасний прикладний ШІ, який вже став частиною нашого онлайн-існування, — це не лише технологія, а й особливий тип соціальної практики, за допомогою якої людина змінює свою природу. Такі практики П. Слотердаjk називає антропотехнологіями (2009). Сучасний прикладний ШІ є чимось більшим, ніж виключно технологічне явище, адже його успішне функціонування у цифровому та фізичному середовищі потребує зворотного зв'язку, який він отримує від людей, впливаючи на «не-технічні» з точки зору традиційної етики аспекти життя: людське повсякдення (життєсвіт), форми комунікації та пізнавальний горизонт. Роль ШІ в нашому житті також визначатиме етику доби планетарного обчислення, адже питання «Що означає бути людиною?» і те, які нормативні характеристики має поняття людини, — це імпліцитний фундамент практичної філософії.

---

гати [machinery]» поняття *техніка* не вживатимемо. У межах цього розмислу я не розглядатиму взаємозв'язок понять *техніка* й *технологія* докладніше.

Детальніше роль ШІ як соціальної практики й поняття антропотехнології я розгляну у шостому та сьомому розділах, а у попередніх розділах окреслю значущі з філософської точки зору риси ШІ як технології, покажу, які антропологічні уявлення втілюють у сучасних ШІ-застосунках, а також проаналізую ці засади з точки зору неінструменталістської філософії техніки.

## 2. «Сильний» і «слабкий» ШІ

Що таке технологія штучного інтелекту? Звернімося до статті «Обчислювальні машини і розум» А. Тюринга, з якої починається історія ШІ в сучасному розумінні. У ній Тюринг [Turing, 1950: p. 434] стверджує, що класичне питання, чи може машина думати, некоректне, і пропонує альтернативний варіант: чи може цифровий комп'ютер продемонструвати таку лінгвістичну компетенцію, яку неможливо відрізнити від людської? Якщо відповідь на перше питання залежить від фундаментальних переконань про те, що таке мислення, свідомість, людина і техніка, то на друге питання можна дати експериментальну відповідь. Вона, згідно з Тюрингом, має спиратися на уявлення «звичайного запитувача» [average interrogator] про те, як розмовляють люди. Таким чином, Тюринг виніс фундаментальні антропологічні й епістемологічні питання за дужки ШІ як прикладної дисципліни, а його підхід, хоч би яким спрощеним він здавався наступним теоретикам ШІ, врешті-решт визначив, які ШІ-застосунки розробляють і впроваджують тепер. Зараз на питання «Що таке штучний інтелект?» більшість інженерів-практиків, мабуть, дадуть хрестоматійну відповідь, що це *комп'ютерна система, яка поводить себе або думає, як людина* [Russel, Norvig, 2020] чи навіть скажуть, що це технологічний «артефакт, який проходить тести на розумові здібності»<sup>2</sup> [Nilsson, 1998]. У цій статті вихідним визначенням штучного інтелекту як технології також слугуватиме дефініція Расела й Норвіга.

За часів Тюринга кібернетика ще не впливала на повсякденне життя безпосередньо, а тому погляд, що розробникам ШІ варто орієнтуватися на «зовнішні» прояви мислення, не здавався проблематичним з точки зору наукового методу. Проте сучасні розробники ШІ вже не теоретики, які позбавлені важелів впливу на людські уми. Ми щодня взаємодіємо з ШІ-продуктами, які створені Google, Facebook, Amazon, Apple, Microsoft, Alibaba, Tencent, Huawei, і перебуваємо під впливом технологій, які значною мірою визначають наш пізнавальний горизонт, наші соціальні зв'язки, наші звички й переконання. Ми взаємодіємо з ШІ, зазвичай цього не усвідомлюючи, оскільки він став тлом нашого повсякденного життя. Між тим, «найгрунтовніші технології — це ті, які стають непомітними» [Weiser, 1999: p. 3].

Не всі системи, які ми називаємо ШІ, покликані імітувати загальні когнітивні здатності дорослої людини. ШІ, який грає в шахи, тим самим проходить лише один зі спеціальних тестів на інтелект. Він являє собою так зва-

<sup>2</sup> Тут і далі переклад з англ. мій (прим. автора)

ний «слабкий» ШІ й тому відрізняється від гіпотетичного «сильного» штучного інтелекту, що має неспеціалізовані когнітивні здатності, як у дорослої людини. Зараз більшість інженерів, залучених до сфер машинного навчання, робототехніки й «великих даних», розробляють «слабкі» ШІ, і саме ці продукти формують сучасну цифрову інфраструктуру. Проте методичні основи ШІ як наукової дисципліни були закладені задовго до того, як ШІ-продукти стали загальнодоступними.

Стартовою точкою ШІ-програми ХХ сторіччя була розробка загального («сильного») штучного інтелекту. 1956 року засновники дисципліни зібралися в Дартмутському коледжі, щоб «за два місяці» дізнатися, як створити комп'ютер, який може «використовувати [природну] мову... розв'язувати проблеми, якими наразі займаються люди, та самовдосконалюватися» [McCarthy et al., 1955]. З огляду на цю мету — якої не досягли ні тоді, ні зараз — виокремилися три загальні парадигми ШІ: логіцистський, пробабілістичний та нейронно-мережевий ШІ<sup>3</sup>. Здавалося б, сучасні «слабкі» ШІ — це здебільшого комерційні продукти, які не мають нічого спільного з повоєнними мріями про людськість, втілену в обчислювальних машинах. Проте ці продукти все-таки створені згідно з методологією, яку було розроблено для «сильного» ШІ. Вона й об'єднує під парасолькою поняття «штучний інтелект» як рекомендаційні алгоритми соціальних мереж, так і системи комп'ютерного бачення у автомобілях Tesla й ШІ-асистентів штибу Google Assistant чи Siri.

Чи існує бодай логічна можливість створити ШІ, який не поступається когнітивним здатностям дорослої людини? Є щонайменше три вектори атаки на мрію про «сильний» ШІ, які досі заслуговують на увагу. Це, по-перше, Пенроузіві (1995) умовисновки з теореми про неповноту математики Геделя (1995), по-друге, «китайська кімната» Серля (1980) й, по-третє, Дрейфусова (1992) апеляція до тілесності як необхідного складника інтелекту. Проте останні досягнення у сфері глибинних нейронних мереж та ШІ-моделей природної мови<sup>4</sup> дають підстави для нового оптимізму серед прихильників «сильного» ШІ. До того ж, практики ШІ звикли ігнорувати зауваги Серля і Дрейфуса, оскільки ці аргументи апелюють до неконективістських, а тому непопулярних серед інженерів ШІ концепцій свідомості<sup>5</sup> чи «лудитської» феноменологічної традиції. Водночас критика Пенроуза, який звертається

<sup>3</sup> Тут і далі йтиметься про нейронно-мережеві ШІ, оскільки засадовою для сучасної цифрової інфраструктури є переважно ця парадигма. Проте варто відзначити, що на практиці ШІ-продукти містять у собі елементи всіх трьох парадигм; нейронно-мережева є головною, але ніколи не єдиною.

<sup>4</sup> Нейронна мережа GPT-3 пише осмислені й цілком читабельні журналістські статті, які легко сплутати з текстами, що їх продукують втомлені копірайтери з плоті й крові. А втім, GPT-3 не розуміє сенсу текстів, які продукує, а лише маніпулює мовою так, щоб автор здавався живою людиною.

<sup>5</sup> Одна з догм сучасної сфери ШІ є такою: «те, що робить мозок, можна на певному рівні розглядати як обчислення». Див.: Charniak, E., McDermott, D. V. (1985). *Introduction to artificial intelligence*. Addison-Wesley.

