

Семінар бажано побудувати за зразком методу "за" і "проти" довкола основного питання Алана Тьюрінга: Чи можуть комп'ютери мислити?

1. Тест Тьюрінга: імітаційна гра, аргументи та контраргументи.
2. Мисленнєвий експеримент «Китайська кімната»: у чому полягає мисленнєвий експеримент, аргументи та контраргументи.

Питання для обговорення:

- Чи є тест Тьюрінга тестом на свідомість?

Не зовсім, адже перевіряє не свідомість, а інтелектуальну поведінку та здібності. Тюрінг відкидає ідею теоретично приходити до висновку чи має машина розуміння, де під машиною має на увазі саме цифрові комп'ютери. Натомість пропонує тест-гру, який визначає, чи поведінка машини не відрізняється від поведінки людини. Аргументація Тьюрінга паралельно з даним тестом є те, що усі людські аспекти такі як емоції, голос, поведінка відкидаються а перевіряються лише інтелектуальні здібності. Також він робить акцент на тому, що головна проблема не апаратних можливостях, а саме в програмуванні. Навіть про людську неформальність, а саме не можливість описати всі ситуації, він каже, що це просто складні до реалізації правила. Він пропонує навчати машину, як дитину, з можливостями вчитися на помилках, ситуаціях, і цим розвивати її інтелектуальну поведінку.

- Чи спростовує «Китайська кімната» тест Тьюрінга?

Я не можу сказати що це спростування, адже тексти наголошують на різних речах, тож скоріш це як критика зі сторони Серля. Тест Тьюрінга каже нам про інтелектуальні здібності/поведінку машини, з акцентом на те, щоб у письмовому спілкуванні людина не могла відрізнити машину від людини (розпізнавання та формулювання людей за певними патернами). Китайська кімната звертає увагу не просто на вміння машини відтворювати правила, а на відсутність у неї справжньої свідомості - тобто уміння усвідомлювати прочитане/сказане й виходити за межі заданих алгоритмів чи шаблонів, що властиве лише біологічному рівню людини. Акцент полягає на тому що поведінка може виглядати розумно, але без всякого розуміння сказаного, що зводить до того, що семантика не ідентична синтаксису

- Що визначає розуміння мови — синтаксис чи семантика?

Розуміння мови залежить передусім від семантики, адже саме вона передає зміст і надає можливість комунікувати. Можна побудувати синтаксично правильне речення, яке буде повним абсурдом й не надасть жодної інформації. Натомість якщо дати змістовну

відповідь на питання, з деякими синтаксичними помилками - це матиме набагато більше сенсу

- Оцінка сучасного статусу тесту Тьюрінга (чи працює тест нерозрізнюваності) та які є альтернативи до нього?

Концепт тесту Тьюрінга досі залишається доволі популярним, як ідея. До прикладу, версії штучного інтелекту такі як GPT, Copilot, Gemini і тд можуть синтаксично вірно формулювати “свої думки”, але чи є вони насправді їхніми? Скоріш вони є масовим досвідом зібраних з найрізноманітніших ситуацій та питань. Але на практиці він сьогодні часто вважається радше філософським експериментом, ніж робочим методом оцінки. Наразі альтернативами до перевірки можуть послужити більш людські фактори такі, як сприйняття, емоція та інші.

Для прикладу перевірки семантичного розуміння:

- Тест Маркуса(ввімкніть якусь комедію ШІ, та потрібно щоб він пояснив де сміятися(перевірка на розуміння іронії, сарказму та юмору))
- Виноград схеми(на честь вченого Террі Винограда): суть у тому що робот має дві відповіді, але речення формулюються таким чином, що без їх семантичного розуміння дати відповідь можна тільки вгадавши
- Який сучасний стан розробки ШІ та які перспективи (чи справдяться прогнози Р.Курцвейля про сингулярність)?

Курцвейл передбачає, що після сингулярності машини будуть здатні **самовдосконалювати себе**, а розвиток технологій стане **експоненційним**. Це момент у майбутньому, коли **штучний інтелект перевищить людський інтелект**, а технологічний прогрес стане **настільки швидким і глибоким**, що люди не зможуть передбачати його наслідки.

Загалом, описані слова Курцвейлом є частково правдивими, адже за останні декілька років ШІ від простих синтаксично правильних речень був швидко вдосконалений і став майже подібним людині у діалозі(семантично правильно формулює речення, та навіть може частково розуміти емоції та інші людські характеристики). Однак чи можна стверджувати, що він робить це свідомо? Ні. Сучасні системи ШІ базуються на великих масивах даних і статистичних моделях, але не має власної свідомості чи самостійного мислення.

Прогнози Курцвейля виглядають більше як гіпотеза, аніж гарантований сценарій. Наразі ми можемо тільки спостерігати за справді експоненційним розвитком технологій, але щодо перевищення людського інтелекту - я думаю, що зарано щось казати.

(Якраз те про що говорив Тьюрінг)

- Трансгуманізм - це спосіб конкурувати з ШІ, перемогти чи інтегруватися з ним?

Трансгуманізм – це рух, який підтримує використання науки та технологій для покращення розумових і фізичних можливостей людини, прагнучи подолати невід'ємні обмеження, такі як старіння, хвороби та смерть, та трансформувати людство в істот з надзвичайно розширеними можливостями, що називаються «постлюдиною»

Трансгуманізм робить акцент на інтеграції, а не на конкуренції чи перемозі. Його головна ідея є використання технологій у покращені та розширені людських можливостей. Замість протистояння ШІ, трансгуманізм розглядає майбутнє, де розвиток ШІ та людини відбувається у тандемі.