

Київський національний університет імені Тараса Шевченка
Філософський факультет
Кафедра теоретичної і практичної філософії

ТЕХНОЛОГІЧНІ МЕЖІ СУЧАСНОГО СПІЛЬНОСВІТУ
TECHNOLOGICAL LIMITS OF CONTEMPORARY COMMON WORLD

Кваліфікаційна робота за спеціальністю 033 філософія
на здобуття освітнього ступеня магістра філософії

Студент-виконавець:
Адаменко Богдан Володимирович

(підпис)
ОС - Магістр

Науковий керівник:
Лактіонова Анна Валеріївна
доктор філософських наук, професор

(підпис)

Допущено до захисту:
на засіданні кафедри історії філософії
протокол №____ від _____ 2022 р.

Зав. кафедри теоретичної і практичної філософії,
доктор філософських наук, професор
Шашкова Людмила Олексіївна _____

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. ФІЛОСОФСЬКИЙ АНАЛІЗ ТЕХНОЛОГІЙ, НА ПРИКЛАДІ ШТУЧНОГО ІНТЕЛЕКТУ	6
1.1 П'ять парадоксів в розробках штучного інтелекту.....	6
1.2 Етика штучного інтелекту.....	12
1.3 Переосмислення концепту моральної відповідальності.....	19
РОЗДІЛ 2. ТЕХНОЛОГІЇ ТА ВІДПОВІДАЛЬНІСТЬ.....	23
2.1. Відповідальність у контексті етики технологій.....	23
2.2. Можлива критика збільшення відповідальності у контексті створення нових технологій.....	29
2.3 Можливості створення ціннісно втілених технологій.....	31
2.4. Втілення цінностей в технологіях. Умови впровадження.....	33
РОЗДІЛ 3. ПОШУКИ МЕЖІ: ВІД ТЕХНОЛОГІЙ ДО ТЕХНОЛОГІЧНОГО ДОВКІЛЛЯ.....	38
3.1 Концептуалізація та аналіз теорій технологічного середовища.....	39
3.2 Перехід від технологічного довкілля до спільносвіту.....	43
ВИСНОВКИ.....	47
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ І ДЖЕРЕЛ.....	51

Актуальність роботи. Зміни у світі будуть тривати до поки ми не прокладемо межі цих змін. Вже зараз нам потрібно бути готовими уявити наслідки нашої технологічної діяльності, яка змінює межі нашого спільносвіту. Технології лиш продовжують посилювати свій вплив на всі сфери людського життя приватні та публічні. Однак, тут і проявляється ключова роль філософів – колективне інтелектуальне обговорення того, який сценарій майбутнього краще обрати. Широка вибірка нових технологій спонукає до невідкладних розмислів направлених на морально етичні запитання, стосовно технологічних артефактів, їх створення, використання, впровадження та впливу. Нашою помилкою було б очікувати, щоб зміни остаточно відбулися, а потім ретроспективно аналізувати наші хиби.

Метою роботи є аналіз і систематизація знань з тематики філософії технологій, а також філософська рефлексія стосовно ідеї та визначення технологій, їх створення, впровадження та впливів на людину. Для досягнення поставленої мети були окреслені наступні **завдання**:

- Провести філософський аналіз технологій, у розрізі розробки “штучного інтелекту”, як технологій в якій найбільше знаходить своє відображення людина.
- Розглянути приклади “парадоксальних” проблем, які несе в собі сфера досліджень “штучного інтелекту” та технологій загалом.
- Узагальнити знання щодо етичної проблематики, яка виникає в контексті розгляду ІІІ як автономної системи. Переглянути “етику даних”, “машинну етику”, а також концепцію відповідальності в межах створення технологій на основі ІІІ.
- Проаналізувати ідею відповідальності, як визначальну умову створення технологічних артефактів.
- Узагальнити критичні реакції філософів на взаємозалежність відповідальності та технологій.

- Узагальнити знання щодо моральної проблематики, яка виникає в контексті філософії технологій; дослідити можливість та умови втілення цінностей в технологіях.
- Концептуалізувати та проаналізувати теорію технологічного довкілля та окреслити умови меж сучасних технологій.

Об’єктом дослідження є філософський доробок дослідників теми філософії технологій.

Предметом дослідження є розгляд ідей відповідальності та межі в проблематизації викликів, які провокує сучасні технології.

Методи дослідження. У ході даного дослідження були використані такі загальнонаукові та філософські методи:

- Аналітичний метод, як один із головних методів даної роботи, що допоміг в історико-філософській реконструкції ідеї “штучного інтелекту”, а також основних підходів та критик даної сфери досліджень.
- Герменевтичний метод, що забезпечив глибше проникнення в сутність досліджуваної теми, її тлумачення та інтерпретації на основі першоджерел з філософського осмислення ІІІ.
- Компаративний метод, застосування якого було зумовлене зверненням до доробку більш ніж як одного дослідника, що обумовило необхідність порівняння та аналізу різних позицій та підходів.

Використання вказаної методології дало змогу комплексно проаналізувати та систематизувати отримані в ході роботи результати задля чіткого і глибинного розкриття теми дипломної роботи.

Ступінь опрацювання. Наголошуючи на актуальності даного дослідження, слід зазначити, що в українському філософському середовищі ідеї “технологічної межі”, “технологічних довкіль” та “втілених цінностей” є новими. Починаючи власне дослідження відповідальності та філософії технологій, наш послідовний історико-ідейний аналіз був заснований на лініях міркувань таких сучасних філософів технологій як: Ендрю Фінберг

(Andrew Feenberg), Мерел Нурман (Merel Noorman), Вінсент Мюллер (Vincent C. Müller), Карла Мітчема (Carl Mitcham), Курта Баєрца (Kurt Bayertz), Пітера-Пол Вербека (Peter-Paul Verbeek) ідеї якого були переосмислені в контексті ідей Клауса М. Маєр-Абіха (Klaus Meyer-Abich).

Сформульовані завдання і загальна логіка дослідження зумовили **структуру даної роботи**, яка складається зі вступу, трьох розділів, висновків та списку використаної літератури та джерел (54 найменування).

Загальний обсяг дипломної роботи становить 49 сторінок (без списку використаної літератури та джерел)

РОЗДІЛ 1. ФІЛОСОФСЬКИЙ АНАЛІЗ ТЕХНОЛОГІЙ, НА ПРИКЛАДІ ШТУЧНОГО ІНТЕЛЕКТУ

Технологія навколо якої ми будемо будувати актуальний філософський аналіз – це штучний інтелект (ШІ). Для цього пропонуємо робоче визначення. “Штучний інтелект” — це технологія, заснована на роботі комп’ютерних програм, що обмежено імітують деякі когнітивні здатності людини без їх розуміння.

Людина у своїй творчості завжди намагалася наслідувати Бога. ШІ — це технологічний інструмент, який не лише полегшує життя, а й допомагає порозуміти над людиною, над тією роллю, яку вона відіграє в створенні ШІ. Саме аналізуючи розвиток ідеї штучного інтелекту проглядаються зміни в уявленнях про людину, її мислення та свідомість. Якщо слідувати тому, що людина є скупим відображенням Бога, то ШІ – це технологія, в якій найкраще проглядається сама людина.

1.1 П’ять парадоксів в розробках технологій штучного інтелекту

Визначення ШІ як технології відкриває для нас поле для глибокої концептуальної проблематизації. Для цього ми залучаємо напрацювання з критичної теорії технологій (criticism of technology), сформульовані та узагальнені американським філософом Ендрю Фінбергом в його роботі “Десять парадоксів техніки” (Ten Paradoxes of Technology) [13]. Імплементуючи наші попередні напрацювання стосовно досліджень ШІ в теорію Е. Фінберга, ми спробуємо окреслити п’ять парадоксів в розробках штучного інтелекту, узагальнивши та доповнивши оригінальні ідеї автора.

1. Парадокс частини й цілого. Е. Фінберг описує даний парадокс в розумінні того, що складні цілісності на позір походять від своїх частин, але в реальності частини віднаходять своє походження у цілісності, до якої вони належать. Порівнявши дану думку з ідеями описаними в працях різних авторів [про це див. 16, 21, 23] ми бачимо, що дослідники ШІ зіткнулися з даним парадоксом в межах двох проблем.

Перша проблема проявила себе в рамках обчислювальної теорії свідомості, згідно з якою складні думки набувають свого значення із комбінацій простих компонентів. Слідуючи за положеннями даної теорії, ці компоненти формують цілісну думку, з'єднуючись певним чином разом. Проте розглядаючи лише частини, дослідження в межах Символьного ІІІ, лишали позаду цілісність, яка оформлює та надає змісту таким компонентам. Розглядаючи лише синтаксис, дослідники ІІІ нехтували семантикою. До прикладу Дж. Хаугеланд писав: “Якщо ви потурбуєтеся про синтаксис, семантика сама про себе потурбується” [32]. Проте ігнорування походження синтаксису з семантики призвело до проблем в межах досліджень ІІІ, які яскраво виражені в критиці Джона Серля [11].

Друга проблема в розробках ІІІ, яку ми вбачаємо в межах даного парадоксу, пов'язана з теорією конекціонізму. Дана теорія, яка розвинулася в межах когнітивних наук та тісно пов'язана зі сферою розробки ІІІ, розглядає та моделює такі вищі когнітивні здатності людини, як , наприклад, “бачення”, “слух” чи “навчання” через ідею того, що вони є результатом взаємодії нейронів в мозку людини, і що їх можна точно відтворити, створивши відповідну штучну (комп'ютерну) нейронну мережу [47]. Задля розкриття глибинних передумов даної проблеми слід звернутися до концептуальних основ нейронаук, які й стимулювали розвиток та заклали зі свого боку теоретичну базу для побудови нейронних мереж. Для цього варто згадати критичну працю “Філософські основи нейронаук” (Philosophical foundations of neuroscience) авторства філософа Пітера Гакера (Peter Hacker) та нейронауковця Макса Беннетта (Max Bennett) [31].

Філософські дослідження концептуальних основ нейронауки, яке проводять автори, направлене на розкриття та прояснення концептуальних істин, які являються умовою змістовності описання когнітивних нейронаукових відкриттів та теорій. В центрі їх дослідження знаходяться концепції “мереологічної помилки” (mereological fallacy) та концептуальної плутанини (conceptual confusion) в нейронаукових дослідженнях.

Випадки мереологічної помилки та випадки концептуальної плутанини мають спільну основу: беззмістовне приписування психологічних атрибутів мозку. Причиною цього, як пишуть М. Беннетт і П. Гакер, є засилля в когнітивних нейронауках Картезіанської концепції дуалізму. Перші два покоління нейронауковців двадцятого століття були загальновідомими Картезіанськими дуалістами, серед них, до прикладу, Чарлз Шеррінгтон (Charles Sherrington), Джон Екклс (John Eccles) та Вайлдер Пенфілд (Wilder Penfield). Третє покоління залишило структуру Декарта, відкинувши субстанційний дуалізм, проте зберігши структурний. “Зараз нейронауковці приписують мозку такий самий набір ментальних предикатів, що й Декарт розуму, зберігши Декартове уявлення про зв'язок між думкою і дією, досвідом та його об'єктами” [31, 130-131].

Сприйняття ідей Декарта легітимізувало приписування психічних атрибутів мозку. Прикладами такого приписування можуть слугувати ідеї Френсіса Кріка (Francis Crick) про те, що в мозку відбувається багаторівнева інтерпретація візуальних сцен, Джеральда Едельмана (Gerald Edelman), що мозок категоризує, розрізняє та об'єднує все у глобальну карту, Коліна Блекмора (Colin Blakemore), в теорії якого знання знаходиться в нейронах, які оцінюють ймовірність зовнішніх подій, а мозок своєю чергою отримує ці знання шляхом індуктивного розмірковування. Також варто згадати теорію Джона Захарі Янга (John Zachary Young), згідно з якою мозок ставить питання, шукає відповіді та будує гіпотези. Гіпотези ж формуються через сигнали зі сітківки ока [31, 16-17].