до математичної логіки, залишається «надто теоретичною», бо вимагає від інженерів *формального* розуміння того, як влаштований їхній ШІ.

Інженери, які розробляють ШІ-застосунки, зазвичай такого розуміння не мають, бо нейронно-мережева парадигма дає змогу створювати лише *несемантичний* штучний інтелект. Нейронно-мережевий ШІ складається з математичної моделі й даних, на основі яких «навчається» алгоритм; інженерам давно відомо, що складний алгоритм, що працює з маленькою вибіркою, даватиме гірші результати, ніж примітивний алгоритм, який може опрацювати великий обсяг даних [Domingos, 2012]. Проте чим складніше ШІ і чим більші масиви даних, тим рідше дослідникам вдається простежити причинно-наслідкові зв'язки між даними, якими «годують» нейронну мережу, і висновками, яких вона дійшла. Попри простоту математики, яка описує алгоритм, зв'язки між вузлами нейромережі надто численні й складні для того, щоб жива людина могла їх інтерпретувати.

### 3. «Чорна скринька»

На відміну від класичних комп'ютерних програм, нейронно-мережевий ШІ є *непрозорим*. «Висновки» (вихідні дані) ШІ не завжди можна передбачити з огляду на «засновки» (вхідні дані). Отже, з'ясувати *post factum*, чому складна ШІ-модель дійшла того чи того висновку, практично неможливо. Саме тому «глибокі» нейронні мережі часто описують метафорою «чорної скриньки» [Buhrmester et al., 2019]. Інженери, які тренують такі нейронні мережі, дедалі більше орієнтуються на інтуїцію і неформальні ноу-хау, про які не розповідають в університетах, — на так звану «чорну магію» глибинного навчання [Anand et al., 2020]. Крім того, результати дослідів, які публікують у наукових журналах найбільші дослідницькі інституції у сфері ШІ, українською складно верифікувати. Наприклад, тренування GPT-3 коштувало щонайменше 4 мільйони доларів машинного часу [Li, 2020].

Колись Д. Денет розглядав ШІ як власне-філософію, як «найбільш абстрактне дослідження можливості розуму чи знання» (1978). Проте найпомітніших успіхів ШІ-наука досягнула тоді, коли рухалася у протилежний бік. Загального ШІ, який би відтворював розумову діяльність людини<sup>6</sup>, поки ніхто не створив. Натомість маємо низку ШІ-технологій, які вкрай успішно розв'язують проблеми, якими раніше опікувалися тільки люди, позаяк, по-перше, спеціалізуються на строго виокремлених типах діяльності (ШІ, який розпізнає рукописи, не може керувати авто, а мовна модель не може писати музику), по-друге, обробляють колосальні обсяги інформації неймовірно швидко й, по-третє, розв'язують проблеми новими, недоступними людям способами: приміром, мовні моделі, які генерують на вигляд осмислені тексти, насправді сполучають слова без розуміння їхніх значень

<sup>6</sup> Нейронна мережа — *схематичний* аналог нервової тканини. Вона покликана відтворювати (на дуже загальному рівні) будову мозку, а не продукти його діяльності: свідомість, розуміння тощо.

[Li, 2020]. Інакше кажучи, досягнення у сфері прикладних ШІ насправді пов'язані з *відмовою* науковців створювати штучний розум, ідентичний натуральному. Етичні ж проблеми, пов'язані з впливом ШІ-застосунків на людину, не завжди бентежать їх розробників.<sup>7</sup> Головним дороговказом для створення цих застосунків є ефективність у розв'язанні поставлених ціле-раціональних завдань, більшість з яких своєю чергою спрямовані на збільшення прибутків компанії-розробника. Надалі у статті йтиметься виключно про такі ШІ.

У західній філософсько-антропологічній традиції «людське питання» зазвичай розглядали як щось, що стосується ідей, досвіду, категорій розуму, норм, цінностей тощо, — речей, які можемо назвати *інтерналіями*, тобто внутрішніми властивостями людини, які вона спостерігає за допомогою рефлексії як суб'єкт. Проте щоб створювати сучасний ШІ з його згаданими вище особливостями, необхідно, усупереч цій традиції, орієнтуватися на людські екстерналії, тобто зовнішні характеристики людини як одного з можливих об'єктів наукового спостереження. Головною екстерналією, за допомогою якої Тюринг пропонував налаштувати ШІ на «людиноподібність», є загальноприйнята й загальнозрозуміла повсякденна мова. Примітно, що у статті Тюринга не знайшлося місця для пояснення того, хто такий «звичайний запитувач», що має верифікувати людськість. Проте апеляція до «звичайного запитувача» має сенс лише тоді, коли повсякденні уявлення про спілкування, на які пропонував орієнтуватися Тюринг, є непідконтрольною екстерналією, тобто незалежним мірилом людськості, на які не можуть вплинути розробники штучного інтелекту. Чи мають сучасні розробники ШІ до своїх послуг такі по-справжньому незалежні мірила?

#### 4. Від повсякдення до людської природи

«Єдиний пізнаваний світ — це світ нашого повсякденного життя», — проголосив Е. Гусерль 1906 року<sup>8</sup>. Феноменологічне мислення, що спрямоване «назад до речей», закріпило повсякденний досвід у центрі уваги європейських філософів XX сторіччя як первинну гносеологічну даність, як спільний знаменник людськості, а також плацдарм для чергового стрибка у трансцендентальне. Гусерля не цікавила антропологічна специфіка матеріального побуту, і доки не з'явилася «Критична теорія», питання «До яких речей і чому саме до них?» звучало б як невдалий жарт. Неспеціальні досвіди та типовий побут цікавили його не самі по собі, а як «чорна скринька», що постачає всім і кожному спільні докритичні очевидності, з яких складено наш життєсвіт.

<sup>7</sup> Див.: НАО, К. (2020, December 4). We read the paper that forced Timnit Gebru out of Google. Here's what it says. *MIT Technology Review* (<https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/>).

<sup>8</sup> Цит. за: Горц, А. (2008). Знание, стоимость и капитал. К критике экономики знаний (Окончание). *ЛОГОС*, 64, с. 6.

Життєсвіт є «суб'єктивним» проявом повсякдення. Останнє, втім, може бути предметом не лише феноменології, а й соціології. «Повсякденне життя не існує саме по собі, а виникає завдяки процесам „оповсякденення“, яким опираються процеси „подолання повсякденності“» [Вальденфельс, 1991: с. 41]. Повсякденне визначають через відмінність від експертного (яке існує у, наприклад, функціональних підсистемах суспільства Лумана) та екстраординарного. Повсякдення — це «звичне, впорядковане і близьке» [Вальденфельс, 1991: с. 45]. Це та царина життя, яку поділяють люди з їх плюралізмом компетенцій, звичок та світоглядів. Повсякдення — це *спільне*, і тому гарантує можливість порозуміння.

І справа не лише в тому, що повсякденний досвід, визначений певними культурними особливостями, формує той чи той історичний тип суб'єктивності. Повсякдення — це мовна сфера. Для Ю. Габермаса сфері повсякдення відповідає комунікативна дія, спрямована на порозуміння, — на протиположну стратегічній, спрямованій на цілераціональну мету дії, локусом якої є ринки, урядові організації та інші «системні» сфери [Bohman, Rehg, 2017]. Пізній Л. Вітгенштайн розглядав повсякденну мову не як плутанину винятків із граматичних правил, а як простір для розв'язання філософських проблем. Г. Патнем у «Значенні значення» (1975) звертається за схожою схемою до легітимаційної сили повсякденної мови. Він пропонує поняття лінгвістичного поділу праці: окрема людина не може володіти повною словниковою компетенцією, а гарантами значення різних понять є носії різних компетенцій, з яких і складається спільнота мовців. Патнем обстоює семантичний екстерналізм: значення перебуває поза мовцем. Навіть свідомість є для нього екстерналією — вона «розміщена» поза суб'єктом.

