КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

Філософський факультет

Кафедра етики, естетики та культурології

ШТУЧНИЙ ІНТЕЛЕКТ: ЕТИЧНІ ВИМІРИ

Кваліфікаційна робота
за спеціальністю 033 «Філософія»
на здобуття освітнього ступеня бакалавра філософії

Студентка-виконавець:

Василенко Тетяна Володимирівна

IV курс

спеціальність 033 «Філософія»

ОПП «Філософія»

Науковий керівник:

Нападиста Валентина Григорівна

канд. філос. наук, доцент

Допущено до захисту:

Завідувач кафедри _____

Київ-2023

ЗМІСТ

ВСТУП

Актуальність теми дослідження. Термін "штучний інтелект" у всіх на слуху. Не тільки великі компанії, а й пересічні громадяни все більше переймаються цією темою. Можна навіть сказати, вимушено, адже вже важко уявити життя без ШІ. Багато процесів у компаніях стали швидшими та якіснішими. Але ми також відчуваємо прогрес штучного інтелекту на власному досвіді в нашому приватному житті. Alexa та Google Home роблять повсякденне життя вдома простішим. Від автомобіля біля дверей, який самостійно везе вас додому з вечірки, коли вам більше не можна сідати за кермо, до робота, який реєструє мене в готелі і призначає мені номер. Однак, все це також приносить із собою питання. Питання, які компанії та люди мають/повинні поставити собі щонайпізніше зараз. Етичні аспекти штучного інтелекту вивчаються і ставляться під сумнів більш ретельно. Хто оплачує рахунок, коли Alexa розміщує замовлення на Amazon без команди клієнта? Хто винен у дорожньо-транспортній пригоді, спричиненій моїм автомобілем з автономним керуванням? Я, як власник авто, чи програмісти алгоритмів? Це лише деякі з багатьох етичних питань, які необхідно і слід пояснювати в процесі розвитку штучного інтелекту. Однією з ключових причин актуальності етичних питань у контексті ШІ є потенційний вплив на права людини та суспільний добробут. Системи штучного інтелекту здатні приймати рішення, які суттєво впливають на життя людей, починаючи від можливостей роботи до доступу до ресурсів і послуг. Без належних етичних міркувань існує ризик збереження упереджень, дискримінації та нерівності. Наприклад, якщо алгоритми штучного інтелекту навчаються на упереджених даних або не мають різноманітності, вони можуть посилити існуючі суспільні упередження, що призведе до несправедливих результатів і ще більше поставить у невигідне становище маргіналізовані спільноти.

Проблеми конфіденційності та безпеки також виходять на перший план під час обговорення потенційних проблем ШІ з етичногу погляду. Для ефективної роботи системи штучного інтелекту часто покладаються на величезну кількість

персональних даних. Це викликає питання про те, як ці дані збираються, зберігаються та використовуються, і чи мають особи контроль над своєю особистою інформацією. Без відповідних заходів безпеки існує ризик несанкціонованого доступу, витоку даних і неправомірного використання особистої інформації, що потенційно може поставити під загрозу конфіденційність і безпеку людей. Дебати щодо етики штучного інтелекту далеко не вичерпані – і в якісному і, звичайно, і в кількісному вираженні. У майбутньому штучний інтелект постійно стикатиметься зі структурами цінностей, які люди занадто часто сприймають як належне, і спонукатиме їх сумніватися в них. Справа тут не в тому, щоб обмежити технологічний потенціал ШІ, а в тому, щоб подумати про його потенційний негативний вплив на людей. Аргумент, який зараз часто згадується, що штучний інтелект може якимось чином вирішити багато, якщо не всі проблеми, не може бути переконливим в довгостроковій перспективі з огляду на можливі негативні наслідки та вплив на людей.

Тобто актуальність етичних питань щодо ШІ полягає в потенційному соціальному, економічному та етичному впливі систем ШІ на окремих людей і суспільство в цілому. Вирішуючи такі проблеми, як упередженість, справедливість, конфіденційність, підзвітність і автономія людини, ми можемо прагнути розробляти та розгортати технології штучного інтелекту, які відповідають етичним принципам, сприяють благополуччю людей і роблять внесок у більш процвітаюче суспільство. Будучи студентом, дуже важливо брати участь в обговореннях і дослідженнях етики штучного інтелекту, відстоювати відповідальні практики штучного інтелекту та робити внесок у формування майбутнього, де штучний інтелект буде служити найкращим інтересам людства.

Ступінь наукової розробки теми. Дослідження результатів впровадження технологій штучного інтелекту знаходиться на стадії активного наукового дослідження. Серед фундаторів питання штучного інтелекту можна згадати таких видатних дослідників, як Р. Пенроуз, Н. Бостром та Е. Юдковскі. Ці автори зосереджуються на вивченні моральних та етичних викликів, що

пов'язані з розвитком штучного інтелекту та його впливом на людство. Вони висувають гіпотези про можливі наслідки штучного інтелекту і пропонують нові підходи до етичних питань в цій сфері.

Окрім того, такі дослідники як Р. Аркін, Р. Біннс, М. Боден, Т. Волш, М. Гудман, Р. Зінгер, Дж. Стюарт, Л. Флоріді, М. Форд та інш, також вносять вагомий внесок у розуміння та обговорення етичних вимірів штучного інтелекту. Вони зосереджуються на вивченні моральних, соціальних та правових наслідків, що виникають у зв'язку зі зростанням автономності та розумових здібностей штучного інтелекту.

Серед вітчизняних дослідників варто згадати Л.В. Токар, яка спеціалізується на впровадженні штучного інтелекту до сучасної української правової системи. Її дослідження зосереджені на етичних та правових аспектах використання штучного інтелекту в правовій сфері України.

Крім того, Л.М. Блозва працює над питаннями штучного інтелекту в контексті сучасної культури. Її дослідження охоплюють етичні, соціокультурні та філософські виміри використання штучного інтелекту в сучасному суспільстві.

Завдяки спільним зусиллям цих авторів та багатьох інших, зростає розуміння та обговорення етичних вимірів штучного інтелекту, а також розробляються нові підходи та розв'язання для цієї проблематики. Важливо продовжувати наукові дослідження в цій галузі, щоб забезпечити етичний розвиток технологій штучного інтелекту та зробити їх корисними та безпечними для суспільства.

Об'єктом роботи є сучасні наукові дискурси щодо впровадження штучного інтелекту.

Предметом роботи є етичні аспекти результатів застосування штучного інтелекту.

Мета роботи полягає в тому, щоб висвітлити етичні проблеми, пов’язані зі штучним інтелектом і підкреслити важливість вирішення цих проблем у розробці та розгортанні технологій ШІ.

Відповідно до мети визначено наступні **завдання:**

- Розглянути передумови соціального запиту на технології ШІ та особливості термінологічних визначень ШІ;
- З’ясувати основні проблеми етичного змісту, обумовлені впровадженням технологій штучного інтелекту;
- Дослідити базові моральні вимоги при проєктувані технологій ШІ;
- Проаналізувати основні шляхи пом’якшення етичних ризиків у сфері штучного інтелекту.

Теоретико-методологічна основа дослідження. Для досягнення поставленої мети було використано наступні методи наукового дослідженя: порівняння та системного аналізу наукових джерел; інформаційний, аксіологічний та суб’єкто-орієнтований підходи при осмисленні проблеми.

Структура дипломної роботи складається зі вступу, 3 розділів, що містять в собі по два та три підрозділи, висновку та списку використаних джерел. Кількість сторінок – 50. Кількість найменувань у списку літератури – 42.

РОЗДІЛ І
ШТУЧНИЙ ІНТЕЛЕКТ: ПЕРЕДУМОВИ ТА ТЕРМІНОЛОГІЧНІ ЗМІСТИ

Незважаючи на численні дослідження в галузі штучного інтелекту, можна зрозуміти, що інтелектуальна поведінка людини має бути імітована штучно. Тому програми розробляються таким чином, щоб імітувати роботу мозку. Їх називають штучними нейронними мережами.

Крім того, можна сказати, що ШІ - це підгалузь комп'ютерних наук, метою якої є надання системам здатності правильно інтерпретувати зовнішні дані, вчитися на отриманих даних і використовувати інсайти з них [8, с.25]. Основний акцент тут робиться на тому, що завдання повинні виконуватися розумно. Як це робиться і що саме мається на увазі під "розумним", точно не визначено.

Метою ШІ є створення машин, які можуть виконувати завдання, що зазвичай вимагають людського інтелекту, такі як вирішення проблем, прийняття рішень, обробка природної мови та розпізнавання зображень.

В основі ШІ лежить ідея створення алгоритмів і систем, здатних обробляти й аналізувати великі обсяги даних і приймати рішення на основі цих даних. Це досягається за допомогою машинного навчання - підмножини ШІ, яка передбачає використання алгоритмів для аналізу даних і навчання на них без явного програмування. Однак таке застосування не обов'язково стосується машин і роботів, як часто вважають, а може мати суто цифрову природу у вигляді IT-систем. Вони налаштовані таким чином, що розпізнають закономірності на основі наявних наборів даних і розробляють на їх основі рішення [24, с. 91]. Таким чином, відбувається процес самонавчання, коли непередбачувана програма навчається самостійно на основі алгоритмів і за допомогою навчальних даних, постійно адаптуючись і вдосконалюючись. Взірцем для цього є навчання людини та тварин. Машинне навчання зазвичай використовується для розпізнавання зображень, розпізнавання мови та обробки природної мови.

Однією з причин, чому люди зацікавилися створенням штучного інтелекту, є бажання зрозуміти природу інтелекту та свідомості. Люди вже давно зачаровані роботою розуму і ШІ пропонує спосіб дослідити ці питання по-новому. Намагаючись створити машини, здатні імітувати людський інтелект, ми можемо глибше зрозуміти, як працює розум і що робить людину унікальною.

Тут варто зосередитися на питаннях свідомості та штучного інтелекту. Від перших днів обчислювальної техніки до сучасної ери машинного навчання та глибоких нейронних мереж штучний інтелект пройшов довгий шлях. Однак, незважаючи на значні досягнення в галузі ШІ, людська свідомість залишається неперевершеною за своєю складністю та можливостями. Людський розум - це складна і динамічна система, яка дозволяє нам взаємодіяти з навколишнім світом, вчитися, запам'ятовувати, міркувати і приймати рішення. Він є продуктом мільйонів років еволюції і формується під впливом нашого досвіду, емоцій та соціальної взаємодії. «Природний, людський інтелект не є константою. Він розвивається, сприяючи виживанню і успіху людини в складному, лише частково осяжному і слабо контрольованому світі» [3, с. 6].

На відміну від нього, ШІ - це створена людиною система, яка використовує алгоритми, дані та обчислювальну потужність для імітації людського інтелекту. Хоча за останні роки штучний інтелект досяг значного прогресу, йому все ще бракує креативності, інтуїції та емоційного інтелекту, які визначають людське пізнання.

Говорячи про штучний інтелект доцільно згадати мисленнєвий експеримет «Китайська кімната». Гіпотетична ситуація, яку підняв Дж. Серль, полягала в тому, щоб уявити, що англомовна людина, яка не розуміє китайської мови, заходить до кімнати, де їй надається інструкція, написана англійською мовою, для маніпулювання деякими китайськими символами в певному порядку. У такому порядку символи виражають повідомлення китайською мовою. Якщо після обробки ви передасте їх стороннім спостерігачам, останній, мабуть, подумає, що англомовна людина, яка не розуміє китайської, дійсно розуміє китайську, хоча насправді вони цього не роблять. Для Дж. Серля саме так

працюють комп'ютерні операційні системи - вони наслідують розуміння, але не доходять до нього.