Спираючись на ці та інші приклади, автори заявляють, що в подібному приписуванні психологічних атрибутів мозку немає жодного сенсу, бо, по-перше, в нас немає концептуальної основи для подібного роду приписування, й, по-друге, використовуючи терміни психології, нейронауковці змінюють їх вихідні значення, що й веде до концептуальної плутанини. “Ми здатні розпізнати, коли одна людина ставить запитання й коли інша на нього відповідає. Проте чи є в нас концепція того, як це для

мозку запитувати чи відповідати на запитання?” Як наслідок, в розробках нейронауковців може виникнути мереологічна помилка. “Мереологія — це логіка стосунків між частиною і цілим. Помилку нейронауковців, що полягає в приписуванні атрибутів певним частинам живої істоти, які логічно можуть бути застосовані тільки до живої істоти як цілого, ми називаємо “мереологічною помилкою” в нейронауках.” [31, 23]. Форма даної помилки була вказана ще Арістотелем в його вченні про душу: “Говорити, що душа гнівається, це те ж саме, що сказати, що душа займається ткацтвом чи будує будинок. Бо ж краще, напевно, не говорити, що душа співчуває або ж вчиться, або ж розмірковує, а говорити, що людина це робить душею” [1]. М. Беннетт і П. Гакер називають це “принципом Арістотеля”. Тож базою для розробки їхньої концепції слугує “принцип Арістотеля”, доповнений поглядами Л. Вітгенштайна стосовно того, що “тільки про людину і про те, що схоже (поводить себе як) на людину, можна сказати: у неї є відчуття; вона бачить; сліпа; чує; глуха; є свідомою чи несвідомою” [40, para. 281].

Узагальнюючи, ми бачимо, що слідом за нейронауковцями, конекціоністи хибують у схожий спосіб. Розгляд нейронних активностей в мозку не здатен прояснити діяльність цілісної людини як такої. Так само як і спеціалісти з розробки нейронних мереж, які вдають, що звертаються до концепцій роботи людського мозку, насправді лиш редукують когнітивні здатності людини до роботи нейронних мереж. Оманливі імплікації у сферах розробки ШІ лиш посилюють парадокс частини та цілого. З першого парадоксу більшою мірою висновуються всі наступні.

2. Парадокс очевидного. Даний парадокс походить від попереднього. Його загальне формулювання за Е. Фінбергом: найочевидніше є найбільш прихованим. Наприклад, після тесту Тюрінга здавалося очевидним, що коли ми створимо комп'ютерну програму здатну проходити ТТ, то ми створимо інтелектуальну машину. Але це “очевидне” породило ряд дискусій, доповнень та інтерпретацій оригінального ТТ, тож очевидне постійно

вислизає з нашого поля зору і потребує розгляду в ширшому контексті проблематики.

3. Парадокс походження. Часом здається, що технології не пов'язані зі своїм минулим. В буденному житті вони постають для нас самодостатніми й зазвичай ми не маємо жодного уявлення про те, як вони з'явилися та розвивалися. Як пише Е. Фінберг: “Адекватне пояснення будь-якого пристрою полягає у простежуванні каузальних зв'язків між його частинами”. І таке пояснення є згідно з автором хибним і, однозначно, недостатнім. Парадокс походження формулюється наступним чином: за всім раціональним стоїть забута історія. Однак, теперішнє засліплює бачення минулого, якщо не докладати критичних зусиль для прояснень. Присутність минулого у теперішньому видається нам непримітною. Освітлені вказівники “Вихід” над дверима сприймаються нами як даність. Їх появу пояснює забута трагічна історія, яка сталася в 1903 році в Чиказькому театрі “Ірокез”, коли 600 людей загинули, намагаючись знайти вихід під час пожежі. Зазвичай ми не замислюємося над походженням таких буденних речей. В нашому розумінні вони перебувають там, де їм і місце. Та це лиш часткове пояснення, яке для філософів є недостатнім. “Штучний інтелект” теж для більшості постає самодостатнім. Проте незнання історико-концептуального розвитку ШІ (й інших технологій) стає причиною хибних уявлень та теорій, що стане очевидним в наступних підрозділах.

4. Парадокс рами. Даний парадокс походить від попереднього. Його формулювання є наступним: не ефективність пояснює успіх, а успіх пояснює ефективність. Хоча, на перший погляд, це здається контрінтуїтивним, бо технології ми розглядаємо через критерій успішності їх застосування. Ефективність технологій є мірилом їх цінності. Е. Фінберг обґрунтовує даний парадокс звертаючись до історії технологій. Більшість технологій, які ми воліємо називати ефективними, є наслідком початкового вибору, який надав успішній технології привілейованого статусу стосовно її альтернатив. Цей вибір здійснюють соціальні актори, що мають достатню владу для здійснення

даного вибору, при цьому до нього застосовуються різні критерії: економічні, соціальні, політичні, іноді технічні. Парадокс рами звертає увагу на те, що ефективність не пояснює присутність певних технологій в нашому житті. Це може нам прояснити лише вивчення історії технологій.

5. Парадокс дії. Е. Фінберг виводить його метафорично, трактуючи третій закон Ньютона, згідно з яким будь-яка дія породжує рівнозначну протидію. І як зазначає автор: “Кожна наша дія повертається до нас у формі зворотного зв’язку з Іншим”. Парадокс постулює те, що діючи, ми стаємо об’єктами дії, і те, що люди здатні впливати лише на ту систему, частиною якої вони є самі. Тобто ми перебуваємо під дією технологій, які самі й створили. Технології приховують від нас три *взаємозумовленості технологічної дії*: побічні ефекти, зміни у значенні нашого світу та зміни у нашій ідентичності.

Повертаючись до ІІІ, в контексті такого впливу актуальним постає питання Джуліана Фрідленда (Julian Fridlend), яке він формулює наступним чином: “Як нам розробляючи (та використовуючи) розумні машини уберегти себе від уподібнення цим машинам?” Тож звернемося до міркувань Д. Фрідленда стосовно того, як технології на основі ІІІ можуть впливати на людину [30].

Застосунки, які встановлені на наших смартфонах чи комп’ютерах, все частіше розробляються на основі технологій машинного навчання. По ходу того, як ми делегуємо багато задач ІІІ, ми стаємо більш пасивними, менш рефлексивними та менш відповідальними за результат. Найпростіші приклади. Нам потрібно кудись піти, ми відкриваємо карти, і програма буде нам маршрут. Один маршрут. Можливо, цей маршрут і є оптимальним, та це вже не ми вибираємо шлях, яким іти. Так, це усуває фактор стресу, нам не потрібно комунікувати з іншими людьми та запитувати дорогу. Начебто нічого страшного, та це призводить до декваліфікації та погіршення орієнтації в просторі. Проте, навчаючись орієнтуватися в новому некомфортному середовищі, налагоджувати в ньому взаємодії, ми отримуємо

контроль, можливо й побічний, та все ж над власною свідомістю. Другий приклад — це рекомендаційні системи підбору фільмів. ІІІ підбирає нам релевантні фільми, засновані на наших уподобаннях. Ми не витрачаємо часу на таку рутинну справу, але це погіршує наше “інтелектуальне тертя” та відповідальність за вибір. Додамо, що всі застосунки розробляються для того, щоб утримувати нашу увагу та направляються на те, щоб задіювати наш автономний розум, а не рефлексивний, як це розрізняє Деніел Канеман (Daniel Kahneman) [31]. Людина технологій, як зазначає Д. Фрідленд, стає менш критичною, легше піддається впливам, та найгірше – “оштучнюючи” себе, нівелює поняття відповідальності [30]. Тож хибно вважати, що ми можемо впливати на світ без наслідків для себе. Варто згадати й про приклад, який наводить Е. Фінберг для підсилення тези, що технології здатні змінювати значення нашого світу. Залізничі, автомобільні та літні радикально змінили наше досвідне переживання відстані. Просторові координати нашого життя, які ми маємо на увазі під словами “близько” та “далеко”, абсолютно відмінні від тих, які були до появи даних винаходів. Більш детально дану проблематику ми розглянемо в РОЗДІЛІ 3.

1.2 Етика штучного інтелекту

П’ять парадоксів технологій, які ми сформулювали в попередньому підрозділі на основі ідей Е. Фінберг, відкривають поле для більш детального розгляду його окремих пунктів. Штучний інтелект як технологія матиме значний вплив на розвиток людства в майбутньому. Хибно говорити про комп’ютерну революцію, яка більшою мірою застосовує ІІІ як інструмент, так, неначе вона вже відбулася. Людство знаходиться лише на початку осмислення та застосування ІІІ. Останнє своєю чергою підіймає ряд фундаментальних питань про те, як нам використовувати дану технологію, які ризики можуть виникнути і як нам їх попередити. Такі технології як ядерна зброя чи автомобілі спричинили етичні та політичні дискусії, як правило лише після того, як завдали шкоди. Проте, приступаючи до розгляду

етичних питань та проблем у сфері ШІ, ми знаходимося в ситуації певної невизначеності. В нас не було чітких прецедентів для того, щоб вибудувати теоретичний підхід до етики ШІ, тому розгляд даної проблематики характеризується можливістю уникнення проблем у майбутньому, хоча вплив даної технології відчутний вже зараз. Для прояснення етичної проблематики у сфері досліджень та застосування технологій ШІ, ми залучаємо напрацювання з етики ШІ описані Вінсентом Мюллером в роботі “Етика штучного інтелекту та робототехніки” (“Ethics of Artificial Intelligence and Robotics”) [37].

Сучасна розробка систем на основі технології ШІ у своїй основі послуговується конекціоністськими моделями нейронних мереж. Дослідники полишили спроби побудови комплексно завершеного ШІ та перейшли до створення ШІ для виконання чітко визначених завдань. З цією метою створюються нейронні мережі під певні цілі. Розробка такого ШІ пов’язана з технологією машинного навчання, яке потребує збору та обробки даних. Тож в даному контексті варто розглянути ряд важливих проблемних моментів.

Дані та ШІ. Як вже зазначалося в попередньому підрозділі, застосунки для смартфонів та комп’ютерів розробляються на основі методів машинного навчання, які засновані на роботі нейронних мереж, які здатні видобувати патерни із заданих наборів даних. З використанням таких методів “навчання” фіксується закономірність в даних і вони позначаються таким чином, що це здається корисним для рішень, які приймає система. Використовуючи додатки, шукаючи інформацію в Інтернеті та навіть просто пересуваючись вулицями, ми мимовільно продукуємо дані для роботи подібних систем. Одна з етичних проблем ШІ полягає в тому, що обробка даної інформації відбувається з метою маніпулювання нашою поведінкою. Найбільше зараз це відчутно в онлайн-рекламі. Збираючи інформацію про нашу “поведінку” в Інтернеті та на її основі “навчаючи” нейронні мережі, маркетологи створюють “dark patterns”. Це свого роду лабіринти, ціллю яких є привести нас до покупки певного товару або ж заплутати нас так, щоб ми більше часу

використовували застосунки. Такого роду маніпуляції є бізнес-моделлю в ігровому бізнесі, але зараз вони активно використовуються в дизайні цифрових технологій. Окремо постає проблематика збору та використання подібних даних, що окреслюється темою “етики даних” (data ethics).

Методи машинного навчання ШІ використовують для свого “навчання” велику кількість даних. Це означає, що часто існує компроміс між конфіденційністю та правами на використання даних й кінцевою технічною якістю продукту. Останнє означає, що з однієї сторони якість персоналізованої реклами, яка показується нам, або ж точність побудови маршруту картами нерозривно пов’язана з тими даними, які збираються програмами для покращення подібного роду роботи. Це впливає на послідовну оцінку практики порушення конфіденційності. Дана тема тісно пов’язана з політикою і правом, що, на жаль, не є темою даного дослідження. Проте коротко зазначимо, що в даній проблематиці є свої злети та падіння: громадянські свободи та захист прав особистості знаходяться під впливом лобістів, спецслужб та інших державних та недержавних організацій. Державні та бізнес-суб’єкти збільшили свої здатності втручатися в приватне життя і маніпулювати людьми з допомогою технології штучного інтелекту. Й, на жаль, продовжуватимуть це робити для просування своїх власних конкретних інтересів, якщо це не буде стримуватися політикою спрямованою на інтереси суспільства в цілому. Зазначимо, що в Європейському союзі лиш нещодавно запровадили “Загальний регламент про захист даних”, що посилив захист конфіденційності.