Фокус на буденності, безперечно, відповідав духові ХХ сторіччя. Ні на кого конкретно не спрямований «постав» [Gestell] індустріальної техніки [Хайдеггер, 1993b: с. 229] визначив те, як масове суспільство вдягається, чим харчується, як відпочиває тощо. Загальна шкільна освіта для всіх чоловіків і, що безпрецедентно, жінок, гомогенізувальний вплив радіо й телебачення, транспортні засоби, які «зробили далеке недалекоим» [Хайдеггер, 1993a: с. 322] типізували матеріальний побут більшості й уніфікували буденну мову настільки, що її можна розглядати як щось позасуб'єктне.

На початку статті вжито поняття «людська природа». Для потреб цієї розвідки його варто уточнити. Я розглядаю людську природу як подвійний концепт<sup>9</sup>, що вказує на фундаментальні характеристики людини двома способами одночасно. «Зовнішня» людська природа — це те, чим люди відрізняються від іншого сушого: корів, столів тощо. Це *описова* частина концепту. З того часу, як Руссо на прикладі «шляхетних дикунів» показав, що визна-

<sup>9</sup> Тобто концепт, який є описовим і нормативним водночас [Reuter, 2019]. Подвійних концептів — чимало. Наприклад, *мистецтво* є подвійним концептом тому, що означає як «будь-що, що можна побачити в галереї», так і «справжнє мистецтво на протиположну безталанній імітації» [Liao et al., 2020].

чальні характеристики людини змінюються від епохи до епохи, поняття природи людини вказуватиме не тільки на вроджені риси людей як двоногих тварин, а й на риси, яких вони набувають. Одне слово, «зовнішня» природа динамічна: люди використовують різні технології, вигадують мови й об'єднуються у держави.

«Внутрішня» людська природа — це ті властивості, які ми *приписуємо* всім людям: унікальність, раціональність, наявність свідомості, креативність тощо. Такі властивості можуть бути як набутими в певний період історії (унікальність особистості), так і натуральними (наявність свідомості у сенсі трансцендентальної єдності апперцепції).

Чи існує сутність людської природи? Відповідь на це питання вже не є першочерговою. Для феноменології та філософських теорій, що пов'язані з нею, відмова від пошуків первинної есенції людського ества була безпрограшною стратегією. Приміром, Гайдегер підходить до питання людини з боку первинності буття: воно визначає людину тому, що людина до нього *долучена* як пастир, що відповідає на його поклик. Для англо-американської традиції поняття сутності людини також радше чуже. Ось тому, наприклад, Патнем передовсім говорить про мовну прагматику; вона важливіша, ніж уявлення окремих людей про значення тих чи тих понять. Якщо ми характеризуємо суспільне й мовно-комунікативне як *виключно людське*, то остаточна відповідь на питання «сутності» людини непотрібна, оскільки говорити про комунікативну сферу зручніше. Тому-то далі я матиму на увазі під природою людини *сукупність дескриптивних і прескриптивних характеристик сучасної людини, якими ми послуговуємося в повсякденні, відрізняючи живих людей від тварин чи речей і свідомо спілкуючись із людьми як людьми*. Приміром, моє визначення сучасної природи людини передбачає те, що кожна людина має тіло й «теорію свідомості» [theory of mind], але воно не базується на уявленні, що людина має стабільну сутність.

Ми не можемо визначити, що робить людину людиною, якщо обмежимося розглядом її повсякдення, адже воно стало філософською проблемою лише тоді, коли розмови про сутність людини — щось, що могло би зв'язати дескриптивні й прескриптивні частини поняття природи людини — вийшли з моди. Це означає: *щось, що використовує природну мову й долучене до повсякденних людських практик, не обов'язково має відповідну людську природу* (остання має прескриптивні аспекти, які безпосередньо не зафіксовані у повсякденні). Таке тривіальне твердження допоможе зрозуміти, чим може бути ШІ, який імітує людину.

Сучасні ШІ-застосунки «зважають» не тільки на те, як розмовляють люди у текстовій, аудіо- чи відеоформі, а також і на те, як вони повсякденно взаємодіють з матеріальним побутом (це робить, приміром, автопілот чи робот-пилосос). Проте та засада, з огляду на яку Тюринг запропонував «калібрувати» ШІ стосовно до мовної компетенції, визначила, як сучасні «слабкі» ШІ імітують людськість в усіх сферах.



Перед тим, як розглянути, чим саме є Тюрингова «гра в імітацію», зауважимо, чим вона *не* є. По-перше, гра в її оригінальній формі не є достатнім тестом на «розумність» (цього не стверджує і сам Тюринг). У подальшому виявилось, що тест Тюринга (так подальші коментатори називали його «гру») має низку практичних недоліків, тож «дослідники ШІ присвятили мало уваги проходженню тесту Тюринга [в оригінальній формі]» [Russell, Norvig, 2020: p. 3]. Свою популярність тест Тюринга здобув як мисленнєвий експеримент, який з усією очевидністю вказує на засадові проблеми ШІ. По-друге, позиція Тюринга, окреслена у його статті, не є строго науковою. Він присвячує окремих розділ телепатії та іншим паранормальним явищам, реальність яких, мовляв, «підтверджують статистичні дані» (посилань на ці дані у нього, звісно, немає), і пропонує провадити свою гру в спеціальних «антителепатичних кімнатах». Тут читач опиняється перед вибором: або прийняти, що математик Тюринг вірив у «привидів» і «злих духів» [Turing, 1950: p. 440], або визнати, що найбільш радикальні тези його статті були гумористичним перебільшенням. У такому разі не дивуватимуть як його безмежний оптимізм щодо створення в недалекому майбутньому «машин, які мислять», так і низка інших химерних фрагментів, наприклад про те, як він карає дитину-комп'ютер [Ibidem, p. 452].

Вихідний постулат його статті — це «беззмістовність» питання про те, чи можуть машини думати. Відповідно, за допомогою «гри в імітацію» Тюринг показує, на яких *ще* засадах можна будувати ШІ, якщо відмовитися від поняття «мислення» й водночас не прив'язувати ШІ до інших нормативних характеристик людської природи: переживання емоцій, креативності тощо. Самі фундатори дисципліни ШІ, до речі, не вважали, що поняття мислення для них нерелевантне, а історія фундаментальних досліджень ШІ прямо пов'язана з біологією, філософією та психологією. Мислення радше виявилось проблемою, успішно розв'язати яку вони не змогли, а після «зими ШІ» 1970-х років, коли фінансування досліджень Збройними силами США суттєво скоротилося, а в університетах перестали з'являтися нові робочі місця для теоретиків ШІ, фундаментальні розробки з ШІ, на думку М. Мінські, так і не відновилися у необхідному обсязі [Singularity.FM, 2013]. Нині ж, як стверджує Мінські, методична база дисципліни зосталася та сама, що й півсторіччя тому, адже розвиток «слабких» ШІ на неї принципово не вплинув.

Отже, якщо проблема комп'ютерного мислення не розв'язана для потреб «сильного» ШІ, то творці «слабкого» ШІ цю фундаментальну нерозв'язаність просто ігнорують. Філософські засади «гри в імітацію» Тюринга залишаються засадовими для розробки сучасного прикладного штучного інтелекту. Тому сучасні ШІ-застосунки орієнтуються на окремі описові характеристики людської природи, які формують сферу повсякденного життя — своєрідний спільний знаменник, який об'єднує всіх людей, і не орієнтуються на прескриптивні риси людської природи, хоч би якими вони були.

