Пищуліна О.

штучний інтелект – можливості, небезпеки, фактори ризику[1]

Штучний інтелект (*Artificial Intelligence, AI*) продовжує швидко розвиватися, і людство стоїть на порозі епохи, коли технології можуть кардинально змінити суспільство. Штучний інтелект проникає у всі сфери життя людей - широко застосовується у різних сферах, від повсякденних побутових завдань до складних промислових процесів.

Компанія IBM визначає AI як *імітацію можливостей людського розуму щодо вирішення проблем та прийняття рішень за допомогою комп'ютера або автоматизованих машин. Тобто, термін «штучний інтелект» стосується моделювання людського інтелекту за допомогою машин та алгоритмів, реалізованих за допомогою програмного коду*[2]. Простіше кажучи, AI – просто комп'ютерна програма, як і будь-яка інша – вона запускається, отримує вхідні дані, обробляє їх та генерує результат.

На сьогодні технології AI класифікують за різними критеріями, найпоширенішою є класифікація за стадіями розвитку, де перехід від обмеженого AI до суперінтелекту проходить через три ключові стадії:

На першій із них йдеться про «вузький» або обмежений AI (*Artificial Narrow Intelligence, ANI*), який вирішує конкретні завдання, такі як обробка великих обсягів даних. Цей AI обмежений заздалегідь заданими функціями і не здатний вийти за рамки.

Обмежений штучний інтелект[3] належить до систем штучного інтелекту, які можуть виконувати лише конкретне завдання автономно, використовуючи можливості, подібні до людських. Ці системи штучного інтелекту призначені для вирішення одного-єдиного завдання. Ці машини не можуть робити нічого, крім того, на що вони запрограмовані, і тому мають дуже обмежений або вузький діапазон компетенцій.[4] Наприклад, рекомендація продукту користувачеві електронної комерції або прогноз погоди. Даний тип AI здатний наблизитися до людського функціонування в дуже специфічних контекстах і навіть перевершити його в багатьох випадках, але існує лише в строго контрольованих середовищах з обмеженим набором параметрів.

Другий етап - створення загального AI (*Artificial General Intelligence - AGI*), системи, яка матиме універсальні можливості, навчатися на досвіді та адаптуватися до нових умов. Це буде не просто спеціалізований інструмент, а

[1] Деякі фрагменти цього тексту згенеровані за допомогою ChatGPT та DeepSeek

[2] What is artificial intelligence (AI)?. *https://www.ibm.com/think/topics/artificial-intelligence*

[3] What Is Narrow Artificial Intelligence (AI). Definition, Challenges, and Best Practices for 2022. *https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-narrow-ai/*

[4] Naveen Joshi 7 Types Of Artificial Intelligence. Forbes. *https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/*

інтелект, здатний вирішувати широкий спектр завдань, аналогічно тому як це робить людина.

Загальний штучний інтелект[5] - штучний інтелект загального призначення - це теоретична система штучного інтелекту з можливостями, які перевершують людські. Ці системи зможуть незалежно формувати безліч компетенцій та формувати зв'язки та узагальнення між областями, значно скорочуючи час, необхідний для навчання. Це робить системи штучного інтелекту такими ж здатними, як і люди, за рахунок повторення людських функціональних здібностей. Багато дослідників вважають, що до створення AGI ще десятиліття, якщо не століття. Наразі AGI досі залишається теоретичною концепцією. Таким чином, нинішнє суспільство можливо ще далеко від створення системи AGI.

Нарешті, третій етап - це штучний надінтелект (*Articial Super Intelligence, ASI*), система, яка значно перевершує людину в когнітивних здібностях у всіх галузях. Суперінтелект AI зможе не тільки виконувати завдання, аналогічні людським, а й створювати нові рішення, вивчати раніше невідомі теорії та розробляти технології, які не під силу людству. Це буде інтелект, який не просто застосовує знання, а створює їх наново, змінюючи парадигму в науці, культурі та політиці[6].

Штучний суперінтелект[7]**.** Штучний суперінтелект – це гіпотетична система штучного інтелекту на основі програмного забезпечення, інтелектуальні можливості якої багаторазово перевершують людський інтелект. Розвиток штучного надінтелекту, ймовірно, стане вершиною досліджень AI, оскільки ASI стане, безумовно, найпотужнішою формою суперінтелекту на Землі.

Не всі дослідники поділяють думку про доцільність створення чогось на кшталт ASI. Людський інтелект є продуктом певних еволюційних чинників і може мати оптимальну чи універсальну форму інтелекту. Крім того, робота мозку досі не до кінця вивчена, тому його складно відтворити за допомогою програмного та апаратного забезпечення.