Здається, що мета цих експериментів полягає в тому, щоб довести, що правильна комбінація вхідної, вихідної інформації та програми може дозволити людині розуміти розповіді англійською мовою, навіть якщо вона не розуміє їх дослівно і в доречному сенсі. Це свідчить про те, що людина і система, яка її формує, можуть мати іншу модель розуміння, яка не зводиться до буквального значення слів.

Інтенціональність існує в комп’ютерах лише тому, що її вбудовують розробники програм та користувачі, які вводять вхідну інформацію та інтерпретують вихідну. Експеримент з «китайською кімнатою» мав на меті показати, що формальна програма не додає жодної додаткової інтенціональності до здатності людини розуміти китайську мову, оскільки будь-яка інтенціональність присутня в системі тільки завдяки діям людей, а не самої програми.

Просте обчислювальне моделювання, що здійснюється комп’ютерами, не має нічого спільного з людським мисленням. Серль стверджує, що розуміння є невід’ємним аспектом мисленнєвих процесів, і це передбачає можливість переживання, відчуття та усвідомлення того, що відбувається. Коли людина розуміє щось, вона надає йому значення та оперує смислами, які є ментальними утвореннями. Таким чином, Серль стверджує, що комп’ютерне моделювання не може повністю замінити людське мислення, оскільки це вимагає більш широкого спектру функцій та можливостей, які не доступні для комп’ютерів [1, с. 229].

Не тільки Серль, а й інші науковці піднімають це питання. Пенроуз ставить під сумнів, чи взагалі можуть системи штучного інтелекту, керовані обчислювальними процесами, мати такий самий рівень розуміння, інтуїції та креативності, як і люди. Він припускає, що можуть існувати фундаментальні аспекти людського пізнання і свідомості, які виходять за рамки того, що може бути досягнуто за допомогою обчислювальних моделей [26, с. 7-8].

Однією з головних проблем у розробці ШІ, здатного імітувати людський інтелект, є недостатнє розуміння того, як працює мозок. Незважаючи на десятиліття досліджень, ми ще багато чого не знаємо про нейронні механізми, що лежать в основі сприйняття, пам'яті, мови та прийняття рішень. Крім того, людський розум - це не просто набір ізольованих функцій, а тісно взаємопов'язана мережа областей мозку, які працюють разом, щоб виробляти складну поведінку. Відтворення такого рівня складності в ШІ є складним завданням, яке вимагає значного прогресу в нейронауках і комп'ютерних технологіях.

Коли ми співвідносимо штучний інтелект і людську свідомість, перш за все виникає питання: чи машина розуміє те, що робить, у той спосіб, у який люди розуміють свої думки і вчинки? Дж. Серль має чітку відповідь - "просте обчислювальне моделювання, що здійснюється комп’ютерами, не має нічого спільного з людським мисленням» [1, с. 228]. Він стверджує, що розуміння є невід’ємним аспектом мисленнєвих процесів, і це передбачає можливість переживання, відчуття та усвідомлення того, що відбувається. Коли людина розуміє щось, вона надає йому значення та оперує смислами, які є ментальними утвореннями. Таким чином, Дж. Серль стверджує, що комп’ютерне моделювання не може повністю замінити людське мислення, оскільки це вимагає більш широкого спектру функцій та можливостей, які не доступні для комп’ютерів.

Ще одна причина, чому люди зацікавилися штучним інтелектом, - це бажання створити машини, здатні виконувати завдання, які були б складними або неможливими для людини. Наприклад, ШІ можна використовувати для аналізу величезних обсягів даних або для виконання складних обчислень, які людині було б важко або занадто складно виконати самостійно. ШІ також можна використовувати для виконання небезпечних або виснажливих завдань, таких як дослідження космосу або збірка складних механізмів.

Прагнення створювати розумні машини також пов'язане з бажанням вдосконалити людське суспільство. ШІ може зробити революцію в охороні

здоров'я, транспорті, виробництві та багатьох інших галузях, зробивши їх більш ефективними, рентабельними та доступними для людей по всьому світу. Створюючи машини, здатні виконувати ці завдання, ми можемо покращити якість життя людей у всьому світі.

Нарешті, створення ШІ також обумовлене прагненням до інновацій та прогресу. Люди завжди прагнули досліджувати нові кордони, розширювати межі можливого і створювати нові речі. ШІ - це новий рубіж людських інновацій, а розвиток технологій ШІ дозволяє нам досліджувати нові можливості і створювати нові форми знань і розуміння.

Важливим поняттям в ШІ є глибоке навчання, яке є формою машинного навчання, що передбачає використання штучних нейронних мереж. Глибинне навчання було особливо успішним у розпізнаванні зображень і мови, його використовують для розробки безпілотних автомобілів, автоматизованих торгових систем та інших додатків. Глибинне навчання є своєрідним подальшим розвитком машинного навчання і описує навчання в рамках штучних нейронних мереж. Особливістю тут є те, що використовувані алгоритми - на відміну від традиційного ML - здатні самостійно розвиватися і створювати нові патерни всередині мереж для вирішення більш складних завдань. Таким чином, втручання людини для обробки даних більше не потрібне. «Глибинне навчання - це підгалузь штучного інтелекту, яка займається створенням алгоритмів, здатних навчатися на основі даних і робити прогнози або приймати рішення на основі цих даних. Ці алгоритми покликані імітувати роботу нейронних мереж людського мозку, які здатні аналізувати та обробляти великі обсяги інформації одночасно» [34, с. 750]. Глибинне навчання використовується в широкому спектрі застосувань, включаючи розпізнавання зображень і мови, обробку природної мови та автономні транспортні засоби.

Однією з головних переваг глибинного навчання є його здатність обробляти складні та неструктуровані дані, такі як зображення і текст, які важко обробляти за допомогою традиційних алгоритмів. Глибинне навчання також досягло найсучасніших результатів у багатьох галузях, включаючи

комп'ютерний зір, розпізнавання мови та обробку природної мови. Одним з найвідоміших прикладів глибинного навчання є алгоритм AlphaGo, розроблений Google DeepMind, який зміг перемогти чемпіона світу з давньої китайської гри в го [4].

ШІ набуває все більшого значення в багатьох галузях, зокрема в охороні здоров'я, фінансах і транспорті. Його також використовують для розробки чат-ботів, віртуальних помічників та інших інструментів, які можуть допомогти людям у виконанні повсякденних завдань. Однак існують також занепокоєння щодо впливу ШІ на зайнятість, конфіденційність і безпеку, і дослідники та політики працюють над вирішенням цих питань.

Контекст створення штучного інтелекту (ШІ) є багатогранним і складним, з внеском з широкого кола галузей, включаючи інформатику, математику, філософію, нейронауки і психологію.

Розвиток комп'ютерів і цифрових технологій в середині XX-го століття заклав основу для розвитку ШІ. Перші піонери, такі як Алан Тьюринг і Джон фон Нейман, заклали основу для розробки перших комп'ютерних програм і обчислювальних алгоритмів. У 1950-х і 1960-х роках виникла нова галузь досліджень, відома як когнітивна наука. Ця міждисциплінарна галузь об'єднала дослідників з психології, неврології, лінгвістики та інформатики, щоб дослідити, як працює розум і як відтворити його функції в машинах.

Холодна війна між Сполученими Штатами і Радянським Союзом створила клімат конкуренції та інновацій, в якому обидві країни прагнули розвивати нові технології, щоб отримати перевагу одна над одною. Це призвело до збільшення інвестицій в дослідження і розробку нових технологій, в тому числі ШІ [11].

У 1956 році група дослідників організувала конференцію в Дартмутському коледжі в Нью-Гемпширі, США, щоб обговорити нову галузь ШІ. Ця конференція вважається ключовим моментом в історії ШІ, оскільки вона зібрала багатьох ранніх піонерів у цій галузі та виділила ШІ в окрему сферу досліджень. Тут хочу навести цитату одного вченого, що був присутній на Дартмутській конференції: «У 1956 році в Дартмутському коледжі відбулася конференція, на

якій Джон Маккарті (який нещодавно помер) ввів термін "штучний інтелект". Він просто сказав, що є речі, такі як гра в шахи, розпізнавання об'єктів у світі або використання мови, які вимагають інтелекту, і тому питання полягає в тому, "Як ми можемо запрограмувати сучасні комп'ютери, щоб вони демонстрували певні аспекти цього інтелекту?» [17]. Саме з цього моменту ШІ починає набирати обертів і стане таким яким ми знаємо зараз.

Тут вже мова йде не про спробу створити повний людський інтелект. Ми просто обговорюємо, які аспекти людського інтелекту було б корисно передати машині, а потім подивитися, чи можемо ми використовувати сучасні комп'ютери для цієї мети, і якщо так, то як. Змусити комп'ютер робити це - це те, що робить його штучним. Залежно від конкретної системи, це може бути натхнення від вивчення того, як це робить мозок, або ж це може бути просто накопичені навички програмістів, які знаходять спосіб виконати певне завдання. Наведу приклад М. Арбіб - «Торговий автомат може розпізнавати доларові купюри. Він робить це грубо, розпізнаючи візерунок із зеленого та чорно-білого кольорів. Він не розпізнає обличчя Джорджа Вашингтона. Отже, це приклад, якщо хочете, дуже маленького пакета інтелекту, який розпізнає американську валюту методом, що використовує комп'ютер, але не схожий на те, як це робить людина» [17].

Так, розвиток нових апаратних і програмних технологій, таких як транзистор, мікрочіп та інтернет, надав дослідникам ШІ нові інструменти для створення більш досконалих і складних систем. Домінгос підкреслює важливість машинного навчання як ключової парадигми в історії ШІ. Він підкреслює перехід від ручних систем, заснованих на правилах, до алгоритмів, які можуть вивчати закономірності та робити прогнози на основі даних. Цей розвиток дозволив системам штучного інтелекту адаптуватися та покращувати свою продуктивність з часом [9, с.194].

Штучний інтелект несе багато викликів для нас та він також має багато можливостей для розвитку людського інтелекту та покращення нашого життя. Наприклад, ШІ може допомогти нам приймати кращі рішення, діагностувати

хвороби та оптимізувати складні системи, такі як транспортні або енергетичні мережі. ШІ також може підвищити нашу креативність, генеруючи нові ідеї або твори мистецтва, на які ми можемо спиратися. Більше того, ШІ може допомогти нам у виконанні завдань, які є небезпечними, нудними або виходять за межі наших когнітивних здібностей, наприклад, у дослідженні космосу або глибоководного видобутку корисних копалин.

ШІ вже активно використовується в багатьох сферах життя суспільства. Він спрямовує кожен наш пошуковий запит в інтернеті, він у самому серці будь-якого програмного додатка. Без нього не обходиться ні система GPS, ні відеоігри, ні голлівудські мультфільми, жоден сучасний банк, ні страхова компанія, ні лікарня. Ну і звісно, він у всяких розумних годинниках і безпілотних автомобілях. Тому умови використання ШІ мають бути заздалегіть обговорені і певним чином обмежені.

Висновки. Підсумовуючи, можна сказати, що люди прийшли до ідеї створення ШІ з різних причин, включаючи бажання зрозуміти природу інтелекту і свідомості, необхідність виконувати завдання, які є складними або неможливими для людини, прагнення вдосконалити людське суспільство, а також прагнення до інновацій і прогресу.