Одними з центральних питань “етики даних” чи “етики великих даних” (big data ethics) постає проблема “непрозорості” та “упереджень”. Як вже зазначалося вище, більшість систем ШІ використовують методи машинного навчання, в основі яких лежать нейронні мережі, які “видобувають” патерни із заданих наборів даних. З використанням таких методів “навчання” фіксуються закономірності в даних і вони маркуються таким чином, що це здається корисним для рішень, які приймає система, однак програміст

насправді не завжди знає, які саме закономірності в даних використовувала система. “Навчаючись”, система розвивається на основі нових даних, які вона видобуває, або ж через внутрішній зворотний зв'язок (система встановлює власний ступінь релевантності різних закономірностей, які пасують або не пасують для її подальшої роботи). Тобто така програма постійно змінює модель власної роботи. Це означає, що результати “непрозорі” для програмістів і для людей, які використовують застосунки на основі технології ШІ. Крім того, якість роботи програми залежить від якості даних, з якими вона “працює”. Якщо програма на вході мала “упереджені” дані, наприклад дані поліції про колір шкіри підозрюваного, то результати роботи програми будуть значною мірою упереджені. Деякі дослідники стверджують, що етичні проблеми з ШІ є результатом технічних “ярликів”, які використовує ШІ. Упередженість, як правило, проявляється, коли людина формує неправильні судження, тому що на неї впливають характеристики, які насправді не мають відношення до питання (чи проблеми), яке розглядається. Одна із таких форм упередження — це когнітивна особливість людини, яка часто не проявляється явно. Окрім соціального феномену наукового упередження, когнітивна система людини, як правило, піддається різним видам “когнітивних упереджень”, наприклад “підтверджувальне упередження” (confirmation bias). Люди схильні інтерпретувати інформацію так, щоб вона підтверджувала їх власні переконання в те, у що людина вже вірить. Виникає питання, чи сама система ШІ може мати такі когнітивні упередження. На нашу думку, не може. Це напряму пов'язане з творцем такої системи й з його власними переконаннями. Детальніше про те, як технології втілюють цінності та упередження ми розглянемо у підрозділі 2.4. Втілення цінностей в технологіях. Умови впровадження.

Варто зважати, що будь-який набір даних є “неупередженим” лише для одного роду проблем, тому просте створення набору даних пов'язане з небезпекою, що воно може бути використаним для іншого типу проблем, в якому може проявитися упередження саме для такого типу. Машинне

навчання на основі таких даних не тільки не зможе розпізнати упередження, а й може систематизувати й автоматизувати його в “історичне упередження”. Проблема з такими системами знаходиться не лише в “упереджених даних”, а й в тому, що люди починають надмірно довіряти подібним системам. Технічні зусилля направлені на уникнення систематичних упереджень в системах штучного інтелекту перебувають на ранній стадії розгляду.

III як автономна система. Існує декілька визначень “автономії” в контексті обговорення автономних систем. Сильне визначення використовують у філософських дебатах, в яких автономію визначають основою індивідуальності та відповідальності [31]. В даному контексті відповідальність означає автономію, але не навпаки. Тому ми можемо говорити про системи, які мають певні ступені автономії, не піднімаючи питання відповідальності. Слабке, або ж більш технічне визначення поняття “автономія” означає, що система в деякому ступені автономна у відношенні людського контролю [36]. Тут віднаходиться паралель з проблемами “упередженості” та “непрозорості” в ШІ, оскільки автономія стосується питання хто контролює і хто відповідальний. Питання автономії підіймає питання не лише технічних аспектів побудови автономних систем, а й питання права та безпеки.

Розглянемо приклад автономної системи на базі ШІ, яка наразі знаходиться серед найбільш обговорюваних, а саме автономні (безпілотні) транспортні засоби. Як очікується, автономні транспортні засоби можуть великою мірою зменшити шкоду, яку наразі завдають водії. Мова тут йдеться не лише про забруднення навколишнього середовища та логістику, найважливішим в даному контексті являється те, що приблизно один мільйон людей гине щорічно на дорогах. Постає питання, як повинен себе “поводити” автономний автомобіль, і як слід розглядати питання ризиків та відповідальності при побудові таких систем. В даному контексті часто розглядають відому “Проблему вагонетки”, при дослідженні теми відповідальності ШІ. Однак, “Проблема вагонетки” не передбачає опису

реальних етичних проблем і їх вирішення з допомогою “правильного” вибору. Дана проблема використовується як теоретичний інструмент для дослідження етичних інтуїцій та теорій. Проте, як зазначає В. Мюллер, насправді перед водіями на дорогах не постає такої дилеми [37]. Більш розповсюдженими етичними проблемами водіння є перевищення швидкості, ризикований обгін, недотримання дистанції та інші, що постають в межах класичних проблем, які пов’язані з переслідуванням власних інтересів. Все це підпадає під дію правових норм. Таким чином, програмування автомобілів згідно з “правилами”, а не для “досягнення максимальної корисності” зводиться до стандартної проблеми програмування етичних машин. Проблемним тут постає не лише вибір програміста, якою етикою послуговуватися (кантіанською чи утилітаристською), а й тема розробки “машинної етики” (Machine ethics) та проблема визначення “штучного морального агента” (Artificial moral agents).

Машинна етика — це етика для машин, для “етичних машин”. Для машин як суб’єктів, а не як таких, що використовуються людиною в якості об’єктів. На сучасному етапі досліджень даної теми не зрозуміло, чи вона має охоплювати всю етику ШІ, чи бути її частиною. Експлікуючи подібні міркування, сумнівним постає висновок, що коли машини діють етично вагомим способом, то для цього потрібно розробити машинну етику. Деякі дослідники, зокрема Стів Торранс (Steve Torrance), використовують ширше визначення машинної етики, яка в їхньому розумінні пов’язана із забезпеченням того, щоб “поведінка” машин у відношенні до користувачів, а також, можливо, щодо інших машин була етично прийнятною [48, 115]. Дане визначення також може включати питання безпечності продукту. Інші автори хоча й виглядають доволі амбіційними, однак послуговуються вузьким визначенням даного поняття. Зокрема Віргінія Дігнум (Virginia Dignum) пише про те, що ШІ у своїх міркуваннях має бути здатним враховувати соціальні цінності, моральні та етичні міркування, а також має зважувати відповідні ціннісні пріоритети, які мають різні зацікавлені сторони в різноманітних

мультикультурних контекстах [25]. Також В. Дігнум наголошує на прозорості систем ШІ, а також на важливості того, щоб система самостійно пояснювала власні міркування. В наявних дискусіях, присвячених машинній етиці, формується припущення, що машини здатні, в певному сенсі, бути етичними агентами, такими, що є відповідальними за власні дії, тобто визначатися як “автономні моральні агенти” (autonomous moral agents). Якщо співвіднести машинну етику та теорію морального агента, то в певному субстанційному сенсі таких агентів можна назвати “штучними моральними агентами”, що мають права та обов’язки. Проте ми з вами маємо розуміти та пам’ятати, що саме програмісти чи користувачі систем ШІ приймають моральні рішення.

Прихильники теорії “автономних моральних агентів” стверджують, що машинна етика виходить за рамки того, як інженери усвідомлюють цінності, які вони закладають при створенні продуктів на основі ШІ. Джеймс Мур (James Moor) для того, щоб додатково визначити, що означає для комп’ютера приймати етичні рішення, проводить розрізнення трьох типів етичних агентів: неявні етичні агенти (implicit ethical agents), явні етичні агенти (explicit ethical agents) та повні етичні агенти (full ethical agents) [35].

Перший тип агентів — це комп’ютери, в основу роботи яких закладена етика розробника. Ці агенти запрограмовані (сконструйовані) так, щоб дотримуватися норм і цінностей контексту, в якому вони розроблені чи будуть використовуватися. Явні етичні агенти — це комп’ютерні програми, як визначає Д. Мур, що можуть на основі етичної моделі визначити, що було б правильним вчинити в тій чи іншій ситуації, враховуючи певні дані на вході. Етична модель може бути заснована на традиційних етичних теоріях, таких як кантіанська чи утилітаристська, в залежності від уподобань творців програми. Явні етичні агенти здатні “приймати етичні рішення” від імені своїх творців. Визначення такого роду агентів більшою мірою подібне до визначення “автономних моральних агентів”. Повні етичні агенти можуть формувати етичні судження та обґрунтовувати їх так само як і люди. Д. Мур зазначає, що хоча в цей момент розвитку технологій немає такого ШІ, який би

можна було назвати повністю етичним агентом, однак це постає питанням, чи існує взагалі можливість його створення в майбутньому.

1.3 Переосмислення концепту моральної відповідальності

Експлікуючи попередні міркування, наявним постає завдання перегляду визначення відповідальності. Традиційні філософські дискусії щодо моральної відповідальності були зосереджені на людському компоненті моральної дії. Моральна відповідальність стосується дій людини, її намірів та наслідків. Узагальнюючи, людина або група людей несуть моральну відповідальність, коли їх вільна дія має морально значливий результат. Приписування моральної відповідальності встановлює зв'язок між людиною (групою людей) і тим, що може впливати на її дію. Людину (групу людей) дії якої спричинили певні наслідки, називають агентом. Людину (групу людей), які знаходяться під впливом дій агентів, називають пацієнтами [42]. Встановлення зв'язків між агентом та пацієнтом у моральному відношенні може здійснюватися ретроспективно і перспективно. Тобто, приписування відповідальності залучає пояснення того, хто був винен, і кого потрібно покарати. У розгляді перспективи, ми визначаємо обов'язки та дії людини, яких вона має дотримуватися чи виконувати у майбутньому. Проте нечітким залишається визначення того, при яких обставинах варто приписувати моральну відповідальність. Дана ситуація спричинила дискусії не лише у відношенні моральної відповідальності щодо діяльності людей. Продукти, в основу роботи яких закладена технологія ШІ, викликали ряд нових питань. Вище були розглянуті питання “автономії” та “непрозорості” в роботі ШІ, коли проблематичним постає визначення того, хто відповідальний — творець чи програма. Алгоритми, які закладаються в нейронні мережі, через внутрішню “непрозору” для самого програміста роботу системи постійно змінюються самою системою, тож програмісти стають нездатними обґрунтувати результати, яких “самотійно” досягла програма. Це спонукає дослідників говорити про певний ступінь автономії таких систем, аж до

приписування їм відповідальності. В даному контексті варто розглянути ідеї Дональда Готтербарна (Donald Gotterbarn), який пропонує переосмислити те, як приписується моральна відповідальність [31], а також їх рецепцію в роботі Мерел Нурман [42].

Беручи до розгляду спеціалістів з обчислювальної техніки, Д. Готтербарн прослідковує можливості обходу питання їх відповідальності знайшовши тих, хто може її понести за них. Д. Готтербарн вбачає дану можливість у двох поширених хибних уявленнях щодо відповідальності. Перша хиба стосується того, що обчислення розглядають як етично нейтральну практику. Слідуючи думці Д. Готтербарна, хибною є думка, що технологічні артефакти й методи їх створення є етично нейтральними. Варто враховувати широкий контекст, в якому дані технології використовуються чи можуть бути використаними. Якщо інженер буде знаходитися лише в рамках створення нової технології, то таке вузьке фокусування може мати фатальні наслідки. Д. Готтербарн наводить як приклад програміста, завданням якого було написати програму для Рентген апарату, що мала підіймати та опускати даний пристрій. Зосередившись лише на написанні програми, і не врахувавши обставини й можливі наслідки її роботи, діяльність програміста спричинила смерть людини. Д. Готтербарн наголошує, що комп'ютерні спеціалісти несуть моральну відповідальність за врахування непередбачуваних обставин, навіть якщо це не є основною частиною їх роботи. Розробка та використання технологічних артефактів — це моральна діяльність, і вибір одного інженерного рішення, а не іншого, має реальні наслідки [31].

