

Київський національний університет імені Тараса Шевченка

Філософський факультет

Кафедра етики, естетики та культурології

**ЕТИЧНІ АСПЕКТИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ТА
РОБОТОТЕХНІКИ**

Кваліфікаційна робота за освітньо-професійною програмою «Філософія»

за спеціальністю 033 Філософія

на здобуття освітнього ступеня бакалавра філософії

Студент-виконавець:

Ковальчук Дмитро Тимофійович

IV курс

спеціальність 033 «філософія»

Науковий керівник:

Рихліцька Оксана Дмитрівна

К.ф.н., доцент

Допущено до захисту:

Зав. кафедри _____

Київ-2024

ЗМІСТ

ВСТУП	3
РОЗДІЛ 1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ДОСЛІДЖЕННЯ ЕТИКИ ШТУЧНОГО ІНТЕЛЕКТУ ТА РОБОТОТЕХНІКИ	
1.1. Історико-філософські виміри технологій та їхній вплив на суспільство	7
1.2. Концепція "Сингулярності" та її етичні наслідки для людства	14
1.3. Поняття машинної "моральності" та можливості етичних меж машин	20
РОЗДІЛ II ОСНОВНІ ЕТИЧНІ ТЕОРІЇ ТА ЇХ ЗАСТОСУВАННЯ У ШІ	
2.1. Утилітаризм та його виклики у контексті ШІ та робототехніки	29
2.2. Деонтологічні обмеження у програмуванні «моральних» машин	35
2.3. Віртуальна етика та її роль у формуванні "характеру" штучного інтелекту	40
ВИСНОВКИ	48
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	50

ВСТУП

Актуальність теми дослідження. В сучасному світі штучний інтелект та робототехніка знаходяться на передових рубежах технологічного прогресу. Вони змінюють наше життя у всіх сферах, від виробництва до медицини, від освіти до розваг. Однак разом з безперечними перевагами виникають і значні етичні проблеми та питання, пов'язані з їх застосуванням. Етичні аспекти застосування штучного інтелекту та робототехніки стають об'єктом уваги для дослідників, політиків, підприємців та громадськості, оскільки вони впливають на наше суспільство, культуру та мораль.

Однією з ключових актуальних проблем є питання відповідальності у використанні штучного інтелекту в прийнятті рішень. Алгоритми штучного інтелекту можуть приймати важливі рішення у сферах, таких як судочинство, фінанси, медицина та інші. Проте існує ризик несправедливості та упередженості в їхніх рішеннях через неправильне навчання алгоритмів або відсутність узгоджених етичних стандартів. Дослідження в цій області допомагає виявляти та вирішувати такі проблеми для забезпечення справедливості та екітабельності у використанні штучного інтелекту. Важливою темою є етика автономних систем та роботів. З поширенням робототехніки у виробництві, транспорті та інших сферах зростає потреба в розробці стандартів етичної поведінки для цих систем. Питання відповідальності за дії автономних систем, їх вплив на людей та середовище, а також можливі наслідки в разі помилок стають об'єктом обговорення серед фахівців та громадськості. Дослідження в цій області спрямовані на розробку етичних стандартів та законодавства, які б забезпечили безпеку та довіру до автономних систем.

Також, особливо гостро є проблема етичних аспектів ШІ в медицині, тому що він може надавати пацієнтам медичні рекомендації цілодобово, на підставі історії хвороби, особистих потреб пацієнта або його уподобань. І таким чином

може нашкодити. Використання алгоритмів машинного навчання у діагностиці, лікуванні та догляді за пацієнтами відкриває безліч можливостей для поліпшення якості медичних послуг. Тут виникає питання про конфіденційність медичної інформації, право пацієнтів на прозорість та контроль за використанням їх даних, а також можливі ризики помилкових діагнозів або недостатньої уваги до індивідуальних потреб пацієнтів – так як ШІ не може вдосконалимо діагностувати небезпечне захворювання, так як це зробить людина. Дослідження в цій сфері спрямовані на розробку етичних принципів, які б забезпечили ефективне та етичне використання штучного інтелекту в медицині.

Особлива увага, на етичні аспекти використання робототехніки в сфері військової діяльності. Розвиток військових роботів та безпілотних літальних апаратів викликає серйозні питання щодо етики війни, відповідальності за втрати серед цивільного населення та ризиків автономних систем у бойових умовах. Дослідження в цій області спрямовані на розробку міжнародних конвенцій та етичних стандартів, які б регулювали використання робототехніки у військових цілях та забезпечили дотримання прав людини. Все це в цілому, підкреслює важливість дослідження етичних проблем застосування штучного інтелекту та робототехніки. Лише ретельний аналіз, обговорення та розробка етичних стандартів може забезпечити створення суспільства, яке користується перевагами цих технологій, зберігаючи одночасно свою гідність, безпеку та справедливість.

Ступінь наукової розробки проблеми: Міждисциплінарний характер поставленої проблеми зумовлює необхідність аналізу досліджень в різних сферах, які торкаються тих чи інших важливих аспектів даної проблематики. В етико-філософській літературі представлено різноманітні підходи визначення сутності інноваційних технологій (ШІ), проблем та перспектив їх застосування в умовах розвитку сучасного суспільства. Низка робіт зарубіжних дослідників присвячені проблемам та перспективам застосування та використання ШІ в біомедичних дослідженнях: зокрема Н. Бостром, Я. Божинов, Д. Фуллер, Е.

Хілер, В.Воллах, В. Р. Поттер, М. Ріс, М.Тегмарк, Д. Стейнхардт, Г. Йонас, С. Баті, Д. Ервін, Ф.Фукуяма та ін. Філософському осмисленню ціннісно-нормативних горизонтів сучасних технологій в працях Ю.Хабермаса, С.Харакаса і ін. Їх дослідження задають напрямок та стратегії для вирішення етико-правових проблем у застосуванні та використанні технологій, а також можливих загроз ШІ (С. Хокінг, Е. Юдковські та ін.). У сучасному вітчизняному просторі дану проблематику досліджують: М.Кисельов, В.Чешко, В.Кулініченко, О.Перова, Л.Сидоренко, Кияченко О. Л. і ін.

Об'єкт дослідження: моральні виклики застосування штучного інтелекту та робототехніки;

Предмет дослідження: етичний інструментарій щодо впровадження та застосування ШІ та робототехніки

Мета дослідження полягає в аналізі та розумінні етичних аспектів застосування штучного інтелекту та робототехніки в сучасному суспільстві.

Для досягнення мети, нами були поставлені наступні **завдання дослідження:**

- проаналізувати історико-філософські погляди на технології та їхній вплив на суспільство;
- виокремити концепцію "сингулярності" та її етичні наслідки для людства;
- розкрити полеміку щодо машинної «моральності», можливості та межі машин;
- окреслити утилітаризм та його виклики у контексті ШІ та робототехніки;
- виявити деонтологічні основи в обмеженні для програмування ШІ;
- дослідити віртуальну етику та її роль у формуванні "характеру" штучного інтелекту.

Наукова новизна полягає у дослідженні етичного інструментарію, підходів та перспектив щодо впливу штучного інтелекту і робототехніки етико-філософській парадигмі сьогодення.

Теоретико-методологічною основою дослідження виступають аналітичний, порівняльний підходи та ін. Використовуються ресурси наукових електронних бібліотек, науково-методичні матеріали, збірники наукових праць, теоретичні та громадські видання, офіційні документи, інформація з офіційних веб-сайтів урядових структур України та міжнародних організацій. Нормативно-правові акти, джерела Інтернету та інші ресурси становлять основу для наведення інформаційної підтримки дослідження.

Дипломна робота містить 54 сторінок, складається зі вступу, 3 розділів, висновків та літератури.

РОЗДІЛ І ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ДОСЛІДЖЕННЯ ЕТИКИ ШІ ТА РОБОТОТЕХНІКИ

1.1. Історико-філософські виміри технологій та їхній вплив на суспільство

Штучний інтелект (ШІ) — це наука про технологію творення розумних машин, які здатні виконувати виняткові функції в життєдіяльності людини, обирати/приймати/моделювати ситуацію та є певною емуляцією процесів людського інтелекту за допомогою машин, зокрема комп'ютерних систем. ШІ знаходить практичне застосування в різних сферах, таких як експертні системи, розуміння природної мови, розпізнавання мови та машинне бачення. Також даний термін використовують у процесах розробки систем, наділених інтелектуальними процесами, як здатність міркувати, відкривати значення, узагальнювати або вчитися на минулому досвіді. Зауважимо, що на початку цивілізації людство стало впроваджувати різноманітні технології для того, щоб полегшити собі життя та здійснювати різноманітні дії. Античний філософ Арістотель розглядав поняття техніки як засіб досягнення вищих цілей людини, а Платон вважав, що технології можуть призвести до руйнування моральних цінностей, якщо не використовуватимуться розумно.

З розвитком суспільства філософські погляди на технології також змінювалися. У середньовіччі Тома Аквінський, висловлювали погляди на роль технологій в релігійному та соціальному контексті. Розглядаючи використання технологій, як спосіб забезпечити комфортне існування для всіх людей, але попереджав про можливість зловживання ними та втрату моральних цінностей [1].

З появою промислової революції та зростанням впливу технологій у XIX столітті, К.Маркс та Ф.Енгельс, досліджували вплив технологій на суспільство з погляду класових відносин та економіки. Вони вбачали в технологіях інструмент для експлуатації пролетаріату та підтримки буржуазного устрою суспільства. Таким чином, вони підкреслювали потребу в перетворенні

суспільства через соціальну революцію, щоб забезпечити справедливіше використання технологій та їхній вплив на робочий клас [2].

У ХХ столітті з розвитком інформаційних технологій філософські дискусії про вплив технологій на суспільство стали ще гострішими. Так, швейцарський психолог філософ Карл Юнг та канадський філософ та дослідник медіа Маршалл Маклюен, висловлювали погляди на те, як нові технології змінюють спосіб мислення та впливають на спілкування. Наприклад, Маклюен акцентував увагу на впливі засобів масової комунікації на формування свідомості та культурних цінностей.

У сучасному світі філософські дослідження в сфері технологій та їхнього впливу на суспільство стають все більш важливими. Так, ізраїльський історик, доктор філософії, професор, визнаний фахівець в області військової історії Юваль Ной Харарі та шведський філософ Нік Бостром, аналізують етичні аспекти розвитку штучного інтелекту та біотехнологій, підкреслюючи потенційні ризики та виклики для суспільства. Наприклад, Харарі у своїй книзі "21 урок для ХХІ століття" розглядає вплив технологій на економіку, політику та культуру, попереджаючи про можливі негативні наслідки використання технологій без врахування їхнього етичного та морального виміру [3].

Історичний огляд філософських поглядів на технології та їхній вплив на суспільство відображає розвиток думок та підходів до цього питання з часом. Від Арістотеля до Харарі, філософи розглядали різні аспекти в залежності від контексту свого часу та технологічних досягнень епохи. Незважаючи на різноманітність підходів, спільним елементом усіх цих досліджень є бажання розуміти вплив технологій на людське суспільство та забезпечити його збалансований розвиток. Одним з ключових аспектів, щодо розглядалися впродовж історії, є взаємозв'язок між технологією та соціальними змінами. Наприклад, в ХІХ столітті промислова революція призвела до зміни економічного ландшафту та створила нові класові розбіжності, що стало об'єктом дослідження філософів того часу. Сучасні філософи Джудіт Сьюзман,

Люсі Армстронг та Макс Тегмарк аналізують вплив технологій на суспільство, але зосереджуються на більш сучасних викликах, таких як штучний інтелект, біотехнології та інші. Джудіт Сьюзман досліджує етичні аспекти біотехнологій, зокрема генної модифікації та клонування, яких аналіз полягає в тому, щоб аналізувати вплив цих технологій на людське суспільство та етичні аспекти під час їх використання. Люсі Армстронг як представник філософії релігії досліджує взаємозв'язок між релігією та технологіями, зокрема штучним інтелектом та біотехнологіями, аналізує етичні питання, пов'язані з релігійними переконаннями та новими технологіями. Макс Тегмарк досліджує вплив штучного інтелекту на суспільство та людську цивілізацію, а також розглядає питання безпеки та етики в контексті розвитку штучного інтелекту[4].