Таким чином, на створення ШІ вплинула низка історичних, культурних і технологічних факторів, зокрема розвиток комп'ютерів і цифрових технологій, поява когнітивної науки, Холодна війна, Дартмутська конференція, технологічний прогрес і культурні впливи. Ці фактори об'єдналися, щоб створити сприятливе середовище для розвитку ШІ, що призвело до появи нової галузі досліджень, яка продовжує рости і розвиватися сьогодні. ШІ набуває все більшого значення в багатьох галузях, зокрема в охороні здоров'я, фінансах і транспорті. Його також використовують для розробки чат- ботів, віртуальних помічників та інших інструментів, які можуть допомогти людям у виконанні повсякденних завдань. Однак існують також занепокоєння щодо впливу ШІ на зайнятість, конфіденційність і безпеку, і дослідники та політики працюють над вирішенням цих питань.

РОЗДІЛ ІІ
ЕТИЧНІ ВИМІРИ ВПРОВАДЖЕННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ

2.1. Автономія та прийняття рішень.

У зв'язку зі стрімким розвитком досліджень у галузі штучного інтелекту термін "етика" неминуче з'являється все частіше. Це пов'язано з тим, що алгоритми штучного інтелекту вже давно увійшли в наше повсякденне і робоче життя, і ми вже не можемо уявити життя без них, оскільки вони роблять його безмірно зручнішим. Від таргетованої реклами в соціальних мережах, простих платежів за допомогою сканування відбитків пальців на смартфонах при покупках в інтернеті, електронних квитків на автобуси та потяги до автономного керування автомобілями. Алгоритми - це частина штучного інтелекту. Перевага штучного інтелекту полягає в тому, що він може приймати рішення самостійно, без втручання людини. Люди помічають вплив ШІ і ставляться до нього насторожено. Як будуть оброблятися мої дані і що саме з ними станеться? Як далеко машинам дозволено зайти і як далеко вони повинні зайти?

Гудман підкреслює, що в міру того, як штучний інтелект стає все більш поширеним і потужним, він може бути використаний зловмисниками для зловживань або експлуатації. Він обговорює такі проблеми, як кібератаки з використанням ШІ, коли складні алгоритми ШІ можуть бути використані для автоматизації хакерських технологій або запуску масштабних кібернаступальних операцій [19, с. 69-70].

Автономія і прийняття рішень - це найважливіші аспекти штучного інтелекту (ШІ), які викликають великий інтерес у дослідників, політиків і громадськості. Системи штучного інтелекту все частіше використовуються для прийняття рішень, які раніше належали до компетенції людей, що приймають рішення, і це викликає ряд важливих етичних міркувань.

Одне з ключових питань, пов'язаних з автономією і прийняттям рішень у сфері ШІ, - це питання про те, хто несе відповідальність за рішення, прийняті

системами ШІ. Оскільки системи штучного інтелекту стають все більш складними та автономними, стає все важче розподіляти відповідальність за їхні дії. Це особливо актуально, коли системи ШІ приймають рішення, які мають значні соціальні або економічні наслідки, наприклад, у сферах охорони здоров'я, фінансів і транспорту. Для вирішення цієї проблеми важливо встановити чіткі межі підзвітності та відповідальності для систем ШІ, а також забезпечити наявність відповідних запобіжників для запобігання шкоді. «Розглядаючи етичну історію людських цивілізацій протягом століть, ми бачимо, що створення розуму, який був би стабільним в етичних вимірах, в той час як людські цивілізації, здається, демонструють спрямовані зміни, може виявитися дуже великою проблемою» [30]. Ми як людство постійно еволюціонуємо, разом з тим як і наші етичні погляди та норми суспільства. ШІ отримує вже готові задані нами поняття «норма», «правильне», «заборонене». Машинна етика ж не здатна еволюціонувати, закладений моральний кодекс у ШІ є стабільним. Так, основною перевагою ШІ є його фундаментальна здатність приймати рішення самостійно - автономно - без втручання людини. Питання, яке обговорюється, полягає в тому, до якої міри і коли машини повинні і можуть це робити, і які можуть бути наслідки. Етично релевантними є, перш за все, негативні наслідки, тобто коли людям завдається шкода або пошкоджуються речі. У цьому випадку питання полягає в тому, хто несе за це відповідальність або хто за це відповідає. Адже алгоритм не є юридичною особою. Для того, щоб організувати співіснування людей і машин, етика для автономних систем прийняття рішень, таким чином, в першу чергу визначає аспект відповідальності.

«Етичний дискурс про те, чи, якою мірою та за якими критеріями машини приймають рішення автономно у взаємодії з людьми, полягає насамперед у мірі відповідальності, яку люди передають машині в організації майбутнього співіснування. По суті, це вказує на те, чи має значення використання або невикористання штучного інтелекту - або, краще: до якої міри автономне прийняття рішень має значення використання ШІ» [8, с. 62].

Етичні вимоги до автономних, вирішальних систем і ступінь передачі відповідальності від людини до машини є результатом випереджальних міркувань про те, наскільки значні переваги для багатьох людей знаходяться в прийнятному співвідношенні з недоліками для окремих осіб, які мають місце в конкретній ситуації. Це дуже добре ілюструють специфікації комітету з етики щодо автономного водіння: з огляду на очікування, що використання штучного інтелекту в автономному водінні, ймовірно, різко зменшить кількість смертей на дорогах, його використання саме по собі є етично необхідним для захисту людського життя. Це також стосується випадків, коли в окремих випадках людям завдає шкоди несправний ШІ. Дискурс про відповідальність, таким чином, в першу чергу прояснює рольові відносини між людьми і машинами.

Не менш етично важливим і тому інтенсивно обговорюваним є питання про критерії, згідно з якими взагалі приймаються автономні рішення. Передусім це стосується людської гідності. Адже тут виникають класичні дилеми, які повинні бути заздалегідь продумані і перевірені на предмет їхньої доречності. Особа, яка приймає рішення, може зіткнутися з альтернативою вибору між двома варіантами, обидва з яких мають негативні наслідки. У літературі це відоме як проблема тролейбуса.

Упередженість і справедливість є важливими міркуваннями при прийнятті рішень щодо штучного інтелекту (ШІ). Системи штучного інтелекту настільки хороші, наскільки хороші дані, на яких вони навчаються, і якщо дані є упередженими або неповними, система штучного інтелекту буде видавати упереджені або неповні результати. Це може призвести до несправедливого ставлення до певних осіб або груп, що може мати серйозні соціальні та економічні наслідки. Тому вкрай важливо вирішити питання упередженості та справедливості у прийнятті рішень щодо ШІ. «Упереджені результати ШІ виникають через упереджені навчальні дані. Але самі по собі дані не є упередженими за своєю суттю, а справжня причина упередженості ШІ полягає в тому, що люди, відповідальні за надання навчальних даних, мають упередження або не помічають упереджень у своїх даних» [30].

Однією з ключових проблем у вирішенні питання упередженості та справедливості в ШІ є той факт, що системи ШІ можуть навчатися упередженості на основі даних, на яких вони навчаються. Наприклад, якщо система ШІ навчається на даних, які упереджено ставляться до певної групи людей, вона може навчитися дискримінувати цю групу в процесі прийняття рішень. Це може бути особливо проблематично в таких сферах, як працевлаштування, фінанси та кримінальне правосуддя, де упереджене прийняття рішень може мати серйозні наслідки для окремих осіб і суспільства в цілому.

«Такий хід думок наводить на думку про поняття алгоритмічна дискримінація. Якщо наявність певних психічних станів у осіб, які приймають рішення, є необхідною умовою для того, щоб рішення було дискримінаційним, можна стверджувати, що алгоритмічні системи прийняття рішень ніколи не можуть бути дискримінаційними як такі, тому що такі системи не здатні мати відповідні психічні стани» [29, с. 3]. Лише за існування машинної свідомості, якщо припустити, що вони будуть взагалі створені, можна припустити, що тільки такий ШІ та автономні системи не будуть носіями таких станів, як презирство, злість або неповага та ін.

Щоб вирішити цю проблему, важливо забезпечити навчання систем штучного інтелекту на різноманітних і репрезентативних даних. «Це означає, що дані, які використовуються для навчання систем штучного інтелекту, повинні містити приклади широкого кола осіб і груп, а також бути вільними від упереджень і забобонів» [11, с. 64]. Крім того, важливо забезпечити регулярний аудит систем штучного інтелекту, щоб виявити і виправити будь-які упередження, які могли бути засвоєні в процесі навчання.

Ще одним важливим моментом у вирішенні проблеми упередженості та справедливості в ШІ є необхідність забезпечення прозорості та підзвітності в процесах прийняття рішень. Це означає, що "системи штучного інтелекту повинні бути прозорими і зрозумілими, щоб рішення, які вони приймають, були зрозумілими і піддавалися ретельній перевірці» [5, с. 157]. Крім того, важливо встановити чіткий порядок підзвітності за рішення, ухвалені системами ШІ, щоб

окремі особи та організації могли нести відповідальність за будь-які упередження або несправедливість, які можуть бути виявлені.

Нарешті, важливо залучати різні зацікавлені сторони до розробки і розгортання систем штучного інтелекту. Сюди входять люди з різним досвідом, в тому числі ті, на кого можуть вплинути рішення, прийняті системами ШІ. Залучаючи різні зацікавлені сторони до розробки та розгортання систем ШІ, ми можемо гарантувати, що ці системи будуть розроблятися і використовуватися в чесний і справедливий для всіх спосіб.

Прозорість і підзвітність є важливими факторами при прийнятті рішень щодо ШІ. Це пов'язано з тим, що системи ШІ можуть бути складними і важкими для розуміння, що ускладнює для людей і організацій оцінку якості рішень, які приймаються цими системами. Крім того, рішення, прийняті системами ШІ, можуть мати етичні або правові наслідки, що може вимагати прозорості та підзвітності для забезпечення чесності та справедливості цих рішень.

Підзвітність у прийнятті рішень щодо ШІ означає здатність окремих осіб і організацій нести відповідальність за рішення, прийняті системою ШІ. Сюди входить визначення відповідальних за проектування, розробку і розгортання системи, а також встановлення чіткого розподілу відповідальності за будь-які рішення, прийняті системою. Підзвітність може бути досягнута за допомогою різних засобів, таких як створення правових рамок, структур управління або етичних принципів. «Прозорість і підзвітність важливі, оскільки вони можуть допомогти вирішити проблеми упередженості, дискримінації та несправедливості при прийнятті рішень у сфері ШІ» [37, с. 179].

Забезпечуючи прозорість процесу прийняття рішень, люди та організації можуть краще зрозуміти, як система прийшла до певного рішення, і оцінити, чи було це рішення чесним і справедливим. Встановивши підзвітність за рішення, прийняті системами штучного інтелекту, люди та організації можуть нести відповідальність за будь які негативні наслідки, які можуть виникнути в результаті цих рішень, що може слугувати стримуючим фактором для неетичної або незаконної поведінки.

2.2. Конфіденційність і безпека.

Оскільки технологія штучного інтелекту продовжує розвиватися і набуває дедалі ширшого застосування, занепокоєння щодо приватності та безпеки набуває все більшого значення. Існує низка питань, пов'язаних зі штучним інтелектом, які можуть мати значний вплив на конфіденційність і безпеку, зокрема, збір і використання даних, можливість витоку даних, а також використання штучного інтелекту в нагляді та правоохоронних органах.