Друга хиба пов'язана з тим, що питання відповідальності виникає лиш тоді, коли намагаються визначити винного, якщо щось пішло не так. Комп'ютерні спеціалісти умовно прийняли модель відповідальності, яка стосується протиправних дій. Така модель дає змогу ухилитися від відповідальності. Відстань між розробником та наслідками використання технології, яку він створив, може слугувати аргументом на користь того, що

немає безпосереднього причинно-наслідкового зв'язку. Також розробники можуть апелювати, що вони були частиною команди, і їх вклад є незначним, або ж їх можливості були обмеженими та диктувалися компанією. Згідно з Д. Готтербарном, така модель спонукає спеціалістів дистанціюватися від відповідальності та звинувачення. Дані хиби стосуються особливого уявлення про відповідальність, в якому акцент робиться на уникненні звинувачень та звільненні від відповідальності. Експлікуючи, Д. Готтербарн позначає це як “негативну відповідальність” і відділяє її від “позитивної відповідальності”. Остання своєю чергою характеризується необхідністю враховувати наслідки своєї діяльності стосовно інших (людей). В контексті позитивної відповідальності, професіоналізм інженерів полягає у тому, скільки зусиль вони докладають задля мінімізації небажаних наслідків. Основна увага тут приділяється не уникненню відповідальності за безвідповідальну поведінку, а тому, що потрібно зробити на благо майбутнього. Д. Готтербарн спонукає комп'ютерних спеціалістів прийняти позитивну концепцію відповідальності, бо саме в ній підкреслюються обов'язок практиків сфери технологій враховувати наслідки їх діяльності та зведення до мінімуму можливої шкоди. Творці технологій несуть моральну відповідальність за те, щоб створений ними продукт не завдавав шкоди.

Попри те, що програміст, створивши нейронну мережу, може бути не здатним пояснити, як програма прийшла до певних висновків, хибно розглядати такий технологічний артефакт як “автономного морального агента”. Причиною “непрозорості” роботи систем ШІ є людина. Відповідальність цілковито і повністю лежить на творцях систем ШІ. В даному контексті важливо згадати моральний імператив для технологічної діяльності, сформульований філософом Анатолієм Єрмоленком: “Чини так, щоб максима твоєї технічної дії, проекту, винаходу, технології та виробництва, які впроваджуватимуться у життя, а також наслідки та побічні наслідки, що здогадно впливатимуть з їх застосування, могли б бути без примусу прийняті всіма, кого це стосується, в процесі аргументації, в якому

не залишався б поза увагою жодний раціональний аргумент!” [4, 326] Також А. Єрмоленко, погоджуючись з іншими класичними авторами комунікативної практичної філософії, етики відповідальності [про це див. 5], наголошує на принципі універсалізації (“для всіх”, в тому числі й для майбутніх поколінь), щодо технологій. Однак аргументація залежить від контексту, в якому її наводять. Наявними постають питання обізнаності користувачів та творців не лише стосовно внутрішньої технічної роботи програм ШІ, а й ідейних передумов створення таких артефактів.

Штучний інтелект відображає саме ті цінності та цілі, які в нього закладають творці. Він буде діяти й вже діє згідно з тими цілями, для яких був створений. Прихильники концепції “автономних моральних агентів” значною мірою хибують, коли в “непрозорості” системи вбачають її самостійність. Проте саме “непрозорість” роботи деяких нейронних мереж має спонукати нас до розмислів, для яких цілей їх застосовувати, якщо потім не буде змоги пояснити причинно-наслідкові зв’язки. Етика та відповідальність тут відіграє важливу роль не для систем ШІ, а для тих, хто їх створює, а також керується результатами роботи подібних програм для прийняття рішень. Людиноцентричність має знаходитися в центрі таких побудов, тому варто уникати будь-якої невизначеності при поширенні та впровадженні подібних технологій. Делегуючи деяку важливу мисленнєву роботу ШІ, варто зважати на те, що небезпека полягає не в тому, що ШІ стає “розумнішим” і його предикативні здатності в деяких питаннях перевершують можливості людини, а в тому, що люди очікують від ШІ більшого, аніж те, на що він здатний, і можуть покладатися на програму в ситуаціях, в яких цього краще уникнути. Завжди варто пам’ятати про “штучність” таких програм, що вказує на їх інструментальність. Штучний інтелект — це технологічний інструмент. Як для будь-якого інструменту, мають бути розроблені правила використання, а творцями має бути сприйнята “позитивна відповідальність” стосовно створення подібних технологій.

РОЗДІЛ 2. ТЕХНОЛОГІЇ ТА ВІДПОВІДАЛЬНІСТЬ

Між розвитком технологій та еволюцією ідеї відповідальності вочевидь існує взаємозв'язок. Це не нова думка, а одним із першим хто описав цей взаємозв'язок був американський філософ Карл Мітчем. За його словами, цей взаємозв'язок розвивався ще від часів початку промислової революції та розвитку ліберальних демократій [34]. “Розширені стосунки” між технологіями та відповідальністю, за словами німецького філософа Курта Баєрца є константою в людській історії. У своїй роботі він пише, що історію людства варто розглядати, як процес постійного збільшення відповідальності [19]. Технології розширили можливості влади людини над природою та над іншими людьми. Дане штучне розширення та дистанціювання, створило ідейну пустоту, яку і заповнила відповідальність. Коли технології розширюють можливості людини, вони дають їй змогу робити те, що раніше здавалося не реальним. Цей уявний важіль контролю, потребував точки опори – нею і стала відповідальність. Мерель Нурман, щоб проілюструвати цей приклад, звертається до цитати із масової культури: “З великою силою, приходить і велика відповідальність”. Впровадження різноманітних технологій впливає на те, як ми наділяємо щось, або ж когось, відповідальністю.

Даний розділ буде присвячений розгляду ідеї відповідальності у контексті етики технологій, а також критиці меж покладання відповідальності на інженерів, в залежності від зростання кількості нових технологій.

2.1. Відповідальність у контексті етики технологій

Термін “відповідальність” й часто зустрічається в літературі присвяченій технологіям та інноваціям, однак його використання в обговореннях має загальний характер та песимістичні прогнози, щодо того чи покладатимуть інженери на себе більше відповідальності, при збільшенні кількості нових технологій. Наприклад у своїй роботі Ханс Йонас (Hans Jonas), говорить що технології вимагають етики, в якій відповідальність має

стати імперативом та центральним поняттям, однак одразу додає негативної конотації, акцентуючи, що це пов'язано з тим, що вперше в історії людство може знищити себе самостійно [5]. Звісно ж ця відповідальність стосується в першу чергу інженерів та розробників, однак те якою “*владою над винаходом*” їх звично наділяють бачимо ще у Жака Еллюля (Jacques Ellul). Він описував інженерів, як верховних жерців, які турбуються про технології, але не мають над ними контролю. Подібні думки, як можна переконатися, призвели до того, що все частіше в колах інженерів стає більш популярною концепція негативної відповідальності, як її описав Дональд Готербарн [31]. Як було описано вище, її суть можна зрозуміти у протиставленні з позитивною відповідальністю, яка характеризується необхідністю враховувати наслідки діяльності стосовно інших при розробці технологічних артефактів. В контексті позитивної відповідальності, професіоналізм інженерів полягає у тому, скільки зусиль вони докладають задля уникнення небажаних негативних наслідків. Основна увага тут приділяється не уникненню відповідальності за безвідповідальну поведінку, а тому, що потрібно зробити на благо майбутнього. Можливі причини чому все ж інженери стають на позиції негативної відповідальності, буде прояснено в наступному підрозділі.

Важливо одразу визначити, про яку відповідальність іде мова в контексті філософії технологій. Якщо розглядається вимога *інженерної прозорливості*, з оцінкою майбутніх можливих наслідків винаходу – це часто, але не завжди, про моральну відповідальність; якщо ми розглядаємо поняття відповідальності, яке направлене на ситуацію минулого, то більшою мірою ми маємо справу з юридичними прецедентами та юридичною відповідальністю [43]. Як в перспективі, так і у ретроспективі відповідальність направляється на побудову соціальних відносин між людьми, а також між людьми та інституціями. Відповідальність встановлює очікування між людьми стосовно того, як виконуються певні зобов'язання та

обов'язки, при цьому формується механізм як виправляти чи заохочувати певну поведінку.

Уявимо, що з'явиться компанія-х, яка займатиметься виробництвом роботів. Очікуватиметься, що така компанія має розробити ефективні заходи безпеки, які не дозволять роботам шкодити людям. Якщо компанія не виправдає цих очікувань їй буде потрібно заплатити за збитки або ж понести будь-яка передбачене покарання. Покарання або ж перспектива покарання, можуть підштовхнути компанію приділяти більше часу системній безпеці, надійності, дизайну та ризикам, які можуть виникнути через виробництва роботів. Однак це може стати причиною того, що компанія стане шукати шляхи дистанціювання та уникнення вини. Ідея відповідальності, як міжособистісних відносин та очікувань, робить акцент на практиках притягнення когось до відповідальності [26].

Як було сказано раніше, відповідальність направлена у майбутнє пов'язана з обов'язками та очікуваннями. Людині притаманно брати на себе відповідальність у різних обставинах. Наприклад, водій сідає за руль автомобіля, цим самим бере на себе певні зобов'язання, перед іншими учасниками руху. Цього ж він очікує і від інших. В ситуації з інженерами, ми природно очікуємо від них певних внутрішніх рефлексій, стосовного того, яке інженерне рішення приймається та які наслідки воно може мати. Це не завжди про юридичну відповідальність, хоча в деяких ситуаціях її не уникнути. Також це стосується певної *моральної налаштованості* та спрямованості, яка закріплюється в посадових інструкціях, правилах чи професійно-моральних кодексах. Своєю чергою коли відповідальність аналізується з огляду дії чи події у минулому, відбувається розгортання причинно-наслідкових зв'язків, для того, щоб визначити хто винний у формальному чи не формальному відношенні.

Даний поділ на моральну та юридичну відповідальність, хоча й має у своїй основі багато внутрішньо подібних аспектів, та все ж стає корисним коли ми намагаємося розмірковувати про відповідальність загалом. Будь-який

поділ сприяє детальному аналізу взаємозв'язків, що у випадку розгляду технологій (ШІ зокрема) та їх творців (користувачів) стає визначальним, бо дає змогу побачити внутрішній розподіл відповідальності між кожним з учасників активності. Конкретні практики та соціальні структури, які використовують для приписування відповідальності, забезпечують певну підзвітність людей, і впливають на наше ставлення до технологій. В такому контексті варто звернутися до ідей Гелен Ніссенбаум (Helen Nissenbaum). Розглядаючи приписування відповідальності у контексті технологій, можуть виникати певні труднощі, які більшою мірою пов'язані з організаційними та культурними передумовами в яких створюється чи впроваджуються технології. Г. Ніссенбаум припускає, що це також залежить від рівня того, як ми розуміємо як працюють певні артефакти. Прикладом тут слугують комп'ютерні технології. Те як люди розуміють та уявляють природу, можливості та обмеження обчислень впливає на відповідальність тих, хто створює і в підсумку використовуватиме технології [41].

Експлікуючи попередні міркування, відповідальність – це про взаємовідношення, про те, як ми можемо діяти та діємо, та які обов'язки від нас вимагають або ж ми самостійно покладаємо на себе. Беручи за основу напрацювання М. Нурман [43] можна виділити наступні компоненти у відношеннях відповідальності:

- *дієвець* (the agent) – це людина (група людей), яка виконує певну дію;
- *пацієнт* (the patient) – це людина (група людей), яка знаходиться під впливом дії, або ж якось на неї впливає;
- *форум* (the forum) – це людина (група людей), яка визначає хто має нести відповідальність. Це може бути окремий самостійний дієвець, який виражає поведінку інших дієвців. Також це може бути суддя чи громадськість;
- *механізм примус* (the enforcement mechanism) – це засіб винесення та реєстрації критичних оцінок дій, як покарання, так і похвали, на основі загальноприйнятого стандартизованого правового механізму.

Різні поєднання даних компонентів формують різницю видів відповідальності, яку ми очікуємо від різних відношень. Традиційно, саме поняття відповідальності у моральній філософії застосовується лише до людини. Зокрема, в ліберальних демократіях вважають, що основою моральної відповідальності є автономність та свобода волі. Це є визначальним і відрізняє нас від машин та тварин. В даній роботі це береться за основу, тому що як було розглянуто раніше, відповідальність означає автономію, але не навпаки. Саме тому ми можемо говорити про технологічні артефакти¹, які мають певні ступені автономії. Наприклад деякі системи штучного інтелекту, можна вважати автономними, але з цього не слідує відповідальність самої системи, як самостійного дієвця. В цьому контексті існує декілька визначень автономії, наприклад технічне визначення автономії передбачає, що система в деякому ступені автономна у відношенні того, як людина здатна контролювати подібну систему. Тут віднаходиться паралель з проблемами автономії у стосунку технологій розроблених на основі штучного інтелекту.

