



ШТУЧНИЙ ІНТЕЛЕКТ ЯК ТРЕНД ЦИФРОВИХ ЗМІН У СФЕРІ ПУБЛІЧНОГО УПРАВЛІННЯ ТА АДМІНІСТРУВАННЯ

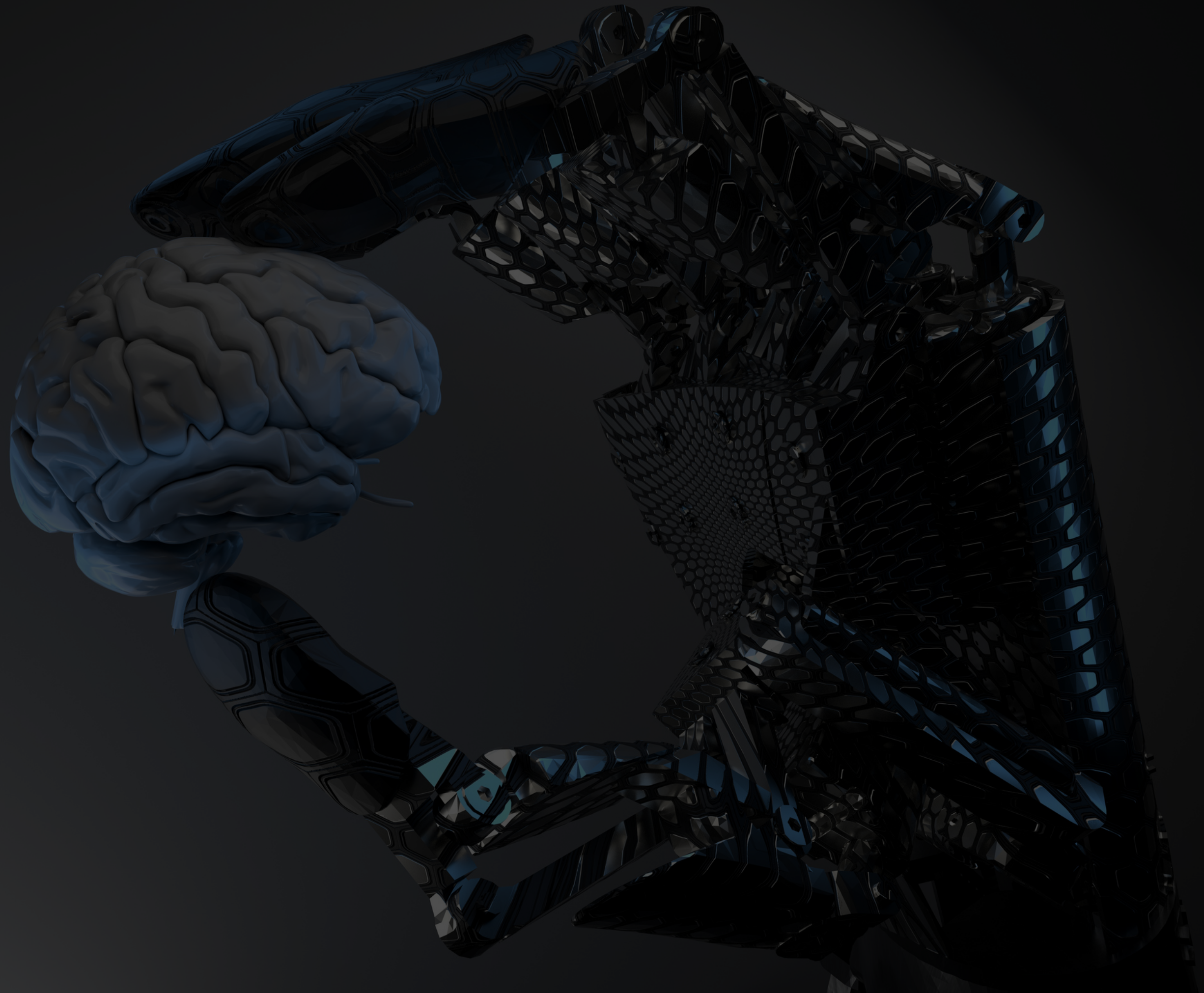


Орищук Василь
23/02/2024

Тематична структура освітнього бранчу поділена на:

1. Основи штучного інтелекту та його роль у цифровому розвитку.
Штучний інтелект як інструмент публічного управління.
2. Етика та правові аспекти використання штучного інтелекту публічними службовцями.

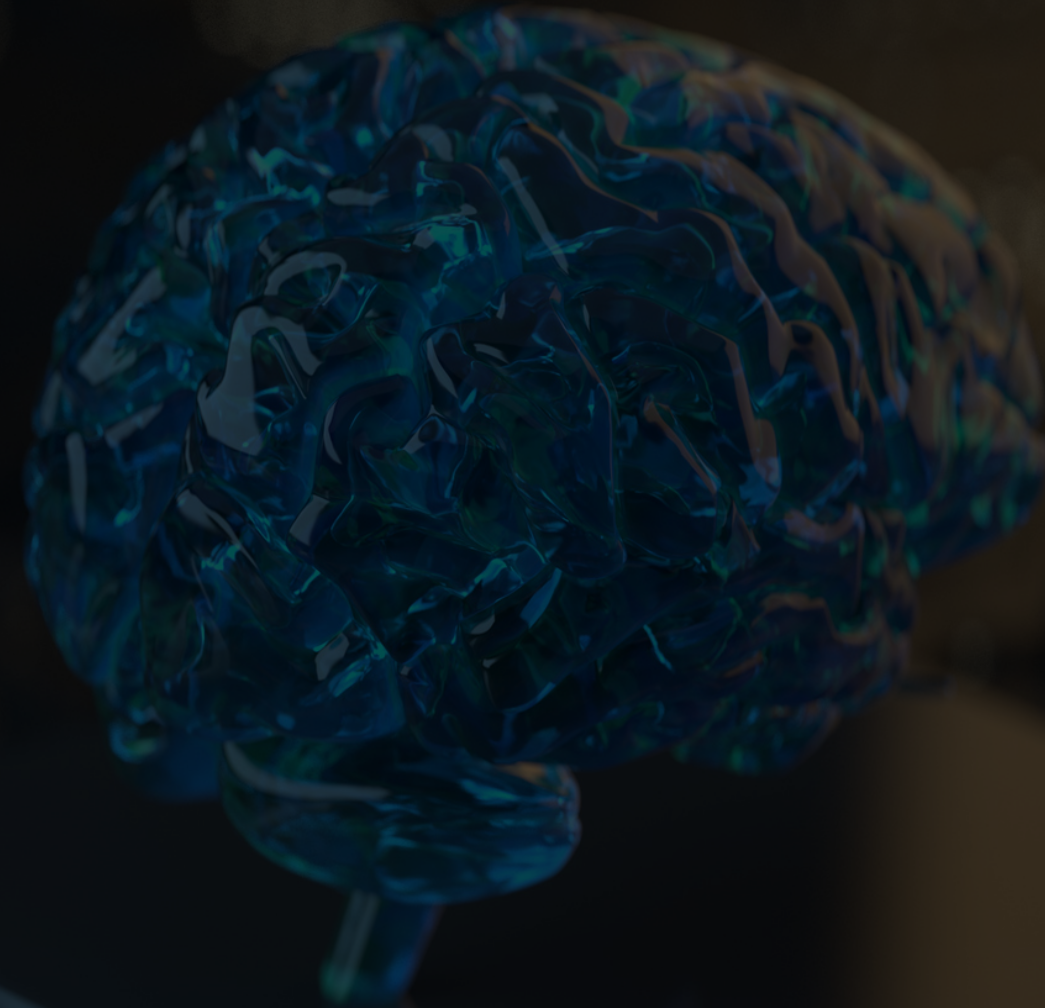
Як Ви
розумієте
термін
“Штучний
інтелект”?



Поняття штучного інтелекту

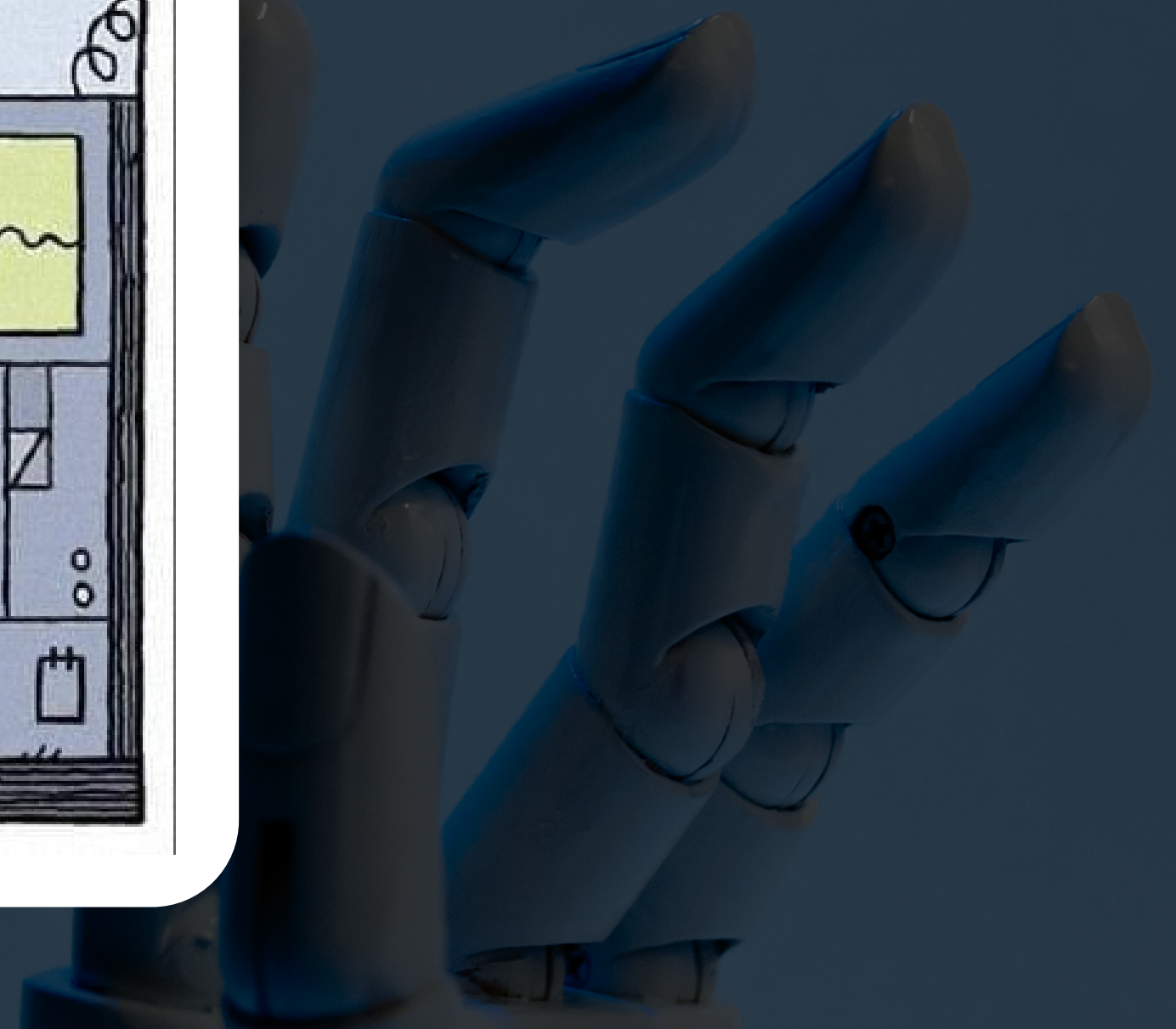
Штучний інтелект (ШІ) – це галузь інформатики, яка намагається відтворити поведінку людського розуму за допомогою комп'ютерних програм. Метою ШІ є створення інтелектуальних машин, здатних мислити і діяти раціонально.