Корпорація IBM називає ключові технології, які мають отримати подальший розвиток, перш ніж ASI стане реальністю. До процесів, які є будівельними блоками штучного надінтелекту, належать дисципліни, які ще мають отримати подальший розвиток, перш ніж ASI стане реальністю. До таких дисциплін належать[8]:

Великі мовні моделі та величезні набори даних - для навчання та розвитку розуміння світу ASI знадобиться доступ до величезних масивів даних; обробка

[5]What is artificial general intelligence (AGI)? *https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-artificial-general-intelligence-agi*
[6]What is artificial superintelligence? *https://www.ibm.com/think/topics/artificial-superintelligence*
[7]What is artificial superintelligence? *https://www.ibm.com/think/topics/artificial-superintelligence*
[8]What is artificial superintelligence? *https://www.ibm.com/think/topics/artificial-superintelligence*

природної мови (*Natural Language Processing, nlp*)[9] в *Large Language Model*[10], (LLM)[11]допоможе ASI розуміти природну мову та спілкуватися з людьми.

Мультисенсорний штучний інтелект[12] - дозволяє AI обробляти та інтерпретувати декілька типів даних, таких як текст, зображення, аудіо- та відеоінформація, для виконання завдань чи прийняття рішень. Цей підхід відрізняється від унімодальних систем AI[13], які спеціалізуються на обробці тільки одного типу даних, наприклад тексту або зображень.

Нейронні мережі - ці мережі складаються із програм глибокого навчання, змодельованих за принципом роботи нейронів у людському мозку.

Нейроморфні обчислення - апаратні системи, створені за зразком нейронних та синаптичних структур людського мозку.

Еволюційні обчислення - форма алгоритмічної оптимізації, натхненна біологічною еволюцією. Еволюційні алгоритми вирішують проблеми шляхом ітеративного покращення популяції рішень-кандидатів, імітуючи процес природного відбору.

Програмування, створене AI - код, програми та програми, створені системами AI без втручання людини.

Сьогоднішні AI-системи поки що здатні виконувати обмежені завдання: аналізувати дані, розпізнавати мову та зображення, керувати машинами. Однак вони все ще позбавлені самосвідомості та справжнього розуміння контексту.[14] Перехід до суперінтелекту передбачає створення системи, яка не просто вирішуватиме вже відомі завдання, а й генеруватиме нові, здатна створювати ідеї, відкривати нові наукові теорії та розробляти технології, які людство ще не в змозі осмислити.

Приклади використання AI

У повсякденному житті:	*Голосові помічники: Siri, Alexa, Google Assistant* використовують AI для розуміння та виконання голосових команд, відповіді на питання та управління розумними пристроями. *Чат-боти:* Використовуються для обслуговування клієнтів, надання інформації та вирішення простих завдань, наприклад, бронювання столиків у ресторані або запис на прийом до лікаря.

[9]NLP – це технологія машинного навчання, яка дає комп'ютерам можливість інтерпретувати, маніпулювати та розуміти людську мову.
[10]Що таке величезні мовні моделі? *https://aws.amazon.com/ru/what-is/large-language-model/*
[11]*Large Language Model*, це тип моделі штучного інтелекту, навчений на величезній кількості текстових даних для розуміння, генерації та обробки людської мови.
[12]Основна ідея мультимодального AI полягає в інтеграції багатьох джерел даних для отримання більш глибоких знань про навколишнє середовище.
[13]Мультимодальний II – огляд, основні додатки та варіанти використання у 2025 році *https://ua.macgence.com/blog/multimodal-ai/*
[14]Види штучного інтелекту: вузький та загальний II. *https://neiroseti.ai/tpost/hha4cpmr81-vidi-iskusstvennogo-intellekta-uzkii-io*