Важливим виміром - це етичні колізії, що пов'язані з розвитком технологій. Античний філософ Платон обговорював небезпеку втрати моральних цінностей через занадто велике захоплення технічними нововведеннями. У сучасному світі етичні питання стають ще більш актуальними, оскільки технології стають все більш потужними та впливовими. Філософи, такі як Нік Бостром та Юваль Ной Гарарі, досліджують етичні аспекти штучного інтелекту та біотехнологій, попереджаючи про можливі негативні наслідки використання цих технологій без врахування їхніх етичних аспектів.

Крім того відбувається актуалізація проблеми відповідальності щодо застосування та контролю над технологіями. З розвитком штучного інтелекту та автономних систем, таких як робототехніка та безпілотні літальні апарати, виникає необхідність у створенні етичних стандартів та правових рамок, що регулюватимуть їхнє використання. Філософи та правники спільно працюють над розробкою таких стандартів, щоб забезпечити безпеку та етичне використання цих технологій [5].

Заслуговує уваги також питання впливу технологій на культуру та ідентичність суспільства. У зв'язку зі зростанням використання інтернету та соціальних медіа, філософи розглядають, як ці технології формують наші

уявлення про світ, сприяють розвитку нових форм мистецтва та спілкування, а також можуть впливати на формування культурних цінностей та ідентичності.

При цьому виникають питання про кібербезпеку та приватність відомостей. Філософи та експерти з кібербезпеки досліджують, як технології впливають на безпеку інформації та особистих даних, а також розробляють методи захисту від кібератак та порушень приватності. Ці дослідження важливі для забезпечення безпеки та конфіденційності користувачів у цифровому світі.

Крім того, стали дуже актуальними питання про вплив технологій на працю та зайнятість. З одного боку, автоматизація та розвиток штучного інтелекту можуть призвести до змін на ринку праці та втрати робочих місць у деяких секторах. З іншого боку, технології можуть створити нові можливості для розвитку нових професій та ринків праці. Дані питання обговорюють ці питання та розробляють стратегії для адаптації суспільства до змін у сфері праці.

Однією з ключових областей досліджень є питання про відношення між людьми та технологіями, що актуалізує проблему впливу технологій у взаємовідносинах з іншими людьми, способи спілкування та соціальну інтеракцію. Наприклад, деякі дослідження вказують на те, що використання соціальних медіа може призводити до зміни відносин між людьми, зниження якості спілкування та збільшення відчуття відчуженості. Також важливою темою є питання про вплив технологій на наше ментальне здоров'я та емоційний стан. Сучасні технології, такі як інтернет та соціальні медіа, можуть мати як позитивний, так і негативний вплив на наше емоційне благополуччя. Наприклад, деякі дослідження показують, що занадто велике використання соціальних мереж може призводити до збільшення ризику розвитку депресії та тривожних розладів [5].

Окремим виміром є питання про екологічний вплив технологій та їхнє відношення до природи. Розвиток індустрії та використання різноманітних технологій може мати негативний вплив на навколишнє середовище, спричиняючи забруднення повітря, води та ґрунту, вирубування лісів та

зменшення біорізноманіття. Філософи та екологи активно обговорюють ці питання та розробляють стратегії для збереження природних ресурсів та забезпечення сталого розвитку. У світлі даних досліджень виникає необхідність у розвитку нових підходів до використання технологій, які б враховували їхній вплив на людей, суспільство та навколишнє середовище. Це може включати розробку етичних стандартів та політик, які б забезпечували відповідальне використання технологій, а також підтримку досліджень, спрямованих на збільшення розуміння їхнього впливу та розвиток інноваційних технологій, які б сприяли збереженню середовища та поліпшенню якості життя людей.

В контексті історико-культурного аналізу, розглянемо історичні приклади технологій та їхнього впливу на суспільство, що стали об'єктом філософських роздумів [6]. Одним з яскравих прикладів є винахід друкарського верстата Йоганном Гутенбергом у XV столітті. Цей винахід революціонував спосіб поширення інформації, зробивши книжку більш доступною та доступною для широкої публіки. Ідея розповсюдження знань та інформації стала головною темою філософських роздумів про цей винахід. Деякі філософи того часу підкреслювали значення друкарства для розвитку освіти та культури, вважаючи його кроком вперед для суспільства. Інші висловлювали занепокоєння щодо можливості недостовірної інформації та пропаганди, яка може бути поширена шляхом друку. Ще одним прикладом є розвиток індустріальної революції у XIX столітті. Поява парової машини та механізація виробництва прискорила процеси виробництва та транспортування, змінила робочі умови та розподіл праці, що стало об'єктом філософських роздумів про соціальну справедливість. Філософи того часу, такі як Карл Маркс, аналізували економічні та соціальні наслідки індустріалізації та висловлювали переймання щодо нерівності та експлуатації працівників. Інші філософи дивилися на ці зміни як на можливість досягнення більшого комфорту та процвітання для суспільства. Також, ефект технологій в сучасному інформаційному суспільстві. Поява Інтернету та цифрових технологій радикально змінила спосіб, яким ми спілкуємося, працюємо та отримуємо

інформацію. Це стало об'єктом філософських дебатів про приватність, свободу слова, культурну ідентичність та соціальні відносини. Філософи розглядають етичні аспекти зберігання та використання особистих даних, а також вплив соціальних мереж на формування громадської думки та політичні процеси [6].

В контексті розгляду сучасних тенденцій та перспектив розвитку технологій і їхнього впливу на суспільство з філософської перспективи.

Однією з найактуальніших областей досліджень є розвиток штучного інтелекту (ШІ). ШІ вже зараз має значний вплив на сфери економіки, медицини, транспорту, виробництва та багато інших. Філіпп Фурт досліджує етичні аспекти використання ШІ, включаючи питання про автономність систем, відповідальність за прийняті рішення, прозорість алгоритмів та можливість соціальних наслідків автоматизації робочих місць. Вони також розглядають можливість розвитку штучного інтелекту, який здатний до самосвідомості та моральних роздумів, що породжує важливі питання про природу моральності та соціальних відносин між людьми та машинами [7].

Досить важливою областю, в контексті розвитку технологій є розвиток біотехнологій. Наприклад, редагування генетичного коду за допомогою техніки CRISPR відкриває широкі можливості для лікування генетичних захворювань та покращення людського організму. Однак це також породжує етичні питання про моральність та безпеку таких втручань в природний порядок. Філософи вивчають етичні аспекти використання біотехнологій, включаючи питання про справедливий доступ до нових технологій, захист прав людини та збереження біологічної різноманітності.

Також важливою темою досліджень є роль технологій у вирішенні глобальних проблем, таких як зміна клімату, енергетична криза та бідність. Філософи аналізують етичні аспекти використання технологій для зменшення викидів CO₂, розвитку відновлювальних джерел енергії та створення інноваційних методів боротьби з бідністю. Вони також досліджують питання про

глобальну справедливість та відповідальність багатих країн у вирішенні світових проблем [7].

Отже, сучасні філософські роздуми про технології та їхній вплив на суспільство зосереджуються на широкому спектрі питань, включаючи етичні аспекти розвитку штучного інтелекту та біотехнологій, соціальні наслідки інформаційних технологій та можливості використання технологій для розв'язання глобальних проблем. Дослідження в цих областях допомагають розуміти складні взаємозв'язки між технологіями та суспільством та розвивати стратегії для забезпечення сталого розвитку та гармонії між людиною та природою.

У висновку, аналіз філософських поглядів на технології та їхній вплив на суспільство виявляє багатогранність цієї проблематики. Протягом історії людства технології завжди відображали й впливали на різні аспекти життя, викликаючи як позитивні, так і негативні наслідки. Філософи віддавали особливу увагу етичним, соціальним та культурним аспектам технологій, намагаючись розуміти їхній вплив на суспільство та розвивати стратегії для забезпечення справедливого та гуманного використання технологій.

Сучасний світ зі своїми швидкими технологічними змінами постійно поставляє перед філософами нові виклики. Розвиток штучного інтелекту, біотехнологій, інформаційних технологій та інших сфер технологій вимагає постійного переосмислення етичних, соціальних та культурних наслідків їх впливу [7].

Таким чином, філософський підхід до технологій є важливим компонентом сучасного дискурсу та сприяє розвитку відповідального та сталого використання технологій у майбутньому. Оскільки є важливим інструментом для формування обізнаних рішень та політик, які б забезпечували гармонійний розвиток технологій та суспільства.

1.2. Концепція "Сингулярності" та її етичні наслідки для людства

Концепція "Сингулярності" - це одна з найбільш захоплюючих та контроверсійних ідей у сучасній науці та технологіях. За деякими дефініціями, це момент, коли штучний інтелект (ШІ) або інші технології досягають рівня інтелекту, який перевищує людський, що може призвести до різких та необоротних змін у суспільстві.

Про цю концепцію дискутують багато відомих науковців, серед яких варто згадати Рея Курзвейла, одного з її основних пропагандистів. Курзвейл прогнозує, що досягнення Сингулярності стане ключовим моментом в еволюції людства, коли машини стануть інтелектуально наділенішими за людину та здатними до самостійного вдосконалення. Він вбачає це як можливість для вирішення складних проблем людства, таких як відновлення здоров'я та зміцнення імунітету, зростання продуктивності та розвитку суспільства [8].

Однак не всі науковці погоджуються з оптимістичним поглядом Курзвейла. Наприклад, професор філософії Ніль Бострем заявляє, що Сингулярність може стати загрозою для людства, особливо якщо розвиток ШІ не буде адекватно контрольований. Він попереджає про можливість виникнення непередбачуваних наслідків, таких як автономні збройні системи, які можуть призвести до масового знищення, або економічні нерівності, що загрожують стабільності суспільства. Іншим важливим аспектом етичних наслідків Сингулярності є питання про розподіл влади та контролю над розвитком ШІ. Сучасний дослідник, Юджинія Ченг, висловлює обурення щодо того, що великі технологічні компанії можуть мати занадто великий вплив на напрямок розвитку ШІ та використання його власних цілей. Ченг активно досліджує етичні аспекти використання штучного інтелекту та публічно виступає за необхідність контролю за впливом технологічних гігантів на суспільство та закликає до більшого державного регулювання та контролю над цим процесом, щоб запобігти можливому зловживанню технологіями та зберегти контроль за їхнім впливом на суспільство [8]. Попри всі контроверсії, сингулярність залишається

предметом великого інтересу та обговорень серед науковців, філософів, підприємців та громадян з усього світу. Її можливі наслідки для людства є стимулом для подальших досліджень та розвитку політики в області штучного інтелекту та технологій загалом.