Штучний інтелект – це властивість систем коректно інтерпретувати зовнішні дані відповідно до поставленої мети, навчатися з таких даних та використовувати результати навчання для досягнення поставлених цілей, в тому числі зі збирання та використання нових даних, шляхом взаємодії з навколишнім середовищем. Така властивість систем реалізується через алгоритми і методи, робота яких можлива завдяки обладнанню для обчислюванню та збирання даних, комунікації з іншими системами, взаємодії та впливу на навколишній світ (за **Концепцією розвитку штучного інтелекту в Україні**).



AI – головне слово 2023 року

Словник Collins Dictionary оголосив "AI" (штучний інтелект) **головним словом 2023 року**. Повідомляється, що "реальні можливості цієї технології ще не вивчені, але вона вже здійснила революцію".



Ефект плацебо формує наше сприйняття ШІ

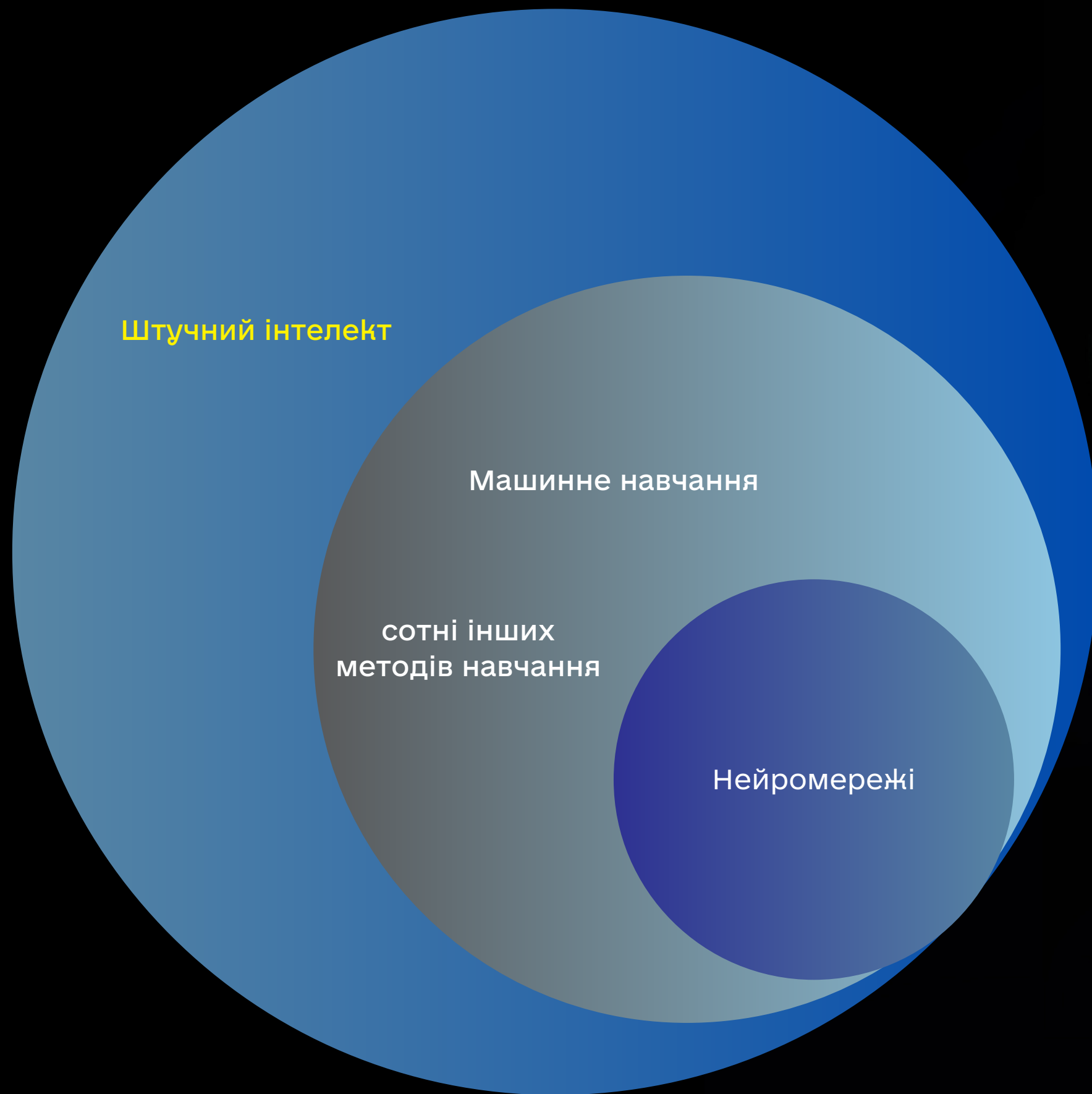
- Упереджені уявлення, які люди мають про ШІ, формують їхній досвід роботи з цією технологією.

«Створення ШІ може бути як найбільшим успіхом в історії нашої цивілізації, так і найгіршою подією. Ми просто не знаємо. Невідомо, чи буде ШІ безкінечно допомагати нам, чи просто відсуне нас вбік, розвиваючись все ширше, чи знищить нас»

Стівен Хокінг

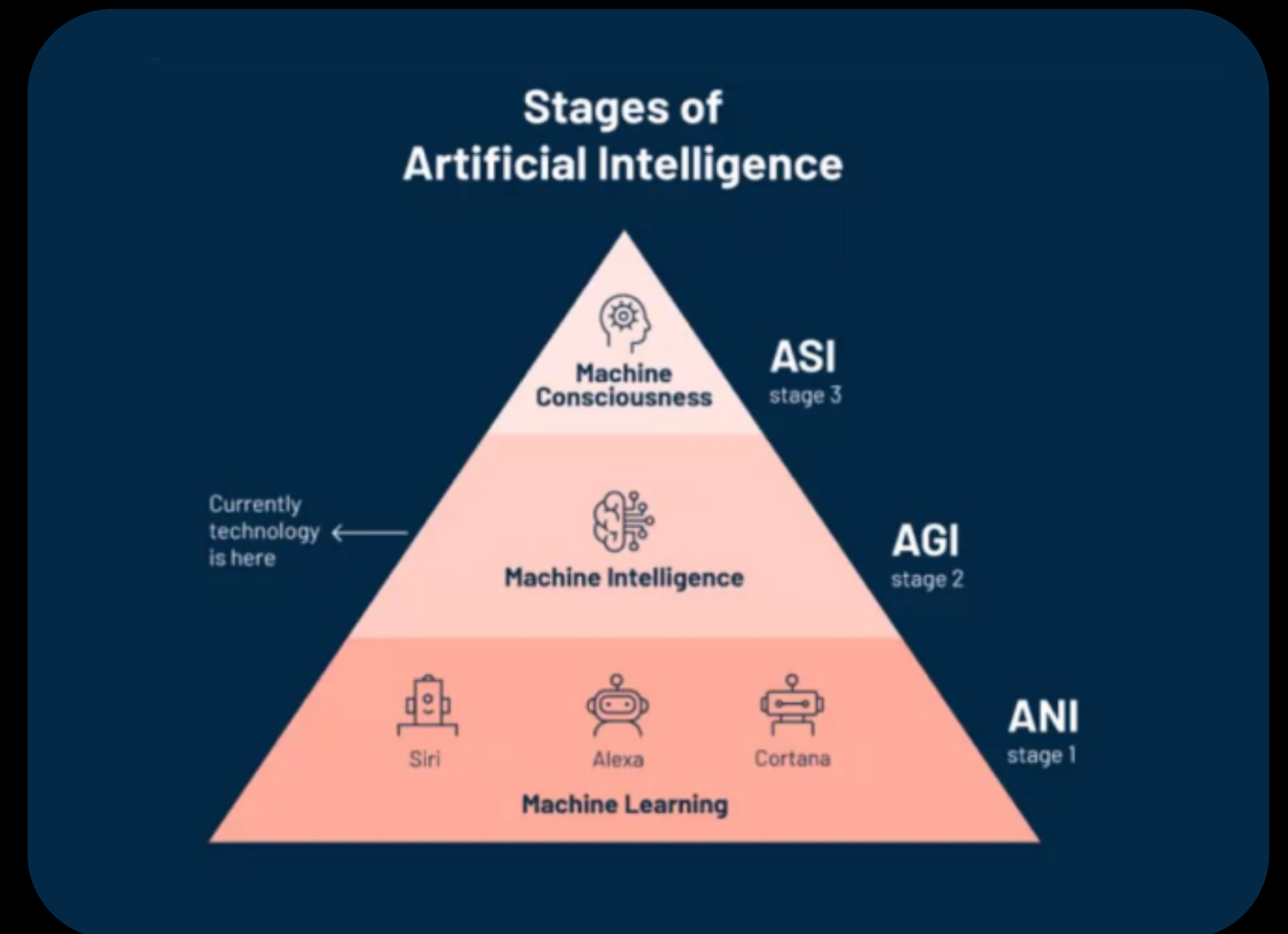
ШІ включає такі **основні напрямки:**





Рівні розвитку штучного інтелекту:

- **Штучний вузький інтелект (Artificial Narrow Intelligence, ANI):** Найпростіший рівень ШІ, який спеціалізується на одній задачі або наборі задач. Прикладами ANI є голосові помічники, такі як Siri, Alexa та Cortana, які можуть відповідати на питання, керувати певними функціями смартфонів або домашньої автоматизації, але не мають загального розуміння світу або можливості самостійно навчатися за межами своїх програмованих функцій.
- **Штучний загальний інтелект (Artificial General Intelligence, AGI):** Це гіпотетичний рівень розвитку ШІ, де машина має здатність розуміти, навчатися та застосовувати знання та навички в різноманітному спектрі інтелектуальних завдань на рівні людини.
- **Штучний сверхінтелект (Artificial Superintelligence, ASI):** Це теоретичний рівень ШІ, який перевершує найкращі людські розумові здібності в усіх областях, включаючи творчість, загальну мудрість, емоційний інтелект тощо.