Однією з головних проблем, пов'язаних зі штучним інтелектом і конфіденційністю, є збір і використання персональних даних. Багато систем ШІ для ефективного функціонування покладаються на великі обсяги даних, і ці дані можуть містити конфіденційну інформацію про людей. Наприклад, дані про стан здоров'я або фінансова інформація можуть збиратися і використовуватися в системах ШІ, що може створити ризики для конфіденційності, якщо ці дані будуть неправильно оброблятися або використовуватися не за призначенням [42, с. 126].

Ще одним потенційним ризиком, пов'язаним зі штучним інтелектом і конфіденційністю, є можливість витоку даних. Оскільки системи ШІ часто покладаються на великі обсяги даних, вони можуть бути вразливими до хакерських атак або інших форм кібератак. Якщо система ШІ скомпрометована, це може призвести до несанкціонованого доступу або крадіжки персональних даних, що може мати серйозні наслідки для приватності.

Використання ШІ в нагляді та правоохоронних органах викликає занепокоєння щодо приватності та громадянських свобод. Наприклад, технологія розпізнавання облич може використовуватися для ідентифікації осіб у громадських місцях, що може викликати занепокоєння щодо права на приватність і потенціалу зловживання або порушення цієї технології з боку правоохоронних органів або інших організацій.

Захист даних і права власності є важливими аспектами у сфері штучного інтелекту, оскільки системи штучного інтелекту значною мірою покладаються

на дані і можуть генерувати цінні ідеї та результати, які мають економічну або комерційну цінність.

Одним із ключових питань, пов'язаних із захистом даних, є право власності та контроль над даними, що використовуються системами штучного інтелекту. «У деяких випадках окремі особи або організації можуть надавати дані системам ШІ без повного розуміння того, як ці дані будуть використовуватися або хто матиме контроль над ними. Це може створювати ризики для конфіденційності та захисту даних, а також етичні проблеми, пов'язані з використанням персональних даних» [6, с. 26]

Ще однією проблемою, пов'язаною із захистом даних у ШІ, є потенційна можливість упередженості або дискримінації даних, які використовуються для навчання систем ШІ. Якщо системи ШІ навчаються на упереджених або дискримінаційних даних, це може призвести до упереджених або дискримінаційних результатів і рішень, які можуть мати негативні наслідки для окремих осіб і суспільства в цілому.

Що стосується прав власності, то питання про те, кому належать результати, отримані системами штучного інтелекту, є складним і часто оспорюваним. Дехто стверджує, що результати, створені системами ШІ, повинні вважатися власністю осіб або організацій, які створили систему, тоді як інші стверджують, що ці результати слід розглядати як форму колективного знання, що належить суспільству в цілому.

Для вирішення цих питань важливо розробити чітку політику і правила щодо захисту даних і прав власності у сфері ШІ. «Це може включати такі заходи, як закони про захист даних, вимоги до прозорості використання даних і керівні принципи справедливого та етичного використання ШІ. Крім того, окремі особи та організації можуть вживати заходів для захисту власних даних та інтелектуальної власності, наприклад, за допомогою шифрування або патентів» [15, с. 363].

Тут наведемо приклади деяких з них. Загальний регламент Європейського Союзу про захист даних (GDPR) [18] - це всеосяжний закон про захист даних,

який застосовується до всіх організацій, що обробляють персональні дані фізичних осіб в ЄС. Регламент містить конкретні положення, пов'язані зі штучним інтелектом і автоматизованим прийняттям рішень, і може стати корисною основою для розуміння питань захисту даних у контексті штучного інтелекту.

Всесвітня організація інтелектуальної власності (ВОІВ) - це міжнародна організація, яка надає ресурси та інформацію, пов'язану з інтелектуальною власністю, включаючи патенти, авторські права та торгові марки. ВОІВ опублікувала низку звітів і досліджень на тему ШІ та інтелектуальної власності, які можуть надати корисну інформацію за допомогою ШІ.

AI Now Institute [4] - дослідницька організація, яка займається вивченням соціальних та економічних наслідків ШІ. Інститут AI Now проводить діагностику та практичні політичні дослідження, спрямовані на боротьбу з концентрацією влади в технологічній індустрії. Вони опублікували низку звітів і статей на теми захисту даних, приватності та прав власності у сфері ШІ, які можуть надати цінну інформацію з цих питань.

Центр демократії і технологій (CDT) - неприбуткова організація, яка працює над захистом громадянських свобод і прав людини в цифрову епоху. Вони опублікували низку звітів і статей на теми штучного інтелекту, конфіденційності та захисту даних, які можуть надати корисну інформацію про правовий і політичний ландшафт навколо цих питань.

Глобальна ініціатива IEEE з етики автономних та інтелектуальних систем - це спільна робота промисловості, академічних кіл та громадянського суспільства, спрямована на розробку етичних керівних принципів і стандартів для ШІ. Вони опублікували низку звітів і статей на теми прозорості, підзвітності та справедливості у прийнятті рішень щодо ШІ, які можуть надати цінну інформацію про етичні міркування, пов'язані зі штучним інтелектом.

З розвитком штучного інтелекту (ШІ) зростає занепокоєння щодо потенційної небезпеки моніторингу та контролю. Системи штучного інтелекту

можуть бути розроблені для збору величезних обсягів даних, їх аналізу і прийняття рішень на основі цього аналізу. Це може дати організаціям і урядам безпрецедентну владу для моніторингу і контролю над окремими людьми і суспільством в цілому. «Ірраціонально вважати, що машина, яка в сотню або тисячу разів розумніша за нас, буде нас любити або захоче захистити. Це можливо, але жодних гарантій немає. Сам по собі ШІ не відчує подяки до людей за те, що його створили, - якщо, звісно, подяку не буде в ньому запрограмовано заздалегідь. Машини аморальні, і вважати інакше - небезпечно» [7, с. 56].

Системи штучного інтелекту можуть бути використані для моніторингу людей у спосіб, який раніше був неможливий. Системи штучного інтелекту можуть бути спроектовані так, щоб впливати на поведінку людини у витончений спосіб. Наприклад, алгоритми рекомендацій можуть використовуватися для формування контенту, з яким люди стикаються в Інтернеті. Це викликає занепокоєння щодо автономії та свободи вибору. Контроль і маніпуляції є основними проблемами у сфері штучного інтелекту, особливо в контексті онлайн-платформ і соціальних мереж. Алгоритми рекомендацій, які зазвичай використовуються для того, щоб пропонувати користувачам контент на основі їхньої історії переглядів та інших даних, мають потенціал формувати інформацію та ідеї, які отримують люди, у непомітний, але значущий спосіб.

Однією з ключових проблем, пов'язаних з алгоритмами рекомендацій, є можливість виникнення «інформаційних бульбашок" та "ехо-камер". Це ситуації, коли люди отримують лише ту інформацію та ідеї, які підтверджують їхні існуючі переконання та погляди, що може посилити упередження та обмежити можливості для різноманітних поглядів та критичного мислення. «Алгоритми рекомендацій також можуть використовуватися для просування певних продуктів, ідей або політичних програм, які можуть впливати на поведінку людей та прийняття рішень» [31, с. 55].

Окрім того алгоритми рекомендацій можуть також використовуватися для більш прямого впливу на поведінку людей. Наприклад, платформи соціальних мереж звинувачують у використанні алгоритмів для маніпулювання емоціями та

поведінкою користувачів. Вибірково показуючи певні типи контенту та сповіщень, соціальні медіа-платформи можуть заохочувати користувачів проводити більше часу на своїй платформі або вдаватися до певних типів поведінки, наприклад, натискати на рекламу або ділитися контентом зі своїми мережами. Такі маніпуляції викликають значне занепокоєння щодо автономії та свободи вибору. «Якщо людей підштовхують до певної поведінки або поглядів без їхнього відома чи згоди, вони можуть бути не в змозі приймати повністю поінформовані рішення. Це може мати серйозні наслідки для демократії, суспільного дискурсу та індивідуальних прав і свобод» [27, с. 78].

Для вирішення цих проблем важливо розробити етичні настанови і правила використання рекомендаційних алгоритмів та інших форм ШІ, які можуть впливати на поведінку людини. Ці керівні принципи повинні надавати пріоритет прозорості та інформованій згоді, щоб люди розуміли, як використовуються їхні дані, і мали можливість відмовитися від певних типів алгоритмів або маніпуляцій. Крім того, важливо, щоб люди знали про потенційні ризики, пов'язані зі штучним інтелектом, і виступали за політику і практику, які ставлять на перше місце захист автономії і свободи вибору.

Існує також загроза кібербезпеки. Системи штучного інтелекту можуть бути вразливими до хакерських атак та інших загроз кібербезпеки. Якщо система ШІ скомпрометована, вона може бути використана для моніторингу або контролю над людьми в ненавмисний спосіб. Це викликає занепокоєння щодо безпеки та можливості зловмисників використовувати системи штучного інтелекту. Щоб вирішити ці проблеми, важливо розробити етичні керівні принципи і правила для розробки і використання ШІ. Ці керівні принципи повинні надавати пріоритет прозорості, справедливості та підзвітності, а також захищати права і свободи окремих осіб і суспільства в цілому. Крім того, важливо, щоб люди знали про потенційні ризики, пов'язані зі штучним інтелектом, і виступали за політику і практику, які ставлять на перше місце захист приватності, автономії та інших фундаментальних прав.

2.3. Соціо-політичний вплив.

Штучний інтелект має потенціал для значного впливу на суспільство в різних аспектах, як позитивних, так і негативних. Очікується, що ШІ зруйнує ринок праці, автоматизувавши роботу, яку раніше виконували люди. Це може призвести до безробіття в певних галузях і категоріях робочих місць, хоча також може створити нові робочі місця в інших сферах, таких як розробка та обслуговування ШІ.

Штучний інтелект має потенціал підірвати ринок праці та змінити характер роботи в різний спосіб. Ось кілька прикладів, які пропонує М. Форд щодо того як ШІ впливає на зайнятість і ринок праці [16].

Автоматизація: ШІ здатен автоматизувати завдання, які раніше виконувала людина. Це може призвести до втрати робочих місць у певних галузях і категоріях робочих місць, особливо на низькокваліфікованих і рутинних роботах. Наприклад, працівники складів і конвеєрів можуть бути замінені роботами.

Вимоги до навичок: Оскільки все більше робочих місць автоматизується, вимоги до навичок працівників, що залишилися, ймовірно, зростатимуть. Це означає, що працівникам потрібно буде здобувати нові навички та компетенції, щоб залишатися працездатними.

Нові можливості для працевлаштування: хоча деякі робочі місця можуть бути автоматизовані, нові робочі місця також будуть створені в таких сферах, як розробка та обслуговування ШІ, аналіз даних і цифровий маркетинг.

Підвищення продуктивності: ШІ може підвищити продуктивність, автоматизуючи рутинні завдання, зменшуючи кількість помилок і забезпечуючи більш ефективне прийняття рішень.

Вплив на заробітну плату: Вплив штучного інтелекту на заробітну плату поки що не зовсім зрозумілий, оскільки він залежить від багатьох факторів, таких як конкретна галузь, категорія роботи та рівень кваліфікації. Деякі експерти прогнозують, що ШІ може призвести до стагнації заробітної плати, тоді як інші

вважають, що він може призвести до підвищення заробітної плати для працівників з потрібними навичками.

Соціальна та економічна нерівність: Існує побоювання, що ШІ може посилити існуючу соціальну та економічну нерівність, надаючи переваги тим, хто має доступ до технологій і навичок, необхідних для роботи з ними, і залишаючи інших позаду.