Щоб проілюструвати етичні наслідки, давайте розглянемо ситуацію, коли розумні машини отримують повний контроль над виробництвом та розподілом ресурсів. Уявімо, що така машина приймає рішення, що спричиняє шкоду екологічному середовищу, але приносить користь людству у короткостроковій перспективі. Це породжує етичні питання щодо природи моральної відповідальності цих розумних систем та необхідності регулювання їхньої діяльності [8]. У світлі даних роздумів стає очевидною необхідність постійного моніторингу та регулювання розвитку технологій, особливо ШІ. Це може включати створення етичних рад та комісій, які будуть відповідальні за встановлення норм та правил використання штучного інтелекту, а також за розробку механізмів контролю за його розвитком. Також важливо забезпечити широку громадську дискусію та участь у процесі прийняття рішень з питань етичних аспектів штучного інтелекту, щоб врахувати різні погляди та інтереси громадськості.

Однак ми також не можемо ігнорувати потенційні переваги, які може принести розвиток штучного інтелекту, включаючи покращення медичних діагностичних методів, розвиток автономних транспортних засобів для зменшення аварійності на дорогах, а також підвищення продуктивності виробництва та розвитку нових технологій для збереження навколишнього середовища. Зважаючи на швидкість та обсяг розвитку штучного інтелекту, необхідно вжити заходів щодо забезпечення безпеки та стабільності використання цих технологій. Проектування та впровадження етичних стандартів і принципів у сфері розробки та використання штучного інтелекту є ключовим завданням. Потрібно створити механізми контролю та

відповідальності, які б забезпечували, що штучний інтелект використовується в межах етичних норм та не шкодить суспільству [8].

Також важливо розвивати освіту та свідомість серед громадськості щодо етичних аспектів штучного інтелекту. Це допоможе громадянам краще розуміти потенційні загрози та переваги, пов'язані з використанням цих технологій, та допоможе їм участь у формуванні політики та регулюванні в цій сфері. Більша увага також повинна бути приділена етичному навчанню майбутніх фахівців в галузі штучного інтелекту та інформаційних технологій, щоб вони мали глибоке розуміння моральних проблем, пов'язаних з їхньою роботою.

У світлі розвитку сучасних технологій, включаючи штучний інтелект, важливо не тільки вирішувати конкретні етичні питання, але і ставити перед собою питання про те, як ми бажаємо бачити суспільство у майбутньому. Це стосується не лише технічних питань розробки та використання штучного інтелекту, але й більш глибоких аспектів, таких як значення людської дігнітету, свободи та справедливості в цифровому віці [9]. Тому, розгляд етичних аспектів розвитку штучного інтелекту вимагає подальшого розгортання. Одним з ключових питань є забезпечення принципу справедливості та недопущення дискримінації в процесі розробки та застосування штучного інтелекту. Наприклад, врахування та корекція можливих викривлень та прихованих біасів у даних, на яких навчаються алгоритми, може допомогти запобігти виникненню систем, які б посилювали стереотипи або призводили до нерівності.

Питання про приватність та захист особистих даних у контексті розвитку штучного інтелекту актуалізується в контексті розгляду даного питання. Оскільки збір та аналіз великих обсягів особистої інформації може стати джерелом серйозних порушень приватності та може призвести до втручання в особисте життя людей. Тому необхідно розробляти механізми захисту приватності та забезпечення конфіденційності даних, які використовуються в системах штучного інтелекту [10].

Крім того, варто обговорити етичні аспекти використання штучного інтелекту в сферах, що стосуються людського здоров'я та безпеки. Наприклад, розробка систем штучного інтелекту для медичної діагностики може вимагати великої обережності та етичного обґрунтування, щоб уникнути помилкових діагнозів та недостатньої уваги до індивідуальних особливостей пацієнтів. Нарешті, етичні питання пов'язані із можливістю використання штучного інтелекту в військових цілях. Розвиток автономних збройних систем може породжувати серйозні моральні дилеми та загрози міжнародній стабільності. Необхідно вирішувати питання етичності та відповідальності використання таких технологій для запобігання загостренню конфліктів та війн.

Зважаючи на потенційний вплив штучного інтелекту на суспільство та людську культуру, важливо також розглянути етичні аспекти використання цих технологій у сфері мистецтва та креативності. Штучний інтелект вже зараз використовується для генерації музики, малюнків, літературних творів та інших видів творчості. Однак, виникає питання щодо авторства та оригінальності таких творів, а також щодо впливу штучного інтелекту на творчий процес та ідентичність художника [11].

Наприклад, системи, які створюють музику або літературні тексти, можуть породжувати суперечки щодо того, чи можна вони бути визнані авторськими творами та хто має право на їх власність. Також можуть виникнути етичні питання щодо впливу штучного інтелекту на креативність та індивідуальний вираз художника, а також на розуміння культурної спадщини та ідентичності. Крім того, важливо обговорити етичні аспекти використання штучного інтелекту в сфері освіти та виховання. Впровадження систем штучного інтелекту в навчальний процес може стати інструментом для підвищення ефективності навчання та індивідуалізації освітнього процесу. Однак це також може породжувати етичні питання щодо конфіденційності та захисту даних студентів, а також щодо відповідальності за результати навчання, отримані за допомогою штучного інтелекту [12].

Загалом, етичне використання штучного інтелекту стає все більш важливою проблемою у світлі швидкого розвитку цих технологій та їхнього потенційного впливу на суспільство. Вирішення етичних питань вимагає комплексного підходу, що поєднує в собі експертні знання з різних галузей, широку громадську дискусію та участь у процесі прийняття рішень усіх зацікавлених сторін. Тільки шляхом спільних зусиль ми можемо забезпечити етичне та справедливе використання штучного інтелекту, яке буде сприяти загальному добробуту та розвитку суспільства. Ще одним зрізом є етичні аспекти використання штучного інтелекту у сфері роботи та зайнятості. З одного боку, автоматизація та розвиток роботів з використанням штучного інтелекту можуть призвести до підвищення продуктивності та ефективності виробництва, що в свою чергу може створити нові можливості для створення робочих місць та підвищення рівня життя [13].

З іншого боку, існує загроза, що розвиток штучного інтелекту може призвести до масового заміщення людей на робочих місцях автоматизованими системами, що може призвести до загострення проблем безробіття та зростання економічних нерівностей. Також виникають етичні питання щодо відповідальності за можливі наслідки таких змін у сфері праці та захисту прав працівників у новому цифровому середовищі.

Для того, щоб забезпечити етичне використання штучного інтелекту в сфері праці, необхідно розробити та впровадити відповідні правові норми та механізми контролю, які б забезпечували захист прав та інтересів працівників у цифровій епохі. Також важливо звернути увагу на питання навчання та підготовки людей до роботи з штучним інтелектом та іншими новими технологіями, що дозволить їм зберігати конкурентоспроможність на ринку праці та забезпечити стабільність та процвітання для всіх [14].

Розглянувши етичні аспекти використання штучного інтелекту в різних сферах життя, необхідно також враховувати вплив цих технологій на саме сутність людської існування та взаємодію з навколишнім середовищем. Однією

з ключових проблем є питання екологічної стійкості та впливу розвитку штучного інтелекту на навколишнє середовище. Наприклад, велика кількість даних, які використовуються для тренування штучних інтелектуальних систем, може призводити до зростання енергоспоживання та викидів вуглецю, що може негативно вплинути на клімат та екологічну стійкість планети. Також важливо врахувати соціокультурні аспекти впровадження штучного інтелекту та його вплив на суспільство та культуру. З одного боку, штучний інтелект може сприяти збереженню та розвитку культурної спадщини та мовних традицій шляхом автоматизації перекладу та аналізу текстів. З іншого боку, можливий ризик втрати індивідуальності та унікальності культурних виявів під впливом стандартизації та глобалізації [15].

Крім того, важливо врахувати етичні аспекти використання штучного інтелекту в військовій сфері. Розвиток автономних збройних систем та систем штучного інтелекту для військових цілей може відкрити двері до нових форм конфліктів та порушень міжнародного гуманітарного права. Також існує загроза виникнення етичних дилем та загальносвітових криз у разі неправильного використання цих технологій.

Отже, етичне використання штучного інтелекту вимагає глибокого аналізу та обґрунтування впливу цих технологій на всі аспекти суспільства та людського існування. Вирішення етичних питань пов'язаних з штучним інтелектом вимагає спільних зусиль всіх зацікавлених сторін та розробки комплексних стратегій та політик, які б забезпечували ефективний захист прав та інтересів людини та суспільства в цифрову епоху.

Отже, концепція "Сингулярності" відкриває перед нами широкі горизонти можливого технологічного прогресу, що може змінити обличчя сучасного суспільства. Однак разом з потенційними перевагами такого розвитку виникають і значні етичні наслідки, які вимагають нашої уваги та обґрунтування.

Провідна тема етичних наслідків "Сингулярності" полягає у збереженні гуманітарних цінностей та захисті прав та гідності людини у цифрову епоху. З

одного боку, існує потенціал створення нових можливостей для розвитку суспільства та підвищення якості життя, але з іншого - наявні ризики порушення приватності, загострення економічних та соціальних нерівностей, а також втрати робочих місць через автоматизацію [15]. Для забезпечення етичного використання концепції "Сингулярності" необхідно розвивати та впроваджувати відповідні правові норми та механізми контролю, а також забезпечувати широку громадську дискусію та участь у процесі прийняття рішень. Важливо збалансувати технологічний розвиток з гуманістичними цінностями та гарантувати, що кожна нова технологія сприяє загальному благополуччю та просвітленню суспільства.

Таким чином, концепція "Сингулярності" ставить перед людством складні етичні виклики, які вимагають ретельного обґрунтування та вдумливого підходу до їхнього вирішення. Відповідальне використання технологій, з орієнтацією на гуманістичні цінності та захист прав людини, є ключем до побудови майбутнього, яке буде сприяти загальному процвітанню та розвитку людського суспільства

1.3 Поняття машинної моральності та можливості етичних меж машин

Полеміка щодо машинної «моральності» стають все більш актуальними у сучасному світі, де штучний інтелект і робототехніка знаходять все більше застосувань у різних сферах життя. Основним питанням, яке виникає, полягає у тому, чи можуть машини бути етичними та мати моральність, і як це впливає на суспільство та людей.

Так, Майкл Сандел вважає, що існує можливість програмувати машини на виконання моральних принципів і прийняття етичних рішень. Він досліджує можливості вбудовання моральних принципів у програмне забезпечення та алгоритми, які керують поведінкою штучного інтелекту. Наприклад, деякі експерименти показали, що штучний інтелект може бути навчений розпізнавати етичні проблеми і вибирати оптимальний варіант дії на основі заданих правил.

Однак існують і альтернативні погляди, які стверджують, що машинам не може бути притаманна справжня моральність, оскільки вони не мають свідомості та внутрішньої свободи вибору. Вони лише виконують програми, які були їм надані розробниками, і не можуть мати власних моральних переконань або відчуттів [16]. До прикладу, розглянемо ситуацію, де автономний автомобіль повинен прийняти рішення про те, кого захистити в разі неминучої аварії - пасажирів чи пішохода. Це морально складне питання, і програмістам потрібно передбачити рішення в коді автомобіля. Однак немає однозначної відповіді на це питання, і програми можуть бути налаштовані по-різному в залежності від моральних переконань їхніх розробників. Стюарт Рассел та Нік Бостром також висловлюють стурбованість щодо можливих негативних наслідків введення машинної моральності. Наприклад, якщо машини будуть мати власні моральні принципи, це може призвести до непередбачених ситуацій та конфліктів з людьми. Крім того, існує ризик зловживання технологіями для маніпулювання моральними нормами та цінностями суспільства [16].