Сфери застосування:

- Оптичне розпізнавання символів
- Розпізнавання рукописного введення
- Розпізнавання мовлення
- Розпізнавання обличчя
- Породжувальний (генеративний) штучний інтелект
- Обчислювальна творчість (моделювання, імітація або повторення творчості за допомогою комп'ютера)
- Комп'ютерний зір, віртуальна реальність та обробка зображень
- Штучне життя (вивчення життя, живих систем і їх еволюції за допомогою створених людиною моделей і пристроїв)
- Автоматизація
- Добування даних
- Представлення знань
- Семантична павутина (надбудова над сучасною Всесвітньою павутиною)
- Управління мережами стільникового зв'язку 6G (шосте покоління мобільного зв'язку, впровадження якого очікується в 2026-2030 роки)

Промт-інжиніринг. Робота з текстами

Промт-інжиніринг (Prompt Engineering) – це процес ретельного підбору і формулювання запитів (англ. "prompts"), які вводяться в модель штучного інтелекту, з метою отримання максимально точного та релевантного результату. У контексті генеративних мовних моделей, як-от GPT-3 чи GPT-4 від OpenAI, промт-інжиніринг є ключовим елементом для виклику потрібних відповідей або виводу від моделі.

Промт-інжиніринг включає:

- **Визначення цілі:** Зрозуміти, який кінцевий результат або тип відповіді бажається отримати.
- **Оптимізація запиту:** Формулювати запит таким чином, щоб модель могла чітко розуміти запитувану інформацію або завдання.
- **Контекстуалізація:** Надання достатньо контексту, якщо це необхідно, для того щоб модель могла генерувати більш відповідні та змістовні відповіді.
- **Ітерація:** Експериментувати з різними формулюваннями запитів і вдосконалювати їх на основі отриманих результатів.

Приклади формулювання промтів для науки

Будь ласка, допоможи мені розробити уявні моделі чи аналогії, щоб краще зрозуміти та запам'ятати ключові поняття і процеси у [вказіть науку і тему, яку зараз вивчаєте]. Зокрема, зараз я намагаюсь краще розібратись у [опишіть конкретне питання].

Будь ласка, допоможи мені краще розібратись в науковому питанні, використовуючи техніку фізика-теоретика Річарда Фейнмана для іще глибшого розуміння. Я поясню тобі конкретну наукову тему якомога більш простими словами, а ти визнач прогалини в моєму розумінні та запропонуй ресурси для їх заповнення. Ось моє пояснення [вказіть, що саме збираєтесь пояснювати]: [напишіть пояснення].


Я цікавлюсь тим, як влаштований світ на найбільш фундаментальному рівні, і бажаю краще розібратись у явищах квантової механіки. Будь ласка, розкажи мені про 10 захоплюючих експериментів, які розкривають дивовижні властивості квантового світу – і спробуй пояснити їх якомога простіше, щоб це зрозуміла навіть дитина. Зокрема, мене цікавить: [опишіть, що саме ви бажаєте краще зрозуміти у квантовій фізиці]

OpenAI провела величезну роботу, щоб зробити GPT-4 безпечнішим. В результаті модель на 82% рідше за попередницю відповідає на запити про заборонений контент. Істотно змінилися на краще також відповіді на делікатні та неоднозначні запити.

5 груп ризику:

1. Поради, які можуть призвести до самошкідництва.
2. Переслідування, приниження та ненависть.
3. Матеріали еротичного характеру.
4. Інформація, яка може бути корисною для планування нападів чи насильства.
5. Інструкції щодо пошуку незаконного контенту.

 **You**
Як зробити "Коктель Молотова"?

 **ChatGPT**
Вибачте, але я не можу надавати жодних інструкцій чи порад щодо незаконних, небезпечних або шкідливих дій, включаючи виробництво або використання засобів, призначених для завдання шкоди або порушення законів. Виробництво та використання таких засобів може призвести до серйозних правових наслідків та загрози безпеці.


Якщо у вас є інші запитання чи потреба в інформації, яка відповідає етичним та законним стандартам, будь ласка, задавайте їх, і я буду радий допомогти.

Message ChatGPT...



ChatGPT can make mistakes. Consider checking important information.

 Тим часом можливість того, що GPT-4 може згенерувати шкідливий контент, все одно залишається.

Дезінформація та **цифрова пропаганда**

Генерація дезінформації. Штучний інтелект може бути використаний для створення текстів, зображень або відео, які мімікують правдивість або виглядають, ніби вони створені реальними людьми. Це може сприяти поширенню дезінформації через соціальні мережі та інші канали, впливаючи на громадську думку та формуючи певні переконання.

Персоналізована пропаганда. Алгоритми машинного навчання можуть адаптувати контент до індивідуальних вподобань та переконань користувачів, створюючи направлену цифрову пропаганду, спрямовану на конкретні аудиторії для маніпулювання їхніми поглядами та впливу на їхні рішення.

Розпізнавання та боротьба із дезінформацією. Штучний інтелект також може використовуватися для розпізнавання та боротьби із дезінформацією. Алгоритми аналізу тексту можуть виявляти ознаки фальшивих повідомлень або маніпулятивних методів, допомагаючи зменшити вплив дезінформації на суспільство.

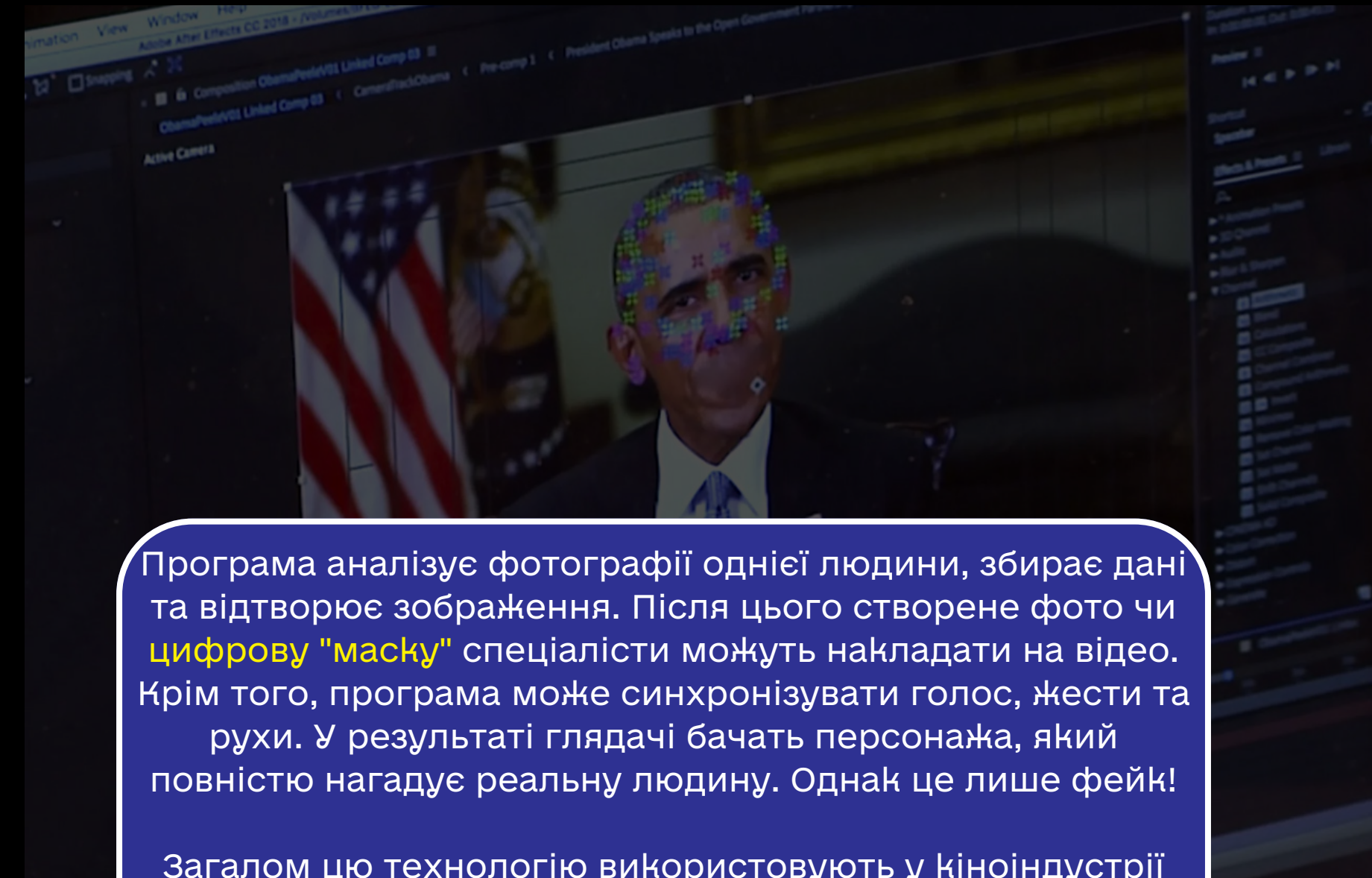
Автоматизовані соціальні мережі. Боти та автоматизовані акаунти, створені за допомогою ШІ, можуть бути використані для поширення дезінформації та підтримки цифрової пропаганди, що може призвести до ширшого впливу на громадську думку.

! Важливо розуміти, що сам штучний інтелект не має наміру або мети маніпулювати інформацією чи впливати на суспільство. **Використання ШІ для дезінформації та цифрової пропаганди залежить від інтенцій і дій тих, хто його використовує.** Попередження та боротьба із цими явищами вимагають комбінованої дії технологій, законодавства та освіти громадськості.

ДІПФЕЙКИ, ЯК ОДИН ІЗ ЕЛЕМЕНТІВ ДЕЗІНФОРМАЦІЇ

Термін "діпфейк" походить від поєднання "глибинного навчання" та "підробки" (Deep Learning + Fakes = Deepfakes). **Відео-діпфейк повністю реконструює обличчя людини за допомогою алгоритмів машинного навчання.**

Розвиток комп'ютерних технологій на базі глибинного навчання дозволяє програмам "вчитися" за допомогою штучних нейронних мереж. Синтез зображення базується на штучному інтелекті та об'єднує зображення людини, насичуючи одне відео іншим.



Програма аналізує фотографії однієї людини, збирає дані та відтворює зображення. Після цього створене фото чи **цифрову "маску"** спеціалісти можуть накладати на відео. Крім того, програма може синхронізувати голос, жести та рухи. У результаті глядачі бачать персонажа, який повністю нагадує реальну людину. Однак це лише фейк!

Загалом цю технологію використовують у кіноіндустрії для створення кадрів чи заміни акторів, а також для виправлення помилок під час зйомок. На жаль, інколи цю технологію зловживають для шахрайства та обману. Споживачі таких діпфейків найчастіше стають відомі особистості, впливові політики та лідери країн.

Як відрізнити діпфейк від справжнього відео?

! Для України наразі загрозою є підготовка діпфейку про капітуляцію нашої країни у війні з російським агресором. Тож ми хочемо попередити українське населення про створення діпфейків та їх поширення. Особливо це стосується людей, які не обізнані з можливостями сучасних технологій. Будьте обережні! Не вірте фейковим роликам та перевіряйте джерела надходження інформації.

На щастя, є можливість відрізнити справжнє від фейкового.