Варто повернутися до думки Г. Ніссенбаум стосовно того, як нерозуміння роботи комп'ютерних програм провокує питання хто контролює і хто відповідальний, якщо ми маємо справу з проблемою чорної коробки (black box problem). Питання автономії підіймає питання не лише технічних аспектів побудови автономних систем, а й питання права та безпеки. Як покласти відповідальність на технологію, якщо її внутрішня робота є для нас не зрозумілою.

Розглядаючи людську автономію, погоджуються, що це визначальна умова відповідальності, однак з деякими уточненнями. За умови, якщо на нас не здійснюється зовнішні впливи, які змушують нас діяти проти нашої волі – кожна наша дія та вибір, покладає на нас відповідальність жити з наслідками нашого вчинку. Це умова нашої свободи, – людина не може

¹ Термін технологічні артефакти буде надалі використовуватися, у визначенні продуктів технологій, які були створені з певною функцією дієвцем технологічної активності.

перекладати власну відповідальність на іншого. Однак окрім автономії існує ще дві умови приписування моральної відповідальності: каузальний зв'язок між поведінкою дієвця та наслідком його дії, а також вимога до дієвця передбачати можливі наслідки його дій [31]. Тобто з цього слідує, що приписування моральної відповідальності залежить від того, чи здатна людина свідомо змінювати хід подій у світі, при цьому передбачаючи можливі наслідки власних дій. В іншому випадку, було б не правильно звинувачувати когось у шкоді, яка була заподіяна не умісно. За таких умов варто наполягати на імперативі для технологічної діяльності [4].

Поняття відповідальності найчастіше знаходить своє відображення в інженерній етиці та етичних кодексах. Останні актуалізують три типи відповідальності інженерів:

- чесне та добросовісне виконання своїх професійних обов'язків;
- відповідальність перед роботодавцем та клієнтами;
- відповідальність перед суспільством та спільнотою.

Г. Ніссенбаум зазначає, що в більшості ситуацій доволі складно говорити про відповідальність одного окремого інженера, бо якщо ми говоримо про його відповідальність перед роботодавцем, то такі ієрархічні та ринкові обмеження впливають на рішення, які приймає інженер у своїй роботі. Причинно-наслідкові зв'язки між дією агента та наслідками його діяльності розмиваються, бо він включений у довгий ланцюг, що передбачає залучення до розробки великої кількості інших дієвців. Труднощі приписування індивідуальної відповідальності стають причиною “Проблеми багатьох рук” (Problem of many hands). Згідно з Г. Ніссенбаум це лиш одна з чотирьох основних перепон до прийняття та приписування відповідальності у сучасному суспільстві. До цих перепон авторка відносить також те, що комп'ютерну помилку розглядають як таку, яку не можливо уникнути у роботі великих комп'ютерних систем. Проблема проявляється в тому, що ці два фактори слугують основою аргументації проти ідеї розгляду створення технологічних артефактів як відповідальної та ціннісно навантаженої

практики.

Автономність, свобода волі, та вимога передбачати можливі наслідки дій, – це три умови приписування відповідальності, які відкривають доволі широке предметне поле для дебатів, стосовно обґрунтування відповідальності загалом. Фактом залишається, що саме дієвець розміщується у центрі уваги, а можливі дебати радше розглядатимуть умови, які впливають на приписування йому відповідальності моральної чи юридичної.

2.2. Можлива критика збільшення відповідальності у контексті створення нових технологій

У класичних теоріях філософії технологій, як можна переконатися вище, є поширеною теза згідно з якою відповідальність збільшується пропорційно з появою нових технологій. Вимога до дієвців технологічної активності передбачати можливі наслідки їхніх дій, продукує ситуацію, в якій як пише К. Баєрц ми несемо відповідальність за все що ми очікуємо, але не зможемо одноосібно контролювати. Згідно з його словами, подобається нам чи ні, але дороги назад у рай безвідповідальності більше немає [19, 146].

Однак такий підхід може супроводжуватися спробами уникнути відповідальності. Бо (а) розробники технологічних артефактів зазвичай працюють у команді, й (б) дистанція яка виникає між творцем та наслідками використання пацієнтом технології, може слугувати аргументом того, що немає чітких причинно-наслідкових зв'язків. Створення чогось відкриває багато різних альтернативних шляхів розвитку майбутнього, кожен з яких може бути актуалізований, але в цей момент ще не актуальний. За умови, якщо ми не розуміємо того, як взагалі відкриваються такі можливості, доволі тяжко описати те, які точні наслідки можуть мати наші дії.

Це може спровокувати запитування, а навіщо нам взагалі технології, якщо ми не маємо чіткого бачення наслідків їх можливих впливів, при тому що кожна нова технологія збільшує наші відповідальності за її створення,

впровадження та використання. Якщо кількість відповідальності збільшується пропорційно з тим як з'являються нові технології та рухається прогрес, то з точки зору людських обмежень, в якийсь момент кількість відповідальності та обов'язків може перевищити наші реальні можливості з ними справлятися. Ця межа є непрозорою, й вона вимагає зупинки прогресу, або ж уникнення відповідальності. Що перший, що другий сценарій є для людства не припустимим, тож виникає запитання, як залишаючись прозорливим щодо наслідків технологічних дій та не впадаючи в стан регресу брати на себе відповідальність. Відповідь на це питання знаходимо у роботі Майкла Кленка (Michael Klenk) та Мартіна Сенда (Martin Sand) "Спадок Прометея: відповідальність та технології" [46].

На початку автори пропонують розглядати технології, як відповідь на моральні запити. Людина по своїй природі не досконала, тому вона шукає та створює інструменти для того, щоб боротися з хворобами, голодом, стихійними лихами, тобто з усім тим що загрожує добробуту. Століттями вчені, інженери, новатори, брали на себе відповідальність та обов'язок, створювати інструменти для розширення людських можливостей у боротьбі за виживання нас як виду. Хоча звісно інновації працюють як і на глобальному рівні, так і на локальному рівні розв'язання буденних проблем та завдань. Важко захищати абсолютність таких прагнень інженерів, бо все ж деякі винаходи направлені не завжди на благо, а, наприклад, супроводжуються комерційними інтересами чи вимогами ринку.

У питаннях про технології М. Кленк та М. Санд висувують вимогу завжди ставити питання відповідальності на перше місце. При цьому вони рекомендують при створенні чогось нового, направлено на покращення життя людей, запитувати чи дійсно для досягнення таких цілей необхідна саме нова технологія. У світі постійного прогресу, ми звикли реагувати на виклики питанням, яка нова технологія нам потрібна, а не навіщо і для кого. Перед тим як приступати до створення нових артефактів, а також при створенні нових технологій, необхідно спочатку проаналізувати потреби та

проблеми людей, провести ранжування загроз та викликів з якими люди зіштовхуються у відношенні їхньої значливості. Подібний аналіз проблем, може показати, нетехнологічні розв'язання проблем, хоча початково ми розглядаємо технології та інновації єдиними способами подолання проблемних ситуацій.

Підхід авторів важко назвати критикою відповідальності, радше це підхід за яким, не передбачається що моральні дилеми вирішуються засобами лише технічними. Технології перебувають в динаміці й даний рух провокує проблеми та виклики, вирішення яких покладається на все нові технологічні інновації. Від цього виникає парадокс дії, де ми як дієвці та пацієнти технологій потрапляємо в коло взаємозалежностей. Людина створює технології, які вирішуючи одну проблему, провокують іншу, яка вимагає створення нових технологій, для яких необхідна людина. Слідуючи логіці М. Кленка та М. Сенда, на кожному з етапів відбувається умовний приріст відповідальності, як того вимагають класичні праці присвячені відповідальності, прикладом тут можуть слугувати ті ж праці Х. Йонаса, К. Мітчема та К. Баєрца. Автори виходять тут на вагоме запитання в доволі оригінальний спосіб. Навіщо нам створювати технології, лиш через те що цього вимагає постійний прогрес. Їхня вимога тут направлена на створення форуму, у класифікації М. Нурман, на якому філософи, інженери, законодавці та громадяни мають визначити обов'язки які ми покладаємо на технології, шляхом визначення колективних потреб та обов'язків, а також засобів їх здійснення.

2.3 Можливості створення ціннісно втілених технологій

Вихідним питанням цього розділу є “Чи здатні технології втілювати цінності?”. Для того що відповісти на це запитання варто розглянути три підходи у контексті сучасної філософії технологій [27].

Інструменталізм. Даний підхід стверджує, що технології вільні від цінностей, тому що це лише інструмент в руках людини. Лише використання

цього інструменту людиною визначає чи дана технологія служить цінностям, чи віддаляє нас від них. Наприклад Національна стрілецька асоціація США, взяла за свій лозунг слова “Зброя не вбиває людей, люди вбивають людей”. Однак вбити людину з гвинтівки набагато простіше ніж звичайним олівцем, і з очевидності, якщо у дім вриваються злодії, поведінка власника стрілецької зброї відрізнятиметься, від людини без зброї. За інструменталістського підходу зброя є лише засобом, який не залежить від тих цілей, для яких його може застосувати споживач. Інструментальний підхід до технологій означає її нейтральність стосовно різних можливостей та варіантів застосувань. Розуміння світу за такого підходу механістичне, його можна і потрібно використовувати та контролювати без розгляду внутрішніх причин чи мети.

Субстантивізм. За такого підходу передбачається, що технології самостійно покладають для себе цінності, і люди не можуть на це вплинути. В межах даної теорії стверджується, що технології втілюють в собі такі цінності як ефективність. Проблема такого підходу в тому, що він ігнорує той факт, що люди впливають на технології, як при розробці, так і при їх використанні. Субстантивізм у філософії технологій означає, що при виборі технології для використання, ми не лише здійснюємо вибір з точки зору ефективності, а й підлаштовуємося під цінності закладені у технології.

Інтеракціонізм. Підхід відстоює думку, що цінності створюються та впроваджуються у взаємодії між людиною та технологіями. Мова тут йде, як про розробку, так і про використання технологій. Тобто за такого підходу відповідь на вихідне питання даного розділу буде позитивною. Однак відкритим залишається питання, про які цінності іде мова. Підхід інтеракціонізму є найбільш ціннісно орієнтованим з вище перелічених. Цінності за такого підходу перебувають у постійній динаміці, і залежать від того як ми впливаємо чи взаємодіємо з технологією. Що важливо, людина тут здатна впливати і як на створення технології, так і на її використання. Це називають людиноцентричним підходом у філософії технологій.

2.4. Втілення цінностей в технологіях. Умови впровадження.

Для інженерів існує спокуса розглядати технології у ціннісно-нейтральному відношенні, вважаючи їх просто інструментальними об'єктами. Дослідження зв'язків між технологіями та їх впливами на суспільство показує, що таке сприйняття технологій є повністю оманливим. Дія завжди провокує протидію. Технології втілюють у собі цінність, і вже навіть це мотивує нас до їхнього використання. Крім того, вони також формують нові типи нашої поведінки, цим самим формуючи очікування та актуалізуючи для нас нові цінності. Іншими словами, технології стають своєрідним посередником (медіатором, як ми побачимо нижче) між сприйняттям, досвідом, практиками та нормами [50].

Першим і найбільш звичним для нас прикладом в цьому контексті може слугувати “лежачий поліцейський”, який як артефакт протидіє певному типу людської поведінки. Інший приклад того, які цінності закладаються у технологічні артефакти, але в більш негативній конотації знаходимо у статті Ленгдона Віннера (Langdon Winner) “Чи є в артефактів політика?” [53]. У цій статті Л. Віннер стверджує, що відомий американський архітектор Роберт Мозес спеціально спроектував естакади над бульварами Лонг-Айленда заниженими, щоб автобуси з Нью-Йорка не могли доїхати до пляжів. Причиною такого інженерного рішення було те, що Р. Мозес був расистом, і зробив пляжі досяжними лише для білого середнього класу. Це яскрава ілюстрація того, як цінності, політичні погляди та влада, оформлені в технологіях, можуть обмежувати чи спонукати дії людей, при цьому формуючи для людей замкнене технологічне докільля (детально про це див. РОЗДІЛ 3). Цим самим Л. Віннер підсумовує, що цінності (в оригіналі автор говорить про політику) іноді інтенційно перетворюються у технологічні артефакти. Але ціннісні аспекти технологій, не завжди є такими очевидними. Іноді вони закладаються у сам дизайн таким чином, що технологічний артефакт стає неявним відбитком світогляду свого творця. Ба більше у світі інфраструктур та великих соціотехнічних систем, процес розробки стає

продовженням соціоінституційних звичок та практик, які постійно адаптуються до нових ідей та технічних можливостей.