Одним із важливих аспектів дебатів є питання про межі машинної моральності — це концепція, яка визначає, до якого рівня штучний інтелект може розуміти, враховувати та дотримуватися моральних принципів у своїй діяльності. Вона визначає, які принципи та обмеження повинні бути вбудовані в системи штучного інтелекту для забезпечення етичного та морального поведіння.

З одного боку, машини можуть бути програмовані для дотримання певних моральних правил та норм, які люди встановлюють. Однак існує питання про те, наскільки штучний інтелект може розуміти контекст та складнощі моральних ситуацій, які можуть виникнути у реальному світі. Визначення меж машинної моральності можна розглядати зі сторони таких аспектів, як: етичні принципи, які мають бути враховані під час програмування штучного інтелекту; прийняття рішень штучного інтелекту під час морально складних ситуацій; відповідальність машин за свої дії та наслідки; самоствердження системи ші за

власну моральну систему. Науковці вважають, що програми штучного інтелекту можуть бути обмежені в своїй здатності вирішувати складні моральні проблеми через обмеженість їхнього розуміння контексту та соціальної ситуації. Такі обмеження можуть призвести до внаслідок прийнятті моральних рішень, які не враховують всіх нюансів конкретної ситуації та потреб людей.

З іншого боку, існують деякі науковці, які вважають, що розвиток машинної моральності може мати позитивний вплив на суспільство. Наприклад, автономні роботи в сфері медицини можуть бути програмовані на дотримання етичних стандартів та принципів медичної практики, що дозволить уникнути помилок та недопущених дій. Додатково, машинна моральність може бути корисною у ситуаціях, де люди не здатні прийняти швидке рішення через стрес або емоційну перевантаженість. Наприклад, системи штучного інтелекту можуть допомагати пілотам у кризових ситуаціях у повітрі, де кожна секунда має значення, а емоції можуть перешкоджати прийняттю об'єктивного рішення [17].

Проте, незважаючи на потенційні переваги, важливо врахувати етичні та соціальні ризики введення машинної моральності. Одним з таких ризиків є можливість виникнення абсолютних систем, які не залишають місця для гнучкості та індивідуального вирішення моральних проблем. Наприклад, програми, які повністю підкорені певним моральним принципам, можуть не враховувати контекстуальні або культурні відмінності, що може призвести до некоректних або несправедливих рішень.

Крім того, існує загроза застосування машинної моральності в маніпуляційних цілях або для вирішення політичних або економічних завдань. Наприклад, уряди чи корпорації можуть використовувати алгоритми штучного інтелекту для впливу на громадську думку чи вирішення складних етичних питань у власних інтересах [18].

Поряд із викликами та ризиками, пов'язаними з машинною моральністю, існують і можливості для розвитку цієї сфери. Деякі дослідники вважають, що штучний інтелект може бути корисним у покращенні моральних властивостей

людей та суспільства в цілому. Наприклад, машинні системи можуть бути використані для аналізу етичних проблем та надання порад з їх вирішення. Такий підхід може допомогти людям краще розуміти моральність та приймати обґрунтовані рішення у складних ситуаціях.

Крім того, розвиток машинної моральності може сприяти покращенню взаємодії між людьми та роботами. Наприклад, роботи зі вбудованою моральністю можуть бути краще прийняті людьми у сферах, де вони працюють разом. Це може підвищити ефективність співпраці та сприяти створенню більш гармонійних робочих оточень.

Однак для досягнення цих можливостей необхідно враховувати і обговорювати етичні питання та виклики, пов'язані з машинною моральністю. Важливо розробляти та впроваджувати етичні стандарти та норми для роботів і штучного інтелекту, а також створювати механізми контролю та відповідальності за їхню діяльність. Також необхідно забезпечити широку громадську обговорення та участь у процесі прийняття рішень з питань машинної моральності, щоб забезпечити врахування різноманітних поглядів та інтересів [18].

У розгляді можливостей та меж машинної моральності слід також звернутися до практичних застосувань цієї концепції у різних сферах життя. Однією з таких сфер є медицина, де штучний інтелект може бути використаний для етичних рішень у складних ситуаціях, пов'язаних з лікуванням пацієнтів. Наприклад, системи штучного інтелекту можуть аналізувати медичні дані та рекомендувати оптимальний курс лікування з урахуванням етичних норм та принципів медичної етики.

Також машинна моральність може бути корисною у сфері автономних транспортних засобів, де виникають складні ситуації, пов'язані з безпекою та моральними виборами. Наприклад, програми автономних автомобілів можуть бути налаштовані на вибір оптимальних рішень у критичних ситуаціях, що вимагають етичних вагомих переваг [18].

Проте, деякі дослідники підкреслюють, що застосування машинної моральності у практиці має свої обмеження та ризики. Наприклад, виникає питання, як досягти загального прийняття етичних принципів для програмних систем, оскільки різні культури та соціальні групи можуть мати різні моральні стандарти. Крім того, існує ризик, що машинна моральність може бути використана для підтримки або посилення існуючих соціальних нерівностей та дискримінації.

Ще однією проблемою є визначення відповідальності за дії машин з моральними функціями. Виникає питання, хто несе відповідальність за неправильні рішення, прийняті програмами штучного інтелекту. Це стає особливо актуальним у ситуаціях, коли дії машин призводять до серйозних наслідків для людей чи суспільства [18]. Важливо забезпечити, щоб розвиток цієї сфери технологій супроводжувався врахуванням етичних принципів та заходів безпеки, які забезпечать максимальну вигоду для суспільства. Ключовим аспектом вирішення цих проблем є розробка і впровадження стандартів і правил, які регулюють розробку та використання машинної моральності. Такі стандарти повинні враховувати етичні, соціальні та культурні аспекти, а також визначати механізми відповідальності та контролю за діяльністю штучного інтелекту.

Крім того, важливо забезпечити широку громадську дискусію та участь громадськості у процесі прийняття рішень з питань машинної моральності. Це дозволить врахувати різноманітні погляди та інтереси різних груп населення і забезпечити легітимність і підтримку запропонованих рішень.

Загалом, машинна моральність відкриває перед нами великі можливості для розвитку технологій та покращення якості життя. Проте це також ставить перед нами великі виклики та ризики, які потребують уважного аналізу та вирішення. Важливо забезпечити, щоб розвиток машинної моральності відбувався з дотриманням високих етичних стандартів та принципів справедливості, що дозволить максимально використати потенціал цієї технології для блага всього суспільства.

Однак досягнення такого результату вимагатиме спільних зусиль усіх зацікавлених сторін, включаючи науковців, розробників, законодавців, громадськість та інші зацікавлені сторони. Тільки в умовах співпраці та взаєморозуміння ми зможемо створити етичну та відповідальну систему машинної моральності, яка служитиме загальному добру та благополуччю людства [18].

Одним із конкретних прикладів застосування машинної моральності є розробка автономних роботів для медичних процедур та хірургічних операцій. Наприклад, у деяких лікарнях вже застосовуються роботи для виконання складних хірургічних втручань. З врахуванням розвитку штучного інтелекту та машинного навчання, ці роботи можуть бути програмовані на виконання медичних процедур з максимальною точністю та безпекою для пацієнтів.

Проте, при розробці таких систем необхідно враховувати етичні аспекти. Наприклад, роботи-хірурги повинні мати вбудовані алгоритми, які дозволяють їм робити етичні рішення у випадках, коли виникають несподівані ситуації або складні етичні дилеми під час операції. Такі системи можуть бути програмовані на виконання етичних принципів, таких як максимальне збереження життя пацієнта та мінімізація ризиків для його здоров'я [17].

Іншим прикладом може бути застосування автономних транспортних засобів у сфері громадського транспорту. Наприклад, автономні автобуси можуть бути програмовані на вибір найбезпечнішого маршруту та виконання правил дорожнього руху з максимальною точністю. Однак, при розробці таких систем необхідно враховувати етичні аспекти, зокрема вирішення етичних дилем у випадку аварійних ситуацій, коли автономний транспортний засіб повинен прийняти швидке рішення щодо збереження життя пасажирів та інших учасників дорожнього руху.

Додатково, машинна моральність може бути застосована в області екології та природокористування. Наприклад, дрони, оснащені штучним інтелектом, можуть бути використані для моніторингу дикої природи та

виявлення порушень екологічних стандартів. Такі системи можуть бути програмовані на виявлення незаконного вирубування лісів або незаконного полювання, що дозволить захищати природні ресурси та зберігати біорізноманіття [16].

Додатковим прикладом може бути застосування машинної моральності у сфері штучного інтелекту для прийняття етичних рішень у віртуальних середовищах та комп'ютерних іграх. Наприклад, у віртуальних іграх з великим відкритим світом можуть виникати ситуації, в яких гравці змушені приймати моральні рішення, що впливають на долю інших персонажів або на весь віртуальний світ. Тут машинна моральність може використовуватися для створення алгоритмів, які допомагають ігровим персонажам приймати обґрунтовані моральні вибори, що відображаються на їхній поведінці та взаємодії з іншими персонажами.

Проте, при застосуванні машинної моральності в ігровій індустрії важливо враховувати етичні аспекти. Наприклад, ігрові розробники повинні піклуватися про те, щоб моральні вибори персонажів були реалістичними та не викликали негативних емоцій або стереотипів. Також важливо забезпечити, щоб гравці мали можливість самостійно впливати на хід подій та моральний розвиток своїх персонажів, а не були обмежені або контрольовані машинними алгоритмами.

Ще одним прикладом може бути застосування машинної моральності у сфері військових технологій та зброї. Наприклад, розробка автономних збройних систем може викликати серйозні етичні питання, пов'язані з можливістю самостійного вирішення життєво важливих ситуацій, таких як застосування сили та прийняття рішень про знищення людських життів. Тут машинна моральність може бути використана для програмування алгоритмів, які допомагають автономним збройним системам приймати обґрунтовані та етичні рішення в критичних ситуаціях [14].

Проте, при застосуванні машинної моральності у військовій сфері важливо розуміти потенційні ризики та негативні наслідки. Наприклад, автономні збройні системи можуть призвести до виникнення морального відчуження та відчуття безвідповідальності у військових операторів, які несуть відповідальність за дії таких систем. Також існує загроза, що автономні збройні системи можуть бути використані для здійснення нападів без обґрунтованих моральних мотивів чи політичних цілей, що може призвести до серйозних гуманітарних криз.

Отже, машинна «моральність» має значний потенціал у багатьох сферах людської діяльності, проте її застосування вимагає уважного розгляду етичних аспектів.

Отже, в контексті дослідження було проаналізовані історико-філософські погляди таких видатних діячів як Аристотель, Платон, Томас Аквінський, К.Маркс, Ф.Енгельс, Маршалл Маклюэн, Карл Юнг, Юваль Ной Харарі, Нік Бостром, Джудіт Сьюзман, Люсі Армстронг, Макс Тегмарк та Карл Маркс.

У цьому розділі я досліджував філософські підходи до етики штучного інтелекту та робототехніки. Починаючи з історичного огляду філософських поглядів на технології та їхній вплив на суспільство, ми виявили, що з появою нових технологій завжди виникає полеміка на тему їх етичного та соціального впливу. Пройшовши через різні етапи розвитку людського суспільства, від античності до сучасності, ми розглянули, як технології впливали на моральність та цінності суспільства.