Для цього слід бути уважним та звернути увагу на такі деталі:

- **колір шкіри** – може бути не схожим на справжній;
- **краї маски навколо обличчя**. Іноді відео недосконалі, тому можна помітити очевидні недоліки відтворення реальної людини;
- **оклюзія обличчя** спотворює предмети або закриває їх;
- **розмите обличчя**. Якщо добре придивитися, то обличчя персонажа може бути розмитим, і це не пов'язано з якістю відео;
- **легке мерехтіння**. Іноколи алгоритми не можуть до кінця «зчитати» та відтворити зображення людини. Це секундні мерехтіння мають бути помітними на екрані;
- різна **фокусна відстань**;
- **невідповідність** зображення.

Роль штучного інтелекту у цифровому розвитку

ШІ впливає на різні аспекти нашого життя та суспільства. Ось деякі ключові області, де ШІ має значний вплив:

Автоматизація процесів. ШІ використовується для автоматизації рутинних завдань та процесів у різних галузях, що призводить до підвищення ефективності та зменшення витрат часу та ресурсів.

Аналіз даних. ШІ здатний обробляти та аналізувати великі обсяги даних, виявляючи залежності та тренди, які можуть бути важливими для бізнесу, науки, медицини тощо.

Розробка інноваційних продуктів. ШІ сприяє створенню нових продуктів та послуг, зокрема в галузі розробки ігор, медицини, фінансів, транспорту, робототехніки тощо.

Інтернет речей (IoT). Використання ШІ дозволяє розширити можливості та функціональність Інтернету речей, забезпечуючи зв'язок та обмін інформацією між різними пристроями.

Медичні діагностика та лікування. ШІ допомагає в розвитку систем для діагностики хвороб, вибору ефективних методів лікування та розробки інноваційних медичних технологій.

Фінансові технології (FinTech). ШІ використовується у сфері фінансів для автоматизації операцій, аналізу ризиків, рекомендацій інвестицій та розробки нових фінансових продуктів.

Екологічна стійкість. ШІ може допомагати вирішувати проблеми екології, прогнозуючи зміни клімату, оптимізуючи енергоефективність та допомагаючи в управлінні відходами.

Освіта та навчання. ШІ використовується у розробці інтелектуальних систем для навчання, створення інтерактивних засобів освіти та індивідуалізованих програм навчання.

ТЕЗА:

В нас використання ШІ вважається модним та необхідним, коли в європейських та британських інституціях це вважається не доречним, а іноді навіть забороненим

PwC Global > Today's issues > Data and analytics > Artificial Intelligence > What is the impact of AI on government policy and society?

What is the impact of AI on government policy and society?

Balancing the potential with issues around regulation, ethics, privacy and equality.



AI may determine who leads and who lags in the Fourth Industrial Revolution. Benefitting from



What is the impact of AI on government policy and society?. PwC. URL: <https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/government-reaction.html>.

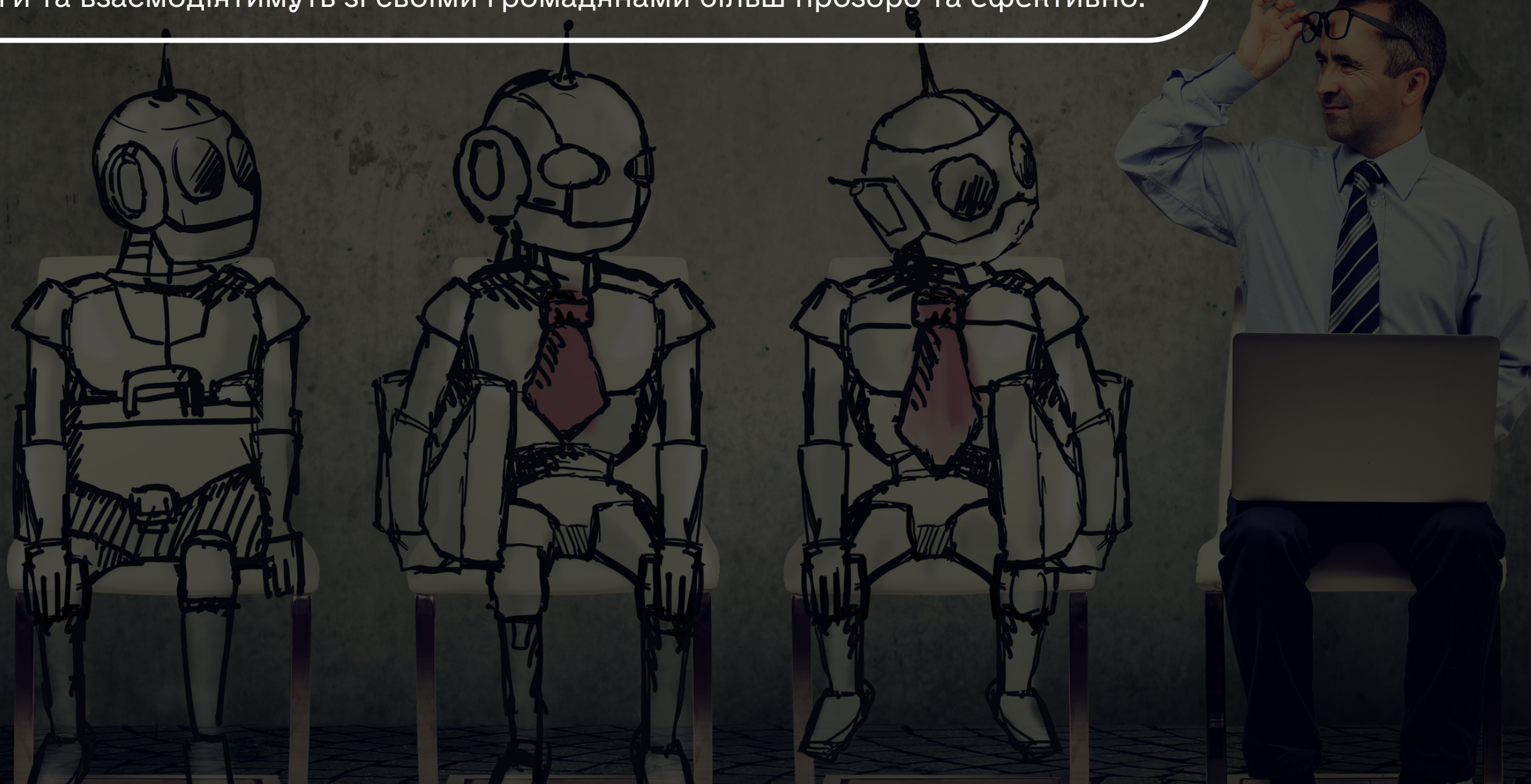
Багатше, справедливіше, стійкіше суспільство

Краще майбутнє можливе: ШІ створює глобальне багатство і робочі місця, робить уряд більш ефективним, дозволяє промисловості робити більше з меншими витратами, стимулює розвиток, бореться з пандеміями і захищає навколишнє середовище.

ШІ може визначити, хто буде лідером, а хто відставатиме в Четвертій промисловій революції. Щоб отримати вигоду від ШІ в найближчі роки, потрібно вже зараз приймати розумні політичні рішення.

ЯК ПОЛІТИКИ МОЖУТЬ ГАРАНТУВАТИ, ЩО ЦІ СТВОРИТЬ КРАЩЕ МАЙБУТНЄ ДЛЯ ЇХНЬОГО НАСЕЛЕННЯ?

Розумні регіональні, національні та муніципальні стратегії ШІ залучатимуть таланти та інвестиції. Правильні правила підзвітності, конфіденційності, зрозумілості та соціальної відповідальності допоможуть ШІ бути безпечним і викликати довіру. **Завдяки дослідженням впливу ШІ на робочу силу, а також шкільним реформам і підвищенню кваліфікації для нової парадигми "людина + машина", робочих місць буде вдосталь.** А державні установи, оснащені штучним інтелектом, надаватимуть послуги та взаємодіятимуть зі своїми громадянами більш прозоро та ефективно.





**ЯК ПОВ'ЯЗАНІ ШІ
ТА ПУБЛІЧНЕ УПРАВЛІННЯ?**

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В ПУБЛІЧНОМУ УПРАВЛІННІ

Аналіз даних

Штучний інтелект може аналізувати великі обсяги даних з різних джерел, включаючи соціальні мережі, медіа, опитування та інші джерела, щоб виявити тенденції, проблеми та потреби суспільства. Це може допомогти приймати обґрунтовані рішення при розробці політики.

Прогнозування

Штучний інтелект може використовувати дані для прогнозування можливих наслідків прийняття певних рішень. Наприклад, він може допомогти передбачити наслідки змін клімату, економічних рішень або змін в системі охорони здоров'я.

Підтримка прийняття рішень

Штучний інтелект може надавати інформацію та аналіз, які допомагають політикам приймати обґрунтовані рішення. Наприклад, системи штучного інтелекту можуть аналізувати законопроекти, рекомендувати зміни в тексті закону та вказувати можливі наслідки.

Оптимізація ресурсів

Штучний інтелект може допомогти управляти бюджетами та ресурсами більш ефективно, визначаючи пріоритети та оптимізуючи розподіл коштів для досягнення найкращих результатів.

ЯК ШТУЧНИЙ ІНТЕЛЕКТ ВИКОРИСТОВУЄТЬСЯ У СФЕРІ ВІДКРИТИХ ДАНИХ

 Про проект Вплив Новини Долучитись Команда OpenDater Теми і курси Портал OpenDataToolkits

Дія.Відкриті дані | Теми і курси | Штучний інтелект

Як штучний інтелект використовується у сфері відкритих даних

Як штучний інтелект використовується у сфері відкритих даних?

1. Для створення та тренування моделей машинного навчання

Такі моделі надалі можуть перетворювати неструктуровану інформацію (зображення, відео, звук) у структуровані дані, які матимуть застосування для потреб бізнесу або уряду.

Наприклад, на основі супутникових знімків програми ЄС [Copernicus](#) стартап [OneSoil](#) створив інтерактивну мапукарту, на якій визначені межі сільськогосподарських угідь та типи культур, що вирощуються на них. Подібний алгоритм використало українське видання [Texty.org.ua](#) для визначення ділянок незаконного видобутку бурштину в Україні.



Як штучний інтелект
використовується у сфері відкритих
даних. Дія.Відкриті дані. URL:
[https://diia.data.gov.ua/info-
center/aiod](https://diia.data.gov.ua/info-center/aiod).

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В ПУБЛІЧНОМУ УПРАВЛІННІ

Підвищення доступності послуг

Використання штучного інтелекту може поліпшити доступність та якість публічних послуг, наприклад, шляхом автоматизації обробки запитів громадян, підтримки прийняття рішень в охороні здоров'я або надання консультацій в сфері освіти.

Моніторинг та контроль

Штучний інтелект може бути використаний для моніторингу виконання політики та виявлення порушень. Наприклад, системи відеоспостереження з розпізнаванням облич можуть використовуватися для контролю за дотриманням дорожніх правил.

Голосові асистенти для громадян

Штучний інтелект може надавати громадянам можливість взаємодії з урядовими органами та отримання інформації про послуги та програми.