## 5. Цифрове повсякдення

Чим далі від Гусерля й Тюринга й ближче до епохи планетарного обчислення, тим вагомніше те, від якого уявлення про людське повсякдення ми відштовхуємося, навчаючи ШІ «людиноподібності». Цифрові об'єкти — які «набирають форми на екрані чи ховаються у серверній частині комп'ютерної програми» та «складаються з даних чи метаданих» [Hui, 2016: р. 36] — посідають у нашому повсякденні не менш значне місце, ніж столи, про які полюбляв розмірковувати Мерло-Понті, або чашки Гайдегера. Проте, на відміну від останніх, цифрові об'єкти не є предметом досвіду або інтуїції у строго феноменологічному розумінні [Hui, 2016: р. 36]. Тоді до якої саме даності повсякдення звертаються сучасні розробники «слабких» ШІ?

Ми завжди взаємодіємо з «слабким» ШІ «на його території»: спілкуємося з ним у чатах, а не віч-на-віч, спілкуємося з ним на загальні теми, позаяк це те, чого найлегше навчити ШІ на величезних базах даних, що містять приклади мововжитку мільйонів користувачів, ніяк не пов'язаних між собою. Ці «великі дані»<sup>10</sup>, здавалося б, репрезентують спільне, неунікальне повсякдення користувачів і можуть бути незалежним мірилом людськості, яке потрібно ШІ-інженерії.

Утім, уявлення про те, що дані — це щось, що генерують самі користувачі інтернет-платформ, але в подальшому його відчужують і привласнюють платформи, є хибним. Дані (і це відомо ще з часів Гайзенберга) створюють ті, хто спостерігає, їх не генерує предмет спостереження. Унікальне життя кожного громадянина Мережі «репрезентоване» сукупністю типових data points: поштова адреса, рекламний ID, списки контактів, відвіданих локацій, банківських транзакцій тощо. Чому саме такі data points, а не якісь інші? Характер того, що про нас відомо умовному Google, визначений не лише засадничою прозорістю нашого цифрового життя для зовнішніх спостерігачів чи нашою готовністю ділитися інформацією про себе. Зміст «зібраних» масивів даних визначений тим, які саме дані необхідні Google для своїх бізнес-потреб (тому в пріоритеті будуть дані, які допоможуть продавати рекламу), якими методами користуються дата-науковці, що працюють в цій компанії, і тим, яким чином впорядковані ці дані. Тому метафора спостереження, яку використовують при обговоренні так званого «капіталізму платформ», відволікає увагу від більшої проблеми. Дані про повсякденне життя не репрезентують його в строгому сенсі. Їх зміст легітимізує побудову цифрового повсякдення за правилами, які були визначені (свідомо чи ні) творцями платформ ще до того, як дані були створені у необхідній кількості.

<sup>10</sup> «Великі дані [big data] відрізняються від звичайних даних, які обробляють комп'ютери, особливо великими обсягами (навіть для новітнього устаткування), особливою швидкістю (тобто дані генеруються у швидкому темпі; наприклад, email-транзакції) та особливою складністю для обробки» [Madden, 2012]. Під останнім мається на увазі, що до даних не прив'язано достатньо семантики — вони «сирі».

Тому якщо сучасний ШІ і наблизився до того, щоб вдало імітувати людську поведінку в цифровому просторі, то це сталося не тому, що його розробники зафіксували «суть» повсякденного життя у програмному коді. Це сталося тому, що ми довірливо розглядаємо цифрове середовище із його довільною архітектурою як середовище повсякденного спілкування, але поводимося у ньому за правилами, відмінними від повсякденно звичних. Якщо врахувати те, що ШІ-застосунки — це не лише «користувачі» цього середовища нарівні з людьми, а й структурний елемент архітектури цифрового простору, стане очевидно, що це не ШІ імітує «природну» повсякденну поведінку, а повсякденна поведінка стає обмеженою такими формами, які можуть імітувати ШІ-застосунки, і починає відтворювати ці форми. Алгоритми машинного навчання не лише «навчаються» розуміти людей, а й дисциплінують своїх користувачів.

Відтак сучасне повсякдення, яке значною мірою є цифровим, не може бути незалежним «орієнтиром» для творців ШІ, адже воно схильне підлаштовуватися під конкретні особливості ШІ-алгоритмів. «Можна запідозрити, що є прямий зв'язок між кінцем теорії та виникненням програмних засобів, оскільки програмні засоби — це настільки ж лінгвістична технологія, наскільки й технологічний тип мови», — припускає Б. Братон [Bratton, 2015: р. 5]. Тепер і предмет дослідження ШІ-науковця, і спосіб, і його результат — це відрізок однієї і тієї самої технологічної й, врешті, математичної мови.

«Створювати замість пояснювати». Цей квазімарксистський підхід до науково-технологічної діяльності якнайліпше втілено у притаманній сфері ШІ нерозривності між теорією і практикою, між імітацією повсякденного спілкування й коригуванням його особливостей. *Синергія між відтворенням повсякденної поведінки людей та програмуванням їх повсякденної поведінки, притаманна прикладним ШІ-застосункам, є фундаментальною рисою цієї технології*, і водночас характеризує прикладний ШІ як щось більше, ніж цілераціональна технологія.

## 6. Непрозорість техніки

Якщо ефективність сучасних прикладних ШІ — це самовтілюване прощство, то чому їх вплив на наше повсякдення не є загальновідомим?

У «Десяти парадоксах техніки» Е. Фінберг стверджує, що «більшість наших [філософських] ідей щодо техніки, ґрунтованих на здоровому глузді, була хибною» [Фінберг, 2016: с. 259]. Філософія техніки виникла у ХІХ сторіччі як філософія історичного типу техніки, панівного на той час: органопроєкція Е. Капа — яскравий приклад інструменталістської філософії техніки. У межах інструменталістської парадигми техніка — це «просто інструмент, за допомогою якого людство як вид задовольняє свої потреби» [Feenberg, 2006]. Цей погляд досі є «стандартною точкою зору» [Ibidem, р. 6] більшості.

Фундаментальним для інструменталістської парадигми є уявлення, що *людина в змозі впливати на навколишнє середовище за допомогою технічних*

засобів так, щоб ці засоби не впливали на неї саму. Цей погляд, своєю чергою, вимагає поділу світу на технічну та нетехнічну царини, де першій відповідає цілераціональність, ефективність і тому подібне, а другій — ціннісна раціональність, культура та, загалом, більшість питомих філософських питань, включно з людською природою.

«Враження, що ми незалежні від об'єктів, на які впливаємо технічно, може виникнути лише в тому разі, якщо відповідна зона дії визначена [надто] вузько» [Фінберг, 2016: с. 270]. Фокусуючись на очікуваних результатах техніки, інструменталізм виносить за дужки не лише імплікації того факту, що будь-яка дія у світі, навіть розглянутому з натуралістичних позицій, є взаємодією, а й те, що інструментальні технології за визначенням завжди функціують у певному середовищі, на яке вони впливати не покликані, але наявності якого вони потребують.<sup>11</sup> Також цьому підходу властива сліпота до історичних причин з'яви та поширення певних технологій. Звично вважати, що саме ефективність певної технології робить її успішною, проте «не ефективність пояснює успіх, а успіх пояснює ефективність» [Ibidem, р. 266]. Те, що певна технологія розповсюджена, інколи має політичні, макроекономічні чи культурні підстави, сама ж ефективність технології qua технології може при цьому залишатися низькою<sup>12</sup>.