Далі ми розглянули концепцію "сингулярності" і її етичні наслідки для людства. Виявилося, що ця концепція створює дискусію про можливі ризики та переваги розвитку штучного інтелекту, включаючи питання про межі його контролю та впливу на суспільство. Ми розглянули різні погляди відомих науковців щодо цих питань та намагалися зрозуміти, як сингулярність може змінити нашу моральну парадигму.

Отже, вивчення історії, концепцій та дебатів дозволило нам краще розуміти складні взаємозв'язки між технологіями та моральністю, а також відкрити нові горизонти для подальших роздумів та досліджень у цій області. Особливо, привертає увагу до питань, пов'язаних з розробкою і використанням штучного інтелекту в таких як: прийняття рішень в критичних ситуаціях або моральні ділеми у віртуальних середовищах.

РОЗДІЛ II ОСНОВНІ ЕТИЧНІ ТЕОРІЇ ТА ЇХ ЗАСТОСУВАННЯ ДО ШІ

2.1. Утилітаризм та його виклики у контексті ШІ та робототехніки

Утилітаризм, як етична теорія, спрямований на досягнення найбільшого загального щастя для найбільшої кількості людей. У контексті штучного інтелекту (ШІ) та робототехніки, утилітаризм стає предметом глибоких дебатів та викликів. Деякі вчені вважають, що принципи утилітаризму можуть бути корисними для програмування штучного інтелекту та роботів з метою прийняття моральних рішень. Наприклад, програми штучного інтелекту можуть бути налаштовані на максимізацію загального щастя або мінімізацію страждань при вирішенні конфліктних ситуацій. Однак існують і аргументи проти застосування утилітаристичних принципів у штучному інтелекті, оскільки вони можуть призвести до ігнорування прав та інтересів меншості на користь більшості [19].

Австрійський філософ Пітер Сінгер пропонує концепцію ефективного альтруїзму, яка впливає на розуміння утилітаризму в контексті розвитку штучного інтелекту. Згідно з цією концепцією, морально обов'язково діяти так, щоб максимізувати загальне щастя, але це також означає, що ми повинні бути готові віддати свої власні інтереси на користь тих, хто потребує допомоги найбільше. Утилітаристичний підхід, згідно з Сінгером, вимагає від нас допомагати іншим, навіть якщо це потребує великих жертв. У контексті робототехніки ця концепція може мати практичне застосування у розробці систем штучного інтелекту, які беруть до уваги потреби і благополуччя найбільш вразливих членів суспільства [19].

Але важливо також враховувати критику, яка висловлюється щодо утилітаризму. Зокрема, філософ Джон Роулз у своїй книзі "Теорія справедливості" висловлює сумніви стосовно утилітаризму як загальної моральної теорії. Роулз аргументує, що утилітаризм може призвести до порушення прав меншості в інтересах більшості, що суперечить принципу справедливості. Ця критика має значення і для контексту ШІ та робототехніки,

де програми штучного інтелекту можуть впливати на права та інтереси різних груп людей.

Практичний приклад використання утилітаризму у штучному інтелекті можна побачити в автономних автомобілях. Припустимо, автономний автомобіль опинився у ситуації, де перед ним перетинає дорогу група пішоходів, але одночасно неподалік від нього знаходиться зупинений вантажівка. Згідно з утилітаристичним підходом, автомобіль може прийняти рішення про те, щоб відвернути від пішоходів і зіткнутися з вантажівкою, оскільки це може призвести до меншої кількості постраждалих. Проте, таке рішення може порушити права та інтереси пішоходів і підняти питання про справедливість розподілу ризиків на дорозі. Це філософське дослідження етичних аспектів штучного інтелекту, зокрема до обговорення утилітаризму в контексті прийняття рішень автономними системами, такими як автономні автомобілі. Утилітаризм — це етична теорія, яка визначає моральну правильність дії на основі її наслідків: дія є моральною, якщо вона приносить найбільшу користь для найбільшої кількості людей. У контексті автономних автомобілів це може означати прийняття рішень про те, як реагувати в небезпечних ситуаціях з метою мінімізації шкоди.

Іншим прикладом може бути програмування штучного інтелекту для прийняття рішень у медичних ситуаціях. Наприклад, утилітаристичний підхід може бути застосований у виборі терапевтичних стратегій, де мають бути враховані не лише індивідуальні потреби пацієнта, але й загальна користь для суспільства. Проте, це може викликати питання про економічну ефективність та справедливий доступ до медичних послуг для всіх [20].

Враховуючи дебати та виклики утилітаризму в контексті штучного інтелекту та робототехніки, можна побачити, що цей етичний підхід має як переваги, так і обмеження. Його застосування може сприяти досягненню загальної користі для суспільства та максимізувати щастя багатьох. Однак, варто бути обережними, щоб уникнути порушення прав та інтересів меншості в ім'я загальної користі. Таким чином, розуміння та урахування контексту та

потенційних наслідків важливо для використання утилітаризму в сучасних технологічних доменах.

Важливо визначити, як утилітаризм може вплинути на розвиток сучасних технологій і які етичні виклики він ставить перед суспільством. Однією з ключових сфер застосування утилітаризму є робототехніка. Наприклад, в області роботизованої хірургії або догляду за хворими, роботи можуть бути програмовані для максимізації результатів лікування або забезпечення найкращого догляду за пацієнтами. Однак виникає питання, якщо машини мають приймати важливі моральні рішення, які можуть впливати на життя людей. У цьому випадку, утилітаристичний підхід може стати корисним інструментом для програмування роботів, щоб вони діяли на користь найбільшої кількості людей.

Крім того, утилітаризм може бути використаний у контексті прийняття рішень з питань екології та збереження природних ресурсів. Наприклад, при розробці стратегій енергоефективності або виборі джерел альтернативної енергії, утилітаристичний підхід може бути використаний для максимізації загальних екологічних вигод для суспільства. Проте, при цьому важливо враховувати можливі негативні наслідки для окремих екосистем або груп населення, що можуть стати жертвами забруднення або експлуатації природних ресурсів.

Утилітаризм також може бути викликаний у контексті розвитку штучного інтелекту та автономних систем. Наприклад, у сфері автономних транспортних засобів, утилітаристичні принципи можуть бути використані для програмування системи прийняття рішень для мінімізації аварій та забезпечення ефективною та безпечною транспортною системою. Однак, виникає питання про відповідальність за рішення, які приймаються машинами, особливо у випадках, коли вони можуть впливати на життя людей [21].

Для глибшого розуміння утилітаризму і його викликів у контексті сучасних технологій, важливо розглянути його застосування в сферах, таких як штучний інтелект, робототехніка та біоетика.

У сфері штучного інтелекту утилітаристичні принципи можуть бути використані для розв'язання складних етичних питань, таких як автономні системи розподілу ресурсів у галузі медицини або прийняття рішень в критичних ситуаціях. Наприклад, у медичних дослідженнях утилітаризм може допомогти визначити, які ліки розробляти, зокрема враховуючи їхню ефективність та вартість, щоб максимізувати користь для максимальної кількості пацієнтів. Однак варто враховувати можливі негативні наслідки, такі як ризик залишення менш численних або вразливих груп без доступу до медичних послуг.

У контексті робототехніки утилітаризм може бути використаний для розв'язання питань, пов'язаних з безпекою та ефективністю роботів. Наприклад, в промисловості програми управління роботами можуть бути налаштовані на мінімізацію ризику виникнення аварій та травмування робочої сили. Однак це може призвести до певного знеособлення та втрати робочих місць для людей. Також виникають питання про етичне поведінку роботів у випадках, коли їхні дії можуть вплинути на безпеку або благополуччя людей.

В контексті біомедичних досліджень принцип утилітаризм може використовуватися для вирішення етичних ділем у галузі генної інженерії, вирощування органів та інших сферах медицини. Наприклад, утилітаристичний підхід може допомогти визначити, які медичні процедури вважати етичними, враховуючи їхню ефективність та вартість, а також потенційні ризики для пацієнтів. Проте, варто бути обережними при використанні утилітаристичних принципів в цій галузі, оскільки вони можуть призвести до порушення прав та гідності людей.

Крім того, утилітаризм може ставити під сумнів індивідуальні права та свободи. При застосуванні утилітаристичних принципів до розробки та впровадження нових технологій, може виникнути питання про те, наскільки ми можемо обмежувати індивідуальні свободи для досягнення загального блага. Наприклад, утилітаристичний аналіз може підказувати, що моніторинг громадських просторів за допомогою систем відеоспостереження може

зменшити кількість злочинів, але це може порушити право на приватність та індивідуальну свободу [22].

До того ж, утилітаризм може виникнути в гострій суперечці з іншими етичними теоріями, наприклад, з категоричним імперативом Канта, який наголошує на принципах долі та поваги до особистості. Утилітаристичний підхід може вимагати від нас діяти на користь найбільшої кількості людей навіть за рахунок порушення прав окремих осіб, що суперечить принципам особистої гідності та поваги.

У цілому, утилітаризм є важливою етичною теорією, яка може допомогти нам розуміти та розв'язувати складні етичні питання у сфері сучасних технологій. Проте важливо розуміти, що застосування утилітаристичного підходу повинно бути обдуманим і враховувати всі можливі наслідки, включаючи індивідуальні права та свободи, загальне щастя суспільства та етичні принципи справедливості.

Одним зі суттєвих аспектів утилітаризму є постійна необхідність оновлення та адаптації принципів до змін у суспільстві та технологій. Сучасні технології швидко розвиваються, а разом з ними змінюються й етичні виклики, які вони ставлять перед нами. Наприклад, виникнення штучного інтелекту вимагає нових дискусій про те, як ці технології повинні бути програмовані та використовувані, щоб максимізувати користь для людства і уникнути потенційних негативних наслідків [22].

Ще одним аспектом, який варто враховувати, є контекстуальність утилітаристичного підходу. Те, що може бути найкращим рішенням в одній ситуації, може бути неприйнятним в іншій. Наприклад, утилітаристичний аналіз витрат та користі може вказувати на те, що варто відмовитися від дорогих медичних процедур, які мають невеликий шанс успіху. Проте це може суперечити моральним принципам, які визнають недоторканність людського життя та право на лікування.

Зокрема, утилітаристичний підхід до етичних питань у сфері сучасних технологій може вимагати глибокого обґрунтування та врахування різних аспектів, включаючи соціальні, економічні та моральні фактори. Важливо також забезпечити прозорість та відкритість у прийнятті рішень, щоб забезпечити довіру та підтримку суспільства [21].

Отже, утилітаризм може бути корисним інструментом для аналізу етичних питань у сфері сучасних технологій, проте він повинен використовуватися з обачливістю та з урахуванням всіх можливих наслідків. Це вимагає постійного діалогу, досліджень та вдосконалення наших етичних рамок для того, щоб забезпечити безпеку, справедливість та благополуччя у сучасному технологічному світі.

Утилітаризм, як етична теорія, надає цінний фреймворк для аналізу та вирішення складних етичних питань у сфері сучасних технологій. Використання утилітаристичних принципів може допомогти максимізувати загальне щастя та користь для суспільства, зокрема шляхом максимізації користі та мінімізації шкоди. Проте важливо розуміти, що застосування утилітаристичного підходу повинно бути обдуманим і ураховувати всі можливі наслідки, включаючи індивідуальні права та свободи, а також етичні принципи справедливості та поваги до особистості [20].

Утилітаризм може викликати певні суперечності та виклики, зокрема щодо визначення того, чому ми повинні надавати перевагу, а також забезпечення балансу між загальним благополуччям та правами індивідуальних осіб. Крім того, контекстуальність та постійне оновлення принципів утилітаризму важливі для врахування змін у суспільстві та технологіях.