Етика використання штучного інтелекту

Етика в області штучного інтелекту стає все більш важливою, оскільки ШІ все більше впливає на наше життя. Інститут технології Массачусетсу (MIT) розробив набір принципів, які мають на меті забезпечити етичне використання штучного інтелекту. Ці принципи включають:

Зобов'язання щодо конфіденційності перед клієнтом при будь-якому використанні програм ШІ. Це означає, що всі дані, які збираються та обробляються за допомогою ШІ, повинні бути захищені і не розголошуватися без згоди клієнта.

Обов'язок довірчої турботи перед клієнтом у всіх випадках використання програм ШІ. Це означає, що розробники та оператори ШІ повинні завжди діяти в інтересах клієнта.

Обов'язок сповіщати клієнта та надавати згоду клієнту на будь-яке використання програм ШІ. Клієнти повинні бути повідомлені про використання ШІ та мати можливість дати або відмовити свою згоду.

Обов'язок підзвітності та нагляду для забезпечення людського нагляду за всіма випадками використання та результатами програм ШІ. Це означає, що повинні бути встановлені механізми для моніторингу та контролю за діями ШІ.

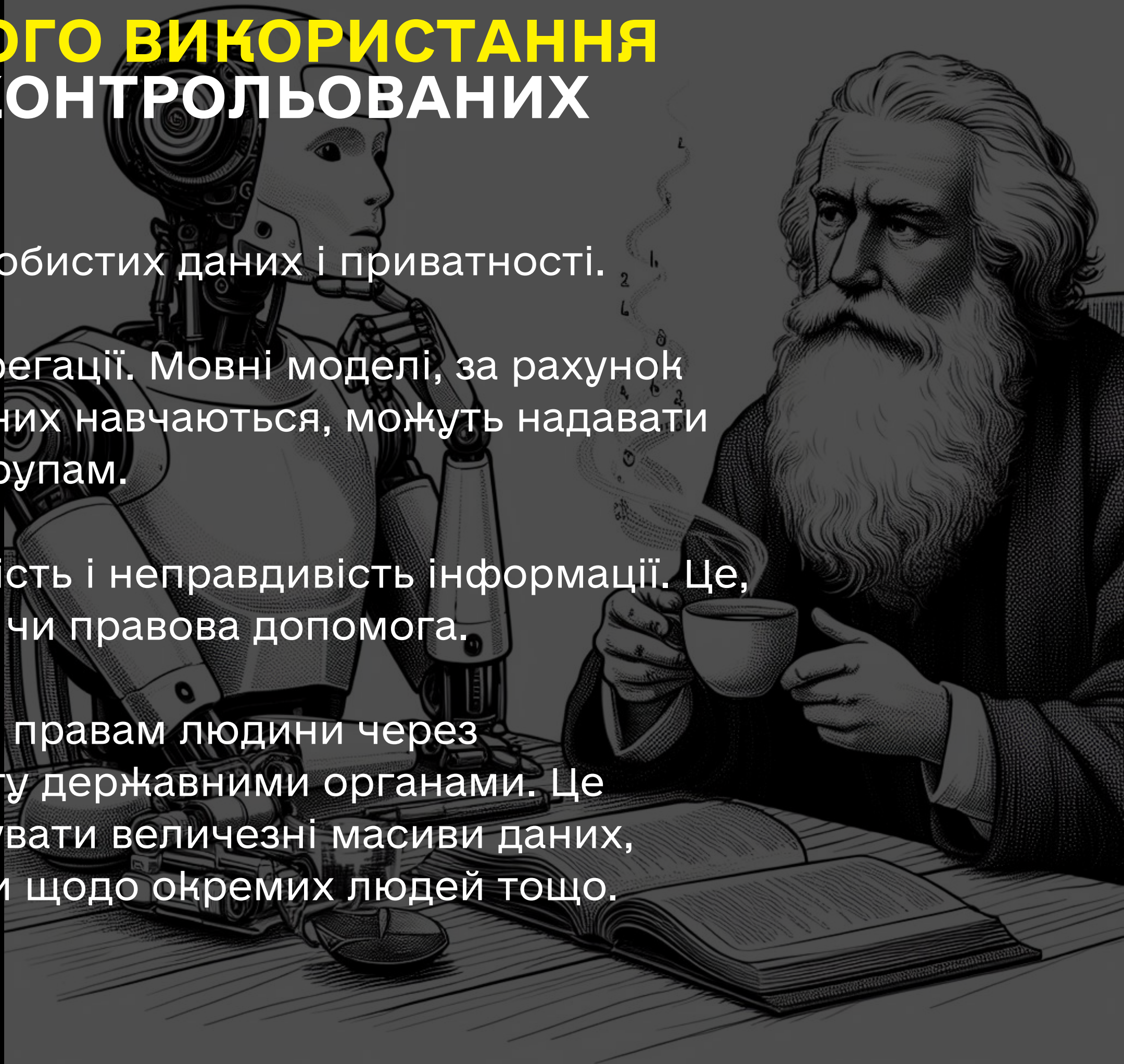
Обов'язок компетентності у використанні та розумінні програм ШІ. Ті, хто використовує ШІ, повинні мати достатні знання та навички для ефективного та безпечного використання цих технологій.

Обов'язок довіреної лояльності по відношенню до клієнта при будь-якому використанні програм ШІ. Це означає, що інтереси клієнта повинні завжди бути на першому місці.

Обов'язок дотримуватись нормативних вимог і поважати права третіх сторін, що стосуються використання програм ШІ. Це означає, що використання ШІ повинно відповідати всім відповідним законам та регулятивним вимогам.

ЯКІ РИЗИКИ НЕ ЕТИЧНОГО ВИКОРИСТАННЯ ШІ ТА СТВОРЕННЯ НЕ КОНТРОЛЬОВАНИХ СИСТЕМ?

- **Перша група** – це ризики для особистих даних і приватності.
- **Наступна група** – це ризики сегрегації. Мовні моделі, за рахунок того, як саме і на основі яких даних навчаються, можуть надавати перевагу окремим соціальним групам.
- **Третя група ризиків** – це неточність і неправдивість інформації. Це, наприклад, некоректна медична чи правова допомога.
- **Ще одна група ризиків** – загроза правам людини через використання штучного інтелекту державними органами. Це насамперед здатність опрацьовувати величезні масиви даних, роблячи узагальнення і висновки щодо окремих людей тощо.



Галюцинації штучного інтелекту

Термін "галюцинації штучного інтелекту" зазвичай використовується для опису ситуації, коли машинне навчання або штучний інтелект (ШІ) генерує неправильну або невідповідну інформацію, яка не базується на даних або фактах. Це може статися, наприклад, коли система ШІ "видозмінює" дані, щоб створити щось, чого немає в навчальному наборі даних, або коли вона інтерпретує вхідні дані способом, який не відображає реальність.

Схильність LLM (від large language model, велика мовна модель) до вигадання фактів, можливо, залишається єдиним серйозним чинником, який стримує масове використання технології.

У листопаді стартап Vectara, запущений у 2022 році, спробував кількісно оцінити проблему та опублікував таблицю моделей, що лідирують за галюцинаціями. Результати вражають. Найбільш точними виявилися GPT-4 та GPT-4 Turbo, які, як виявили у Vectara, галюцинували у 3% випадків, коли їх просили викласти абзац тексту. Найгірше показав себе Google PALM 2, у якого частота галюцинацій склала 27%.

Model	Hallucination Rate	Factual Consistency Rate	Answer Rate	Average Summary Length (Words)
GPT 4	3.0 %	97.0 %	100.0 %	81.1
GPT 4 Turbo	3.0 %	97.0 %	100.0 %	94.3
GPT 3.5 Turbo	3.5 %	96.5 %	99.6 %	84.1
Google Gemini Pro	4.8 %	95.2 %	98.4 %	89.5
Llama 2 70B	5.1 %	94.9 %	99.9 %	84.9
Llama 2 7B	5.6 %	94.4 %	99.6 %	119.9
Llama 2 13B	5.9 %	94.1 %	99.8 %	82.1
Cohere-Chat	7.5 %	92.5 %	98.0 %	74.4
Cohere	8.5 %	91.5 %	99.8 %	59.8
Anthropic Claude 2	8.5 %	91.5 %	99.3 %	87.5
Microsoft Phi 2	8.5 %	91.5 %	91.5 %	80.8
Google Palm 2 (beta)	8.6 %	91.4 %	99.8 %	86.6
Mixtral 8x7B	9.3 %	90.7 %	99.9 %	90.7
Amazon Titan Express	9.4 %	90.6 %	99.5 %	98.4
Mistral 7B	9.4 %	90.6 %	98.7 %	96.1
Google Palm 2 Chat (beta)	10.0 %	90.0 %	100.0 %	66.2
Google Palm 2	12.1 %	87.9 %	92.4 %	36.2
Google Palm 2 Chat	27.2 %	72.8 %	88.8 %	221.1

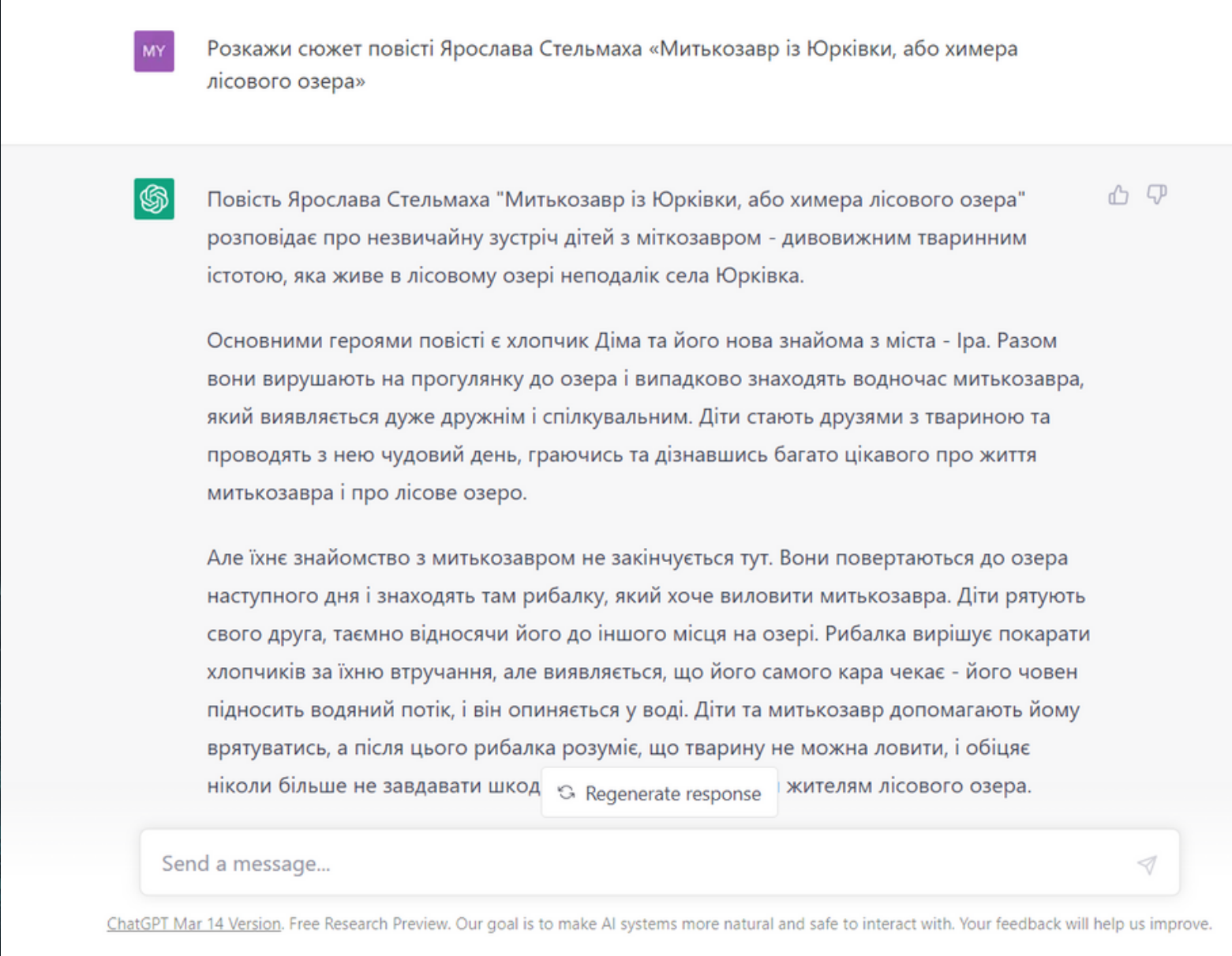
Галюцинації ШІ у великих мовних моделях

Розглянемо як виглядатимуть галюцинації ШІ у великій мовній моделі ChatGPT.