Якщо ми беремо за основну мету технологій, створення ідеального, чи як мінімум кращого, світу, то варто закладати в технології не тільки інструментальні цінності, такі як ефективність, простоту використання чи надійність, але й моральні та політичні цінності, дотримання яких ми вважаємо основними для людства. Наприклад для ліберальних демократій такими цінностями можуть слугувати свобода, справедливість, довіра, прозорість та інші. З філософської точки зору ці ідеї є самодостатніми, однак як це перетворити на зрозумілу інженеро-орієнтовану вимогу.

В академічному середовищі ми систематично розмірковуємо про цінності, що слугує основою знань, які потрібно передати людям, що займаються дизайном технологій. В цьому контексті дослідники, зокрема Г. Ніссенбаум, пропонують серед параметрів оцінки добросовісності інженера розглядати засвоєння останнім філософських рефлексій про цінності та мораль [41]. За таких умов роботу технічного спеціаліста, або ж команди спеціалістів можна поділити на три режими: технічний, філософський, емпіричний.

На першому рівні відбувається звернення до їх професійних якостей та знань, для розробки технологічних механізмів. Це базова вимога професійності. Після того, як було зроблене відкриття, або ж знайдене вдале інженерне рішення, відбувається перехід у філософський режим. Ключова вимога такого режиму полягає у пошуку відповідей на питання про проходження цінностей, які виникли під час створення технологічного артефакту, аналізу їхнього концептуального значення і сили. В залежності від конкретного проекту буде виникати більше додаткових запитань щодо окремих цінностей, їх походження та ролі. Подібне запитування допоможе розробникам більш усвідомлено ставитися до своєї роботи, узгоджуючи власні наявні знання з глобальним інтелектуальним контекстом. Втілення цінностей в технологічний артефакт напряму залежить від усвідомлення та

розуміння конкретної цінної моделі.

Високий рівень розуміння термінів та ідей, – це важливий місток між цінностями та конкретними інженерними рішеннями. Для філософів тут з’являється додаткове завдання, розробити детальний та конкретний аналіз філософських понять для застосування в контексті розробки нових технологій. Як зазначають автори [8] не зважаючи на те, що технології зазвичай активно створювалися у ліберальних демократіях, де очікувано вищий рівень публічної філософії та політичної думки, однак це не спричинило втілення цінностей в технологіях. Острах викликає той факт, що популяризація деяких філософських або ж навіть наукових ідей маючи спрощений та доступний характер не формує цілісного та системного знання, на основі якого інженери б могли формувати свій світогляд, та включати такі ціннісно навантажені ідеї в основу своїх технологічних побудов. Така напівосвіта може створити хибну ілюзію обізнаності, однак за цими не слідуватиме глибока рефлексія.

Філософський режим запитування та досліджень, сприяє втіленню цінностей у технологіях. Він закладає базу для обґрунтування вибору певних цінностей, якими керуються інженери при своїй роботі, чим впроваджує їх в теоретичні основи побудов технологічних артефактів. Знання класичних теорій філософії політики та філософії моралі, дозволяє творцям технологій зрозуміти, чому деякі цінності актуалізуються в різних історико-культурних контекстах. Це дає змогу зрозуміти, на яких ідейних основах стоїть як сам інженер, так і користувачі його продукту.

Емпіричні дослідження це третій режим роботи, який відіграє таку ж визначальну роль для втілення цінностей в технології, як філософський та технічний рівні. Емпіричні дослідження тут не лише доповнюють філософські дослідження того, які цінності важливі та мають бути впроваджені в конкретному технічному завданні. Вони стають основним засобом для систематичного розв’язання питань про успішність (чи не успішність) спроб втілити цінності. Емпіричні дослідження стають ще одним

шляхом для визначення цінностей, які варто враховувати. Навіть якщо філософи сформулюють базовий набір універсальних людських цінностей, через те що розвиток технологій продукує нові цінності в різних контекстах, технології вимагатимуть застосування ширших наборів цінностей та засобів для їх з'ясування. Потрібно не забувати, що цінності та їх вага залежить від культури, історії, національності, етнічної чи релігійної приналежності. Врахувати дані обставини важливо, бо за умови коли дизайнер вибирає серед багатьох технологічних рішень декілька основних, ранжуючи їх лише по релевантності для поточного завдання, він може спровокувати різного роду упередження. Для розв'язання проблем такого порядку деяким дослідникам однієї філософії стає недостатньо, тож практики сфери технологій звертаються за альтернативою, а саме до емпіричних досліджень. Своєрідний вихід до людей допомагає краще дослідити конкретні групи населення для яких цінності й втіляться в дизайні. Це може також допомогти визначити, чи справді конкретний технологічний артефакт втілює задані цінності, а також прояснити наслідки та дійсні способи застосування конкретної технології.

Звісна річ, кожен з описаних режимів роботи команди розробників працює лише в постійному взаємозв'язку з іншими. Через те, що технології постійно перебувають в динаміці, розробникам, окрім того, що варто працювати в трьох режимах роботи одночасно, варто також враховувати, що нові знання отримані при одному з режимів вимагають ревізії вихідних пунктів на яких формувалися початкові гіпотези. В попередніх роботах ми зазначали, що в час коли технології перебувають у неймовірно швидкому русі вперед, варто робити осмислені зупинки, щоб аналізувати пройдений шлях та звірятися з цінностями та цілями, які беруться за основу в наших теоретико-практичних побудовах. Якщо наше ставлення до технологій не інструменталістське, то ми маємо час для того, щоб зробити осмислену зупинку в роботі, та перейти від рівня технологічних інновацій на рівень відповідальних інновацій.

В даному розділі було розглянуто ідею відповідальності, як визначальну умову створення нових технологічних артефактів. Особі чи групі осіб, що відповідальна за створення нових технологій відповідальність може бути приписана перспективно та ретроспективно. В кожному з випадків відповідальність є невіддільною від сфери технологій. Якщо ми погоджуємося на світ наповнений новими технологіями, ми погоджуємося брати на себе відповідальність за їх розробку, впровадження та застосування. З великою силою приходить велика відповідальність, і з огляду актуальних філософських досліджень сфери штучного інтелекту, людство хоча й знає про можливі загрози безвідповідального створення та застосування технологій, та насправді не усвідомлює масштабів можливих викликів. І хоча в межах наших досліджень ми не розглядаємо детально нанотехнології та нейротехнології, та все ж вони мають та матимуть вагомий соціальний та етичний вплив у майбутньому. Створення нових технологій змушує замислитися про те, яким може бути наше технологічне майбутнє. Чи дійсно розвиток технологій якісно вплине на наше життя, і взагалі чи співмірні загрози з можливими покращеннями. Не впадаючи у страхи та антиутопічні сценарії, ми маємо усвідомити, що технологічний розвиток залежить від нас, – це наша відповідальність, яким буде наше майбутнє. Ми можемо брати відповідальність та робити вибір, які технології і як розвивати. Якщо вважати абсурдною ідею про зупинку розвитку, то варто замислитися які цінності закладаються у розробку нових технологічних артефактів і яка межа простягання самих технологій в їхньому впливі на людину.

РОЗДІЛ 3. ПОШУКИ МЕЖІ: ВІД ТЕХНОЛОГІЙ ДО ТЕХНОЛОГІЧНОГО ДОВКІЛЛЯ

Ми говоримо про технології звертаючись до ідеї спільносвіту, щоб закласти межі створення технологій, з якими б світ, водночас був кращим ніж без них, і водночас залишався у своїй цілісності. Те, як відбувається розвиток технологій наразі – це формування відносно замкнених середовищ, які “виривають” людину зі спільносвіту, і поміщають в штучне *технологічне довкілля*.

Створюючи технології інженери зазвичай розглядають власну роботу занадто вузьку. Це не про спеціалізацію роботи, це про не усвідомлення меж самих технологій. Повернемося до прикладу Д. Готтербарна з Рентген апаратом. Технологія стала нести загрозу тоді, коли програміст звів свою роботу лише до *цїлі* написання програми, не розглянувши свою роботу у широкому масштабі.

Допоки ми не встановимо *технологічних меж*, які ми не маємо переступати, доти творці технологій не знатимуть іншого масштабу, окрім лиш технологічного. Тобто для технологій потрібно встановити мету та межі, які не легітимізуються ними самими. Технологія не має виправдовуватися лише технологічністю, позірною інноваційністю чи суто економічною доцільністю. Технології варто помістити в межі (*культури створення технологій*), які визначатимуть, що є благом для розвитку людства, а що ні.

Ми відповідальні за технології, які створюємо, й за світ, як цілісність, на який ми маємо технологічний вплив. Якщо ми беремо відповідальність, як ключову умову створення технологій, нам не потрібно окремо наділяти технології автономією. Аналогічно, щоб говорити про безпеку створення та використання ШІ, не потрібно наділяти комп’ютерні програми свідомістю. Натомість потрібно наповнити технологічний дискурс філософськими ідеями, зокрема принципами етики відповідальності при створенні нових технологій, які як ми переконалися в попередньому розділі втілюють в собі цінності.

3.1 Концептуалізація та аналіз теорій технологічного середовища

Для того, щоб концептуалізувати роль та визначення *технологічного довкілля* ми побудуємо свій шлях через актуальні ідеї філософії технологій. Саме поняття технологічного довкілля не є новим, однак його наявне наповнення не вдається нам достатнім, бо воно нехтує найголовнішим – людською відповідальністю за технологічну дію.

Актуальна розвідка даної теми буде базуватися на фундаментальній статті “Технологічна екологія: концептуалізація технологій як медіаторів довкілля” (Technological Environmentality: Conceptualizing Technology as a Mediating Milieu). Центральне завдання даної статті визначити, як технології стають нашим середовищем (перехід до довкілля автори роблять лише в кінці статті), бо за сучасного розвитку технологій нам потрібне ширше розуміння впливу таких середовищ на людські когнітивні здатності та дії [17, 323].

Будинки, офіси, школи, парки та міста ніколи не були нейтральним фоном для процесів людського становлення: люди створили ці середовища підтримки, однак такі середовища “створюють” і людей. В біо-артефактних середовищах (bio-artifactual environments) люди організовують свій повсякденний досвід, підтримують свою рутину та тілесні звички, а також вибудовують свої когнітивні здатності. В цьому сенсі, як зазначають автори, багато людських здатностей, когнітивних та не когнітивних, завжди залежали від того, як люди створювали (розробляли) своє середовище для підтримки своїх активностей.

Сучасні інженерні процеси супроводжуються тим, що “загортають” нашу повсякденність в цифрову текстуру, щоб зробити наше середовище більш зовнішньо розумним [17, 324]. Розумні технології воліють стати непомітними для нас, стерти *межу* між технологічним середовищем та людською взаємодією з ним та в ньому (див. підрозділ 2.2 Етика штучного інтелекту). Однак через те, що технологічне середовище, як ми це визначаємо, є радше замкненим довкіллям, ми опиняємося, застосовуючи

термінологію Дона Ігде (Don Ihde) в технологічних коконах (technological cocoons).

Автори пропонують новий тип екологічних технологій (environmental technology) – це Активні Технологічні Середовища (Active Technological Environments). Тобто технологій не розглядаються, як німий та стабільний фон для наших дій, вони стають активно залученими до взаємодії з людьми та іншими матеріальними об'єктами, які формують наш простір. Активні Технологічні Середовища підіймаються на вищий рівень і формують *Технологічне Середовище (Technological Environments)*, яке стає характеристикою людського існування у світі Активних Технологічних Середовищ. Концепція Технологічного Середовища показує, що наше середовище при зустрічі з технологічними середовищами починає набувати характеристик останнього, що веде до нових технологічних умов існування homo feber. Технологічні середовища зазвичай розглядають в межах двох теорій: Постфеноменологічної (теорія розроблена американським філософом технологій Доном Ігде) та Теорії Матеріальної Залученості (Material Engagement Theory). Коротко розглянемо суть даних теорій, та причини за якими ці теорії не є достатніми.

Постфеноменологічний підхід. Початково Постфеноменологія фокусується на індивідуальному досвіді у стосунку до індивідуальних технологій у формі “Я-технології-світ”. Тобто технологія тут постає очевидним медіатором. Однак сучасні технології, через свою “прихованість”, все менше стають медіатором, а більше самим світом. Тому автори хочуть розширити постфеноменологічну рамку, для того , щоб показати, що не лише людина інтенціонально направлена на технологічне середовище, а й середовище “інтенціонально” направлене на людину [17, 329].