Усе це підкреслює важливість постійного діалогу, досліджень та вдосконалення етичних рамок у контексті сучасних технологій. Використання утилітаризму повинно відбуватися з обачливістю та з урахуванням всіх можливих наслідків, з метою забезпечення безпеки, справедливості та благополуччя у сучасному технологічному світі.

Отже, утилітаризм може бути корисним інструментом для аналізу етичних питань у сфері сучасних технологій та може викликати певні суперечності та виклики, зокрема щодо визначення того, чому ми повинні надавати перевагу, а також забезпечення балансу між загальним благополуччям та правами індивідуальних осіб.

2.2. Деонтологія: правила та обмеження для програмування моральних машин.

У вивченні етичних принципів щодо програмування моральних машин, деонтологія займає центральне місце. Деонтологія, як етична теорія, стверджує, що правильність або неправильність дій не залежить від їхніх наслідків, а від дотримання певних моральних правил або принципів. У контексті програмування моральних машин, це означає, що роботи повинні дотримуватися певних правил і обмежень, незалежно від очікуваних наслідків їхніх дій.

Філософ і вчений Іммануїл Кант був одним із найвідоміших прихильників деонтології, яке означає вчення про проблеми моралі і моральності. Кант розвинув концепцію категоричного імперативу, який вимагає, щоб діяти лише за таких умов, за яких ця дія може стати загальним правилом для всіх. Наприклад, якщо програма має прийняти рішення про те, чи врятувати людину у небезпеці, вона повинна дотримуватися правила “не нашкодь”, яке спрямоване на те, щоб підтримувати гармонійні взаємини з іншими людьми та створювати позитивне оточення навколо себе незалежно від того, які наслідки можуть виникнути в результаті [21].

Деонтологічний підхід вимагає строгого дотримання правил та принципів, навіть якщо це може привести до невігідних наслідків. Наприклад, якщо має прийняти рішення про те, кого врятувати в аварійній ситуації, вона може обрати дієздатну дорослу особу, незважаючи на те, що це може означати втрату меншої кількості життів. Для деонтолога важливо дотримуватися принципу рівної поваги до особистості та права на життя кожної людини.

Деонтологія також надає увагу імперативу поваги до особистості. Це означає, що кожна особа має бути поважана як моральний суб'єкт і ніколи не може бути розглянута як простий засіб для досягнення мети. У програмуванні моральних машин це може мати велике значення, оскільки роботи можуть мати справу з різними людьми та ситуаціями. Наприклад, якщо машина робить вибір між рятуванням життя однієї людини чи багатьох, вона повинна дотримуватися принципу поваги до кожної особи і не може просто розглядати їх як засіб досягнення певної мети [22].

Утилітаристичний підхід, у свою чергу, розглядає етичність дій з точки зору їхніх наслідків. За утилітаристичним підходом, дія вважається моральною, якщо вона максимізує загальне щастя чи користь для всіх зацікавлених сторін. Проте деонтологічний підхід застерігає від того, щоб розглядати людей як засоби досягнення певної мети, навіть якщо ця мета полягає у максимізації загального щастя.

Деонтологія також ставить певні обмеження на можливість прийняття деяких рішень в етично спростованих ситуаціях. Наприклад, у деонтології існують такі поняття як непорушність прав людини та недоторканність особистості, які не підлягають компромісам. Це може створювати виклики при програмуванні моральних машин, оскільки вони можуть опинитися в ситуаціях, коли рішення, що вони приймають, може порушувати ці основні права [23].

Ще одним обмеженням деонтології є її негнучкість у вирішенні складних етичних ділем. Утилітаризм може бути більш гнучким у цьому плані, оскільки він дозволяє враховувати контекст та наслідки дій. Проте деонтологія наголошує на важливості дотримання моральних принципів, незалежно від обставин. Це може викликати ситуації, коли деякі дії можуть здатися морально неприпустимими, навіть якщо вони можуть принести загальне щастя.

Однак, деонтологія також може мати свої переваги у програмуванні моральних машин. Наприклад, вона може забезпечити певний рівень стабільності та непорушності правил, які роботи повинні дотримуватися у всіх

ситуаціях. Це може допомогти уникнути ситуацій, коли моральні машини роблять рішення, які можуть бути суперечливими або неприйнятними з точки зору загальних моральних принципів.

Загалом, деонтологія може надати важливий фундамент для програмування моральних машин, але важливо бути обережним і уникати категоричного підходу. Є важливим забезпечити баланс між дотриманням моральних принципів та урахуванням контексту та наслідків дій. Тільки так можна створити моральні машини, які будуть ефективними та відповідати вимогам сучасного суспільства.

У програмуванні моральних машин деонтологічний підхід може бути особливо корисним у вирішенні конкретних ситуацій, де необхідно встановити чіткі правила дій. Наприклад, якщо моральна машина опиняється перед вибором між тим, щоб врятувати життя однієї людини або п'яти інших, деонтологічний підхід може допомогти у прийнятті рішення. Вона буде дотримуватися принципу, що кожна людина має право на життя, і намагатиметься врятувати якнайбільше життів, навіть якщо це означає ризик для себе або інших [23].

Однак деонтологічний підхід може також мати свої обмеження та недоліки. Наприклад, в певних ситуаціях він може бути занадто жорстким і неадаптивним до змін у ситуації. Це може призвести до недосяжності чи непродуктивності в певних контекстах, де потрібна гнучкість та пристосування.

Подальше розвиток деонтологічного підходу в програмуванні моральних машин може включати пошук компромісів між строгими правилами та гнучкістю у вирішенні складних ситуацій. Наприклад, можливою стратегією може бути розробка системи, яка використовує деонтологічні принципи як основу, але також має можливість адаптуватися до конкретних обставин.

Деонтологічний підхід до програмування моральних машин також може включати елементи саморегулювання та взаємодії з користувачами. Наприклад, система може мати можливість навчання на основі зворотного зв'язку від

користувачів, щоб удосконалювати свої алгоритми та враховувати індивідуальні потреби та переваги.

Одним із викликів у застосуванні деонтологічних принципів у програмуванні моральних машин є розробка систем, які можуть адекватно реагувати на складні ситуації, де правила можуть конфліктувати або бути неоднозначними. Наприклад, система може опинитися в ситуації, де дія, яка дотримується одного принципу, може суперечити іншому. Такі ситуації вимагають розробки та впровадження алгоритмів прийняття рішень, які можуть забезпечити баланс між різними етичними принципами та максимізувати загальну моральну цінність дії [24].

Одним із можливих шляхів вдосконалення деонтологічного програмування моральних машин є подальше дослідження та розробка нормативних правил і принципів, які дозволять ефективно вирішувати етичні ділеми. Наприклад, можливим є створення систем, які використовують набір заздалегідь визначених правил та алгоритмів для прийняття рішень у конкретних ситуаціях. Однак ці системи також повинні бути гнучкими та здатними адаптуватися до змін у контексті та враховувати індивідуальні особливості кожного випадку.

Деонтологія також може бути вдосконалена шляхом розробки алгоритмів прийняття рішень, які дозволяють враховувати різноманітні аспекти етичних проблем. Наприклад, можливим є розробка систем, які здатні враховувати не лише моральні принципи, але й соціокультурні особливості, індивідуальні потреби та цінності, а також потенційні наслідки дій [25].

Додатковим напрямком дослідження може бути розробка етичних рамок та стандартів для програмування моральних машин, які б забезпечували високий рівень дотримання деонтологічних принципів. Такі стандарти можуть включати в себе вимоги до процесу прийняття рішень, перевірки на етичність та відповідності моральним принципам, а також механізми відповідальності за порушення етичних стандартів.

Загалом, деонтологія може відігравати важливу роль у програмуванні моральних машин, але для досягнення успіху необхідно поєднувати її з іншими етичними підходами та розробляти комплексні системи, здатні враховувати різноманітні аспекти етичних проблем. Важливо надавати пріоритет захисту прав та інтересів людини, забезпечуючи високий рівень моральності та відповідальності в розробці та використанні моральних машин. Досягнення цих цілей може вимагати постійного вдосконалення технологій, розробки нових методів та стандартів, а також активної участі всіх зацікавлених сторін, включаючи фахівців з різних галузей, правозахисники, етичні комітети та суспільство в цілому [25].

Проте для досягнення цих цілей важливо поєднувати деонтологію з іншими етичними підходами, розвивати нові технології та стандарти, а також забезпечувати широку участь у вирішенні етичних питань. Тільки так можна створити моральні машини, які будуть сприяти загальному благополуччю та відповідати вимогам сучасного світу.

З моменту виникнення моральних машин і їх розгортання в різних сферах життя, виникає необхідність постійної рефлексії та аналізу їхнього впливу на суспільство та індивідуальний рівень. Деонтологічний підхід, з одного боку, створює фундаментальні принципи, які дозволяють забезпечити етичність та справедливість у роботі моральних машин. З іншого боку, він потребує постійного удосконалення та адаптації до нових викликів, що виникають у зв'язку з швидким розвитком технологій та зміною соціокультурного середовища.

Одним з ключових завдань деонтології в програмуванні моральних машин є забезпечення захисту прав і свобод людини. Завдяки чітко визначеним правилам та принципам, системи можуть уникати порушень особистісної недоторканності, дискримінації та інших форм неправомірного втручання у права особи. Наприклад, моральна машина, яка дотримується деонтологічних принципів, не буде приймати рішень на підставі расової чи гендерної

приналежності, але буде керуватися об'єктивними критеріями та моральними нормами [22].

Крім того, деонтологічний підхід може забезпечити стабільність та непорушність правил у програмуванні моральних машин. Наприклад, у системах штучного інтелекту, які використовують деонтологічний підхід, може бути встановлено чіткі правила та принципи, які не підлягають компромісам у будь-яких обставинах. Це дозволяє забезпечити стабільність та передбачуваність у роботі системи, а також зменшує ризик виникнення непередбачених наслідків.

Проте одним із викликів деонтології в програмуванні моральних машин є пошук балансу між дотриманням принципів та урахуванням контексту та потреб користувачів. Наприклад, у складних ситуаціях, коли правила можуть конфліктувати або бути неоднозначними, моральні машини можуть зіштовхуватися з викликом прийняття рішень. Такі ситуації вимагають розробки нових методів та стратегій, які дозволять забезпечити баланс між деонтологічними принципами та реальними потребами та обставинами [33].

Отже, подальший розвиток деонтології в контексті програмування «моральних» машин є підґрунтям у забезпеченні їх ефективності, етичності та відповідності сучасним вимогам. Це вимагає спільних зусиль фахівців з різних галузей, правозахисників та представників суспільства, а також постійного дослідження та інновацій у цій сфері з особливим акцентом на етичній відповідальності.

2.3. Віртуальна етика та її роль у формуванні "характеру" штучного інтелекту

Віртуальна етика відіграє вирішальну роль у формуванні "характеру" штучного інтелекту. Вона охоплює моральні принципи та норми, що стосуються створення та використання віртуальних агентів, програмного забезпечення та інших технологічних систем. Зокрема, розвиток штучного інтелекту (ШІ)

ставить перед нами численні етичні виклики, які потребують уваги з боку науковців, розробників і користувачів [26].

Одним із основних питань у цій сфері є відповідальність за рішення, прийняті штучним інтелектом. Наприклад, в контексті автономних автомобілів, які самостійно приймають рішення про маневрування та швидкість, можуть виникнути ситуації, коли машина має обрати між меншим з двох зол. Такі сценарії ставлять під сумнів здатність ШІ дотримуватись етичних стандартів та викликають дискусії про необхідність програмування моральних принципів в ШІ [32].