Інструменталізм не є філософськи задовільним описом техніки як явища, бо розглядає її надто вузько. Одну з найвпливовіших голістичних філософських концепцій техніки пропонує Ж. Симондон (2020). Появу типового поділу об'єктів на ті, що мають функцію (цеглина), й ті, що мають сенс (картина), він пов'язує з гіломорфічною схемою Аристотеля, за якою речі складені з форми й матерії<sup>13</sup>. Проте картина світу, де парадигмою технічності є майстерня, до якої постачають матеріали, а на виході отримують артефакти, ґрунтована на неадекватних уявленнях про суть роботи й техніки — вона відповідає досвіду вільного громадянина, що спостерігає за незрозумілою йому роботою рабів, але не відповідає досвіду сучасної людини, яка активно залучена до постіндустріального виробництва. Симондонова теорія техніки спирається на поняття *трансіндивідуального колективу*, тобто системи, що складається як з об'єктів, так і з осіб, і визначена технічною діяльністю як комунікативним процесом. Технічна діяльність — парадигма соціальності, а не навпаки, і так було завжди. Навіть гончар взаємодіє з матеріалом й інструментами

<sup>11</sup> Це передусім інші технології, спрямовані на досягнення паралельних людських потреб, та, що суттєвіше, стабільна «нетехнічна» царина, яка уможливує матеріальне існування технічних засобів відповідно до їх специфікації, а також можливість їх осмисленого та коректного використання людьми.

<sup>12</sup> Bitcoin, що був покликаний замінити традиційні валюти, виявився у крайній неефективним у своїй головній ролі, оскільки Bitcoin-транзакції повільні й дорогі порівняно з транзакціями через, приміром, Visa, а сама криптовалюта надто волатильна. Проте ця технологія успішна: Bitcoin — дуже прибутковий актив.

<sup>13</sup> Також див.: [Feenberg, 2006:р. 3]: «[Античні] греки привносять [...] аспекти techné у царину природи і розглядають усю природу в телеологічних категоріях».

комунікативно, дослухаючись до зворотного зв'язку. Людина — не господар машин-рабів, а диригент «технічних ансамблів» [Lindberg, 2019].

Б. Штиглер (1998) йде ще далі й пропонує для опису техніки поняття *третьої ретенції*, тобто колективної пам'яті, яка зберігається у технологіях. Відповідно, технічний прогрес є процесом екзосоматизації знання, а будь-який технічний артефакт *буквально* втілює знання про світ та/або оптимальні способи взаємодії з ним; незалежно від Штиглера це показав А. Горц у «Нематеріальному» (2010). Безперечно, способи «зберігання» екстерналізованого знання різняться від епохи до епохи. Не міф і не книжка, а мережа нині є колективним спогадом і кодексом неформальних знань, а формальне ноу-хау вже давно втілюють у технічних засобах — автоматах, вправніших за ремісників.

Голістичні філософії техніки відмовляються ділити світ на технічну та нетехнічну царини, зв'язують технічність з комунікативністю та знанням — і таким чином об'єднують філософію техніки з антропологією. Це, безперечно, більш чесний погляд на технологічне суспільство, ніж інструменталізм, який у строгому сенсі навіть не філософія техніки, а констатація головного *інженерного принципу* техніки, мати уявлення про який необхідно, але не достатньо для опису техніки як явища. Інструменталізм виносить за дужки джерела існування техніки, які не можна рефлексивно вивести з інженерного принципу техніки. За схемою це нагадує кантівський проєкт, хіба що замість реальності, розглянутої як недоступні розуму речі в собі, інструменталізм розглядає як недоступні для технічного розуму цінності, втілені в технологіях, історичні та культурні контингентності успіху технологій та непрямі наслідки їх використання.

Проте, якщо ми проігноруємо інженерний принцип техніки, то обсяг поняття техніки розмиється настільки, що зміст цього поняття визначити буде непросто. Інструментальні *уявлення* про техніку — невіддільна складова сучасної техніки й запорука її успішності. Тож коли інженери докритично приймають інструменталізм, а користувачі мають аналогічно помилкові очікування від технологій, що з'являються в їх житті, це вже не стільки мета-технологічний контекст, оцінювати який покликана теорія на межі антропології та філософії техніки, скільки принципова риса сучасної техніки.

Якщо технологія створена з огляду на недостатні й хибні уявлення про техніку як суспільне явище й водночас має суспільно значущі наслідки, які залишаються непомітними для її користувачів (бо їхні очікування, як і у інженерів, обмежені інструментальною парадигмою), то я називатиму таку технологію *непрозорою*. Непрозорість — рефлексивна форма непередбачених наслідків інженерної настанови: люди, на життя яких впливає непрозора технологія, не усвідомлюють вплив техніки на них, бо недооцінюють її потенціал. Приміром, поки у паливо для авто додавали свинець, а громадськість, знаючи про цю домішку, не усвідомлювала, що вона шкідлива для здоров'я, ця технологія збільшення ефективності палива була непрозорою.

Загалом саме пелена непрозорості приховує від нас, що ми взаємодіємо з технікою комунікативно (Симондон), а не інструментально, і приховує від нас, що техніка є екстерналізованим знанням чи спогадами про природу чи суспільство (Штиглер).

Ця ж непрозорість притаманна технології ШІ, в якій математичний алгоритм та дані про людей, які він узагальнює, насправді є невіддільними одне від одного. Нейронно-мережевий ШІ екстерналізує спільне знання про повсякдення так, що ми перестаємо контролювати це знання. Здавалося б, сучасний ШІ як відкрита «розумна» машина якнайкраще ілюструє тези Симондона про комунікативний характер техніки та нероздільність буття техніки та людини. Проте розробники ШІ досягають успіху, обмежуючись інструментальною парадигмою та свідомо ігноруючи питання, що стосуються нетехнічних царин («ми створюємо ШІ, але нам байдуже, що таке мислення»). Поки ширший загал теж мислить ШІ в рамках інструментальної парадигми (мовляв, це технологія, яка «розпізнає природну мову» чи «рекомендує цікаві відео»), будь-який вплив ШІ на людину залишатиметься непомітним.

Чи буде, втім, достатньо, якщо ми означимо ШІ як «особливо комунікативну» технологію і зупинимося на цьому? Голістичні філософії техніки мали на меті розширити обсяг поняття техніки, *доповнивши* його чимось, що не вважали технічним до того. Втім, якщо ми розглянемо технологію ШІ з усіма аспектами її непрозорості, виявиться, що *більшість її філософськи значущих рис належать виключно до розширеного обсягу поняття техніки*. Інакше кажучи, робити наголос на тому, що штучний інтелект є передусім технологією, не надто продуктивно. Буде зручніше розглядати ШІ з позицій інших дисциплін — антропології, соціальної філософії тощо, бо принципові особливості ШІ стосуються саме них.

Не заперечуючи його технологічної природи, віднесемо ШІ до категорії *соціальної практики*. Справді, ШІ не лише взаємодіє з людиною, ніби намагаючись приховати, що він технологія, а і є явищем виключно суспільного рівня, і потребує активної залученості членів суспільства, навіть якщо вони це так не осмислюють. Через «великі дані» прикладні ШІ-застосунки «узагальнюють» суспільство загалом, а фокусом їх застосування є великі групи людей, які обирають комунікацію з ШІ, щоб з його допомогою спілкуватися між собою.

## 7. ШІ як засіб самоопанування людини

У «Правилах для людського зоопарку» П. Слотердайк (2009) доводить, що гуманізм — не спроба вирвати природну раціональність людини з пут забобів та міфологій, а навпаки, намагання приборкати первісні деструктивні імпульси, притаманні людській природі, за допомогою «гуманітарної», тобто книжної освіти. Не самозвільнення людства, а його самоприборкання — таку мету ставили перед собою просвітники-гуманісти, які прагнули солідаризувати

співгромадян у любові до «класиків» і «граматизувати» суспільне життя, тобто організувати його за моделлю буржуазного літературного клубу. Практики самоопанування людства Слотердайк і називає антропотехнологіями. Гуманізм був чільною антропотехнологією Модерну, поки не зазнав краху під час Другої світової, проте існують й інші антропотехнології. Перехід людини до осілого життя на світанку цивілізації був першим антропотехнологічним самоопануванням людини; зараз же в арсеналі людства чимало антропотехнологій різного рівня успішності й перевірності на практиці — від платонівського проєкту, викладеного в «Державі», до евгеніки та генної інженерії людини.