Теорія Матеріальної Залученості була представлена американським дослідником Ламбросом Малафурісом (Lambros Malafouris) в книзі “Як речі оформлюють свідомість” (How Things Shape the Mind). Теорія матеріальної залученості пропонує багатий аналіз зв'язків між людиною та її матеріальним

середовищем, пояснюючи як середовище (environment) допомагає задати форму (to shape) нашим когнітивним здатностям, нашій діяльності, та нашій інтенціональності. Замість того, щоб брати дихотомію свідомість-світ за початкову точку, Теорія Матеріальної Залученості підходить до речей, як невіддільних компонентів людської думки. Як приклад, Л. Малафуріс пропонує розглянути гончарну справу. Гончар та глина перебувають в “dance of agency”, як рівні партнери в цьому танці; в циклі гончарства стає неможливо розділити причину і наслідок. Як зазначають автори, Теорія матеріальної залученості, аналізує та охоплює лише малотехнологічні середовища, і не можливо просто екстраполювати її на такі сучасні технології як Інтернет речей чи Навколишній Інтелект (Ambient Intelligence). З чим важко не погодитися, бо зі ступенем інтенсивності впливу технологій на людину радикально змінюється й *межа технологічного впливу* (технологічна межа, як така). Для прикладу, вже зараз літаки, які докорінно змінили текстуру світу у сприйнятті простору, можна розглядати вже як джерела малотехнологічних середовищ. Тобто межа сприйняття змінюється, але не так радикально.

Теорія матеріальної залученості первинно зосереджується на “відчутній матеріальності” (глина, весла, дверні ручки й т.п.), в той час, коли технологічне середовище зазвичай *приховане* від повсякденного феноменального досвіду, і знаходяться за межами “досвідчуваної матеріальності” (experienceable materiality). Як зазначають автори, технологічні середовища є “розумними” і стають “розумнішими”, бо вони все більше можуть *функціонувати* без залучення [нашого] досвіду та уваги.

Ця думка протиставляється нашим уявленням про *технологічне довкілля*, бо тоді не зрозуміло, як вибудовувати нашу відповідальність. Відповідальність за створення технологій, які неявно оформлюють межі нашого довкілля, щоб потім ми не могли діяти вільно в такому середовищі, а узгоджували дії з технологіями.

Теорія матеріальної залученості занадто концентрується на “користувачеві” та “речах в середовищі”, ігноруючи при цьому, як вже зараз зросла здатність технологічного середовища впливати на людського агента для певних цілей (що покладає середовище). Це ілюзія, як зазначають автори, що ми можемо контролювати активні та непомітні технології, так само як наївна віра в те, що ми можемо контролювати нашу природну екологію.

У актуальній роботі ми виступаємо проти того, що це ілюзія. Навпаки ми вказуємо на активно відповідальну залученість людини в її довкілля та спільносвіт. Наприклад, К. Маєр-Абіх, своїми конкретно-практичними прикладами показує, як ми своїми діями можемо позитивно впливати на екологію, або ж мінімізувати шкоду від нашої діяльності [8].

3.2 Перехід від технологічних середовищ до технологічного довкілля

Розглянувши плюси та особливості Постфеноменологічної теорії та Теорії Матеріальної Залученості, які найперше пов'язані з інтенціональністю та дією, автори пропонують власну теорію, бо (а) Постфеноменологія хоча й має фокус на стосунках людина-технології, упускає той момент, коли технології починають грати активну роль в нашому середовищі й стають нашим “світом”. Своєю чергою (б) Теорія матеріальної залученості визнає активну роль матеріального середовища, однак окремо (відірвано) не розглядає технологію з її власною інтенціональністю та дієвістю. Саме тому автори й пропонують об'єднану концепцію *Активного Технологічного Середовища*.

Автори хочуть знайти фреймворк для розуміння технологічного довкілля. Але вони спільно з іншими авторами, яких ми розглянули вище (див. Розділ 1), хибують в тому, що в непрозорості технологій вбачають їхню самотійність (автономію): самотійність технологій та самотійність середовищ. Повторимо, причиною непрозорості технологічних артефактів є людина. Непрозорість роботи таких артефактів, це відповідальність людини, що в роботі даних авторів ігнорується.

В першу чергу, дослідники ставлять перед собою завдання розв'язати проблему концептуалізації дії матеріального середовища у відношенні до медіаційної дії (mediated agency). Запитанням на яке вони хочуть знайти відповідь: “... дієвість є в самій речі чи у відношенні між людиною та річчю?”. Можлива відповідь – це *dance of agency* радше ніж “матеріальна медіація людської дії” (materially mediated human agency): танець в який людина та середовище залучають один одного, і в якому дія людини та дія матеріального середовища тісно переплетені, й разом конституюють практику [333].

Розглянемо взаємодію дієвостей людини та матеріального середовища, ще раз. Та взаємодія, що виникає між ними, це взаємодія *сама по собі*, яку автори концептуалізують як “танець” в якому людська і матеріальна дієвість переплетені, та оформлюють одна одну. Результатом такої взаємодії, є форма людської дії. Концепцію авторів наочно можна зобразити в наступній схемі:

**Human agency + material agency = human agency → organizes practices
= technological environment → human agency**

Автори роблять спробу об'єднання бачення інтенціональності в Постфеноменології та Теорії Матеріального Залучення. В той час коли Теорія Матеріальної Залученості розглядає інтенціональність, як таку що оформлює *взаємодію (interrection)* між людьми та середовищем, для Постфеноменології інтенціональність є *медіатором (mediation)* технологій, з якими взаємодіє людина. Для того, щоб показати, що це одне й теж дослідники використовують концепцію імерсивних стосунків (immersive relation), яку можна подати в такій схемі:

Human ↔ technology/world

Як бачимо, технології збігаються зі світом. Тут технологічна інтенціональність, в подвійному розумінні є перцептивною та діяльнісною інтенціональністю, й має стати центральною в концептуалізації Технологічного Середовища. А для нас це постає центральною темою проблематизації Технологічного Середовища, яке ми розглядаємо як *технологічне довкілля*.

Центральний інсайт, що лежить у постфеноменологічному концепті медіації, – це те що ми маємо розглядати технології не як елемент світу, а як *відношення* людини та світу. В Активному Технологічному Середовищі, технології *стають частиною світу*, і так само виконують роль медіатора.

Технології стають довкіллям. У розумінні авторів, довкілля зв’язує людей та світ інтерактивно “охоплюючи” шляхи якими людина пов’язана зі світом; як вода стає довкіллям для риб, пов’язуючи їх з іншими рибама, рослинами та об’єктами, а також конститує їх як риб, а їхнє середовище, як їхній світ [337]. Ось так *специфічне довкілля* може підмінити нам спільноствіт, в його багатогранності, перетворивши нас на “німих риб”.

Технологічне довкілля впливає на те, як людина відчуває власну належність до світу. В даному контексті варто ще раз окремо звернутися до праці К. Маср-Абіха “Повстання на захист природи: від довкілля до спільносвіту”, який є нашим ситуативним союзником для повернення людини до цілісного буття-в-світі. Наші підходи подібні в тому, як ми розглядаємо питання відповідальності. Однак наш проект не такий поки не такий масштабний, бо з наших первинних вимог турботи про людськість, ми у наших побудовах зупиняємося на рівні антропоцентризму сучасності, хоча й робимо спробу вписати його у межі значно масштабнішого проекту (турботу про людство в цілому). Що нас відрізняє, так це центральна тема, для нас це технології, а не природа. Хоча у відношенні до природи та технологій, та їхніх впливах та ролі ми віднаходимо багато спільного. Не даремно ми

говоримо, не про абстрактне технологічне середовище, а саме технологічне *довкілля*.

Як зазначає перекладач книги К. Маєр-Абіха Анатолій Єрмоленко розмірковуючи про відповідальність: “Це має бути не тільки відповідальність за інших людей, не тільки відповідальність за тварин і рослин, а й відповідальність за всі речі, створені людиною і не створені нею” [171]. У межах нашої роботи ми розглядаємо речі створені людиною й вище ми переконалися наскільки важливо працювати над розумінням цінностей, які ми вкладаємо в технології, які й формують наше довкілля. Як зазначає К. Маєр-Абіх: “...не людина є мірою всіх речей, а все, що нас оточує, є мірою нашої людяності” [96], що при розгляді наявних технологічних довкіль набуває особливої уваги та тривоги. Ті технологічні довкілля, в яких ми перебуваємо задають нам нові міри нашої людяності, й в кожному такому довкіллі творці технологій розглядають власну відповідальність по різному: в контексті цілей, а не мети технологічної дії. Технології фрагментують нашій світ та нашу відповідальність. Однак, відповідальність це не окремі фрагменти зображення у калейдоскопі, це саме погляд "на" і "через" відповідальність. Лише за таких умов ми зможемо створювати технології на благо людства.

ВИСНОВКИ

У ході роботи нами були виконані поставлені на початку завдання. У першому розділі даної роботи ми провели детальний аналіз технологій, на прикладі розробки ШІ. Ми встановили, що ШІ, як і будь-яка інша технологія впливає на людину, бо як зазначає Е. Фінберг: “Кожна наша дія повертається до нас у формі зворотнього зв’язку...”. В такому ключі постає питання, хто відповідальний за таку дію, якщо використовується ШІ. Попри очевидність відповіді, використання нейронних мереж розмило відповідь на дане запитання. Нейронні мережі, які наразі є невіддільною частиною ШІ у внутрішній своїй роботі постають для інженерів “непрозорими”. Такі системи працюють з неймовірною кількістю даних і в алгоритмі їх побудови закладене внутрішнє визначення релевантності певних даних. Система постійно змінює свою роботу, перебираючи та аналізуючи дані на вході, тож відповідь на виході стає неочевидною для творця такої програми. Не можна просто встановити причинно-наслідкові зв’язки поглянувши на те, як програма працювала. Для людини це стає ускладненим, але не неможливим, через величезну кількість внутрішніх операцій. З “непрозорості”, як було побачено, хибно експлікувалося визначення *автономності* таких систем аж до приписування їм відповідальності.

Розглядаючи ШІ як “автономного морального агента”, дослідники почали розробляти етику для подібних програм, щоб встановити, чим для них є прийняття рішень. Подібні міркування значною мірою нагадують ідеї, в яких маніпуляції символами програмою сприймалися за усвідомлене діяння. Ігнорування цілісного характеру людської дії та семантичного характеру мисленнєвої діяльності показує, що історія в дослідженнях ШІ повторюється. В “непрозорості” роботи нейронних мереж починають вбачати можливість свідомості. Така оманлива можливість розмиває межі відповідальності творців подібних програм, які “непрозорість” редукують до автономії та відповідальності. Однак творці цим самим ігнорують власну відповідальність за “дії” подібних програм, забуваючи про “штучність” того, що вони

створюють. Завжди варто пам'ятати, що причиною “непрозорості” системи є сама людина. І хоча тяжко прояснити шлях причинно-наслідкових зв'язків, яким “рухалася” програма, ми точно знаємо, що причиною створення програми була робота людини. У філософському осмисленні важливою постає концепція “*позитивної відповідальності*”, в якій йдеться про моральний обов'язок практиків сфери технологій враховувати та аналізувати можливі наслідки їх діяльності. Створюючи та застосовуючи технології, варто бути прозорливими, робити зупинки для осмислення питання “навіщо” і того, які це буде мати наслідки у майбутньому.

Технологія, як інструмент, у не інструментальному розумінні, розширює межі впливу людини на світ та на саму себе. Встановленню взаємозалежності між розвитком технологій та *межами відповідальності* був присвячений другий розділ. Кожен новий технологічний артефакт формуючи нові інтеракції та дистанції, створює *простір відповідальних практик*. Дієвцем та активним учасником за таких обставин постає творець технологічного артефакту. Відповідальність у такому контексті постає як взаємовідношення дієвця з простором в якому він створює та впроваджує технології. Інші, як учасники цього простору очікують від нього прозорливості щодо можливих наслідків його технічної дії.

На основі праць Е. Фінберга та Є. ван Ден Говена було експліковано три підходи розгляду цінностей у стосунку технологій. Після детального аналізу інструменталістський підхід був відкинтий, як такий, що не відповідає нашим засадничим умовам необхідності приписування відповідальності за технологічну діяльність. Далі на основі підходу інтеракціонізму, ми розглянули варіанти того, як цінності втілюються в технологіях.