Джудіт Саймон, професорка етики інформаційних технологій, стверджує, що "програмування етики у штучний інтелект вимагає не тільки технічних знань, але й глибокого розуміння моральних цінностей, які мають бути інтегровані". Це підкреслює важливість міждисциплінарного підходу у розробці ШІ, де філософи, соціологи та інженери працюють разом для вирішення складних моральних дилем [27].

З іншого боку, виникає питання про приватність та конфіденційність даних, що є особливо актуальним у контексті ШІ, який обробляє великі масиви інформації. Наприклад, системи розпізнавання облич, використовувані в секторах безпеки, можуть порушувати приватність індивідуумів без їх відома чи згоди. Тому важливо розробляти ШІ з урахуванням етичних норм, що захищають особисті дані та не дозволяють їх зловживання [31].

Професор Рон Аркін, який займається робототехнікою, акцентує увагу на необхідності створення "етичних алгоритмів" для роботів, які можуть діяти у відповідності з передбаченими моральними правилами в екстремальних ситуаціях. Він прикладає значні зусилля для впровадження таких алгоритмів в роботів, що виконують пошуково-рятувальні операції або діють у військових умовах. Етичні алгоритми дозволяють мінімізувати ризики для життя людей, зокрема, обираючи тактики, які виключають ненавмисне завдання шкоди цивільному населенню.

Ще одним аспектом віртуальної етики є здатність ШІ вчитися з досвіду та його вплив на моральні рішення, що приймаються машиною. Наприклад, системи штучного інтелекту, що використовують машинне навчання для оптимізації своїх алгоритмів, можуть розвивати непередбачувані звички або приймати рішення, які важко пояснити з точки зору моральності. Це викликає дебати про те, наскільки довіряти таким системам та як контролювати їхню поведінку [27].

Данахер, професор філософії, вважає, що "проблема чорного ящика" у машинному навчанні, коли неможливо зрозуміти, як було прийняте рішення, ставить під загрозу здатність ШІ діяти етично. Це призводить до необхідності створення більш прозорих алгоритмів, які б дозволили аналізувати та перевіряти логіку роботи ШІ [30].

Крім того, розвиток автономних систем викликає питання про "машинні права". Зокрема, деякі науковці, як Френсіс Камм, аргументують, що високорозвинені ШІ можуть мати право на певний рівень захисту чи навіть права, подібні до людських, якщо їхні когнітивні здібності досягнуть відповідного рівня [27].

Освіта та навчання населення також відіграють ключову роль у розумінні та застосуванні віртуальної етики. Широка громадськість повинна бути обізнана про етичні виміри технологій ШІ, щоб краще розуміти потенційні ризики та переваги. Включення курсів з етики ШІ у навчальні програми може сприяти формуванню більш відповідального та обізнаного підходу до технологій серед майбутніх фахівців.

Програми такого типу допомагають студентам розуміти не тільки технічні аспекти штучного інтелекту, але й глибокі моральні дилеми, що можуть виникати при його використанні. Так, Венді Холл, професорка комп'ютерних наук, зазначає, що "інтеграція етичних розглядів у курси з ШІ та робототехніки є критично важливою для розвитку відповідальних технологій" [29].

Прикладом практичного застосування віртуальної етики у штучному інтелекті є робота компаній, що розробляють алгоритми рекомендаційних систем. Ці системи часто використовуються в соціальних мережах та платформах для стрімінгу контенту, де вони аналізують поведінку користувачів для надання персоналізованих рекомендацій. Однак, без належних етичних рамок, такі системи можуть непомітно маніпулювати користувачами, спонукаючи їх до перегляду більшої кількості контенту та, можливо, сприяючи формуванню залежностей [27].

Етичні розгляди також стосуються використання ШІ у медицині. Наприклад, алгоритми ШІ, які допомагають у діагностуванні хвороб, повинні бути точними та безпечними, щоб не завдати шкоди пацієнтам через помилкові діагнози або рекомендації. Етичні проблеми з'являються також тоді, коли ШІ використовується для визначення пріоритетів у лікуванні пацієнтів, що може призвести до моральних конфліктів щодо рівності доступу до медичних послуг.

Для вирішення цих та інших проблем, з'являється все більше ініціатив з регулювання ШІ на міжнародному рівні. Організації, такі як Європейський союз і Об'єднані Нації, розробляють стандарти та рекомендації, які спрямовані на забезпечення безпечного та етичного використання штучного інтелекту. Ці стандарти можуть включати вимоги до прозорості, справедливості, недискримінації та здатності до пояснення прийнятих рішень ШІ. Забезпечення дотримання цих стандартів може допомогти уникнути потенційно небезпечних ситуацій, пов'язаних з неправильним використанням технологій [28].

Важливість залучення громадськості до обговорення етичних аспектів ШІ не може бути недооціненою. Збільшення обізнаності та розуміння викликів, які стоять перед суспільством у зв'язку з розвитком штучного інтелекту, є ключем до формування здорового ставлення до технологій. Прозорість у діяльності компаній, що розробляють та впроваджують ШІ, забезпечує довіру та сприяє більш відповідальному використанню технологій.

На додаток до правових та регуляторних заходів, важливим є створення етичних рад та комітетів, які могли б слідкувати за розвитком та впровадженням штучного інтелекту. Ці органи могли би займатися оцінкою нових технологій, враховуючи їх потенційний вплив на суспільство, та виносити рекомендації щодо їх етичного використання [27].

Підсумовуючи, віртуальна етика у сфері штучного інтелекту є надзвичайно важливою для забезпечення того, що технологічний прогрес не йде на шкоду основним людським цінностям та правам. Інтеграція етичних принципів у розробку та застосування ШІ не тільки сприяє створенню більш безпечних та надійних технологій, але й підтримує моральний розвиток суспільства як цілого. Завдяки спільним зусиллям науковців, розробників, політиків та громадськості можна досягти збалансованого та відповідального підходу до використання штучного інтелекту, який буде відповідати етичним нормам та сприяти загальному благу.

Таким чином, дослідження основних етичних теорій та їхнє застосування до штучного інтелекту (ШІ) та робототехніки, розкривають особливості впровадження технологій. Утилітаризм, який прагне максимізувати загальне благо, стикається з викликами, пов'язаними з передбаченням та оцінкою всіх потенційних наслідків дій ШІ. Це вимагає глибокого аналізу та розвитку алгоритмів (етичного інструментарію), які можуть адекватно реагувати на складні ситуації, зважаючи на різноманітні інтереси та потреби. З іншого боку, деонтологія пропонує рамки для програмування ШІ засновані на чітко визначених правилах та обмеженнях, що дозволяють створювати моральні машини, які функціонують згідно з встановленими нормами. Це забезпечує додаткову безпеку та передбачуваність у поведінці ШІ, але також може призводити до жорсткості в реагуванні на нестандартні сценарії.

Нарешті, віртуальна етика відіграє ключову роль у формуванні "характеру" штучного інтелекту, забезпечуючи, що машини діють не тільки ефективно, але й морально прийнятно. Інтеграція етичних принципів в основу

розробки ШІ може допомогти збалансувати технологічний прогрес з моральними цінностями, забезпечуючи таким чином довіру та прийняття цих технологій суспільством.

Разом ці підходи формують комплексний етичний фундамент, який може вести до більш відповідального та обачного використання ШІ. Важливо продовжувати діалог між розробниками, етиками, правовими експертами та широкою громадськістю, щоб забезпечити, що розвиток штучного інтелекту відповідає не тільки технічним, але й моральним стандартам, підтримуючи сталий розвиток та благо людства.

Отже, утилітаризм, як етична теорія, спрямований на досягнення найбільшого загального щастя для найбільшої кількості людей. У контексті штучного інтелекту (ШІ) та робототехніки, утилітаризм стає предметом глибоких дебатів та викликів. Деякі вчені вважають, що принципи утилітаризму можуть бути корисними для програмування штучного інтелекту та роботів з метою прийняття моральних рішень. Австрійський філософ Пітер Сінгер пропонує концепцію ефективного альтруїзму, яка впливає на розуміння утилітаризму в контексті розвитку штучного інтелекту. Згідно з цією концепцією, морально обов'язково діяти так, щоб максимізувати загальне щастя, але це також означає, що ми повинні бути готові віддати свої власні інтереси на користь тих, хто потребує допомоги найбільше. Враховуючи дебати та виклики утилітаризму в контексті штучного інтелекту та робототехніки, можна побачити, що цей етичний підхід має як переваги, так і обмеження. Його застосування може сприяти досягненню загальної користі для суспільства та максимізувати щастя багатьох. Однак, варто бути обережними, щоб уникнути порушення прав та інтересів меншості в ім'я загальної користі. Таким чином, розуміння та урахування контексту та потенційних наслідків важливо для використання утилітаризму в сучасних технологічних доменах.

Одним зі суттєвих аспектів утилітаризму є постійна необхідність оновлення та адаптації принципів до змін у суспільстві та технологій. Сучасні

технології швидко розвиваються, а разом з ними змінюються й етичні виклики, які вони ставлять перед нами.

У вивченні етичних принципів для програмування моральних машин, деонтологія займає центральне місце. Деонтологія, як етична теорія, стверджує, що правильність або неправильність дій не залежить від їхніх наслідків, а від дотримання певних моральних правил або принципів. У контексті програмування моральних машин, це означає, що роботи повинні дотримуватися певних правил і обмежень, незалежно від очікуваних наслідків їхніх дій.

Філософ і вчений Іммануель Кант був одним із найвідоміших прихильників деонтології, яке означає вчення про проблеми моралі і моральності. Кант розвинув концепцію категоричного імперативу, який вимагає, щоб діяти лише за таких умов, за яких ця дія може стати загальним правилом для всіх.

Деонтологія також надає увагу імперативу поваги до особистості. Це означає, що кожна особа має бути поважана як моральний суб'єкт і ніколи не може бути розглянута як простий засіб для досягнення мети. У програмуванні моральних машин це може мати велике значення, оскільки роботи можуть мати справу з різними людьми та ситуаціями.

Ще одним обмеженням деонтології є її негнучкість у вирішенні складних етичних ділем. Утилітаризм може бути більш гнучким у цьому плані, оскільки він дозволяє враховувати контекст та наслідки дій. Проте деонтологія наголошує на важливості дотримання моральних принципів, незалежно від обставин. Це може викликати ситуації, коли деякі дії можуть здатися морально неприпустимими, навіть якщо вони можуть принести загальне щастя. Додатковим напрямком дослідження може бути розробка етичних рамок та стандартів для програмування моральних машин, які б забезпечували високий рівень дотримання деонтологічних принципів.

Віртуальна етика відіграє вирішальну роль у формуванні "характеру" штучного інтелекту. Вона охоплює моральні принципи та норми, що стосуються

створення та використання віртуальних агентів, програмного забезпечення та інших технологічних систем. Зокрема, розвиток штучного інтелекту (ШІ) ставить перед нами численні етичні виклики, які потребують уваги з боку науковців, розробників і користувачів.

Джудіт Саймон, професорка етики інформаційних технологій, стверджує, що "програмування етики у штучний інтелект вимагає не тільки технічних знань, але й глибокого розуміння моральних цінностей, які мають бути інтегровані". Це підкреслює важливість міждисциплінарного підходу у розробці ШІ, де філософи, соціологи та інженери працюють разом для вирішення складних моральних дилем.