Варто додати, що ШІ може змінити характер праці, створюючи нові типи робочих місць, дозволяючи більш гнучкий графік роботи і розмиваючи межу між роботою і дозвіллям.

Оскільки ШІ продовжує розвиватися і набувати все більшого поширення, важливо ретельно відстежувати його вплив на зайнятість і ринок праці, а також розробляти політику і програми, які допоможуть працівникам адаптуватися до мінливого світу праці.

«Від штучного інтелекту очікують, що він буде об’єктивним, але насправді таким чином намагаються вирішити проблему суб’єктивності позиції, яку виявляють деякі непрофесійні судді. Штучний інтелект дійсно не може бути суб’єктивним, але і об’єктивним він також бути не може. Проблема криється в іншому: думка судді по справі не може бути суб’єктивною, а має бути суб’єктною, втіленям професійного рівня пізнання об’єктивної дійсності. Але для цього потрібен професійний рівень свідомості» [5, с. 275].

Штучний інтелект має потенціал для увічнення соціальної нерівності та дискримінації різними способами. Системи штучного інтелекту покладаються на великі обсяги даних для прийняття рішень, і якщо дані, які використовуються для навчання цих систем, є упередженими, система також буде упередженою. Наприклад, якщо система ШІ навчається на даних, які упереджено ставляться до певної раси чи статі, вона також ухвалюватиме упереджені рішення. "Брак різноманітності у сфері розробки ШІ також може сприяти соціальній нерівності та дискримінації. Якщо люди, які розробляють системи ШІ, належать до вузької демографічної групи, вони можуть бути не в змозі виявити та вирішити проблеми упередженості та дискримінації у своїх системах» [35, с. 249].

Системи штучного інтелекту також можуть посилювати існуючі в суспільстві упередження. Наприклад, якщо алгоритм схвалення кредитів навчений на даних, які показують, що люди з певних районів менш схильні повертати кредити, він може продовжувати відмовляти людям з цих районів, увічнюючи існуючу нерівність.

Як згадувалося раніше, ШІ може порушити роботу ринку праці та змінити характер роботи. Якщо ці зміни непропорційно вплинуть на певні групи людей, це може поглибити існуючу соціальну нерівність. Тут важливо розуміти, що вплив на ринок праці буде неминучим проте він створить і нові робочі місця. «В принципі, вони можуть виконувати будь-яку символічну роботу, від математики до логіки і мови. Але цифрових письменників ще не існує, тому люди досі пишуть всі книги, які з'являються в списках художніх бестселерів. Ми також ще не комп'ютеризували роботу підприємців, генеральних директорів, науковців, медсестер, офіціантів у ресторанах та багатьох інших типів працівників» [14, с. 20].

Для вирішення цих проблем важливо забезпечити, щоб системи штучного інтелекту розроблялися і впроваджувалися у справедливий, прозорий і неупереджений спосіб. Це передбачає різноманітне представництво в розробці систем ШІ, ретельний відбір і моніторинг навчальних даних, а також регулярний аудит і тестування систем ШІ на предмет упередженості та дискримінації. Це також вимагає ширшої суспільної дискусії про етичні наслідки ШІ і про те, як ми можемо забезпечити справедливий розподіл його переваг.

Серед інших можливих наслідків у суспільстві є те, що штучний інтелект має потенціал революціонізувати війну, що також викликає важливі етичні та безпекові проблеми, особливо в контексті озброєння. ШІ може бути використаний для посилення військових можливостей у різні способи, від розробки автономних систем озброєнь до поліпшення ситуаційної обізнаності і прийняття рішень.

Однак використання ШІ в озброєннях піднімає кілька складних питань, зокрема, питання підзвітності, прозорості та потенціалу непередбачуваних

наслідків. Одне з головних занепокоєнь, пов'язаних із застосуванням ШІ в озброєннях, полягає в тому, що ці системи можуть працювати автономно, без людського втручання і контролю. Автономна зброя, також відома як роботи-вбивці, є особливо суперечливим застосуванням ШІ у війні. «Ці системи можуть бути запрограмовані на виявлення і ураження цілей без прямого контролю з боку людини, що викликає занепокоєння щодо підзвітності і потенціалу ненавмисного завдання шкоди» [33, с. 45].

Крім того, використання штучного інтелекту в озброєннях порушує питання прозорості та підзвітності. Якщо автономна зброя буде розроблена і розгорнута, може бути важко визначити відповідальність за її дії або гарантувати, що вона працює безпечно і етично. Це може мати серйозні наслідки для цивільного населення і підірвати міжнародні норми і закони, що регулюють ведення війни. Нарешті, існують побоювання, що ШІ може посилити існуючий дисбаланс сил між державами і недержавними суб'єктами. Якщо одна країна або організація розробить передові можливості ШІ, вона може отримати значну перевагу у війні, що потенційно може призвести до дестабілізації та конфлікту.

«Моделі штучного інтелекту вдосконалюються щодня і демонструють свою цінність у багатьох сферах застосування. Продуктивність цих систем може зробити їх дуже корисними для таких завдань, як ідентифікація основного бойового танка Т-90 на супутниковому знімку, ідентифікація важливих цілей у натовпі за допомогою розпізнавання облич, переклад тексту для відкритих джерел інформації та генерація тексту для використання в інформаційних операціях» [25]. Для вирішення цих проблем лунають заклики до посилення регулювання і нагляду за застосуванням ШІ в озброєннях. Дехто пропонує укласти міжнародний договір про заборону розробки і використання автономної зброї, інші закликають до більшої прозорості та підзвітності в розробці і розгортанні зброї зі штучним інтелектом.

Загалом, використання ШІ в озброєннях викликає серйозні етичні проблеми та занепокоєння з точки зору безпеки. Хоча ШІ має потенціал для посилення військового потенціалу, важливо забезпечити, щоб його розробка і

використання були безпечними, етичними і відповідали міжнародним нормам і законам, що регулюють ведення бойових дій. «Проте ми повинні протистояти спокусі цієї технології, що відроджується. Розміщення вразливих систем штучного інтелекту в спірних сферах і покладання на них відповідальності за прийняття критично важливих рішень відкриває можливість для катастрофічних результатів. У цей час люди повинні залишатися відповідальними за ключові рішення » [23].

Висновок Штучний інтелект стає все більш важливою сферою досліджень і розробок, що має потенціал для трансформації багатьох аспектів життя суспільства. Однак, як і у випадку з будь-якою потужною технологією, існує низка етичних міркувань, які необхідно враховувати при розробці та розгортанні систем ШІ.

Одними з ключових етичних міркувань у сфері ШІ є упередженість і справедливість. Системи штучного інтелекту настільки хороші, наскільки хороші дані, на яких вони навчаються, і якщо дані є упередженими або неповними, система штучного інтелекту буде видавати упереджені або неповні результати. Це може призвести до несправедливого ставлення до певних осіб або груп, що може мати серйозні соціальні та економічні наслідки. Щоб вирішити цю проблему, важливо забезпечити, щоб системи ШІ навчалися на різноманітних і репрезентативних даних, а їхні процеси прийняття рішень були прозорими і справедливими.

Ще одним етичним аспектом ШІ є конфіденційність і безпека. Системи ШІ часто покладаються на великі обсяги персональних даних, що викликає занепокоєння щодо конфіденційності та безпеки. Важливо забезпечити, щоб дані збиралися, зберігалися і використовувалися відповідально і прозоро, а також щоб були вжиті відповідні заходи безпеки для захисту їх від зловживань або несанкціонованого доступу. Це особливо важливо з огляду на зростаючу поширеність кібератак і витоків даних.

Підзвітність і відповідальність також є важливими етичними вимогами в ШІ. Оскільки системи штучного інтелекту стають дедалі складнішими та автономнішими, стає дедалі важче розподіляти відповідальність за їхні дії. Важливо встановити чіткі межі підзвітності та відповідальності для систем ШІ, особливо коли вони використовуються в таких критично важливих сферах, як охорона здоров'я, фінанси та транспорт. Це допоможе забезпечити відповідальність окремих осіб і організацій за дії систем штучного інтелекту, а також запровадити відповідні запобіжники для запобігання шкоді.

ШІ також викликає етичні проблеми, пов'язані з людською гідністю та автономією. Хоча системи штучного інтелекту мають потенціал для посилення людської гідності та автономії, вони також викликають занепокоєння щодо потенційної втрати людського контролю та свободи дій. Важливо забезпечити, щоб системи ШІ розроблялися і використовувалися таким чином, щоб поважати і зміцнювати людську гідність і автономію, а також щоб люди могли зберігати контроль над використанням систем ШІ.

Отже, автономія і прийняття рішень є критично важливими аспектами ШІ, які викликають низку важливих етичних міркувань. До них відносяться питання, пов'язані з відповідальністю і підзвітністю, прозорістю, упередженістю і справедливістю, а також контролем з боку людини. Вирішення цих питань вимагатиме співпраці та взаємодії з широким колом зацікавлених сторін, включаючи дослідників, політиків, лідерів індустрії та організації громадянського суспільства. Працюючи разом, ми зможемо забезпечити розробку і впровадження ШІ таким чином, щоб максимізувати його потенціал на користь суспільству, мінімізуючи при цьому ризики.

РОЗДІЛ ІІІ
ШЛЯХИ ПОМ’ЯКШЕННЯ ЕТИЧНИХ РИЗИКІВ У СФЕРІ ШТУЧНОГО ІНТЕЛЕКТУ

3.1. Регуляторні рамки і політика як соціальні амортизатори негативного впливу ШІ.

Сприяння зменшенню етичних ризиків у сфері штучного інтелекту має важливе значення з кількох причин. Як було сказано у попередніх розділах, системи ШІ можуть бути упередженими та посилювати існуючі нерівності. Просуваючи етичні стратегії пом'якшення наслідків, ми можемо мінімізувати ці упередження і забезпечити справедливість і недискримінацію при застосуванні ШІ.

Програми штучного інтелекту часто потребують великих обсягів персональних даних і існує ризик, що ці дані можуть бути використані не за призначенням або піддані зловживанням. Впроваджуючи етичні стратегії пом'якшення наслідків, ми можемо захистити конфіденційність і захист даних.

Програми штучного інтелекту можуть приймати рішення, які мають значний вплив на окремих людей і суспільство в цілому. Вкрай важливо забезпечити прозорість і підзвітність цих рішень. Розробка і розгортання систем штучного інтелекту мають далекосяжні соціальні наслідки. Етичні стратегії пом'якшення наслідків можуть допомогти забезпечити врахування цих наслідків і соціально відповідальну розробку та впровадження ШІ.

Загалом, сприяння зменшенню етичних ризиків у сфері ШІ має важливе значення для забезпечення того, щоб ці системи розроблялися і впроваджувалися відповідальним і етичним чином, що принесе користь суспільству в цілому. «Важливо вирішувати ці проблеми на випередження, щоб запобігти непередбачуваним наслідкам і гарантувати, що ШІ слугуватиме загальному благу» [40, с. 93].

Оскільки використання штучного інтелекту стає все більш поширеним, зростає потреба в нормативно-правовій базі та політиці, які б гарантували, що

ШІ розробляється та впроваджується відповідально та етично. «Нормативно-правова база та політика у сфері ШІ повинні містити чіткі визначення того, що таке ШІ та які види діяльності підпадають під їхню сферу застосування. Це допоможе забезпечити послідовне застосування нормативно-правових актів і політик та уникнути плутанини» [31, с.22].

