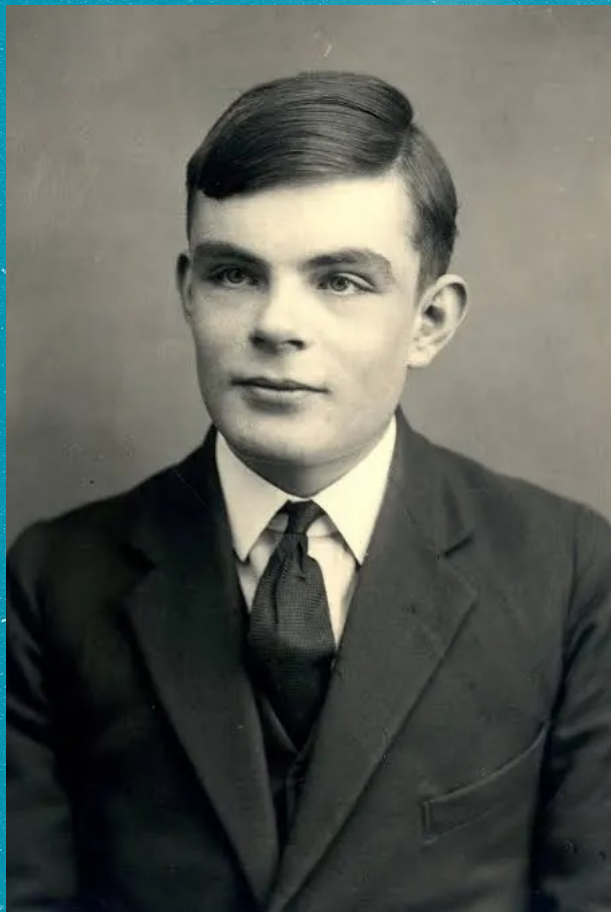


Тест Тьюринга і аргумент "Китайська кімната".



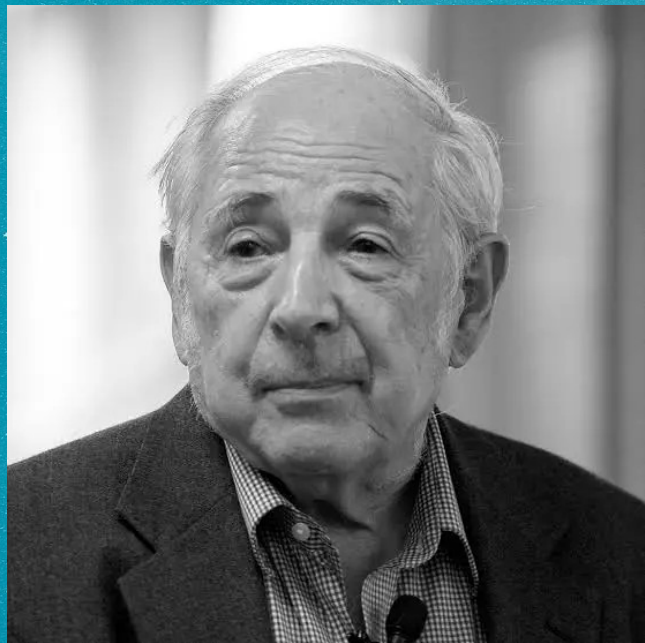
# Тест Тьюринга і аргумент "Китайська кімната". Автори. Алан Тьюрінг



Алан Тьюрінг англійський математик і один із засновників теоретичної інформатики, який першим показав, що процеси мислення можуть бути описані як обчислювальні операції. Він запропонував ідею “універсальної машини”, що здатна виконувати будь-який алгоритм - концепцію, яка лежить в основі сучасних комп’ютерів. У своїй статті 1950 року Тьюрінг сформулював тест, покликаний оцінити, чи можна вважати машинну поведінку інтелектуальною. Його підхід зосереджується на тому, як мислить машина зовні, а не на тому, що відбувається всередині.



# Тест Тьюринга і аргумент "Китайська кімната". Автори. Джон Серль



Джон Серль - американський філософ, який досліджував природу свідомості та ставив під сумнів можливість того, що комп'ютер може "розуміти" так, як людина. У 1980 році він запропонував аргумент "Китайська кімната", спрямований проти ідеї, що виконання програми може бути еквівалентним справжньому мисленню. Серль стверджує, що комп'ютери працюють лише з синтаксисом - формальними символами без значень - і тому не мають семантики та інтенціональності. Його позиція протиставляє поведінковому підходу Тьюринга і підкреслює необхідність внутрішнього розуміння для мислення.



# Тест Тьюринга і аргумент "Китайська кімната". Проблема: Чи можуть машини мислити?

Центральна проблема обговорення полягає в тому, чи може комп'ютер не лише імітувати людську поведінку, а й мати справжнє розуміння, свідомість або інтенціональність. Тьюрінг пропонує оцінювати інтелект машини за її здатністю поводитися так само, як людина, тоді як Серль наголошує, що формальна обробка символів не створює змісту чи розуміння. Таким чином виникає суперечність між поведінковим та внутрішнім підходами до визначення мислення.