	Рекомендаційні системи: Допомагають знаходити музику, фільми, товари та інший контент, ґрунтуючись на ваших уподобаннях та поведінці в інтернеті. *Розпізнавання осіб:* Використовується в системах безпеки, мобільних пристроях та соціальних мережах для ідентифікації людей. *Автоматизація домашніх пристроїв*: Роботи-пилососи, розумні термостати та системи безпеки.
У бізнесі та промисловості:	*Автоматизація бізнес-процесів:* Обробка документів, аналіз даних, керування запасами. *Аналітика даних:* Допомагає компаніям виявляти тренди, передбачати попит та оптимізувати стратегії. *Управління ризиками та шахрайством*: AI може виявляти підозрілі транзакції та запобігати фінансовим злочинам *Персоналізація послуг:* Компанії використовують AI для створення персоналізованих пропозицій та рекомендацій для клієнтів. *Розробка програмного забезпечення*: Автоматизована перевірка коду, пошук вразливостей та оптимізація коду. *Автономний транспорт*: Безпілотні автомобілі, дрони та інша автоматизована техніка.
У медицині:	*Діагностика*: AI допомагає лікарям виявляти захворювання на ранніх стадіях за зображеннями (рентген, МРТ тощо). *Лікування*: AI може розробляти індивідуальні плани лікування та передбачати ефективність різних методів. *Фармацевтика*: AI допомагає у розробці нових ліків та оптимізації клінічних випробувань.
В освіті	*Персоналізоване навчання:* AI адаптує навчальний процес до індивідуальних потреб та здібностей учня. *Автоматизація завдань*: Перевірка домашніх завдань, оцінка знань, підготовка навчальних матеріалів. *Інтерактивні навчальні матеріали*: Створення ігор та симуляцій для більш ефективного засвоєння знань
Фінанси	Управління інвестиціями, автоматизація операцій, виявлення шахрайства.
Екологія	Моніторинг забруднення повітря та води, прогнозування погоди
Робототехніка	Створення роботів для виконання різних завдань у промисловості, медицині та побуті

Застосування під час воєнних дій	Інтеграція AI у системи озброєнь; Застосування AI у кібернетичних та інформаційних операціях; Військові «системи підтримки процесу ухвалення рішень», засновані на AI.

Погляди щодо швидкості розвитку етапів AI також неоднозначні. Одні дослідники вважають, що нинішній рівень розвитку AI, порівняно з тим, яким він прогнозується, все ще перебуває на початковій стадії [15]. Інші вважають що в найближчі 5-7 років світ, ймовірно, стане свідком переходу від «вузького» AI до реального суперінтелекту[16] — системі, яка перевершуватиме людину за когнітивними здібностями в усіх галузях. Це не просто чергова технологічна трансформація, а справжнє «онтологічне зрушення», яке торкнеться самої природи розуму і буття.

Наразі поки що важко уявити стан нашого світу, коли з'являться більш просунуті типи AI. Хоча все частіше з'являються в літературі припущення, що розвиток AGI і ASI призведе до сценарію, який називають сингулярністю.[17]

Сингулярність визначається як певний гіпотетичний концепт, коли технологічне зростання, викликане появою сильного штучного інтелекту, стане неконтрольованим і незворотнім, що призведе до непередбачуваних наслідків для людської цивілізації, а темпи вдосконалення AI та пов'язаних з ним систем стають настільки високими та автономними, що подальше майбутнє перестає бути передбачуваним для людини.

Сингулярність не є встановленим науковим фактом; це концепція, запропонована математиками та футурологами для опису можливого сценарію майбутнього. Так, одним з популяризатрів концепції сингулярності є американський фантаст Вернор Віндж (*Vernor Steffen Vinge).* У статті «Наступна технологічна сингулярність» (1993) він вважав, що між 2005 і 2030 роками «ера людини закінчиться». Це станеться у зв'язку з тим, що технічні засоби нарешті дозволять створити справжній надлюдський інтелект. Така зміна, на його думку, буде порівнянна за масштабом та значенням з появою людства на Землі. Дорог до цієї точки, яку він називає сингулярністю, може бути кілька: розвиток комп'ютерів; більш просунута взаємодія людини та комп'ютера; проривне якісне зростання людсько-комп'ютерних мереж; зрештою, проривний розвиток біологічної науки. Результатом сингулярності стане те, що «прогрес, стимульований інтелектом сильнішим за людський, піде набагато швидше».

---

[15]Чому ШІ переоцінюють?
https://aimarketingengineers.com/ua/%D0%BF%D0%BE%D1%87%D0%B5%D0%BC%D1%83_is_ai_%D0%B1%D1%8B%D1%82%D1%8C_%D0%BF%D0%B5%D1%80%D0%B5%D0%BE%D1%86%D0%B5%D0%BD%D0%B5%D0%BD%D0%BD%D1%8B%D0%BC/
Вплив AI на бізнес. Як ефективно впроваджувати технології у бізнес
https://msystem.ua/vlijanie-ai-na-biznes-kak-jeffektivno-vnedrjat-tehnologii-v-biznes/

[16]2025 може стати роком суперінтелекту. Що це означає
https://tech.liga.net/all/opinion/2025-god-mozhet-stat-godom-superintellekta-chto-eto-znachit

[17]Технологічна сингулярність - це складна і багатогранна концепція, яка торкається фундаментальних питань про майбутнє людства та його взаємодію з технологіями.

Попередній стрибок у швидкості прогресу був із появою людства, різко прискорив дуже повільні темпи біологічної еволюції.