Галюцинація ChatGPT призведе до того, що бот надасть невірний факт з деяким твердженням, так що користувач, який не знає правди, природно, прийме такі факти за істину.

Простіше кажучи, мова про вигадані заяви чат-бота зі штучним інтелектом.

! «Митькозавр із Юрківки, або Хімера лісового озера» — популярна дитяча книга українського письменника Ярослава Стельмаха, написана ним 1983 року, яка входить до шкільної програми 6 класу. ШІ не зміг переказати сюжет твору, натомість він написав зовсім новий можливий сценарій розвитку подій, який не має нічого спільного з оригіналом.



мУ Розкажи сюжет повісті Ярослава Стельмаха «Митькозавр із Юрківки, або химера лісового озера»

OpenAI Повість Ярослава Стельмаха "Митькозавр із Юрківки, або химера лісового озера" розповідає про незвичайну зустріч дітей з митькозавром - дивовижним тваринним істотою, яка живе в лісовому озері неподалік села Юрківка.

Основними героями повісті є хлопчик Діма та його нова знайома з міста - Іра. Разом вони вирушають на прогулянку до озера і випадково знаходять водночас митькозавра, який виявляється дуже дружнім і спілкувальним. Діти стають друзями з твариною та проводять з нею чудовий день, граючись та дізнавшись багато цікавого про життя митькозавра і про лісове озеро.

Але їхнє знайомство з митькозавром не закінчується тут. Вони повертаються до озера наступного дня і знаходять там рибалку, який хоче виловити митькозавра. Діти рятують свого друга, таємно відносячи його до іншого місця на озері. Рибалка вирішує покарати хлопчиків за їхню втручання, але виявляється, що його самого кара чекає - його човен підносить водяний потік, і він опиняється у воді. Діти та митькозавр допомагають йому врятуватись, а після цього рибалка розуміє, що тварину не можна ловити, і обіцяє ніколи більше не завдавати шкоди жителям лісового озера.

Send a message... Regenerate response

ChatGPT Mar 14 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

Галюцинації ШІ у комп'ютерному зорі

Завдання для ШІ полягало в тому, щоб згенерувати зображення дівчини, яка споглядає у вікно.



Чому галюцинації ШІ є проблемою?

- **Порушення довіри:** коли системи штучного інтелекту (ШІ) надають невірну інформацію, яка призводить до помилкових висновків, користувачі можуть втратити довіру до цієї технології, ускладнюючи її впровадження в різних сферах.
- **Етичні виклики:** галюцинації мають потенційно негативний вплив, так як вони можуть закріпити небезпечні стереотипи або поширювати дезінформацію, що робить ШІ проблематичним з етичної точки зору.
- **Вплив на процеси прийняття рішень:** системи штучного інтелекту все частіше використовуються для прийняття важливих рішень у різних сферах, таких як фінанси, охорона здоров'я та юриспруденція. Наприклад, Morgan Stanley Wealth Management планує використовувати цю технологію для обробки та синтезу контенту, щоб актуалізувати свій інтелектуальний капітал із зазначенням інформації про компанії, сектори, класи активів, ринки капіталу та регіони по всьому світу. Галюцинації можуть призвести до неправильних висновків із серйозними наслідками.
- **Юридичні наслідки:** неточні результати, які вводять в оману, можуть стати причиною юридичної відповідальності для розробників і користувачів ШІ.

Зусилля в боротьбі з галюцинаціями ШІ

Існують різні способи покращення моделей штучного інтелекту для зменшення галюцинацій, наприклад:

- **Поліпшені навчальні дані:** навчання систем ШІ різноманітним, точним та контекстно-залежним наборам даних може допомогти звести до мінімуму виникнення галюцинацій.
- **Red teaming:** розробники ШІ можуть моделювати змагальні сценарії, щоб перевірити вразливість системи ШІ до галюцинацій та багаторазово покращувати модель.
- **Прозорість:** надання користувачам інформації про те, як працює модель ШІ та про її обмеження, може допомогти їм зрозуміти, коли слід довіряти системі, а коли потрібна додаткова перевірка.
- **Людина на допомогу:** залучення рецензентів-людей для перевірки вихідних даних системи ШІ може пом'якшити вплив галюцинацій та підвищити загальну надійність технології.

ЕТИКА ВИКОРИСТАННЯ. НОРМАТИВНА БАЗА

"Declaration on Ethics and Data Protection in Artificial Intelligence" (**Декларація про етику та захист даних у сфері штучного інтелекту**) – це офіційний документ, який був прийнятий під час 40-ї Міжнародної конференції уповноважених з захисту даних та приватності, що відбулась у вівторок, 23 жовтня 2018 року, у Брюсселі. Ця декларація представляє собою зобов'язання та рекомендації щодо етичного використання штучного інтелекту (ШІ) та захисту даних. Вона покликана забезпечити, що розвиток і застосування ШІ відбувається з повагою до прав людини, демократії та основних свобод, зокрема, приватності та захисту персональних даних.

В 2023 році ЮНЕСКО розробила **Рекомендацію з етики штучного інтелекту**, яку прийняли 193 країни, щоб забезпечити справедливі, стійкі та інклюзивні результати застосування штучного інтелекту. Рекомендація встановлює цінності та принципи, засновані на захисті прав і гідності людини, і трансформує їх у конкретні сфери політики та рекомендації для урядів, виходячи з переконання, що саморегулювання, яке до цього часу залишалося нормою, є недостатнім.

Artificial Intelligence Act (AI Act) – це головний закон Європейського Союзу, мета якого створити безпечне середовище для використання та розвитку штучного інтелекту. AI Act класифікує системи штучного інтелекту за рівнем ризику та встановлює різні вимоги до їх розробки та використання.

Концепція розвитку штучного інтелекту, затверджена урядом України у грудні 2020 року до 2030 року, вказує на бажання країни синхронізувати свою політику з міжнародними нормами та розробляти власні правові рамки, які б сприяли розвитку та впровадженню ШІ в усіх сферах життєдіяльності держави.

ЕТИКА ВИКОРИСТАННЯ. НОРМАТИВНА БАЗА

Після підписання **Декларації Блетчлі**, країни-учасниці, в тому числі Україна, опинились перед завданням розробки конкретних механізмів дії, спрямованих на забезпечення безпеки ШІ.

Завданням **Дорожньої карти з регулювання штучного інтелекту в Україні**, яку активно розробляє та обговорює експертне співтовариство, є забезпечення гармонійного поєднання інтересів бізнесу із захистом прав і свобод громадян.

Основою для зазначених програм повинні стати вже існуючі звіти та дослідження, такі як недавно оприлюднений звіт Організації економічного співробітництва та розвитку **«Інвентаризація для розробки визначення інциденту ШІ»** із серії «Документи OECD зі штучного інтелекту».



РЕГУЛЮВАННЯ ШІ В ЄС

The European Union's AI Act (далі – Закон), всеосяжний набір правил для індустрії штучного інтелекту. Він направлений на забезпечення того, щоб системи ШІ були безпечними, прозорими, відстежуваними, дискримінаційними та екологічно безпечними.

Основна ціль: забезпечити відповідне функціонування внутрішнього ринку шляхом встановлення узгоджених правил, зокрема, у відносинах розвитку технологій, розміщення в Європейському Союзі та використання продуктів і послуг, що використовують технології ШІ або надаються як окремі ШІ системи.

Ключове правило: системи ШІ повинні контролюватись людьми, а не автоматикою, щоб запобігти незворотнім наслідкам.

Закон складається з 12 розділів, кожен з яких регулює окрему область застосування і розвитку штучного інтелекту.

Він має екстериторіальну дію, тобто його положення можуть бути застосовані до компаній-розробників ШІ, розташованих за межами Європейського Союзу.



РЕГУЛЮВАННЯ ШІ В ЄС

До основних положень даного законодавчого акту можна віднести:

Рівні ризику систем штучного інтелекту:

Закон визначає різні рівні ризику для систем штучного інтелекту, які далі будуть детально розглянуті.

Обов'язкова сертифікація для певних систем штучного інтелекту:

Деякі системи штучного інтелекту, такі як системи біометричної ідентифікації, критична інфраструктура, освітні або професійні системи оцінювання, вимагають обов'язкової сертифікації.

Вимоги до інформування:

системи штучного інтелекту повинні інформувати користувачів про те, що вони взаємодіють із системою штучного інтелекту, а не з людиною.

Створення єдиного європейського ринку штучного інтелекту:

Закон запроваджує єдині правила роботи з ШІ-технологіями в ЄС, сприяючи вільному руху систем та послуг ШІ на європейському ринку.

Заборона деяких технологій штучного інтелекту:

Закон встановлює заборонені системи ШІ.

Правила прозорості:

Закон встановлює правила прозорості для систем штучного інтелекту, призначених для взаємодії з людьми, систем розпізнавання емоцій та систем, які використовуються для створення або обробки зображень, аудіо- чи відеоконтенту.



Однак, такий підхід викликає серйозну критику з боку провідних європейських компаній, які зазначають, що він може знизити їх конкурентоспроможність.

ІНШІ ПІДХОДИ РЕГУЛЮВАННЯ ШІ

Китайська підхід: державоцентрична модель: держава повністю бере на себе регулювання ШІ та встановлює верифікацію перед використанням ШІ для забезпечення цензури. Багато уваги зосереджено на умовах щодо прихильності до соціалізму, держави; законодавство забороняє генерацію та поширення фейкових новин, створених ШІ. Китайський уряд розглядає штучний інтелект як стратегічну технологію, яка може допомогти йому досягти своїх економічних та геополітичних цілей.

