

## Лекція. Проблема штучного інтелекту

### План

1. Концепція штучного інтелекту
2. Критерії наявності інтелекту
3. Етико-філософські виміри проблеми штучного інтелекту

*1. Штучний інтелект – метафорична назва одного з пріоритетних наукових напрямів, що склався у загальному комплексі кібернетичних досліджень із проблем моделювання процесів мислення, інтенсифікації інтелектуальної діяльності через комп'ютеризацію її видів.*

Метафоричний сенс концепту «штучний інтелект» полягає у використанні двох метафор:

- 1) головний мозок за структурною організацією та діяльністю уподібнюється комп'ютеру;
- 2) комп'ютер наділяється мозкоподібними структурами та функціями.

Виділяють декілька блоків науково-філософських питань, пов'язаних із «штучним інтелектом»:

- 1) принципової можливості;
- 2) технічної здійсненності;
- 3) соціально-практичної доцільності.

В основі відповіді на перші питання – теорія машин А. Тьюрінга і теорія формальних нейронних сіток Маккаллока-Піттса, з яких випливає основна теза теорії штучного інтелекту:

«Будь-яка функція мислення, описана логічно-строго і однозначно (з допомогою скінченного числа слів мовою з чіткою семантикою) у принципі може бути реалізована формально нейронною сіткою, а отже, і універсальною машиною Тьюрінга (тобто комп'ютером)».

Зазначену тезу спростовують, виходячи з теорем Курта Геделя (1930) про неповноту («Будь-яка система математичних аксіом починаючи з певного рівня складності або внутрішньо суперечлива або неповна»).

*Примітка: Ідея неможливості логічної завершеності, внутрішньо несуперечливої мови спростовувалася ще давньогрецькими софістами (парадокси «критянина, який бреше», «оплати послуг вчителя риторики» та ін.).*

Водночас, на думку прибічників теорії штучного інтелекту (А.Тьюрінга, В.Глушков), теорема про неповноту працює лише за умови, коли певна інтелектуальна система функціонує ізольовано від зовнішнього світу.

І, навпаки, теорема губить свою силу, коли інтелектуальна система неперервно взаємодіє з актуально нескінченною інформаційною базою.

Залежно від характеру відповіді на питання про принципову можливість штучного інтелекту Дж. Сьорл виділяє 2 гіпотези:

1) «сильного штучного інтелекту»: «відповідно запрограмований комп'ютер з потрібними входами та виходами і буде розумом, у тому смислі, в якому людський розум – розум», по суті, йдеться про творення штучних особистостей. Відмінність між людським розумом та штучним інтелектом – кількісна, тобто комп'ютер здатний мислити так ж як і людина, наприклад, розуміти як людина;

2) «слабого штучного інтелекту»: здійснення заданих інтелектуальних/обчислювальних операцій на підґрунті заданих незмінних правил (приклади: електронні перекладачі, чат-боти, системи розпізнавання облич, безпілотні транспортні засоби та ін.). Відмінність між людським розумом та штучним інтелектом – якісна, тобто комп'ютер здатний рахувати, обчислювати, але не здатний мислити.

Для ілюстрації принципової якісної обмеженості технологій штучного інтелекту Дж. Сьорл використовує мисленневий експеримент «китайська кімната»: Ви не знаєте китайську мову. Уявіть, що Вас розмістили у кімнаті, в якій розставлені кошики з китайськими ієрогліфами. Вам видали правила поєднання цих ієрогліфів відповідно до їх форми. Люди за дверима, які знають китайську мову, будуть передавати Вам у віконце картки з питаннями на китайській мові. У відповідь ви будете їм передавати свої картки відповідно до наданих правил поєднання китайських ієрогліфів. Люди, які знають китайську мову, можуть подумати, що ви її теж знаєте. Але насправді ви просто бездумно виконуєте надані інструкції, так як це робить комп'ютер виконуючи програми, складені програмістами.

## 2. Критерії наявності інтелекту.

Існують різні підходи. Так, пропонують такі критерії як:

- здатність приймати рішення в умовах невизначеності;
- здатність навчатися;
- здатність планувати;
- здатність спілкуватися на природній мові;
- сила волі;
- здатність відчувати, сприймати навколишній світ;
- здатність усвідомлювати себе як окрему особистість;
- здатність розуміти, співпереживати, любити, ненавидіти та ін.;
- поєднання всіх цих здатностей заради досягнення певної цілі.

та ін.

У 1950 р. у статті «Обчислювальні машини та розум» (Computing Machinery and Intelligence) А. Тьюрінг запропонував замінити питання «Чи можуть машини думати» питанням «Чи можуть машини робити те, що робимо ми як мислячі істоти». Щоб продемонструвати цей підхід, був запропонований тест: «Людина взаємодіє з 1 комп'ютером та 1 людиною. На підставі відповідей на питання вона повинна визначити, з ким вона

розмовляє: з людиною чи комп'ютерною програмою. Завдання комп'ютерної програми ввести людину в оману, змусивши зробити невірний вибір».

Випробування вважається успішним, якщо вдалося обдурити хоча би 30% суддів. У 2014 р. чат-бот програма «Женя Густман» на конкурсі Turing Test 2014 Prize в англійському університеті Редінга змогла переконати 33% суддів, що вона справжній одеський 13-річний хлопець.

Появі тесту Тьюрінга передувала книга філософа неопозитивіста Альфреда Айєра «Мова, істина та логіка» (1936 р.), в якій був запропонований алгоритм розпізнавання людини, яка усвідомлює, та машина, яка не усвідомлює: «Єдиною підставою, на якій я можу твердити, що об'єкт, який видається розумним, насправді не розумна істота, а просто дурна машина, є те, що він не зможе пройти один із емпіричних тестів, за якими визначається наявність чи відсутність свідомості».

### 3. Філософсько-етичні виміри проблеми штучного інтелекту

Існують протилежні підходи до розуміння ролі штучного інтелекту у бутті людини та суспільства.

У рамках просвітницької концепції штучний інтелект розглядається як:

- рушій соціального поступу, зростання продуктивності праці, безпеки, гарантування справедливості і народовладдя (наприклад, «електронна демократія»);

- емансипації, звільнення людини від одноманітної нетворчої праці.

На сучасному етапі просвітницька концепція продовжується у трансгуманізмі, представники якого (Нік Бостром, Девід Пірс, Ганс Моравек та ін.) підкреслюють, що стрімкий розвиток штучного інтелекту збільшує еволюційну перспективу людства і веде до «постлюдського майбутнього», зокрема, за допомогою застосування технологій штучного інтелекту для «поліпшення людини»:

- штучні «розумні» органи;
- нейрокомп'ютерний інтерфейс;
- збереження та завантаження свідомості та ін.

Разом із цим прогрес технологій штучного інтелекту змушує переосмислювати поняття «людина» і задумуватися над питаннями захисту прав всіх істот наділених розумом (приклад: Айзек Азімов «Три закони роботехніки», «Двохсотрічна людина» та ін.).

У 1998 р. Н. Бостром та Д. Пірс заснували Всесвітню організацію трансгуманістів (<https://www.humanityplus.org/>).

Н+