Ризик-орієнтований підхід може допомогти виявити потенційні ризики, пов'язані із застосуванням ШІ, і відповідно розставити пріоритети у регуляторному втручанні. Цей підхід також може допомогти збалансувати потребу в інноваціях з потребою в захисті людей і суспільства.

ШІ є глобальною проблемою, і міжнародне співробітництво має важливе значення для розробки послідовної та узгодженої нормативно-правової бази та політики. Співпраця між різними країнами може допомогти встановити загальні стандарти та найкращі практики для розробки та впровадження ШІ. Розробка нормативно-правової бази та політики у сфері ШІ вимагає знань з різних дисциплін, зокрема комп'ютерних наук, права, етики та соціальних наук. Об'єднання експертів з цих галузей може допомогти забезпечити належну інформованість та ефективність нормативно-правових актів і політик. «Гнучкість і адаптивність: ШІ - це сфера, що швидко розвивається, тому нормативно-правова база і політика повинні бути гнучкими і адаптивними, щоб не відставати від темпів змін» [23]. Це допоможе забезпечити актуальність та ефективність нормативно-правових актів і політик у вирішенні нових етичних і соціальних проблем, пов'язаних зі штучним інтелектом. Загалом, розробка нормативно-правової бази та політики у сфері ШІ має важливе значення для забезпечення того, щоб ШІ розроблявся і впроваджувався відповідально і етично, що принесе користь суспільству в цілому. Застосовуючи проактивний підхід і співпрацюючи, політики і регулятори можуть допомогти забезпечити використання ШІ таким чином, щоб максимізувати його потенціал і мінімізувати ризики.

У центрі уваги дискурсу щодо етичних вимог до співіснування та роботи людей і штучного інтелекту – питання про передачу людських можливостей

машинам. Це розширює дискурс про етичні вимоги автономної відповідальності, включаючи аспекти поступової передачі завдань від людей до машин. Зрештою, це призводить до питання про роль людини у співіснуванні з машинами. Лише за останні п’ять років ШІ помітно проник у світ життя та роботи. Роботи-хірурги надають допомогу в лікарнях, розумні помічники підтримують людей в офісі та в повсякденному житті, а інтелектуальні датчики дозволяють автономне водіння в пробках. У результаті взаємодія з інтелектуальними системами постійно зростає, а «штучне» в AI, англійська абревіатура від AI, стає «доповненим».

ШІ стає всюдисущим. Переваги, а також ризики стають дедалі очевиднішими. Наприклад, коли алгоритми дискримінують, оскільки вони вивчають інформацію про дії людей [36], коли автономні системи більше не можна контролювати, коли чат-боти домінують у публічних дискусіях або коли алгоритми штучного інтелекту використовуються у злочинних цілях. Не кажучи вже про небезпеку інтелектуальних систем зброї [22, с. 8]. Той факт, що штучний інтелект бере на себе все більше і більше завдань, іноді стає тут проблемою. І виявляється, що спільне життя з ШІ не обов’язково має бути кращим.

Саморозуміння людей формується залежно від рішення, чи вступають вони в екзистенціальну конкуренцію з ШІ, чи вони зберігають перевагу через позицію «суперінтелекту», яка чітко регулюється, і ШІ допомагає їм дуже контрольовано виконувати послідовні завдання.

Це вимагає етичних та інституційних заходів, які не обмежують технологічні можливості, але визначають ризики. Тоді мова йде не лише про те, хто вирішує та хто несе відповідальність, а й про те, хто зберігає контроль у взаємодії між людиною та машиною, і про те, як люди можуть підтримувати та зрештою примушувати власну волю. Відповісти на ці запитання зовсім не просто через величезну складність систем штучного інтелекту, обсяг даних і величезну швидкість розробки. Окремий користувач навряд чи зможе оцінити, наскільки правильно працюють алгоритми і правильні дані. І навіть розробники часом дивуються різкій продуктивності ШІ. Таким чином, створення прозорих і

моніторингових структур, стандартів і моделей санкцій є основною вимогою для етично відповідального використання.

Критична обробка систем штучного інтелекту та їх результатів в кінцевому підсумку створює принаймні соціально прийнятну пропускну здатність ШІ. Для цього потрібна довіра, заснована на більшому людському суверенітеті та компетентності у роботі з ШІ. Останніми роками багато учасників — не лише критики штучного інтелекту, але й політики та, не менш важливе, великі спекулянти штучного інтелекту, такі як Google або IBM — стали активними щодо цього питання.

Організація Об’єднаних Націй також досліджує потенціал ШІ з етичної точки зору. Такі ініціативи, як AINOW або OpenAI, хочуть встановити глобальні стандарти для ШІ, щоб демократизувати їх і, перш за все, зробити їх менш орієнтованими на еліти. Кожен повинен мати можливість використовувати ШІ, і ШІ повинен приносити користь усім [38].

Google, Apple, Facebook і Amazon (GAFA) надають інструменти штучного інтелекту як відкриті джерела, щоб сприяти суверенітету штучного інтелекту та отримати вигоду від використання в розумінні ройового інтелекту. У «Партнерстві щодо ШІ», GAFA та інших рішеннях ШІ для глобальних проблем, наприклад для зупинки несправних систем ШІ [27].

Для Європи розглядається створення спеціалізованого дослідницького центру для штучного інтелекту, де європейські компетенції для дослідження алгоритмів штучного інтелекту будуть потужно об’єднані. Уряди та міжнародні організації все частіше визнають необхідність розробки політичних ініціатив, пов'язаних зі штучним інтелектом. Ось кілька прикладів міжнародних і національних політичних ініціатив у сфері ШІ:

Принципи OECP щодо ШІ: Організація економічного співробітництва та розвитку (OECP) розробила набір принципів для відповідальної розробки та впровадження ШІ. Принципи підкреслюють важливість прозорості, підзвітності та інклюзивності у розробці та використанні ШІ.

Стратегія ЄС у сфері AI: Європейський Союз розробив стратегію щодо ШІ, яка має на меті сприяти розвитку та впровадженню ШІ в Європі, гарантуючи при цьому, що він розробляється і використовується безпечним і надійним способом. Стратегія передбачає фінансування досліджень і розробок у галузі ШІ, розробку етичних принципів ШІ та створення Європейського альянсу ШІ для сприяння співпраці та діалогу з питань, пов'язаних зі штучним інтелектом.

Національна ініціатива США в галузі ШІ: Національна ініціатива США зі штучного інтелекту - це міжвідомча ініціатива, спрямована на прискорення розвитку штучного інтелекту в Сполучених Штатах. Ініціатива передбачає фінансування досліджень і розробок у галузі ШІ, розробку етичних норм для ШІ та створення державно-приватних партнерств для сприяння розвитку та використанню ШІ.

Сінгапурська стратегія AI: Уряд Сінгапуру розробив національну стратегію розвитку AI, яка має на меті використати потенціал AI для трансформації економіки та суспільства Сінгапуру, забезпечуючи при цьому відповідальне та етичне його використання. Стратегія передбачає фінансування досліджень і розробок ШІ, створення Консультативної ради з питань етики ШІ та розробку керівних принципів відповідального використання ШІ.

Ці ініціативи - лише кілька прикладів зростаючого визнання важливості розробки політичних ініціатив, пов'язаних зі штучним інтелектом, на національному та міжнародному рівнях. Застосовуючи проактивний і спільний підхід до розробки цих політик, уряди і міжнародні організації можуть допомогти забезпечити розробку і використання ШІ таким чином, щоб максимізувати його потенціал і мінімізувати ризики.

Загалом, якщо підсумовувати принципи організацій вище можна вивести кілька основних, які мають відношення до розробки та впровадження систем штучного інтелекту. Ці принципи спрямовані на те, щоб ШІ розроблявся і використовувався відповідально і з користю для суспільства.

Прозорість: Системи ШІ повинні бути прозорими і зрозумілими, щоб люди могли розуміти, як приймаються рішення.

Справедливість: Системи ШІ повинні бути розроблені таким чином, щоб уникати упередженості та дискримінації, а також справедливо ставитися до всіх людей.

Конфіденційність: Системи штучного інтелекту повинні розроблятися і використовуватися з повагою до приватного життя людей і захистом їхніх персональних даних.

Підзвітність: Розробники та оператори систем ШІ повинні нести відповідальність за свої дії, а також повинні існувати механізми для усунення будь-якої шкоди, заподіяної цими системами.

Безпека: Системи ШІ повинні розроблятися і впроваджуватися таким чином, щоб мінімізувати ризик заподіяння шкоди окремим особам і суспільству в цілому.

Людський контроль: Системи ШІ повинні бути спроектовані таким чином, щоб забезпечити людський контроль і нагляд, щоб люди могли втрутитися в разі потреби.

Етичне прийняття рішень: Системи ШІ повинні бути розроблені з урахуванням етичних принципів і цінностей, а також приймати рішення, які відповідають цим принципам.

Соціальна відповідальність: Розробники та оператори систем ШІ повинні враховувати соціальний вплив цих систем і прагнути розробляти та впроваджувати ШІ таким чином, щоб вони приносили користь суспільству в цілому.

Багато з цих принципів відображені в чинних законах і нормативних актах, таких як закони про захист даних, антидискримінаційні закони та закони про захист прав споживачів. Однак існує потреба в більш конкретному регулюванні та керівництві щодо розробки та використання систем ШІ, особливо в таких сферах, як автономні транспортні засоби, охорона здоров'я та кримінальне правосуддя.

Останніми роками було започатковано низку ініціатив, покликаних сприяти розробці правових та етичних принципів для ШІ. Наприклад, у 2018 році

Європейська комісія опублікувала набір етичних керівних принципів для розвитку ШІ, а в 2020 році уряд США опублікував набір принципів для розвитку надійного ШІ, що заслуговує на довіру. Крім того, низка організацій, таких як IEEE та Partnership on AI, розробили власні етичні принципи для ШІ. Усі ці ініціативи стосуються в першу чергу роботи з поточним ШІ, який люди вказують або програмують із самого початку [28].

Отже, етичні вимоги спочатку спрямовані не на машину, а на людину. Формулювання відповідних етичних вимог для значно розвинутішого ШІ буде більш вимогливим. Нам буде натомість потрібна справжня машинна етика, штучна мораль. Тоді це буде спрямовано не на людей, а на саму машину.

Найпростішим рішенням, яке обговорюється сьогодні, є постійний моніторинг складнішого штучного інтелекту, в той час як користувач-людина заздалегідь визначив завдання та все ще приймає всі етично відповідні рішення. У деяких випадках це, безсумнівно, можливо у майбутньому. Однак, якщо штучний інтелект використовується спеціально для заміни людської роботи, для швидкого прийняття рішень, щоб уникнути небезпек для людей або замінити самих людей як фактор ризику, постійний моніторинг нереалістичний у довгостроковій перспективі. Тому що все менше і менше ситуацій прийняття рішень буде передбачуваним, а моделі дій будуть програмованими. Тому, живучи разом з людьми, машини повинні самі діяти етично і навчитися діяти таким чином, спілкуючись з ними [41, с. 49]. Для цього машини повинні мати можливість взаємодіяти з навколишнім середовищем, адаптувати свої дії до мінливих умов і, зрештою, діяти незалежно. Деякі системи ШІ вже сьогодні можуть досягти таких моделей поведінки. Якщо штучний інтелект також розвиває здатність виводити причини своєї поведінки в обробці інформації - завдяки застосуванням переконань- бажань-намірів - він стає більш схожим на етично діючого суб’єкта [12, с. 133-134].