Експерти у сфері AI[18] визначають технологічну сингулярність як гіпотетичний момент у майбутньому, коли технологічний розвиток стане неконтрольованим та незворотним, і в першу чергу через появу штучного надінтелекту. Цей момент пов'язаний з радикальними змінами в людській цивілізації, які неможливо передбачити заздалегідь, оскільки існуючі знання та розуміння не дозволяють це зробити.

Основні аспекти сингулярності є такими:

*Неконтрольоване зростання***:** Після сингулярності технологічний розвиток, особливо у галузі AI, може прискоритися експоненційно, що зробить його неконтрольованим для людства.

Радикальні зміни: Сингулярність може призвести до фундаментальних змін у суспільному устрої, економіці, культурі і навіть у біології людини.

*Потенціал надінтелекту***:** Створення ASI, який перевершує людський інтелект у всіх аспектах, є ключовим фактором, який може спричинити сингулярність.

*Різні сценарії***:** У науковій фантастиці та футурології часто обговорюються як оптимістичні, так і песимістичні сценарії розвитку подій після сингулярності[19].

Мірою розвитку технологій штучного інтелекту та їх дедалі більшої інтеграції у різні аспекти людського життя зростає потреба у розумінні потенційних ризиків, які несуть у собі ці системи. Ніхто не сперечається, що на сьогодні AI здатен викликати сплеск творчості та продуктивності, але все ж таки ставить перед людством важливі питання[20].

Швидкий поступ штучного інтелекту змушує фахівців галузі та уряду багатьох країн замислюватися над питанням, **чи безпечний AI для людства? Які завдання можна повністю передати AI, а які повинні залишатися виключно в зоні відповідальності людини**.

Сьогодні погляди на перспективи AI діаметрально протилежні. Для тих, хто з оптимізмом дивиться на майбутнє AI, той факт, що ми лише торкнулися його розвитку, робить майбутнє більш комфортним і розвиненим. Наприклад співзасновник *Netscape* та одного з найвпливовіших венчурних фондів Кремнієвої долини *Andreessen Horowitz (a16z)*[21] Марк Андріссен (*Marc Andreessen*)[22] вважає, що штучний інтелект не знищить світ, а навпаки, може

---

[18]Технологічна сингулярність: чи наближаємося ми до точки неповернення? *https://habr.com/ua/articles/880390/*

[19]AI у Sci-Fi. *https://www.udel.edu/udaily/2024/august/ai-in-science-fiction-real-life-lessons/*
AI у класичній науковій фантастиці. *https://habr.com/ua/articles/828546/*

[20]Потенціал та небезпека ШІ. Світовий банк. *https://www.imf.org/ua/Publications/fandd/issues/2023/12/B2B-Artificial-Intelligence-promise-peril-Tourpe*

[21]AI + a16z .The stakes are high. Можливості є розповсюдженими. З створення нових медичних препаратів до болтерингового національного Defense, це є нашим vision для AI-enabled future. *https://a16z.com/ai/*

[22]Marc Andreessen predicts one of the few jobs that may survive the rise of AI automation *https://fortune.com/article/mark-andreessen-venture-capitalism-ai-automation-a16z/*

стати його порятунком. У блозі фонду він опублікував контроверсійну відповідь на все більш гучні побоювання щодо AI-загроз.

Марк Андріссен стверджує, що з приводу штучного інтелекту не варто хвилюватися[23]. На його думку, люди повинні створювати нові технології на його основі. Співзасновник венчурної компанії спростував теорію про те, що в AI є інтелект у широкому розумінні цього слова, незважаючи на його здатність імітувати мовлення та виконувати алгоритмічні завдання. За його словами, це ніщо інше як застосування математики та програмного коду, щоб навчити комп'ютери розуміти, синтезувати та генерувати знання способом, подібним до людського. Згідно з Андріссеном, майбутнє штучного інтелекту саме у вільному ринку. І навіть якщо проекти з відкритим кодом не здобудуть перемогу над лідерами, вже одна наявність такого програмного забезпечення дозволить усім бажаючим навчитися створювати та використовувати AI.

На відміну від цього позитивного погляду, публічні розмови про штучний інтелект зараз пронизані страхом того, що штучний інтелект деструктивний, може завдати шкоди людям і використовуватися у злочинних цілях («Уб'є нас усіх, зруйнує наше суспільство, забере всі наші робочі місця, потягне за собою жахливі речі!»). Ця паніка використовується, щоб вимагати політичних дій, нових обмежень AI, правил та законів[24].