Стосунок поняття антропотехнології до поняття технології у «Правилах для людського зоопарку» не прояснений прямо. Для Слотердайка «антропотехнологія» належить до такого антропологічного словника, де питання про те, що таке «справжня» людина, включає питання про ті чи ті засоби, за допомогою яких люди змінюють свою природу. У доповіді «Антропо-технологія» (2014) Слотердайк прояснює свою позицію. Поки суб'єктно-об'єктна дихотомія (наслідок якої — поділ світу на технічну та нетехнічну царини) була філософськи актуальною, техніка як інструмент здавалася нам *домінуванням* активного над пасивним, живого над неживим. Цей погляд був хибним завжди, а по з'яві кібернетики постметафізичний і комунікативний характер техніки проявляється настільки явно, що техніка та людина набувають «спільної ідентичності». У цифровому середовищі все менше йдеться про домінування людини за допомогою інструментів; суб'єкт — той, хто домінує — у цифровому середовищі розчиняється, бо йому нічого протиставити. Як і для Штиглера, для Слотердайка артефакти — «спогади чи рефлексії, які стали об'єктивними». Але, на відміну від Штиглера, він не абсолютизує технічність, а навпаки, інтегрує її у поняття людського<sup>14</sup>. Технічним засобом гуманізму як антропотехнології було, приміром, книгодрукування чи шкільна освіта як система. Проте гуманізм неможливо пояснити виключно через інженерну логіку його технічних засобів. Кінцевим засобом і водночас суттю гуманізму була «граматизація» — соціальна практика, яка потребує технічних засобів, але не звідна до них.

Отже, антропотехнологія — це *соціальна практика недомінантного впливу людей на свою природу*. Звісно, сама розмова про «приборкання людини» небезпідставно здається розмовою про домінацію: хто саме вирішує, якою має бути природа нової людини? Проте, на думку Слотердайка, історичні форми антропотехнологій, зокрема гуманізм, були «безсуб'єктивним біокультурним дрейфом» (2009), тобто здійснювалися людиною без достатньо чіткого усвідомлення, що вона займається самообмеженням. Питання ж, що стосуються біоетики, зараз надто небезпечні, щоб бути поставленими з

---

<sup>14</sup> «Якщо людина існує, вона існує тому, що технологія змусила її еволюціонувати з протолюдини» [Sloterdijk, 2014: p. 16].

необхідною для впливу на людську природу рішучістю. Дійсна антропотехнологія має бути непомітною.

За горизонтом гуманізму, за доби цифрових технологій, які є невідомі-нантними й тому стають невіддільною частиною нашого самоусвідомлення, проблема антропотехнології постає з новою актуальністю. І все ж вона залишається малопомітною, адже здоровоглузові уявлення про техніку поки що заважають нам розрізнити суспільні імплікації технологій штибу ШІ, приховуючи їх пеленою непрозорості. Інженерні особливості прикладного, «слабкого» ШІ (комп'ютерна система, що «мислить» чи «поводиться» як людина у чітко окреслених обставинах), є лише верхівкою айсберга ШІ як суспільного явища. Так само як «класичні» книжки була засобом «граматизації», технологія ШІ уможливорює ШІ як соціальну практику. Саме у соціальній практиці ШІ сконцентрований антропотехнологічний потенціал цифрової епохи. Практика взаємодії зі штучним інтелектом, яка, своєю чергою, тотожна взаємодії з іншими людьми, доступ до яких у цифровому середовищі опосередкований ШІ, є тим невідомі-нантним способом, яким сучасна людина змінює те, чим вона є.

Словом, не дивно, що цифрове середовище, якому властиве деяке нерозрізнення між людиною та алгоритмом, стає антропотехнологічним плацдармом. На відміну від явно антропотехнологічних питань штибу модифікації геному людини, самопанування людини за допомогою штучного цифрового образу своєї природи, який постійно змінюється під впливом культурних, економічних, політичних та інженерних чинників, забезпечене від критики непрозорістю техногенного середовища, у якому і здійснюють це самоопанування. «Слабкі» ШІ в їх сучасному стані — це випадковий, побічний наслідок розвитку платформ як способу організації цифрового простору. Безперечно, творці того чи того ШІ-застосунку не мають на меті «вносити правки» до ідеї людини — вони створюють комерційний продукт, досягають інженерних цілей, та й по всьому. Але от ШІ як загальнолюдська соціальна практика, глобальні риси якої ніким свідомо не визначені, скеровує дрейф людської природи у новому, поки що прихованому технічною непрозорістю, напрямку.

Отже, як саме влаштована соціальна практика штучного інтелекту? ШІ, що спочатку був покликаний відтворювати «природну» поведінку людини в доцифровій повсякденній сфері, наразі відтворює специфічний тип поведінки, а саме те, що люди роблять у цифровому середовищі відповідно до особливостей цього середовища. Цифрове ж середовище — це середовище, організоване за допомогою ШІ. Тому ми можемо результативно взаємодіяти зі «слабким» ШІ тільки відповідаючи зразкам людськості, кодифікованим у ньому. Останні подекуди фіксують «природну» поведінку, а подекуди є зручними для впровадження схематизмами. Відповідно, у *цифровому середовищі людина імітує поведінку штучного агента, а не навпаки*. Таким чином вона закріплює цю поведінку як власну норму, оскільки статус «користувача» може отримувати як людина, так і штучний агент; до того ж інколи складно



знати напевно, з першим чи з другим ми взаємодіємо. Вочевидь, якщо й далі поступ ШІ пов'язуватимуть з тим, що люди підлаштовуватимуть своє повсякдення під ШІ-норми, то з'ява «сильного» ШІ вже не здаватиметься чимось неймовірним.

Людина імітує штучний інтелект, який є техногенним, модифікованим і випадковим відбитком людської природи — так виглядає антропотехнологія цифрової доби. Цим можемо пояснити успіх ШІ-інженерії та ШІ-за-стосунків, якщо нам необхідно водночас зберегти довіру до понять мислення, мови та людини як таких, що мають зміст. Але на питання про те, *якою саме* стане людина, що опановує себе за допомогою ШІ, навряд чи можна задовільно відповісти саме зараз. Щоб розгледіти змістовні наслідки ШІ як антропології, необхідно не лише дочекатися часу, доки емпірика ШІ-впливу людини на людину буде задовільно описана, а й зняти пелену непрозорості, притаманну цифровим технологіям — саме їх непрозорість є активним елементом сучасної антропотехнології. Ці завдання, втім, вимагатимуть не стільки розвитку філософсько-технічної традиції, скільки з'яви нової теоретичної платформи, яка не лише констатуватиме «злиття» людини й техніки, а й міститиме необхідний категорійний апарат, який буде ліпшою альтернативою новоевропейського метафізичного апарату, ґрунтованого на протиставленні суб'єкта та об'єкта, і не виявиться черговою спробою критики метафізики, що не пропонує їй реальних альтернатив.