# Тест Тьюринга і аргумент "Китайська кімната".

## Цитати авторів

Алан Тьюрінг:

1. Про заміну питання “Чи можуть машини мислити?”

“I propose to consider the question, ‘Can machines think?’... Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.”

2. Про критерій поведінки

“The new form of the problem can be described in terms of a game which we call the ‘imitation game.’”

3. Про достатність поведінкової еквівалентності

“If a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.”

Джон Серль:

1. Про відсутність розуміння у програм

“In the Chinese case I have everything that artificial intelligence can put into me by way of a program, and I understand nothing.”

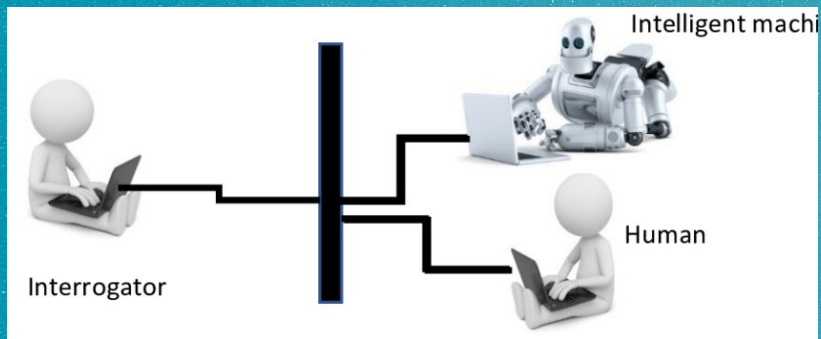
2. Про синтаксис без семантики

“All I do is manipulate formal symbols... but I still understand nothing.”



# Тест Тьюрінга і аргумент "Китайська кімната".

## Тест Тьюрінга (Імітаційна гра)



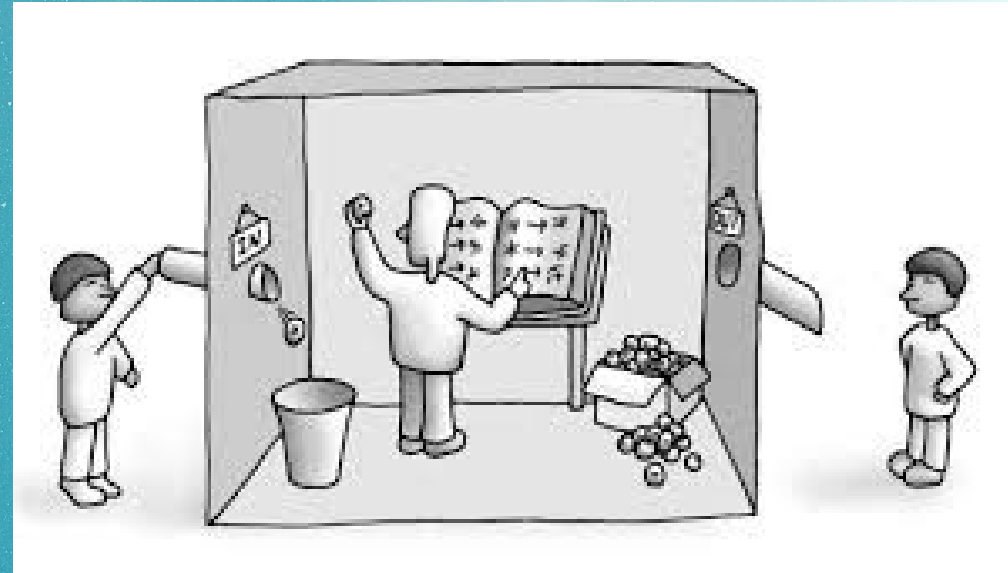
Тест Тьюрінга - це запропонований Аланом Тьюрінгом спосіб визначити, чи може машина “мислити” в поведінковому сенсі. У грі людина-суддя спілкується з двома невидимими співрозмовниками: людиною і машиною. Якщо суддя не може надійно визначити, хто з них машина, вважається, що машина успішно пройшла тест. Суть тесту: якщо поведінка машини невідрізнявана від людської, ми можемо трактувати її як інтелектуальну.



# Тест Тьюринга і аргумент «Китайська кімната».

## Китайська кімната

“Китайська кімната” показує, що машина може ідеально імітувати знання мови, не розуміючи її. Уявіть людину, яка не знає китайської, але отримує правила, як відповідати на китайські символи іншими символами. Зовні її відповіді виглядають правильно, але вона не розуміє змісту жодного слова. Серль робить висновок: виконання програми - це лише маніпуляція символами, а не справжнє розуміння чи мислення.





# Тест Тьюрінга і аргумент "Китайська кімната". Конфлікт поглядів Тьюрінга та Серля

Алан Тьюрінг вважав, що питання “Чи мислить машина”» слід оцінювати за поведінкою: якщо машина здатна імітувати людину так, що її не можна відрізнити в розмові, то це функціонально дорівнює мисленню.