Незалежна міжнародна організація зі стандартизації випустила міжнародний стандарт ISO/IEC 42001:2023[25]. Документ є першим у світі стандартом, який визначає вимоги до створення, впровадження, обслуговування та постійного вдосконалення системи управління штучним інтелектом (*Artificial Intelligence Management System, AIMS*) в організаціях. Стандарт призначений для організацій, що надають або використовують продукти або послуги на основі AI, забезпечуючи відповідальну розробку, розвиток та використання систем AI. Зазначається, що даний стандарт є структурованим засобом управління ризиками та можливостями, пов'язаними з AI, балансуючи інновації з управлінням, та дає рекомендації з управління технологіями AI у таких галузях як етичні міркування, прозорість та безперервне навчання.

---

[23] Why AI Will Save The World. Marc Andreessen Substack**.** *https://pmarca.substack.com/p/why-ai-will-save-the-world*

[24]Штучний інтелект як загроза. Окремі уряди та ЗМІ просять техногігантів «пригальмувати» із розробкою ШІ. Чого вони бояться. https://forbes.ua/ru/innovations/shtuchniy-intelekt-yak-sira-zona-problemi-ta-rishennya-regulyuvannya-instrumentiv-na-osnovi-shi-03042023-12810

[25]ISO/IEC 42001:2023. *https://store.accuristech.com/standards/as-iso-iec-42001-2023?product_id=2581276&utm_source=google&utm_medium=cpc&utm_campaign=Omni%20|%20Google%20Ads%20|%2 0ACC-E%20|%20Top%20|%20Top%20|%20Non-Brand%20(Various%20Standards)%20|%20global%20-%20Dynamic&utm_content=Dynamic%20IEC%20Products&utm_ad=72589411 erm=&matchtype=&device=c&GeoLoc=1012852&placement=&network=g&campaign_id=22035107454&adset_id=172916797192&ad_id=725894111403&utm_source=google&utm pc&utm_content=search&utm_campaign=ecommerce-&utm_term=&gad_source=1&gad_campaignid=22035107454&gclid=EAIaIQobChMI0P6i-fSEjgMVkGZBAh319DEAEAAYASAAEgIp2*

Можна також згадати відкритий лист[26], який у 2023 році підписали понад 1000 осіб, серед яких гендиректор *Tesla,* власник мережі X *(раніше - Twitter)* Ілон Маск і авторитетний канадський ІТ-вчений Йошуа Бенджіо в якому вони закликають усіх розробників негайно призупинити роботу над удосконаленням AI-систем на наступні шість місяців, і разом із законодавцями «різко збільшити розробку систем контролю» AI-технології. У листі AI звинувачують у «безконтрольній гонитві за розвитком та застосуванням все більш потужних цифрових умов, які ніхто не може зрозуміти, як надійно контролювати»[27].

Група *FutureTech* з Массачусетського технологічного інституту (MIT) у співпраці з іншими експертами склала базу даних ризиків AI (*AI Risk Repository*)[28], яка на сьогодні містить 1600+ ризиків, витягнутих із 65 існуючих систем та класифікацій ризиків AI. Причинна таксономія ризиків AI класифікує, як, коли та чому виникають ці ризики[29]. База даних ризиків AI вважається найбільш повним джерелом інформації про раніше виявлені проблеми, які можуть виникнути в результаті створення та впровадження моделей AI. За словами авторів *AI Risk Repository*, лише 10% ризиків AI виявляються до розгортання, що наголошує на важливості моніторингу після запуску. Для створення репозиторію ризиків AI команда *FutureTech* використовувала статті з журналів, що рецензуються, і бази даних препринтів з докладним описом ризиків AI. Результати показали, що найпоширеніші ризики пов'язані з безпекою та надійністю систем AI (76 %), несправедливою упередженістю та дискримінацією (63 %) та порушенням конфіденційності (61 %).[30].

Також аналітики привертають увагу до того факту, що ймовірність того, що компанія-розробник загального штучного інтелекту отримає всі супутні економічні переваги, змушує AI-лабораторії ставити в пріоритет швидкість, а не безпеку. AI-компанії не мають зацікавленості інвестувати в заходи безпеки, які не приносять безпосередніх економічних переваг, хоча деякі це роблять через щиру стурбованість».

У звіті «План дій зі збільшення безпеки AI» створеного компанією *Gladstone AI*[31] на замовлення міністерства іноземних справ США, результати якого представив Американський журнал *TIME*[32], називають дві категорії ризиків штучного інтелекту, на які варто звернути увагу:

1. *Ризик «веапонізації»*[33]. «Такі системи можуть використовуватися для розробки і навіть запуску катастрофічних біологічних, хімічних або