## 8. Висновки

Норма «позначає певні очікування, які генералізуються та універсалізуються» [Єрмоленко, 2019: с. 24]. Ці очікування спираються на факти людського життя, і від цих фактів залежить прескриптивна частина наших уявлень про людську природу. Перехід від спроб створити «сильний» ШІ, який мав імітувати людський інтелект, до створення «слабкого» ШІ, який автоматизує людську роботу й імітує спілкування, став переходом від орієнтації на фактичне до імпліцитної орієнтації на контрфактичне: «якою має бути людина, яка взаємодіє з ШІ». Розробники ШІ, які, спираючись на факти повсякдення, кодифікували людськість, прогавили момент, коли ШІ став впливати на ці факти — на зразки людськості. Тому сучасний «слабкий» ШІ кодифікує не лише описову, а й прескриптивну частину людської природи.

Ця прескриптивна частина, проте, не еквівалентна моральним нормам, і стверджувати, що ШІ впливає на етику безпосередньо, поки зарано. Цифрове середовище — не єдине середовище, в якому ми живемо. До того ж ми недостатньо розуміємо природу зв'язку факту і припису у цьому середовищі — новому, незвичному й непрозорому, адже обґрунтування моральних належностей досі робили з огляду на уявлення про природу людини, сформовані у доцифровий час.

Те, що ми можемо назвати антропотехнологічним приписом ШІ, за внутрішньою логікою цієї сфери буде описом, адже тілесна напередданість

повсякдення, яка забезпечує порозуміння, — це щось, що неможливо помислити у безтілесному середовищі. Лише створивши універсальну онтологію, що описуватиме і тілесне, і цифрове, ми зможемо переоцінити регуляторні принципи сучасних етик чи створити нові.

Уніфікацію онтології можна здійснити, зокрема, на підставі своєрідної антропології техніки чи технічної антропології, яка знімає дихотомію людського і технічного; принаймні, цей напрямок окреслює Слотердайк, та й голістичні філософії техніки можна розвивати схожим чином. Без уніфікованої онтології нам доведеться вдаватися до безнадійних суперечок про те, яке середовище первинне. Хоча природа каузально передує цифровому середовищу, консервативна етика — зокрема та, яку пропонує Йонас — не в змозі розв'язати проблем, які неодмінно виникатимуть у цифровому середовищі. Ідея ж відмовитися від нашої цифрової домівки здається надто фантастичною. Не менш фантастичною, ніж криптогегельянські спроби довести первинність цифрового. Проте оцінити ШІ як антропотехнологію можна навіть наосліп — через поняття відповідальності. Зміщення фокусу з «сильного» на прикладний ШІ супроводжувала відмова від відповідальності розробників ШІ за зміст норм людськості, утілених у ШІ-застосунках як «чорних скриньках». Незалежно від того, наскільки значним є вплив ШІ на природу людини, те, що цей вплив прихований, некерований і випадковий, є практичною проблемою не тільки з погляду будь-якої етичної традиції, а й з погляду здорового глузду.

#### ДЖЕРЕЛА

- Вальденфельс, Б. (1991). Повседневность как плавильный тигль рациональности. У: *Социологос* (сс. 39—50). Прогресс.
- Горц, А. (2010). *Нематериальное. Знание, стоимость и капитал*. Изд. дом. ГУ ВШЭ.
- Єрмоленко, А. (2019). Нормативна когерентність філософського дискурсу. *Філософська думка*, 5, 21—28. <https://doi.org/10.15407/fd2019.05.021>
- Йонас, Г. (2001). *Принцип відповідальності. У пошуках етики для технологічної цивілізації*. Лібра.
- Фінберг, Е. (2016). Десять парадоксів техніки. В: *Антологія сучасної філософії науки, або Усмішка ASIMO* (сс. 258—281). ЛНУ імені Івана Франка.
- Хайдеггер, М. (1993а). Вещь. В: *Время и бытие: Статьи и выступления* (сс. 316—327). Издательство «Республика».
- Хайдеггер, М. (1993б). Вопрос о технике. В: *Время и бытие: Статьи и выступления* (сс. 221—238). Издательство «Республика».
- Anand, K., Wang, Z., Loog, M., van Gemert, J. (2020). Black Magic in Deep Learning: How Human Skill Impacts Network Training. *ArXiv:2008.05981 [Cs]*. <http://arxiv.org/abs/2008.05981>
- Bohman, J., Rehg, W. (2017). Jürgen Habermas. In: E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2017)*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2017/entries/habermas/>
- Bratton, B. H. (2015). *The stack: On software and sovereignty*. <http://site.ebrary.com/id/11206783>
- Buhrmester, V., Münch, D., Arens, M. (2019). Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *ArXiv:1911.12116 [Cs]*. <http://arxiv.org/abs/1911.12116>

- Dennett, D. C. (1978). Artificial intelligence as philosophy and as psychology. *Brainstorms: Philosophical Essays on Mind and Psychology*, 109—126.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55 (10), 78—87. <https://doi.org/10.1145/2347736.2347755>
- Dreyfus, H. L. (1992). What computers still can't do: A critique of artificial reason. MIT Press.
- Feenberg, A. (2006). What is the philosophy of technology? In: J. R. Dakers (Ed.), *Defining Technological Literacy: Towards an Epistemological Framework*. Palgrave-Macmillan.
- Hui, Y. (2016). *On the existence of digital objects*. University of Minnesota Press.
- Li, C. (2020, June 3). *OpenAI's GPT-3 Language Model: A Technical Overview*. <https://lambdalabs.com/blog/demystifying-gpt-3/>
- Liao, S., Meskin, A., Knobe, J. (2020). Dual Character Art Concepts. *Pacific Philosophical Quarterly*, 101 (1), 102—128. <https://doi.org/10.1111/papq.12301>
- Lindberg, S. (2019). Being with Technique—Technique as being-with: The technological communities of Gilbert Simondon. *Continental Philosophy Review*, 52 (3), 299—310. <https://doi.org/10.1007/s11007-019-09466-9>
- Madden, S. (2012). From Databases to Big Data. *IEEE Internet Computing*, 16 (3), 4—6. <https://doi.org/10.1109/MIC.2012.50>
- McCarthy, J., Minsky, M., Rochester, N., Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- Müller, C. J. (2016). Prometheanism: Technology, digital culture and human obsolescence. Rowman & Littlefield International.
- Nilsson, N. J. (1998). *Artificial Intelligence A New Synthesis*. Elsevier Science. <http://kcl.eblib.com/patron/FullRecord.aspx?p=1014253>
- Penrose, R. (1995). Beyond the Doubting of a Shadow A Reply to Commentaries on Shadows of the Mind. *PSYCHE: An Interdisciplinary Journal of Research On Consciousness*, 2.
- Putnam, H. (Ed.) (1975). The meaning of 'meaning.' In: *Mind, Language and Reality* (pp. 215—271). Cambridge University Press. <https://doi.org/10.1017/CBO9780511625251.014>
- Reuter, K. (2019). Dual character concepts. *Philosophy Compass*, 14 (1), e12557. <https://doi.org/10.1111/phc3.12557>
- Russell, S. J., Norvig, P. (2020). *Artificial intelligence: A modern approach* (Fourth edition). Pearson.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3 (3), 417—424. <https://doi.org/10.1017/S0140525X00005756>
- Simondon, G., Adkins, T. (2020). *Individuation in light of notions of form and information*. University of Minnesota Press.
- Singularity.FM (2013). *Marvin Minsky on AI: The Turing Test is a Joke!* (July 12, 2013). <https://www.singularityweblog.com/marvin-minsky/>
- Sloterdijk, P. (2009). Rules for the Human Zoo: A Response to the Letter on Humanism. *Environment and Planning D: Society and Space*, 27 (1), 12—28. <https://doi.org/10.1068/dst3>
- Sloterdijk, P. (2014). Anthro-Technology. *New Perspectives Quarterly*, 31 (1), 12—19. <https://doi.org/10.1111/npqu.11419>
- Stiegler, B. (1998). *Technics and time*. Stanford University Press.
- Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX (236), 433—460. <https://doi.org/10.1093/mind/LIX.236.433>
- Weiser, M. (1999). The Computer for the 21st Century. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3 (3), 3—11. <https://doi.org/10.1145/329124.329126>