В третьому розділі ми констатували, що межею створення технологій має бути межа відповідальності за людство в цілому. Якщо така межа фрагментує світ до *технологічних довкіль*, з яких не відбувається вихід до цілісного розгляду технологій та їх ролі в *глобальній технологічній текстурі*,

ми маємо переглянути доцільність створення, впровадження та використання таких технологій.

Ми переконані що задля створення безпечних технологій, інженери мають об'єднатися з філософами та законодавцями, щоб закласти межі створення технологій. Таких технологій, з якими б світ, був кращим ніж без них. Щоб не завдаючи негативного впливу людству, підтримувалася цілісність нашого світу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ І ДЖЕРЕЛ

1. Аристотель. О душе [Електронний ресурс] / Аристотель // Мысль. – 1976. – Режим доступу до ресурсу: <http://psylib.org.ua/books/arist01/index.htm>.
2. Де Лібера А. INTELLECTUS Інтелект, здатність розуміти, значення, сенс / Ален Де Лібера // Європейський словник філософій: Лексикон неперекладностей / Ален Де Лібера. – Київ, 2011. – (ДУХ І ЛІТЕРА). – С. 96–107.
3. Дрейфус Х. Чего не могут вычислительные машины. Критика искусственного разума / Хьюберт Дрейфус. – Москва, 2010. – 340 с. – (Либроком).
4. Єрмоленко А. Соціальна етика та екологія. Гідність людини — шанування природи. / Анатолій Єрмоленко. – Київ, 2010. – 416 с. – (Лібра).
5. Йонас Г. Принцип відповідальності. У пошуках етики для технологічної цивілізації / Ганс Йонас. – Київ, 2001. – 400 с. – (Лібра).
6. Канеман Д. Мислення швидко й повільно / Деніел Канеман. – Київ, 2017. – 480 с. – (Наш Формат).
7. Кірюхін Д. Дискурси справедливості в історичному контексті / Денис Кірюхін. – Київ: Стилос, 2021. – 202 с.
8. Маєр-Абіх К. М. Повстання на захист природи від доккілля до спільносвіту / Клаус Міхаель Маєр-Абіх. – Київ: Лібра, 2004. – 196 с.
9. Маркофф Д. Homo Roboticus? Люди и машины в поисках взаимопонимания / Джон Маркофф. – Москва, 2017. – 408 с. – (Альпина Паблишер).
10. Рассел С. Искусственный интеллект. Современный подход / С. Рассел, П. Норвиг. – Москва, 2007. – 1408 с. – (Вильямс).
11. Серль Д. Розуми, мозки і програми / Джон Серль // Антологія сучасної аналітичної філософії, або жук залишає коробку / Джон Серль. – Львів, 2014. – (Літопис). – С. 229–255/
12. Фає Е. ІНТЕЛЕКТ, ІНТЕЛІГУВАТИ / Емануель Фає // Європейський словник філософій: Лексикон неперекладностей (Том другий) / Емануель Фає. – Київ, 2011. – (ДУХ І ЛІТЕРА). – С. 81–83.

13. Фінберг Е. Десять парадоксів техніки / Ендрю Фінберг // Антологія сучасної філософії науки, або усмішка ASIMO / Ендрю Фінберг. – Львів, 2017. – (ЛНУ ім. І. Франка). – С. 258–281.
14. Шинкарук В. Філософський енциклопедичний словник / Володимир Шинкарук. – Київ, 2002. – 742 с. – (Абрис).
15. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955 / J. McCarthy, M. Minsky, N. Rochester, C. Shannon. // AI Magazine. – 2006. – №27. – С. 12–14.
16. Arkoudas K. Philosophical foundations / K. Arkoudas, S. Bringsjord // The Cambridge handbook of artificial intelligence / K. Arkoudas, S. Bringsjord. – Cambridge, 2014. – (Cambridge University Press). – С. 34–63.
17. Aydin C. Technological Environmentalty: Conceptualizing Technology as a Mediating Milieu / C. Aydin, M. W. González, P. Verbeek. // Philosophy & Technology. – 2019. – №32. – С. 321–338.
18. Bats J. Mediating technology: how ICT influences the morality of the digital generation [Електронний ресурс] / J. Bats, R. Valkenburg, P. Verbeek // Proceedings of the International Conference on Engineering Design, ICED. – 2013. – Режим доступу до ресурсу: https://www.researchgate.net/publication/277006917_Mediating_technology_How_ICT_influences_the_morality_of_the_digital_generation.
19. Bayertz K. Increasing Responsibility as Technological Destiny? Human Reproductive Technology and the Problem of Meta-Responsibility / Bayertz Kurt // Technology and Responsibility / Kurt Bayertz., 1987.
20. Block N. Troubles with functionalism / Ned Block // Perception and Cognition: Issues in the Foundations of Psychology / Ned Block. – Minneapolis, 1978. – (University of Minnesota Press). – С. 261–325.
21. Bringsjord S. Artificial Intelligence [Електронний ресурс] / Selmer Bringsjord // Stanford Encyclopedia of Philosophy. – 2018. – Режим доступу до ресурсу: <https://plato.stanford.edu/entries/artificial-intelligence/#LogiBaseAISomeSurgPoin>

22. Brooks R. Intelligence Without Reason [Электронный ресурс] / Rodney Brooks // MIT Artificial Intelligence Laboratory. – 1991. – Режим доступа до ресурсу: <https://dspace.mit.edu/bitstream/handle/1721.1/6569/AIM-1293.pdf>.
23. Carter M. Minds and computers: An introduction to the philosophy of artificial intelligence: An introduction to the philosophy of artificial intelligence / Matt Carter. – Edinburgh, 2007. – 222 с. – (Edinburgh University Press).
24. Christman J. Autonomy in Moral and Political Philosophy [Электронный ресурс] / John Christman // Stanford Encyclopedia of Philosophy. – 2015. – Режим доступа до ресурсу: <https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/>.
25. Dignum V. Ethics in Artificial Intelligence: Introduction to the Special Issue / Virginia Dignum. // Ethics and Information Technology. – 2018. – №20. – С. 1–3.
26. Eshleman A. Moral Responsibility [Электронный ресурс] / Andrew Eshleman // The Stanford Encyclopedia of Philosophy. – 2016. – Режим доступа до ресурсу: <https://plato.stanford.edu/entries/moral-responsibility/>.
27. Feenberg A. What Is Philosophy of Technology? / Andrew Feenberg. // Defining technological literacy: Towards an epistemological framework. – 2006. – С. 5–16.
28. Floyd J. Wittgenstein's Diagonal Argument: A Variation on Cantor and Turing / Juliet Floyd // Epistemology versus Ontology. Logic, Epistemology, and the Unity of Science / Juliet Floyd. – Dordrecht, 2012. – (Springer). – С. 25–44.
29. Fodor J. Tom Swift and his procedural grandmother [Электронный ресурс] / Jerry Fodor // Cognition. – 1978. – Режим доступа до ресурсу: <http://www.nyu.edu/gsas/dept/philo/courses/mindsandmachines/Papers/tomswift.pdf>.
30. Friedland J. AI can help us live more deliberately [Электронный ресурс] / Julian Friedland // MIT Sloan Management Review. – 2019. – Режим доступа до ресурсу: https://www.academia.edu/39520151/AI_Can_Help_Us_Live_More_Deliberately.

31. Gotterbarn D. Informatics and professional responsibility / Donald Gotterbarn. // Science and Engineering Ethics. – 2001. – №7. – С. 221–230.
32. Haugeland J. Artificial Intelligence: The Very Idea / John Haugeland. – Cambridge, 1985. – 302 с. – (MIT Press).
33. McCarthy J. What is Artificial Intelligence? [Электронный ресурс] / John McCarthy // Stanford University. – 2007. – Режим доступа до ресурсу: <http://www-formal.stanford.edu/jmc/whatisai.pdf>.
34. Mitcham C. Responsibility and Technology: The Expanding Relationship / Carl Mitcham // Technology and Responsibility / Mitcham Carl., 1987.
35. Moor J. The Nature, Importance, and Difficulty of Machine Ethics / James Moor. // IEEE Intelligent Systems. – 2006. – №21. – С. 18–21.
36. Müller V. Autonomous Cognitive Systems in Real-World Environments: Less Control, More Flexibility and Better Interaction / Vincent Müller. // Cognitive Computation. – 2012. – №4. – С. 212–215.
37. Müller V. Ethics of Artificial Intelligence and Robotics [Электронный ресурс] / Vincent Müller // The Stanford Encyclopedia of Philosophy. – 2020. – Режим доступа до ресурсу: <https://plato.stanford.edu/entries/ethics-ai/#OpacAISyst>
38. Müller V. Philosophy and Theory of Artificial Intelligence 2017 / Vincent Müller. – Leeds, 2018. – 320 с. – (Springer Nature Switzerland AG).
39. Neuroscience and Philosophy. Brain, Mind, and Language / M. Bennett, D. Daniel, H. Peter, J. Searle. – New York, 2009. – 232 с. – (Columbia University Press).
40. Newell A. Computer science as empirical inquiry: Symbols and search / A. Newell, H. Simon. // Communications of the Association for Computing Machinery. – 1976. – №19. – С. 113–126.
41. Nissenbaum H. Embodying values in technology: Theory and practice / H. Nissenbaum, M. Flanagan, D. Howe. // Information Technology and Moral Philosophy. – 2008. – С. 322–353.
42. Noorman M. Computing and Moral Responsibility [Электронный ресурс] / Merel Noorman // Stanford Encyclopedia of Philosophy. – 2018. – Режим доступа

до

ресурсу:

<https://plato.stanford.edu/entries/computing-responsibility/#RetConMorRes>

43. Noorman M. Responsibility and Liability / Merel Noorman // A Citizen's Guide to Artificial Intelligence / Merel Noorman., 2020. – (The MIT Press).

44. Putnam H. Minds and machines [Электронный ресурс] / Hilary Putnam // New York University Press. – 1960. – Режим доступа до ресурсу: <https://philpapers.org/archive/PUTMAM.pdf>.

45. Russell S. Rationality and intelligence / Stuart Russell. // Artificial Intelligence. – 1997. – №94. – С. 57–77.

46. Sand M. Prometheus' Legacy: Responsibility and Technology / S. Martin, K. Michael // Welche Technik? / M. Sand, M. Klenk. – Dresden, 2020. – (Text & Dialog). – С. 23–40.

47. Sun R. Connectionism and neural networks / Ron Sun // The Cambridge handbook of artificial intelligence / Ron Sun. – Cambridge, 2014. – (Cambridge University Press). – С. 108–127.

48. Torrance S. Machine Ethics and the Idea of a More-Than-Human Moral World / Steve Torrance // Machine Ethics / Steve Torrance. – Cambridge, 2011. – (Cambridge University Press). – С. 115–137.

49. Turing A. Computing machinery and intelligence [Электронный ресурс] / Alan Turing // Mind. – 1950. – Режим доступа до ресурсу: <https://academic.oup.com/mind/article/LIX/236/433/986238>.

50. Verbeek P. Materializing morality design ethics and technological / Peter-Paul Verbeek. // Science Technology & Human Values. – 2006. – №31. – С. 361–380.

51. Verbeek P. What things do: philosophical reflections on technology, agency, and design / Peter-Paul Verbeek. – University Park: Pennsylvania State University Press, 2005. – 259 с.

52. Wetzel L. Types and Tokens [Электронный ресурс] / Linda Wetzel // Stanford Encyclopedia of Philosophy. – 2006. – Режим доступа до ресурсу: <https://plato.stanford.edu/entries/types-tokens/>.

53. Winner L. Do Artifacts Have Politics? [Электронный ресурс] / Langdon Winner // The MIT Press. – 1980. – Режим доступа до ресурсу: https://web.media.mit.edu/~ascii/papers/winner_1980.pdf.

54. Wittgenstein L. Philosophical Investigations [Электронный ресурс] / Ludwig Wittgenstein // Basil Blackwell. – 1958. – Режим доступа до ресурсу: <https://static1.squarespace.com/static/54889e73e4b0a2c1f9891289/t/564b61a4e4b04eca59c4d232/1447780772744/Ludwig.Wittgenstein.-.Philosophical.Investigations.pdf>.