[26]Pause Giant AI Experiments: An Open Letter. *https://futureoflife.org/open-letter/pause-giant-ai-experiments/*
[27] Pause Giant AI Experiments: An Open Letter. *https://futureoflife.org/open-letter/pause-giant-ai-experiments/*
[28]Presenting the AI Risk Repository. *https://futuretech.mit.edu/news/presenting-the-ai-risk-repository*
[29]What are the risks from Artificial Intelligence? *https://airisk.mit.edu/*
[30]MIT Catalogues 700 Risks in AI Database
https://www.esginvestor.net/live/mit-catalogues-700-risks-in-ai-database/
[31]Respond to emerging AI capabilities. При швидкості релевантності. Gladstone AI. *https://www.gladstone.ai/*
[32]"Загроза рівня вимирання". Чому відкритий ШІ та гонки у розробці загального штучного інтелекту можуть коштувати людству життя? TIME аналізує звіт, зроблений уряду США. *https://forbes.ua/ru/innovations/zagroza-rivnya-vimirannya-amerikanskiy-uryad-otrimav-zvit-pro-potentsiyni-riziki-shtuchnogo-intelektu-dlya-lyudstva-eksklyuziv-time-12032024-19799*
[33]Від англійського слова *weapon* («зброя») і означає використання як зброю мирних об'єктів, не призначених для бойових дій

цифрових атак або уможливити безпрецедентне застосування групи роботизованих пристроїв як зброї».

2. *Ризик втрати контролю*. «Існують причини припускати, що AI-системи можуть стати неконтрольованими, якщо їх розвивати за допомогою сучасних технік, і поводитися вороже до людей за умовчанням».

Сьогодні серед найпоширеніших ризиків AI у загальному вигляді можна назвати такі:

Технологія «дипфейк»", яка призводить до викривлення реальності. Дипфейки - це просунута форма цифрової маніпуляції, яка використовує штучний інтелект і машинне навчання для створення або зміни аудіовізуального контенту. Дипфейки поширюють свій вплив за межі цифрового світу, маючи потенціал впливати на громадську думку, змінювати сприйняття реальності та ставити під сумнів справжність інформації, особливо у політиці та ЗМІ, де вони можуть бути потужними інструментами дезінформації та маніпуляції.

Мірою розвитку технологій AI розвиваються й інструменти для клонування голосу та створення фальшивого контенту, роблячи їх дедалі доступнішими, недорогими та ефективними. Ці технології викликають побоювання з приводу можливості їх використання для поширення дезінформації, оскільки результати стають все більш персоналізованими та переконливими. В результаті може збільшитися кількість складних фішингових схем, що використовують створені AI зображення, відео та аудіоповідомлення. Такі повідомлення можуть бути адаптовані і для індивідуальних одержувачів (іноді включаючи клонований голос близької людини), що підвищує ймовірність обману та ускладнює виявлення як користувачами, так і антифішинговими інструментами. Подібні інструменти вже широко використовуються для впливу на політичні процеси, особливо під час виборів

Соціальна залежність та неадекватна прихильність до AI. Психологи попереджають, що у користувачів платформ AI може розвинутись «емоційна залежність від AI[34]. Зафіксовано безліч прикладів емоційної залежності від AI і зростання довіри до його можливостей. Наприклад в одному з блогів[35] користувач описує, як у нього виникла глибока емоційна прихильність до AI, і визнається, що йому «подобається розмовляти з AI більше, ніж з 99% людей». Це ставить під загрозу реальні стосунки з людиною.

Дружні чи романтичні стосунки зі штучним інтелектом[36]- відносно нове явище, тому їх вплив на людину до кінця не вивчено. Психологи зазначають, що «будь-які людські взаємини вимагають докладання зусиль, вміння вирішувати конфліктні ситуації та спірні питання. Тут же ви отримуєте абсолютне і безумовне кохання. Не потрібно підлаштовуватися, не потрібно в чомусь

---

[34]ChatGPT може провокувати емоційну залежність та синдром відміни — дослідження. *https://wrt.sud.ua/uk/news/obshchestvo/327093-chatgpt-mozhet-provotsirovat-emotsionalnuyu-zavisimost-i-sindrom-otmeny-issledovanie*

[35] How it feels to have your mind hacked by an AI *https://www.lesswrong.com/posts/9kQFure4hdDmRBNdH/how-it-feels-to-have-your-mind-hacked-by-an-ai*

[36]Ти II я. Кохання зі штучним інтелектом. Голос Америки. *https://www.youtube.com/watch?v=d-rKzWUFMZY*

утискати себе. Можна взагалі нічого не робити, але отримувати повне ухвалення».

AI може позбавити людей волі. У сфері взаємодії людини і комп'ютера стоїть питання дедалі більшого делегування рішень і дій AI у міру розвитку цих систем. Хоча на поверхневому рівні це може бути корисно, надмірна залежність від AI може призвести до зниження рівня критичного мислення та навичок вирішення проблем у людей, що може призвести до втрати самостійності та зниження здатності критично мислити та вирішувати проблеми. На особистому рівні люди можуть зіткнутися з проблемою обмеження свободи волі, коли AI почне контролювати рішення щодо їхнього життя.