ВИСНОВКИ

У цій роботі ми розглянули етичні аспекти застосування штучного інтелекту та робототехніки, починаючи з історичного огляду філософських поглядів на технології та їх впливу на суспільство. Історичний контекст дозволив нам краще зрозуміти, як змінювалася взаємодія людини та технологій та які філософські течії впливали на цей процес. Від античності, коли технології сприймалися як інструменти для покращення життя людини, до сучасності, де штучний інтелект виступає як потенційний учасник культурних та соціальних процесів, ми бачимо, що технології невпинно формують наше суспільство.

Одним із ключових аспектів роботи стала концепція "Сингулярності" – моменту в історії, коли машини здобудуть інтелектуальні спроможності, що можливо перевершать людські, та їх етичні наслідки для людства. Обговорення сингулярності виявилось особливо важливим у контексті етичних дилем, пов'язаних з автономією, етичною відповідальністю та можливістю контролю за штучним інтелектом. Сингулярність ставить перед нами питання про права машин, їх статус у суспільстві та можливості інтеграції штучного інтелекту в соціальні структури.

Полеміка щодо «машинної» моральності, проаналізовані в роботі, розкрили потенційні можливості та межі створення «етичних» машин. Розглядаючи «інструментальність» моралі, що традиційно приписується тільки людям, ми зіткнулися з концептуальними труднощами у її застосуванні до машин. Втім, розвиток алгоритмів, здатних імітувати моральні рішення, вказує на потенційну можливість створення "моральних машин", хоча це також вимагає вирішення складних технічних та філософських питань.

Також в рамках утилітаризму проаналізовано виклики, з якими може зіткнутися ШІ при спробі максимізувати загальне благо. Виявилося, що утилітаризм може надати корисні напрямки для оцінки дій ШІ, особливо коли йдеться про оптимізацію вирішення завдань з численними стейкхолдерами. Однак цей підхід також ставить перед дослідниками та розробниками складні питання щодо оцінки та прогнозування довготривалих наслідків рішень, що приймаються машинами, відповідальність. Утилітаризм вимагає від алгоритмів ШІ здатності аналізувати значний обсяг даних та зіставляти потенційні наслідки різних дій, що може бути технічно складним та етично суперечливим.

В контексті деонтології, ми зосередили увагу на створенні правил та обмежень для програмування машин. Деонтологічний підхід пропонує чітку структуру для програмування ШІ, де визначені правила поведінки можуть допомогти уникнути етичних порушень. Однак цей підхід також може обмежувати гнучкість штучного інтелекту, що може бути критично у складних ситуаціях, де потрібно врахувати нюанси або контекст. Тому, хоча деонтологія забезпечує чіткість та передбачуваність, вона також вимагає гнучкості у вирішенні етичних дилем, що може бути важко втілити в алгоритмах.

Віртуальна етика, зокрема, виявилася ключовим компонентом у формуванні "характеру" штучного інтелекту. Цей аспект нашого дослідження показав, як важливо інтегрувати моральні та етичні розгляди в основу проектування та розробки ШІ. Віртуальна етика не тільки підкреслює необхідність врахування моральних принципів у технологіях, але й акцентує на важливості того, щоб ці системи відображали загальнолюдські цінності, сприяючи таким чином більш гармонійній взаємодії між людиною та машиною.

У сукупності, виконані завдання дослідження дозволили нам глибше зануритись у комплексні етичні питання, пов'язані з розвитком штучного інтелекту та робототехніки та виміри етичної відповідальності за наслідки дій /рішень. Важливо продовжувати цей діалог між розробниками, етиками, законодавцями та громадськістю, щоб забезпечити розвиток цих технологій у

відповідності до етичних норм та цінностей. Зростаюча складність застосування та втілення технологій ШІ та робототехніки вимагає відповідального підходу до інтеграції на користь суспільному благу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аббадія Д. Вивчення ролі штучного інтелекту в академічних дослідженнях. Mind the Graph. Режим доступу: URL: <https://mindthegraph.com/blog/uk/ai-in-academic-research/>
2. Візнюк І. Використання штучного інтелекту в освіті // Сучасні інформаційні технології та інноваційні методики навчання в підготовці фахівців: методологія, теорія, досвід, проблеми. 2021. № 59. С. 14–22. DOI: <https://doi.org/10.31652/2412-1142-2021-59-14-22>
3. Забара І. М. До питання визначення проблематики міжнародно-правової відповідальності за правопорушення при використанні новітніх інформаційнокомунікаційних технологій. Інтернет речей: проблеми правового регулювання та впровадження : Матеріали III наук.-практ. конф. (м. Київ, 21 лист. 2019 р.) // Упоряд. : В. М. Фурашев, С. О. Дорогих, С. Ю. Петряєв. Київ : КПІ ім. Ігоря Сікорського, вид-во «Політехніка», 2019. С. 59–62. URL: http://ippi.org.ua/sites/default/files/maket_ot_20.11.19._1.pdf
4. Каткова Т. Г. Штучний інтелект в Україні: правові аспекти. Право і суспільство. 2020. № 6. С. 46–55. DOI: <https://doi.org/10.32842/2078-3736/2020.6.1.8>

5. Кузьомко В., Бурангулова В., Бурангулова В. Можливості використання штучного інтелекту в діяльності сучасних підприємств. Економіка та суспільство. 2021. № 32. DOI: <https://doi.org/10.32782/2524-0072/2021-32-676>.
6. Мар'єнко М., Коваленко В. Штучний інтелект та відкрита наука в освіті. Фізико-математична освіта. 2023. Т. 38, № 1. С. 48–53. DOI: <https://doi.org/10.31110/2413-1571-2023-038-1-007>
7. Рогожа М. М. Моральні дилеми у галузі штучного інтелекту. Категорії. Мораль : зб. наук. матеріалів за результатами Всеукраїнського круглого столу «Читання пам'яті Івана Бойченка-2019. Людина. Історія. Мораль» (25 жовтня 2019 р.) / редкол.: А. Є. Конверський [та ін.] ; Київський національний університет ім. Т. Шевченка, Інститут філософії ім. Г. С. Сковороди НАН України, Центр гуманітарної освіти НАН України. Київ: Знання України, 2019. С. 45–50. Режим доступу URL: <https://ekmair.ukma.edu.ua/handle/123456789/18081>
8. Турута О. В., Турута О. П. Штучний інтелект крізь призму фундаментальних прав людини. Науковий вісник Ужгородського національного університету. Серія: Право. 2022. № 71. С. 49–54. DOI: <https://doi.org/10.24144/2307-3322.2022.71.7> (дата звернення: 26.06.2023).
9. Яровой Т. С. Можливості та ризики використання штучного інтелекту в публічному управлінні. Economic Synergy. 2023. № 2. С. 36–47. DOI: <https://doi.org/10.53920/ES-2023-2-3>
10. Awasthi S., Soni Y. Empowering education system with artificial intelligence: opportunities and challenges. Shodh samagam. 2023. Vol. 6, No. 1. P. 1–4. URL: <http://www.shodhsamagam.com/admin/uploads/Empowering%20Education%20System%20with%20Artificial%20Intelligence%20%20Opportunities%20and%20Challenges.pdf>

11. AI and Fundamental Rights in Europe (2020) by the Fundamental Rights Agency (FRA).URL: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificialintelligence_en.pdf.
12. AI Watch. Defining Artificial Intelligence 2.0.URL:<https://publications.jrc.ec.europa.eu/repository/handle/JRC126426>.
13. AI, Robotics and European Civil Liability Rules (2019) by the European Parliament Committee on Legal Affairs. URL: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU\(2020\)621926_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf).
14. Álvarez-Machancoses Ó, Fernández-Martínez JL. Using artificial intelligencemethods to speed up drug discovery. *Expert Opin Drug Discov.* (2019) 14:769–77. doi: 10.1080/17460441.2019.1621284
15. Asare J.G. The Dark Side Of ChatGPT. *Forbes*: Jersey City, NJ, USA, 2023. URL: <https://www.forbes.com/sites/janicegassam/2023/01/28/the-dark-side-ofchatgpt/?sh=31f2e08a4799>.
16. Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A, Le S, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Comput Biol Med.* (2019) 109:79– 84. doi: 10.1016/j.compbimed.2019.04.027
17. Basta C., Costa-jussà M. R., Casas, N. Evaluating the Underlying Gender Bias in Contextualized Word Embeddings. *Proceedings of the Workshop on Gender Bias in Natural Language Processing.* 2019. 1. P. 33–39. URL: <https://doi.org/10.18653/v1/W19-3805>
18. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in Obstetrics and Gynecology. *Ultrasound Obstetr Gynecol.* (2020) 56:498–505. doi: 10.1002/uog.22122
19. Ethical Guidelines for Trustworthy AI (2018) by the High-Level Expert Group on AI.URL: <https://www.aepd.es/sites/default/files/2019-12/ai-ethics-uidelines.pdf>. 10. Ethics Guidelines for Trustworthy AI (2019) by the European

Commission's HighLevel ExpertGroup on Artificial Intelligence.URL:
<https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

20. Floridi L., Chiriatti, M. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*. 2020. 30(4). P. 681-694. URL: <https://doi.org/10.1007/s11023-020-09548-1>

21. Getahun H. ChatGPT Could Be Used for Good, But Like Many Other AI Models, It's Rife with Racist and Discriminatory Bias; Insider. 2023. URL: <https://www.insider.com/chatgpt-is-likemany-other-ai-models-rife-with-bias-023-1>

22. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. *PLoS Med*. (2013) 10:e1001413. doi: 10.1371/journal.pmed.1001413

23. Henz P. Ethical and legal responsibility for artificial intelligence. *Discov Artif Intell*. (2021) 1:2. doi: 10.1007/s44163-021-00002-4

24. Hutchinson B., Prabhakaran V., Denton E., Webster K., Zhong Y., Denuyl S. Social Biases in NLP Models as Barriers for Persons with Disabilities. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2020. 58. P. 5491–5501. URL: <https://doi.org/10.18653/v1/2020.acl-main.487>

25. Jingshan Huang, Ming Tan. The role of ChatGPT in scientific communication: writing better scientific review articles. *American Journal of Cancer Research*. 2023. 13 (4). P. 1148-1154. URL: <https://e-century.us/files/ajcr/13/4/ajcr0150104.pdf>

26. Mannes A. Governance, risk, and Artificial Intelligence. *AI Magazine*. (2020) 41:61–9. doi: 10.1609/aimag.v41i1.5200

27. Miller DD, Brown EW. Artificial Intelligence in medical practice: the question to the answer? *Am J Med*. (2018) 131:129–33. doi: 10.1016/j.amjmed.2017.10.035

28. Morley J, Floridi L. An ethically mindful approach to AI for Health Care. *SSRN Electron J*. (2020) 395:254–5. doi: 10.2139/ssrn.3830536

29. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med.* (2020) 260:113172. doi: 10.1016/j.socscimed.2020.113172
30. Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with Artificial Intelligence. *npj Digit Med.* (2019) 2:26. doi: 10.1038/s41746-019- 0103-3
31. Nelson GS. Bias in artificial intelligence. *North CarolinaMed J.* (2019) 80:220–2. doi: 10.18043/ncm.80.4.220
32. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in Health Care. *JAMA.* (2019) 322:2377. doi: 10.1001/jama.2019.18058
33. Perc M. The Matthew effect in empirical data. *Journal of the Royal Society Interface.* 2014.11(98).URL: <https://doi.org/10.1098/rsif.2014.0378>