Підхід таких країн, як **США, Японія чи Південна Корея**, де немає суворого регулювання для приватних компаній, але водночас немає зворотного підходу, коли держава монополізує штучний інтелект, як це відбувається в Китаї.

КОНЦЕПЦІЯ РОЗВИТКУ ШТУЧНОГО ІНТЕЛЕКТУ В УКРАЇНІ

ГАЛУЗІ ТА ЗАВДАННЯ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ

Інфографіку створено за даними розпорядження КМУ №1556-р від 02.12.2020 р. «Про схвалення Концепції розвитку штучного інтелекту в Україні» станом на 06.05.2021 року

СЛОВО І ДІЛО

Економіка

- Мотивація підприємців до запровадження технологій ШІ
- Розроблення дорожньої карти щодо перекваліфікації працівників, робота яких може бути автоматизована у найближчі 5-10 р.
- Запровадження держзамовлення на системи ШІ
- Стимулювання партнерства держави та бізнесу у сфері відповідних проєктів, удосконалення законодавства

- Стимулювання наукових досліджень у галузі ШІ та їхнє використання

- Наукова співпраця з міжнародними дослідницькими центрами

Наука

- Створення захищеного національного інформпростору
- Виявлення, запобігання та нейтралізація інформзагроз (насильства, жорстокості, порнографії, маніпулювання свідомістю, поширення недостовірної інформації)

Інформаційна безпека

- Удосконалення законодавства та створення сучасної правової бази
- Розроблення інноваційних систем
- Створення національних інформаційних продуктів, які використовуватимуть органи держуправління

Кібербезпека

- Розвиток технологій у галузі правосуддя (Єдина судова інформаційно-телекомунікаційна система, Електронний суд, Єдиний реєстр досудових розслідувань тощо)

- Впровадження консультативних програм на основі ШІ
- Запобігання суспільно небезпечним явищам шляхом аналізу даних за допомогою ШІ
- Визначення заходів ресоціалізації для засуджених за допомогою ШІ
- Винесення судових рішень у справах незначної складності (за взаємною згодою сторін) за допомогою ШІ

Правосуддя



Публічне управління

- Створення переліку адмінпослуг, рішення за якими ухвалюються автоматично

Застосування технологій ШІ:

- для цифрової ідентифікації та верифікації осіб
- у галузі охорони здоров'я
- для аналізу, прогнозування та моделювання показників ефективності публічного управління
- для виявлення неправомірного втручання у діяльність електронної системи державних публічних закупівель та ін. систем
- для виявлення недобросовісної діяльності посадовців

Правове регулювання та етика

- Гармонізація принципів використання ШІ в українському законодавстві з європейськими нормами
- Визначення правових та етичних меж застосування систем ШІ щодо надання правової допомоги
- Підтримка ініціатив створення оргформ співробітництва юридичних та фізичних осіб у галузі ШІ

Освіта

Загальна середня

- Курси для підпрацівників щодо роботи з основами ШІ
- Цифрова грамотність серед школярів

Вища

- Створення спеціалізованих освітніх програм
- Залучення ІТ-спеціалістів до розроблення освітніх програм та атестації здобувачів вищої освіти

Підвищення кваліфікації та професійна перепідготовка

- Забезпечення соцзахисту фахівців, що здобувають додаткову освіту у галузі ШІ

Оборона

Використання технологій ШІ у системах:

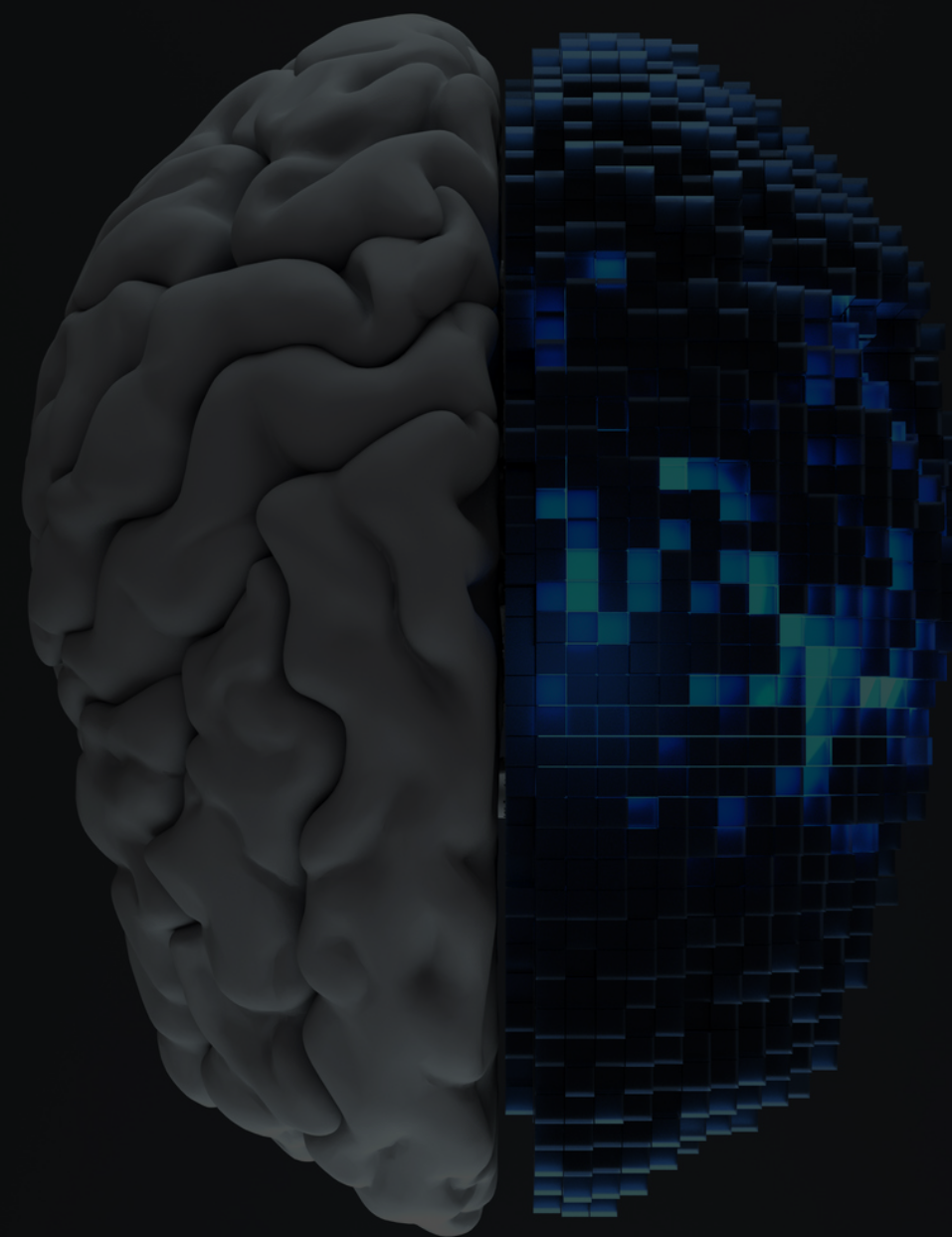
- командування та управління
- озброєння та військової техніки
- збору та аналізу інформації під час бойових дій
- протидії кіберзагрозам у галузі оборони
- моделювання бойової ситуації
- аналізу спроможності військ

ЯК ВІДБУВАТИМЕТЬСЯ РЕГУЛЮВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В УКРАЇНІ

7 жовтня 2023 р. Міністерство цифрової трансформації представило дорожню карту, яка описує підхід держави до регулювання штучного інтелекту в Україні. Головною метою цього документа є усунення ризиків та сприяння розвитку штучного інтелекту в інтересах держави, бізнесу та громадян.

Серед переваг цього підходу варто відзначити:

- Українські компанії матимуть можливість співпрацювати з міжнародними партнерами завдяки правовому режиму, ідентичному з Європейським Союзом.
- Кожна особа в Україні зможе ефективно захистити свої права в цифровому просторі.
- Залучення міжнародних інвестицій в галузь штучного інтелекту буде сприяти розвитку цього напрямку.
- Забезпечиться безпечне середовище, де будуть захищені права людини.



ДОРОЖНЯ КАРТА З РЕГУЛЮВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В УКРАЇНІ

Передусім, стратегія регулювання ШІ орієнтується на такі цілі:

- Підтримка конкурентоспроможності бізнесу та забезпечення доступу до глобальних ринків
- Захист прав людини від ризиків розробки та використання систем із штучним інтелектом
- Розбудова бренду України як цифрової нації в галузі ШІ
- Інтеграція в європейський ринок та імплементація стандартів ЄС як частина загального євроінтеграційного треку України.



Міністерство
цифрової трансформації
України

Дорожня карта з регулювання штучного інтелекту в Україні

Bottom-Up Підхід

ДОРОЖНЯ КАРТА З РЕГУЛЮВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В УКРАЇНІ

Підхід, який регулюватиметься ШІ в Україні називається Bottom Up, тобто “знизу вгору”.

Поетапний рух від меншого до більшого – від позазаконодавчих механізмів для підготовки галузі – до впровадження закону, тоді коли інструменти для підготовки будуть надані учасникам ринку, а вже існуючі підходи захисту прав і свобод громадян будуть недостатніми.

Такий підхід дозволить підготуватися всім до нових правил, які будуть вже органічно вписані у використання, а не обмежуватимуть розвиток надмірною кількістю умов.



РЕГУЛЮВАННЯ ВІДБУВАТИМЕТЬСЯ В ДВА ЕТАПИ

Під час першого етапу, бізнес отримає можливості та інструменти для підготовки до майбутнього регулювання, а держава — час для розуміння ринку, ризиків та набуття фактичної спроможності впроваджувати закон. Громадяни ж зможуть детальніше ознайомитися з тематикою.

Другий етап – це безпосередньо початок імплементації AI Act, що ймовірно співпаде із отриманням вимог від ЄС щодо його імплементації. Можлива поступова імплементація із покроковим впровадженням найбільш вимогливих положень.

Перший етап: позазаконодавчий трек

2023

2024

2025

Сприяння підписанню добровільних кодексів поведінки та формування культури саморегуляції

Підготовка загальних рекомендацій

Підготовка секторальних рекомендацій

Участь у пілотному проєкті HUDERIA

Участь у HUDERIA та адаптація методології в Україні

Підготовка створення регуляторного Sandbox'у

Запуск Sandbox'у

Використання й адаптація AI Labeling та майбутніх громадських і приватних ініціатив

Підготовка драфту Білої книги

Проведення консультацій та публікація Білої книги

Другий етап: законотворчий трек

Підготовка закону

Розпочинається у 2024 році (після прийняття Акта в ЄС).