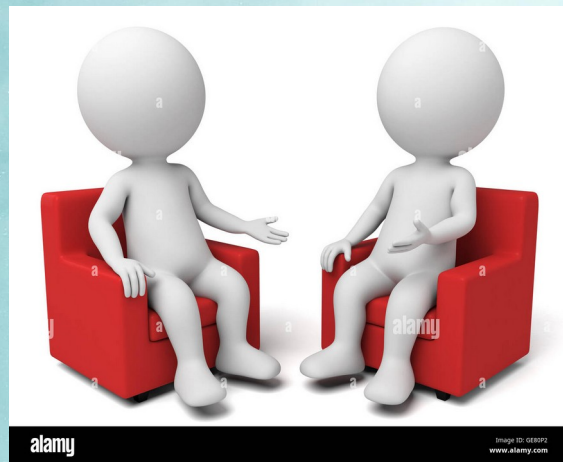
Джон Серль заперечує: імітація ще не означає свідомість, бо машина оперує синтаксисом, але не володіє семантикою — внутрішнім змістом і розумінням. Його “Китайська кімната” показує: можна давати правильні відповіді, нічого не розуміючи. Отже, для Серля імітація  $\neq$  розуміння, а Тест Тьюрінга не є тестом на свідомість.

Systems Reply каже, що «розуміє вся система», але Серль відповідає, що навіть якщо людина запам’ятає всю систему, вона все одно не розумітиме китайську.

Robot Reply пропонує дати машині сенсори та тіло, але Серль заперечує: навіть робот лише переробляє символи, не розуміючи їхнього змісту.

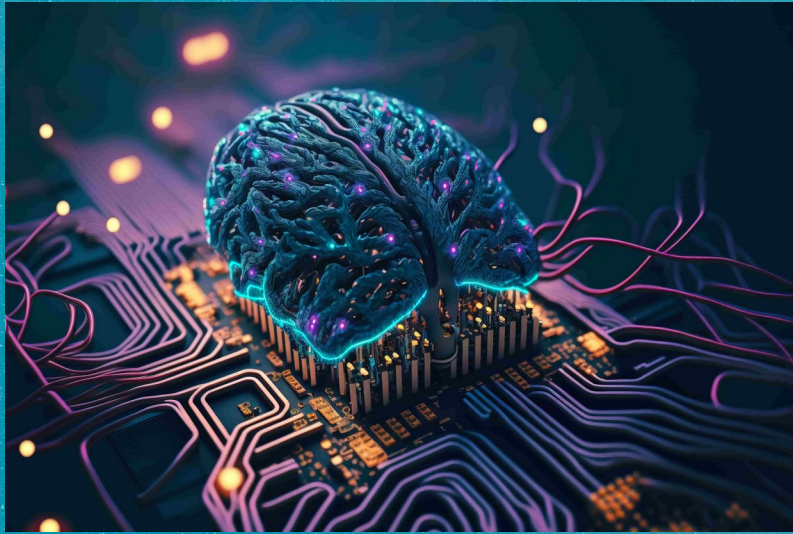
Серль вважає, що справжнє мислення вимагає біологічних каузальних процесів мозку, а не просто виконання програми.

Таким чином, Тьюрінг зводить мислення до поведінки, тоді як Серль - до внутрішнього розуміння.





# III, свідомість і межі тесту Тьюрінга



1. Чи є тест Тьюрінга тестом на свідомість?

Ні. Він перевіряє не свідомість, а лише інтелектуальну поведінку - здатність машини імітувати людину в розмові. Тьюрінг свідомо уникав питання “розуміння”, зводячи оцінку до того, як машина поводить себе, а не що вона усвідомлює.

2. Чи спростовує “Китайська кімната” тест Тьюрінга?

Не спростовує, але критикує: Серль показує, що правильні відповіді ще не означають розуміння. Поведінка може виглядати “розумною”, хоча всередині немає жодної свідомості чи значення.

3. Що визначає розуміння мови: синтаксис чи семантика?

Серль підкреслює: справжній зміст передає семантика. Можна правильно будувати речення за правилами (синтаксис), але все одно не розуміти, про що йдеться.



# Сьогодення ШІ, альтернативи та погляд у майбутнє



1. Який сучасний стан розробки ШІ та які перспективи (чи справдяться прогнози Курцвейля про сингулярність)?

Рей Курцвейл прогнозує технологічну сингулярність - момент, коли ШІ перевищить людський інтелект і зможе самостійно себе вдосконалювати. Сучасні моделі ШІ справді розвиваються дуже швидко, проте вони працюють на статистичних закономірностях і не мають власної свідомості. Тому сингулярність поки залишається гіпотезою: розвиток є експоненційним, але “перевищення людського інтелекту” ще не доведене.

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

Трансгуманізм - це філософія, яка пропонує використовувати технології для розширення фізичних і розумових можливостей людини. Він не бачить ШІ як суперника, а радше як інструмент, з яким людство може співпрацювати і вдосконалюватися разом. Тому трансгуманізм робить акцент не на конкуренції чи “перемозі”, а на інтеграції людини й технологій.