Одержано 12.07.2021

## REFERENCES

- Anand, K., Wang, Z., Loog, M., van Gemert, J. (2020). Black Magic in Deep Learning: How Human Skill Impacts Network Training. In: *ArXiv:2008.05981 [Cs]*. <http://arxiv.org/abs/2008.05981>
- Bohman, J., Rehg, W. (2017). Jürgen Habermas. In: E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2017)*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2017/entries/habermas/>
- Bratton, B. H. (2015). The stack: *On software and sovereignty*. <http://site.ebrary.com/id/11206783>
- Buhrmester, V., Münch, D., Arens, M. (2019). Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *ArXiv:1911.12116 [Cs]*. <http://arxiv.org/abs/1911.12116>
- Dennett, D. C. (1978). Artificial intelligence as philosophy and as psychology. *Brainstorms: Philosophical Essays on Mind and Psychology*, 109–126.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55 (10), 78–87. <https://doi.org/10.1145/2347736.2347755>
- Dreyfus, H. L. (1992). What computers still can't do: A critique of artificial reason. MIT Press.
- Feenberg, A. (2006). What is the philosophy of technology? In: J. R. Dakers (Ed.), *Defining Technological Literacy: Towards an Epistemological Framework*. Palgrave-Macmillan.
- Feenberg, A. (2016). Ten paradoxes of technology. [In Ukrainian]. In: Anthology of the Modern Philosophy of Science, or ASIMO Smile (pp. 258–281). Ivan Franko National University of Lviv. [=Фінбепр 2016]
- Gorz, A. (2010). *Intangible. Knowledge, value and capital*. [In Russian]. PH. SU HSE. [=Торч 2010]
- Heidegger, M. (1993a). The Thing. [In Russian]. In: *Time and Being: Articles and Speeches* (pp. 316–327). PH "Respublika". [=Хайдеггер 1993a]
- Heidegger, M. (1993b). The question of technology. [In Russian]. In: *Time and Being: Articles and Speeches* (pp. 221–238). PH "Respublika". [=Хайдеггер 1993b]
- Hui, Y. (2016). *On the existence of digital objects*. University of Minnesota Press.
- Jonas, H. (2001). *The principle of communication. At the jokes of ethics for technological civilization*. [In Ukrainian]. Libra. [=Йонас 2001]
- Li, C. (2020, June 3). OpenAI's GPT-3 Language Model: A Technical Overview. <https://lambdalabs.com/blog/demystifying-gpt-3/>
- Liao, S., Meskin, A., Knobe, J. (2020). Dual Character Art Concepts. *Pacific Philosophical Quarterly*, 101 (1), 102–128. <https://doi.org/10.1111/papq.12301>
- Lindberg, S. (2019). Being with Technique—Technique as being-with: The technological communities of Gilbert Simondon. *Continental Philosophy Review*, 52 (3), 299–310. <https://doi.org/10.1007/s11007-019-09466-9>
- Madden, S. (2012). From Databases to Big Data. *IEEE Internet Computing*, 16 (3), 4–6. <https://doi.org/10.1109/MIC.2012.50>
- McCarthy, J., Minsky, M., Rochester, N., Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- Müller, C. J. (2016). Prometheanism: Technology, digital culture and human obsolescence. Rowman & Littlefield International.
- Nilsson, N. J. (1998). *Artificial Intelligence A New Synthesis*. Elsevier Science. <http://kcl.eblib.com/patron/FullRecord.aspx?p=1014253>
- Penrose, R. (1995). Beyond the Doubting of a Shadow A Reply to Commentaries on Shadows of the Mind. *PSYCHE: An Interdisciplinary Journal of Research On Consciousness*, 2.
- Putnam, H. (Ed.) (1975). The meaning of 'meaning.' In: *Mind, Language and Reality* (pp. 215–271). Cambridge University Press. <https://doi.org/10.1017/CBO9780511625251.014>
- Reuter, K. (2019). Dual character concepts. *Philosophy Compass*, 14 (1), e12557. <https://doi.org/10.1111/phc3.12557>
- Russell, S. J., Norvig, P. (2020). *Artificial intelligence: A modern approach* (Fourth edition). Pearson.

- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3 (3), 417—424. <https://doi.org/10.1017/S0140525X00005756>
- Simondon, G., Adkins, T. (2020). *Individuation in light of notions of form and information*. University of Minnesota Press.
- Singularity.FM (2013). *Marvin Minsky on AI: The Turing Test is a Joke!* (July 12, 2013). <https://www.singularityweblog.com/marvin-minsky/>
- Sloterdijk, P. (2009). Rules for the Human Zoo: A Response to the Letter on Humanism. *Environment and Planning D: Society and Space*, 27 (1), 12—28. <https://doi.org/10.1068/dst3>
- Sloterdijk, P. (2014). Anthropo-Technology. *New Perspectives Quarterly*, 31 (1), 12—19. <https://doi.org/10.1111/npqu.11419>
- Stiegler, B. (1998). *Technics and time*. Stanford University Press.
- Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX (236), 433—460. <https://doi.org/10.1093/mind/LIX.236.433>
- Waldenfels, B. (1991). Everyday life as a melting pot of rationality. [In Russian]. In: *Socio-Logos* (pp. 39—50). Progress. [=Вальденфельс 1991]
- Weiser, M. (1999). The Computer for the 21st Century. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3 (3), 3—11. <https://doi.org/10.1145/329124.329126>
- Yermolenko, A. (2019). Normative coherence of philosophical discourse. [In Ukrainian]. *Philosophska Dumka*, 5, 21—28. <https://doi.org/10.15407/fd2019.05.021> [=Єрмоленко 2019]

Одержано/Received 12.07.2021

*Mykhailo Bogachov*, Master of Philosophy,  
graduate student of the 3rd year of study, Department of Social Philosophy,  
H.S. Skovoroda Institute of Philosophy, National Academy of Sciences of Ukraine,  
4, Triokhsviatytska St., Kyiv, 01601  
[m.bogachov@gmail.com](mailto:m.bogachov@gmail.com)  
<https://orcid.org/0000-0002-3223-7227>

#### ARTIFICIAL INTELLIGENCE AS AN ANTHROPOTECHNOLOGY

Artificial intelligence is a computer system that thinks or acts like humans. Features of AI systems embody implicit beliefs concerning the human nature that AI developers have. “Strong” AI, which has the general cognitive abilities of an adult, has not yet been created, while “weak” AI is already part of the planetary computation infrastructure. Neural network AI mimics specific types of human behavior, generalizing data about the everyday lives of its users. This AI approach corresponds to the philosophical mainstream of the 20th century, when everyday life was seen as a source of the linguistic and the social pre-given that yields mutual understanding. This approach is also based on the traditional human-machine dichotomy and the corresponding idea that human nature is stable and independent of the technological condition. However, in the post-metaphysical age, when human interaction with technology is communicative rather than instrumental, data on everyday life cannot be an independent paragon of the human nature. AI systems do not only codify the descriptive features of human nature, but also discipline their users, as the digital environment in which everyday data can be collected is already organized by AI. Accordingly, in the digital environment, people are forced to reproduce new norms of behavior, codified by AI, which became one of the forms of human self-mastery, or anthropotechnology. The impact of AI is rarely noted, as the digital environment in which people interact with AI is not organized in a way that is clearly understandable. The anthropotechnological nature of AI is a side effect of the development of platforms, so AI developers rarely take responsibility for the norms embodied in the systems they create.

**Keywords:** *artificial intelligence, anthropotechnology, everyday life, human nature, digital environment, instrumental rationality.*