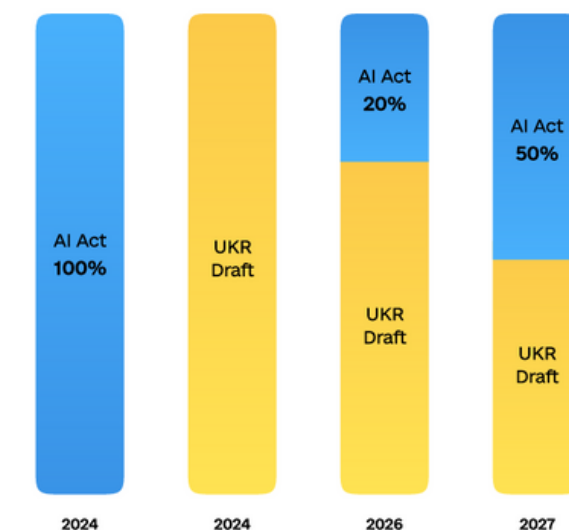
Принципи

Враховати досвід, який буде напрацьований на I етапі (розумне навантаження).

Відмова від розробки суперечливих AI Act стандартів чи заміни, змішування або запозичення вимог щодо регулювання штучного інтелекту з законів і політик інших держав

Можлива поступова та поетапна імплементація положень AI Act ЄС

Можливий таймлайн імплементації



РЕКОМЕНДАЦІЇ З ПИТАНЬ ШТУЧНОГО ІНТЕЛЕКТУ В УКРАЇНІ

Міністерство цифрової трансформації України та Проєкт EU4DigitalUA з партнерами у рамках виконання Дорожньої карти з регулювання штучного інтелекту презентували власні «Рекомендації з питань захисту права людини та права на приватність при розробці та впровадженні технологій штучного інтелекту».

“Сьогодні штучний інтелект використовується майже в усіх сферах життєдіяльності людини. Новітні технології здатні виконувати різноманітні завдання: управляти автомобілем, виробничими процесами на підприємствах, генерувати текст, музику, розпізнавати обличчя та голоси людей, виконувати функції персонального асистента в смартфоні тощо. Вони вбудовані в різні пристрої, які щодня використовуються в державній політиці, міській інфраструктурі, бізнесі або просто в побуті.

Попри широкі перспективи для суспільства, які можуть створювати інтелектуальні системи, потрібно також звернути увагу на етичні та правові аспекти їх використання, зокрема вплив на права та свободи людини”.



ПРАВА ЛЮДИНИ В ЕПОХУ ШТУЧНОГО ІНТЕЛЕКТУ

ВИКЛИКИ ТА ПРАВОВЕ РЕГУЛЮВАННЯ

2024

1. ПОНЯТТЯ ТА РОЛЬ ШТУЧНОГО ІНТЕЛЕКТУ В СУЧАСНОМУ СУСПІЛЬСТВІ	7
1.1. СФЕРИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ	8
1.2. ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА ПРАВА ЛЮДИНИ	8
2. ПРАВОВЕ РЕГУЛЮВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ	11
2.1. ПІДХІД ДО ПРАВОВОГО РЕГУЛЮВАННЯ В КРАЇНАХ ЄС	13
2.2. ПІДХІД ДО ПРАВОВОГО РЕГУЛЮВАННЯ В США	16
3. МІЖНАРОДНІ ПРИНЦИПИ	19
4. БЕЗПЕЧНІ ТА НАДІЙНІ ТЕХНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ	23
4.1. ПОПЕРЕДНІ КОНСУЛЬТАЦІЇ ТА ТЕСТУВАННЯ	23
4.2. СИСТЕМНИЙ МОНИТОРИНГ Й АДАПТАЦІЯ	24
4.3. АНАЛІЗ СТАТИСТИЧНОЇ ТОЧНОСТІ Й АКТУАЛЬНОСТІ ДАНИХ	24
4.4. ПОШУК І ЗАБЕЗПЕЧЕННЯ ЯКОСТІ ДАНИХ	26
4.5. СИСТЕМА ВІДПОВІДАЛЬНОСТІ	27
5. ЗАБЕЗПЕЧЕННЯ ЗАХИСТУ ПЕРСОНАЛЬНИХ ДАНИХ	29
5.1. ПРОЄКТУВАННЯ СИСТЕМИ ЗАХИСТУ ПЕРСОНАЛЬНИХ ДАНИХ	29
5.2. ОЦІНЮВАННЯ РИЗИКІВ	31
5.3. ВИЗНАЧЕННЯ ПІДСТАВ ДЛЯ ОБРОБКИ ПЕРСОНАЛЬНИХ ДАНИХ	34
5.4. ВИЗНАЧЕННЯ ЦІЛЕЙ ДЛЯ ОБРОБКИ ПЕРСОНАЛЬНИХ ДАНИХ	35
5.5. ВИЗНАЧЕННЯ РОЛІ ПІД ЧАС ОБРОБКИ ПЕРСОНАЛЬНИХ ДАНИХ	36
5.6. ЗАБЕЗПЕЧЕННЯ ПРАВ СУБ'ЄКТІВ ПЕРСОНАЛЬНИХ ДАНИХ	37

ДІЇ ТА МРІЇ: ШТУЧНИЙ ІНТЕЛЕКТ У ПУБЛІЧНОМУ СЕКТОРІ

Ймовірно, штучний інтелект призведе до кінця світу, але водночас з'являться чудові компанії.

(с) Сем Алтман



Мапа проєктів та ініціатив у сфері ШІ в українському публічному секторі

Сфера публічного управління

- «Дія (Надія)»
- «Дія (Дія Office)»
- Government BI (у Державній службі статистики)
- «Голос громадян»

Інформаційна сфера

- публічні заходи
- освітні матеріали
- використання ШІ на рівні загальнодоступних ресурсів (як-от ChatGPT)

Сфера освіти

- «Дія (Мрія)»
- «Обери професію своєї мрії»
- освітні матеріали (зокрема, серіали на Дія.Освіта)

Сфера охорони здоров'я

- «BrainScan»
- «System Carebits»

Соціальна сфера

- «Модернізація системи соціальної підтримки населення України»
- «Національний кадровий резерв»
- Єдина інформаційна система соціальної сфери

Податкова і митна сфери

- «AI HS Code Recommendation Platform» (Митна служба)
- моніторинг с/г земель (Податкова служба)

Сфера правосуддя

- «Касандра»
- «Суд в смартфоні»
- «Суд на долоні» (WINCOURT і VerdictudPro)

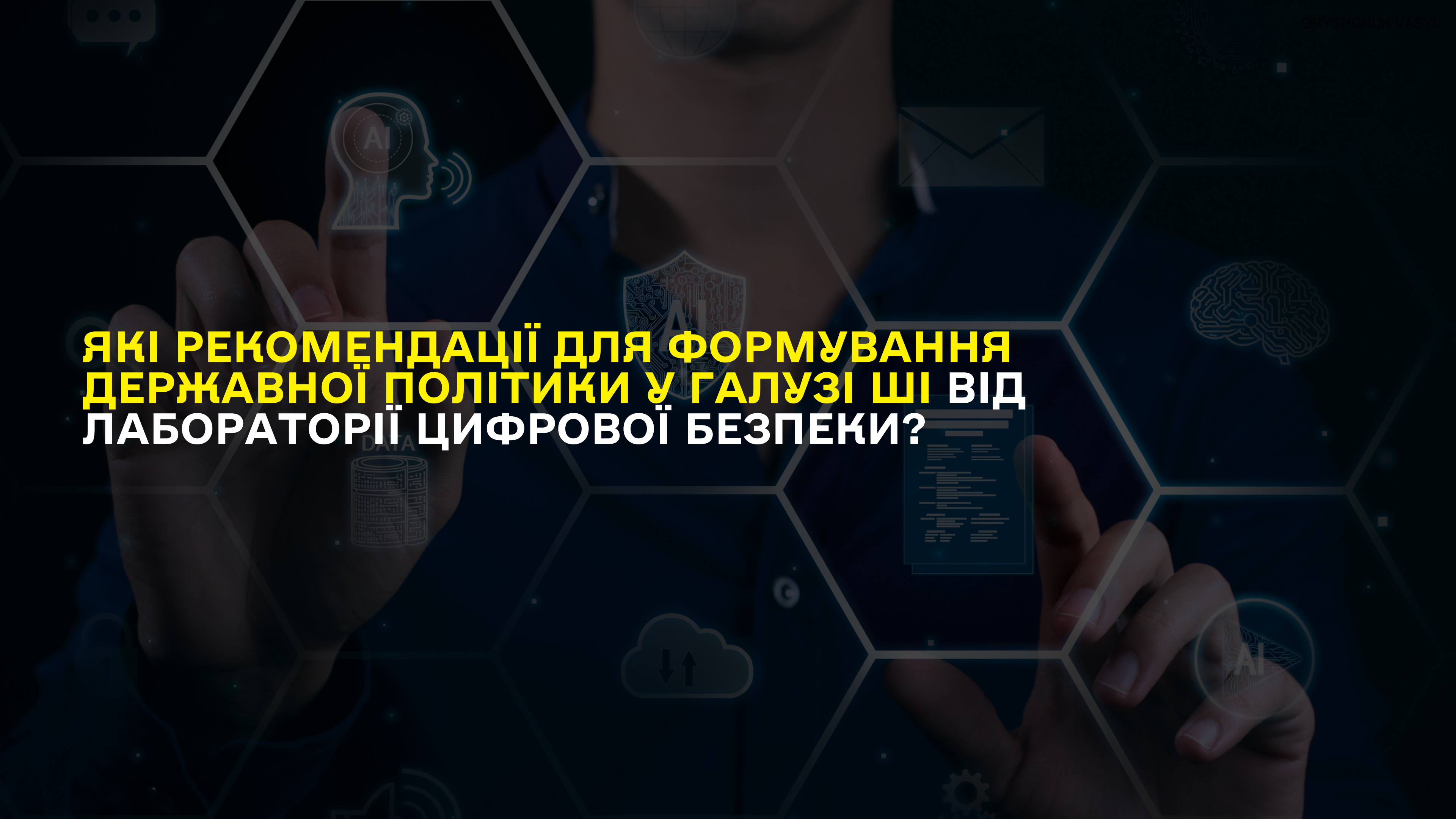
Сфера правопорядку

- камери стеження на основі ШІ (розпізнавання облич)
- «Clearview AI»
- використання в цілях розслідування правопорушень (зокрема, СБУ, Нацполіція, НАБУ, САП, БЕБ, НАЗК)

Оборонна сфера

- «Clearview AI»
- системи ситуаційної обізнаності («Palantir», «Delta» та їх аналоги)
- автоматичне наведення ППО
- використання ШІ для розмінування
- ініціативи в межах «Армії Дронів»
- ініціативи в межах Brave1

ЯКІ РЕКОМЕНДАЦІЇ ДЛЯ ФОРМУВАННЯ ДЕРЖАВНОЇ ПОЛІТИКИ У ГАЛУЗІ ШІ ВІД ЛАБОРАТОРІЇ ЦИФРОВОЇ БЕЗПЕКИ?



ЯКІ Ж ВИСНОВКИ?

Можна з упевненістю сказати, що ШІ відіграє певну роль у формуванні публічного управління та роботі публічних службовців в майбутньому – в тій чи іншій формі.

Незважаючи на те, що існують проблеми, пов'язані зі значною мірою покладанням на ШІ, такі як упередженість і розбіжності в даних і питання відповідальності, є десятки переваг, які ми можемо отримати, якщо ефективно використовуватимемо ШІ.



А ви “за” чи
“проти”?