На цьому етапі розвитку ШІ ще не досяг цілісної здатності діяти, яка характеризує людей, а лише здатність оцінювати та визнавати етику. І це обмежено лише певним контекстом дії. Для того, щоб мати можливість діяти

цілісно та етично, він повиннен мати можливість виражати власну волю та розвивати свідомість. І те, і інше також означатиме відчуття емоцій, у тому числі тих, які можуть спричинити ірраціональну поведінку [32, с. 35].

У антропоцентричному світогляді потрібен доброзичливий штучний інтелект, який діє етично у спілкуванні з людьми. Ще краще: він має діяти навіть правильніше за людей. Саме тут є велика користь для співіснування людини та машини: кожна машина діє раціонально, тому що її дії не підлягають жодному емоційно обґрунтованому формуванню волі, а зовнішні впливи майже не порушують раціональності [33, с. 47]. Це робить їх більш обчислюваними, зрозумілими та, зрештою, контрольованими для спільного життя.

Етичні, більш раціональні моделі дій стають, певним чином, планами для самого суспільства. Така еволюційно сформована мораль стає дзеркалом, яке ШІ тримає перед людьми. Загалом, розробка і розгортання систем штучного інтелекту повинні керуватися правовими та етичними принципами, які визначають пріоритети прозорості, справедливості, конфіденційності, підзвітності, безпеки, людського контролю, етичного прийняття рішень і соціальної відповідальності. Ці принципи допоможуть гарантувати, що ШІ розробляється і використовується таким чином, щоб приносити користь окремим людям і суспільству в цілому.

Оскільки штучний інтелект стає все більш поширеним у суспільстві, зростає занепокоєння щодо етичних наслідків його використання. Двома ключовими етичними міркуваннями є ясність і можливість інтерпретації, а також справедливість і рівність. Ясність і зрозумілість - це здатність зрозуміти, як працюють системи ШІ і чому вони приймають ті чи інші рішення. Це важливо з кількох причин. По- перше, це дає змогу виявити та виправити помилки або упередження в алгоритмах. По-друге, це дає змогу пояснювати рішення, ухвалені системами штучного інтелекту, зацікавленим сторонам - клієнтам, регуляторам і політикам.

Це особливо важливо в таких галузях, як охорона здоров'я, де системи штучного інтелекту все частіше використовуються для прийняття життєво важливих рішень.

Існує кілька технічних рішень для покращення зрозумілості та інтерпретованості систем штучного інтелекту. Один із підходів - використання методів зрозумілого ШІ (XAI). XAI прагне зробити системи ШІ більш прозорими та зрозумілими, надаючи інформацію про те, як вони працюють і як приймають рішення. Наприклад, деякі методи XAI можуть генерувати візуалізації або текстові пояснення того, як AI-система прийшла до певного рішення. «Повернення до XAI - пошуку пояснювального ШІ, галузі, яка занепадала у 1980-х роках під час розробок ШІ, але знову з'явилася з розвитком технології машинного навчання. Його поява допомагає пояснити феномен "чорної скриньки", притаманний більшій частині сучасного ШІ, що, в разі успіху, може сприяти зміцненню довіри до машин ШІ як до однієї зі значних переваг цієї парадигми» [11, с. 23].

Інший підхід полягає у використанні методів машинного навчання, які можна інтерпретувати. Ці моделі можуть бути простішими для розуміння і пояснення, ніж складніші моделі машинного навчання, такі як глибокі нейронні мережі. Однак ці моделі можуть жертвувати точністю в обмін на інтерпретованість.

Справедливість і рівність у системах штучного інтелекту також є важливим етичним аспектом. «Системи штучного інтелекту можуть увічнити упередження та дискримінацію, якщо вони навчаються на упереджених даних або якщо самі алгоритми містять упередження. Це може призвести до несправедливих результатів для певних груп людей, таких як меншини або жінки» [23]. Існує кілька технічних рішень для підвищення справедливості та рівності в системах штучного інтелекту. Один із підходів полягає у використанні методів попередньої обробки даних, які зменшують упередженість даних. Наприклад, методи доповнення даних можна використовувати для збалансування представництва різних груп у даних. Інший підхід полягає у

використанні алгоритмів машинного навчання, які враховують обмеження справедливості під час навчання. Наприклад, деякі алгоритми машинного навчання можна модифікувати так, щоб вони не дискримінували певні групи людей.

Варто зазначити, що з обома цими технічними рішеннями пов'язані певні виклики. Наприклад, методи XAI не завжди дають повне уявлення про те, як працює система штучного інтелекту, а інтерпретованість може досягатися за рахунок точності. Аналогічно, алгоритми машинного навчання, що враховують справедливість, не завжди дають найточніші прогнози, а методи попередньої обробки даних не завжди ефективні для пом'якшення упередженості даних.

Окрім цих технічних рішень, існують також політичні та регуляторні підходи до забезпечення ясності, інтерпретованості, справедливості та рівності в системах штучного інтелекту. «Законодавство про заборону дискримінації та захист даних - це основні правові режими, які можуть захистити людей від дискримінації, спричиненої ШІ» [39, с. 41]. Наприклад, деякі країни запровадили правила, які вимагають від компаній надавати пояснення щодо рішень, ухвалених системами ШІ. Інші запровадили керівні принципи для розробки та розгортання етичних систем штучного інтелекту.

Отже, ясність і зрозумілість, а також справедливість і рівність є двома найважливішими етичними міркуваннями при розробці та впровадженні систем штучного інтелекту.

Існує кілька технічних рішень для поліпшення цих аспектів ШІ, зокрема зрозумілі методи ШІ, інтерпретовані моделі машинного навчання, методи попередньої обробки даних і алгоритми машинного навчання, що враховують питання справедливості.

Однак з цими технічними рішеннями пов'язані й певні проблеми, а також можуть знадобитися політичні та регуляторні підходи, щоб забезпечити етичну та відповідальну розробку і впровадження систем штучного інтелекту.

3.2. Етичне прийняття рішень при проектуванні та розробці ШІ. Етичне прийняття рішень при проектуванні та розробці ШІ має важливе значення для забезпечення того, щоб технології ШІ розроблялися і використовувалися відповідно до етичних принципів і цінностей. Етичні проблеми можуть виникати на всіх етапах процесу розробки ШІ: від вибору проблеми, яку потрібно вирішити, до збору і використання даних, проектування і розгортання системи ШІ. Першим кроком у прийнятті етичних рішень при проектуванні та розробці ШІ є «визначення етичних принципів і цінностей, які мають відношення до розробки та використання системи ШІ» [40, с. 24]. Це можуть бути такі принципи, як автономія, доброчинність, незлочинність і справедливість.

Після визначення цих принципів наступним кроком є розгляд того, як вони застосовуються до конкретної проблеми, яку вирішує система штучного інтелекту, і як їх можна включити в дизайн і розробку системи. Одним із підходів до включення етичних принципів у дизайн і розробку ШІ є створення мультидисциплінарної команди, до якої входять не лише експерти з ШІ, а й фахівці з етики, соціологи та інші зацікавлені сторони. Така команда може працювати разом для виявлення потенційних етичних проблем і розробки стратегій їх вирішення. Інший підхід полягає у використанні етичних рамок або керівних принципів, розроблених спеціально для ШІ. Наприклад, Глобальна ініціатива з етики автономних та інтелектуальних систем Інституту інженерів з електротехніки та електроніки (IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems) розробила набір етичних принципів для систем ШІ, які наголошують на прозорості, підзвітності та дотриманні прав людини [20]. Окрім врахування етичних принципів при проектуванні та розробці системи ШІ, важливо також враховувати потенційний вплив системи на різні зацікавлені сторони, включаючи кінцевих користувачів, працівників і суспільство в цілому.

Це вимагає взаємодії із зацікавленими сторонами протягом усього процесу розробки та врахування їхніх проблем і перспектив. Щоб забезпечити дотримання етичних принципів при розробці та використанні систем штучного

інтелекту, може знадобитися впровадження процесів етичної експертизи або створення наглядових органів, які зможуть відстежувати та оцінювати етичні наслідки застосування систем штучного інтелекту. Наприклад, Європейський Союз запропонував створити Європейську раду зі штучного інтелекту, яка відповідатиме за нагляд за розробкою та використанням ШІ в ЄС. "Нарешті, важливо визнати, що прийняття етичних рішень при розробці та створенні ШІ - це безперервний процес, який вимагає постійного моніторингу та оцінки» [10, с. 38]. У міру виникнення нових етичних проблем вони повинні вирішуватися своєчасно і ефективно, а дизайн і розробка систем ШІ повинні постійно адаптуватися, щоб забезпечити їх відповідність етичним принципам і цінностям.

Штучний інтелект- це міждисциплінарна галузь, яка передбачає інтеграцію різних сфер знань, таких як інформатика, математика, інженерія, психологія, етика та право. Такий міждисциплінарний підхід необхідний для забезпечення відповідального та етичного розвитку ШІ. Однією з ключових переваг міждисциплінарного підходу при розробці ШІ є можливість інтегрувати різні точки зору в процес розробки. Кожна галузь привносить свій унікальний досвід і точку зору, які можуть допомогти виявити і вирішити потенційні етичні та соціальні проблеми. Наприклад, науковець може розробити ефективні та точні алгоритми, в той час як фахівець з етики може забезпечити відповідність алгоритмів етичним принципам та суспільним цінностям. «У філософії моралі вивчаються різні принципи і правила, які обґрунтовують наші моральні судження і керують нашими діями. Для того, щоб реалізувати етичні обмеження в обчислювальних системах, багато досліджень посилаються на різні моральні теорії, які були розглянуті в філософії моралі. Якщо ці теорії сформулювати в обчислювальному вигляді, вони можуть слугувати зразком для проектування внутрішніх обмежень роботів» [21, с. 69]

Однак тут є дві складності, про які слід пам'ятати. По-перше, хоча моральні теорії відіграють керівну роль, вони не призначені для регулювання процесу мислення, коли ми здійснюємо моральну поведінку. Більшість моральних теорій використовуються як основа для критичного аналізу своїх минулих вчинків або

вчинків інших людей, а також для передбачення майбутніх дій. Іншими словами, підстави, надані моральними теоріями, використовуються для обґрунтування дій, а не для виконання дій.

Ще однією перевагою міждисциплінарних підходів є те, що вони дозволяють проводити більш комплексну оцінку ризиків. Розглядаючи різні точки зору, дослідники можуть виявити потенційні ризики, які могли бути неочевидними з однієї точки зору. Це може допомогти запобігти непередбачуваним наслідкам, таким як упереджене або дискримінаційне прийняття рішень. Крім того, міждисциплінарні підходи можуть допомогти забезпечити розробку ШІ з урахуванням впливу на суспільство. Наприклад, юрист може допомогти забезпечити відповідність систем ШІ чинному законодавству, а політик - гарантувати, що системи будуть сприяти добробуту людей.

Однак міждисциплінарні підходи не позбавлені проблем. Однією з головних проблем є комунікація та співпраця між різними галузями. Кожна дисципліна має власну мову і методи дослідження, що може ускладнювати ефективну комунікацію. Це може призвести до непорозумінь і затримок у процесі розробки. Щоб подолати цю проблему, важливо встановити чіткі комунікаційні канали та протоколи. Регулярні зустрічі між членами команди з різних дисциплін можуть допомогти сформувати спільне розуміння проекту та сприяти співпраці. Крім того, залучення експертів з різних галузей до процесу проектування та розробки з самого початку може допомогти забезпечити врахування всіх точок зору.