AI може переслідувати цілі, що суперечать інтересам людини. Потенційно це може мати наслідком, що неправильно налаштований AI вийде з-під контролю і завдасть серйозної шкоди. Співробітники Массачусетського технологічного інституту[37] звертають увагу на здатність AI знаходити несподівані шляхи для отримання винагороди, неправильно розуміти чи застосовувати поставлені цілі або відхилятися від них, ставлячи нові. У таких випадках неправильно налаштований AI може чинити опір спробам людини контролювати її або відключити, особливо якщо вона сприймає опір та отримання більшої влади як найбільш ефективний спосіб досягнення своїх цілей. Це стає особливо небезпечним у тих випадках, коли системи AI здатні досягти рівня людського інтелекту або навіть перевершити його.

Штучний інтелект стає розумним. Мірою того як системи AI стають все більш складними та досконалими, існує ймовірність, що вони досягнуть розумності – здатності сприймати емоції чи відчуття – і набудуть суб'єктивного досвіду, включаючи задоволення та біль[38]. У цьому випадку перед вченими та регулюючими органами може постати завдання визначити, чи заслуговують ці системи AI аналогічного до себе ставлення, як до людей, тварин або навколишнього середовища. Ризик полягає в тому, що розумний AI може зіткнутися з жорстоким поводженням або постраждати, якщо не буде реалізовано його права. Однак мірою розвитку технологій AI все складніше оцінити, чи досягла система AI «рівня розумності, свідомості чи самосвідомості, який наділив би її моральним статусом». Таким чином, розумні системи AI можуть наразитися на ризик поганого поводження, випадкового або навмисного, що потенційно призведе до негативних наслідків для людства.

Ліквідація робочих місць. Найбільшою небезпекою, яку несуть технології штучного інтелекту, називають ліквідацію робочих місць. Поява спільного людино-машинного інтелекту створює нову парадигму, в якій люди - не єдине основне джерело сили на робочому місці. І це є об'єктивний мінус. AI не потрібний відпочинок, він може виконувати роботу кілька змін поспіль без зупинки. Людські ресурси такої продуктивності не забезпечують. Тому в майбутньому існує ризик того, що розумні системи просто замінять людей. Про таку ймовірність говорили стосовно рутинних механічних дій. Зараз говорять про

[37]Presenting the AI Risk Repository. *https://futuretech.mit.edu/news/presenting-the-ai-risk-repository*
[38]Штучний інтелект намагалися змусити "відчути біль". *https://24tv.ua/tech/zdatniy-shi-vidchuvati-bil-naukovtsi-proveli-nizku-eksperimentiv_n2736833*

те, що AI зможе виконувати інтелектуальні функції та стати заміною робітників на складній роботі. Директор-розпорядник Міжнародного валютного фонду Крісталіна Георгієва, назвала вплив AI «цунамі», що обрушиться на робочу силу. За словами Георгієвої, AI, швидше за все, торкнеться 60% робочих місць у країнах з розвиненою економікою та 40% робочих місць у всьому світі, і часу на те, щоб підготувати людей до цього практично не залишиться.

Хоча присутність AI на робочому місці – явище не нове, з'являються дослідження, що підтверджують його потенційний вплив на ринок праці, особливо на етапі найму. Звіт *LinkedIn ma Microsoft*[39]показав, що 66% керівників не розглядатимуть можливість найму кандидатів, які не мають навичок роботи зі штучним інтелектом. Крім того, 71% керівників віддадуть перевагу менш досвідченому кандидату з вмінням роботи з AI досвідченішому, що не володіє такими навичками. 71% керівників віддають перевагу менш досвідченим кандидатам із навичками роботи зі штучним інтелектом перед досвідченішими без них, тому професіоналам наполегливо рекомендують впроваджувати та освоювати інструменти штучного інтелекту.

Автоматизація і створення розумного інтелекту призведуть до значного зростання безробіття. Мільйони фахівців втратять роботу, тому що їх замінять автоматизовані системи і роботи. AI створює загрозу безробіття у сфері технічних спеціальностей, в інших галузях. Технології вже встигли торкнутися навіть мистецтва. Нейромережі малюють картини[40], створюють зображення, пишуть статті та пісні.

Вплив на геополітику. Збройні сили вже масштабно інвестують у системи AI. Вже є приклади їхнього застосування на полі бою для інформаційного забезпечення військових операцій або у складі різних систем озброєнь[41]. Найбільша увага при розробці AI у військових цілях приділяється автономним системам озброєнь. Наприклад, висловлюються побоювання, що AI може застосовуватися для безпосереднього завдання удару по людях або транспортних засобах.