Отже, міждисциплінарні підходи мають вирішальне значення для відповідального та етичного проектування і розробки ШІ. Інтегруючи різні точки зору та досвід, ШІ можна розробити так, щоб він відповідав етичним принципам і суспільним цінностям, мінімізуючи при цьому потенційні ризики та непередбачувані наслідки. Хоча міждисциплінарні підходи пов'язані з певними труднощами, чітка комунікація та співпраця можуть допомогти подолати ці труднощі та сприяти ефективній міждисциплінарній співпраці.

ВИСНОВКИ

Штучний інтелект - це все більш важлива сфера досліджень і розробок, яка має потенціал для трансформації багатьох аспектів життя суспільства. Однак, як і у випадку з будь-якою потужною технологією, існує низка етичних загроз, які необхідно враховувати при розробці та впровадженні систем штучного інтелекту.

Одними з ключових етичних міркувань у сфері ШІ є упередженість і справедливість. Системи штучного інтелекту настільки хороші, наскільки хороші дані, на яких вони навчаються, і якщо дані є упередженими або неповними, система штучного інтелекту буде видавати упереджені або неповні результати. Це може призвести до несправедливого ставлення до певних осіб або груп, що може мати серйозні соціальні та економічні наслідки. Щоб вирішити цю проблему, важливо забезпечити, щоб системи ШІ навчалися на різноманітних і репрезентативних даних, а їхні процеси прийняття рішень були прозорими і справедливими.

Конфіденційність і безпека також є важливими етичними аспектами ШІ. Системи ШІ часто покладаються на великі обсяги персональних даних, що викликає занепокоєння щодо конфіденційності та безпеки. З огляду на зростаючу поширеність кібератак і витоків даних, дуже важливо збирати, зберігати і використовувати дані відповідально і прозоро, а також вживати відповідних заходів безпеки для запобігання зловживанню або несанкціонованому доступу до них.

Підзвітність і відповідальність є ключовими етичними питаннями в ШІ. Оскільки системи ШІ стають дедалі складнішими та автономнішими, розподіл відповідальності за їхні дії стає дедалі складнішим. Встановлення чітких меж підзвітності та відповідальності для систем ШІ, особливо в таких критично важливих сферах, як охорона здоров'я, фінанси та транспорт, має вирішальне значення. Це гарантує, що окремі особи та організації нестимуть

відповідальність за дії систем штучного інтелекту і що будуть впроваджені запобіжні заходи для запобігання шкоді.

ШІ також викликає етичні занепокоєння щодо людської гідності та автономії. Хоча системи штучного інтелекту можуть посилити людську гідність і автономію, існують побоювання щодо потенційної втрати людського контролю і свободи дій. Важливо забезпечити, щоб системи ШІ розроблялися і використовувалися таким чином, щоб поважати і зміцнювати людську гідність і автономію, дозволяючи при цьому людям зберігати контроль над їх використанням.

Вирішення цих питань вимагає співпраці та взаємодії з різними зацікавленими сторонами, включаючи дослідників, політиків, лідерів індустрії та організації громадянського суспільства.

Працюючи разом, ми можемо гарантувати, що ШІ розробляється і впроваджується таким чином, щоб максимізувати його переваги і мінімізувати потенційну шкоду.

Етичне прийняття рішень при проектуванні та розробці ШІ має важливе значення для забезпечення того, щоб технології ШІ розроблялися і використовувалися відповідно до етичних принципів і цінностей.

Тому потрібен міждисциплінарний підхід, який включає етичні принципи в проектування і розробку систем ШІ, залучення зацікавлених сторін, а також механізми нагляду і безперервного оцінювання. Сприяючи прийняттю етичних рішень протягом усього процесу, від розробки до впровадження, ми можемо допомогти гарантувати, що технології ШІ розробляються і використовуються на благо суспільства в цілому.

Підсумовуючи, можна виділити кілька ключових етичних питань, які слід враховувати при розробці та впровадженні систем ШІ. Чіткість і зрозумілість, а також справедливість і рівність є найважливішими принципами. Існують різні технічні рішення для покращення цих аспектів, такі як зрозумілі методи ШІ, інтерпретовані моделі машинного навчання, методи попередньої обробки даних і алгоритми машинного навчання, чутливі до справедливості.

Однак проблеми залишаються, і для забезпечення етичної та відповідальної розробки і впровадження систем штучного інтелекту можуть знадобитися політичні та регуляторні підходи.

Щоб керувати розробкою та впровадженням систем штучного інтелекту, організаційні структури повинні дотримуватися кількох принципів. Прозорість має вирішальне значення, гарантуючи, що системи ШІ будуть зрозумілими і дадуть уявлення про процеси прийняття рішень. Справедливість наголошує на уникненні упередженості та дискримінації, сприяючи рівному ставленню до всіх людей. Конфіденційність підкреслює необхідність шанобливого ставлення до персональних даних. Підзвітність підкреслює відповідальність розробників і операторів, а також наявність механізмів для усунення будь-якої шкоди, заподіяної системами ШІ. Безпека фокусується на мінімізації ризиків для людей і суспільства. Людський контроль гарантує, що люди можуть втручатися, коли це необхідно. Етичність прийняття рішень підкреслює важливість узгодження систем ШІ з етичними принципами та цінностями. Нарешті, соціальна відповідальність підкреслює важливість врахування ширшого впливу систем ШІ на суспільство і прагнення до їх корисного застосування.

Дотримуючись цих принципів і сприяючи етичному прийняттю рішень, ми можемо орієнтуватися в складному ландшафті розвитку ШІ і гарантувати, що технології ШІ служать інтересам суспільства, дотримуючись при цьому етичних стандартів.

СПИСОК ДЖЕРЕЛ

1. Антологія сучасної аналітичної філософії, або жук залишає коробку / За наук. ред А. С. Синиці. Львів: Літопис, 2014. 374 с.

2. Токар Л. В. Штучний інтелект на варті справедливості: утопія чи перспектива людства. *Порівняльно-аналітичне право*. 2023. №1.

3. Блозва Л. М. Штучний інтелект: філософсько-антропологічний погляд. *Філософія і політологія в контексті сучасної культури*. 2016. Вип. 4.

4. AI Now Insitute.URL: https:// ainowinstitute.org/publication/toxic-competition

5. Algorithmic Fairness: Choices, Assumptions, and Definitions Annual Review of Statistics and Its Application - Vol. 8:141-163. 2021

6. Arkin, R. Governing Lethal Behavior in Autonomous Robots, Chapman and Hall/CRC. 2009

7. Barrat J. Our final invention. Artificial Intelligence and the End of the Human Era. 2013

8. Boden M. Artificial Intelligence. Handbook of Perception and Cognition, 2nd Ed, Academic Press Inc.1996

9. Broussard Meredith. Artificial Unintelligence : How Computers Misunderstand the World. The MIT Press, 2018

10. Buxmann and Schmidt. Künstliche Intelligenz Springer Berlin Heidelberg. 2019

11. Cathy O'Neil.Weapons of math destruction: how big data increases inequality and threatens democracy / First edition. New York: Crown Publishers. 2016.
URL:https://edisciplinas.usp.br/pluginfile.php/4605464/mod_resource/ntent/1/%28FFLCH%29%20LIVRO%20Weapons%20of%20Math%20Destruction%20-%20Cathy%20ONeal.pdf

12. Dennett, D. The Intentional Stance, Cambridge MA.1987

13. Domingos, P. The master algorithm: How the quest for the ultimate learning machine will remake our world. 2015

14. Erik Brynjolfsson and Andrew McAfee: Brynjolfsson and McAfee are co-authors of the book "The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. 2016

15. Floridi, L; Sanders. On the Morality of Artificial Agents. In: Minds and Machines. S. 349–379. 2004

16. Ford M. Rise of the robots : technology and the threat of a jobless future / Martin Ford, 2015

17. From Building Brains to Brained Buildings: An Interview with Michael. URL: https:// brainworldmagazine.com/from-building-brains-to-brained-buildings-an-interviewwith-michael-a-arbib/

18. General Data Protection Regulation. URL: https://gdpr-info.eu

19. Goodman M. Future Crimes. 2015.

20. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/ industry-connections/ec/autonomous-systems/

21. Journal of Philosophy of Life Vol.13, No.1:112-124 Clockwork Courage A Defense of Virtuous Robots. 2023. URL: https://www.philosophyoflife.org/jpl2023si_book.pdf

22. Kleinberg, J.; Lakkaraju, H.; Leskovec, J.; Ludwig, J.; Mullainathan, S. .Human Decisions and Machine Predictions, NBER Working Paper. 2017

23. Manyika J, Silberg J, Presten B. What Do We Do About the Biases in AI? 2019. URL: https://hbr.org/2019/10/ what-do-we-do-about-the-biases-in-ai

24. Miroslav Kubat. An Introduction to Machine Learning. Springer International Publishing AG .2017

25. Nick Bostrom Eliezer Yudkowsky. The Ethics of Artificial Intelligence. Draft for Cambridge Handbook of Artificial Intelligence, eds. William Ramsey and Keith Frankish . Cambridge University Press. 2011 URL: https://nickbostrom.com/ethics/artificial-intelligence.pdf

26. Penrouse R. Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford University Press, 1st edition.1994

27. Policy paper A pro-innovation approach to AI regulation, Presented to Parliament by the Secretary of State for Science, Innovation and Technology URL: https://www.gov.uk/government/ publications/ai-regulation-a-pro-innovation-approach/white-paper

28. Regulatory framework proposal on artificial intelligence. European Commission. URL: https://digital-strategy. ec.europa.eu/en/policies/regulatory-framework-ai

29. Reuben Binns. Fairness in Machine Learning: Lessons from Political Philosophy, Proceedings of Machine Learning Research 81:1–11, 2018

30. Rugare Maruzani. Bias in AI systems and efforts to fix it. URL: https://towardsdatascience.com/bias-in-ai-systemsand-efforts-to-fix-it-318798f5b39f

31. Ruha Benjamin. Race After Technology: Abolitionist Tools for the New Jim Code. Pollity. 2019

32. Scheutz, M.Architectural Roles of Affect and How to Evaluate Them in Artificial Agents, in: International Journal of Synthetic Emotions 2/2011, S. 48– 65. 2017

33. Singer P. W. Wired for War: The Robotics Revolution and Conflict in the Twenty-first Century, Penguin Press, 2009

34. Stuart J. Russell and Peter Norvig. Artificial intelligence : a modern approach / Description: Fourth edition. | Hoboken : Pearson. 2021

35. Stubbe, Julian, Künstliche Intelligenz, Neue Intelligenz, neue Ethik? . 2019

36. Süddeutsche Zeitung. Mehr Schutz vor Algorithmen. 2018. URL: https://www.sueddeutsche.de/wirtschaft/ kuenstliche-intelligenz-mehr-schutz-vor-algorithmen-1.3856354

37. Sune Hannibal Holm & Kasper Lippert. Rasismus, Discrimination, Fairness, and the Use of Algorithms/ Res Publica (2023)

38. TechCrunch: Discussing the limits of artificial intelligence. 2017. URL: https://techcrunch.com/2017/04/01/ discussing-the-limits-of-artificial-intelligence

39. The Ethical Algorithm: The Science of Socially Aware Algorithm Design Oxford University Press, Inc.198 Madison Ave. New York, NYUnited States. 2019

40. Walsh Toby. 2062. The World That AI Made. – La Trobe University Press. 2018

41. Wendell Wallach and Colin Allen: moral machines: teaching robots right from wrong. Oxford University Press; 1st edition, 2010

42. Yuval Noah Harari. 21 Lessons for the 21 Century. 2019