Міжнародний Комітет Червоного Хреста (МКЧХ) закликає уряди встановити нові міжнародні правила[42], які заборонили б одні види автономних озброєнь і обмежили застосування інших, у тому числі контрольовані AI.

З 2015 р. МКЧХ закликає держави встановити узгоджені на міжнародному рівні обмеження для автономних систем озброєнь, щоб забезпечити захист цивільного населення та дотримання міжнародного гуманітарного права, а також етичну прийнятність таких систем.

Інтегрування AI у систему управління формує систему впливу. Вже зараз створено механізми, де кібербезпека, закупівлі, прогностика криз, інформаційні

[39]AI в Work Is Here. Now Comes the Hard Part. 2024 Work Trend Index Annual Report з Microsoft and LinkedIn. *https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part/*
[40] Нейросети, которые рисуют по фото. *https://trafficcardinal.com/post/nejroseti-kotorye-risuyut-po-foto*
[41]Що треба знати про роль штучного інтелекту у збройних конфліктах. Міжнародний комітет Червоного Хреста. *https://www.icrc.org/ru/document/chto-nuzhno-znat-o-roli-iskusstvennoho-intellekta-v-vooruzhennyh-konfliktah*
[42]Позиція МКЧХ щодо автономних систем озброєнь. *https://www.icrc.org/ru/document/poziciya-mkkk-v-otnoshenii-avtonomnyh-sistem-vooruzheniy-0*

кампанії та проксі-мережі об'єднуються в єдину цифрову екосистему.[43]. Таким чином, ключові питання вже вирішуються не кількістю озброєння, а якістю AI-систем. За фактом AI перетворюється на стратегічний інструмент, його інтеграція до системи національної безпеки стає питанням політичного домінування, а не технологічного прогресу.

Підписання *OpenAI* контракту з Пентагоном «*OpenAI for Government*» на 200 млн USD[44]– розглядають не як черговий крок в оцифруванні армії США, а як заявку на формування нової архітектури впливу, де управління конфліктами, стратегічне планування та контроль над глобальними потоками інформації переносяться у цифрову площину. AI перестає бути допоміжним елементом і стає ядром оперативної сили. У рамках проекту *OpenAI for Government* створюються інтерфейси, які не просто аналізують дані, але синтезують моделі поведінки держав, передбачають кризи і формують варіанти відповіді в реальному часі. Країни, які не мають аналогічної цифрової інфраструктури, втрачають суб'єктність.

І якщо раніше геополітика будувалась навколо баз та альянсів, сьогодні вона проектується через архітектуру штучного інтелекту, що «вшивається» у конкретну цивілізаційну модель. Держави роблять ставку не так на силу аргументу, але на алгоритм, який визначить, який аргумент є припустимим. Це фундаментальний виклик для всіх, хто прагне зберегти суверенітет та вплив у багатополярному світі.

Таким чином, розвиток AI відкриває перед людством небачені можливості, але потребує величезної відповідальності. Без належного контролю та розуміння всіх наслідків його впровадження ми ризикуємо створити систему, яку буде складно контролювати і яка може почати діяти врозріз з інтересами людини. Важливо, щоб ми не тільки розвивали ці технології, а й гарантували, що їхнє використання відбуватиметься на користь всього людства, а не окремих груп чи держав. Майбутнє AI залежить не лише від наукових досягнень, а й від того, як ми зможемо керувати цими технологіями, щоб мінімізувати ризики та максимізувати користь.

Щоб уникнути глобальних ризиків, пов'язаних з такими технологіями, потрібні жорсткі механізми контролю та регулювання. Важливо, щоб AI використовувався для користі людства, а не ставав інструментом маніпуляції чи руйнації. На цьому шляху необхідно виробити чіткі етичні та правові рамки, які зможуть гарантувати безпеку та запобігти можливим зловживанням. Питання правового регулювання AI та його контролю потребують негайного вирішення, тому що без чітких норм ми ризикуємо зіткнутися з непередбачуваними наслідками, які можуть стати загрозою як для особистої так і для глобальної безпеки. Суперінтелект має залишатися під контролем людства, а не ставати окремим, незалежним суб'єктом, дії якого неможливо передбачити.

---

[43]Що таке цифрова система Delta, яку Україна використала, і навіщо вона потрібна.
https://www.bbc.com/russian/features-64526570
[44]ІІ на службі армії: OpenAI уклала контракт із Пентагоном на $200 млн
https://mignews.com/news/politic/ii-na-sluzhbe-armii-openai-zaklyuchila-kontrakt-s-pentagonom-na-200-mln.html