Philosophy of AI

AIMA Ch 26

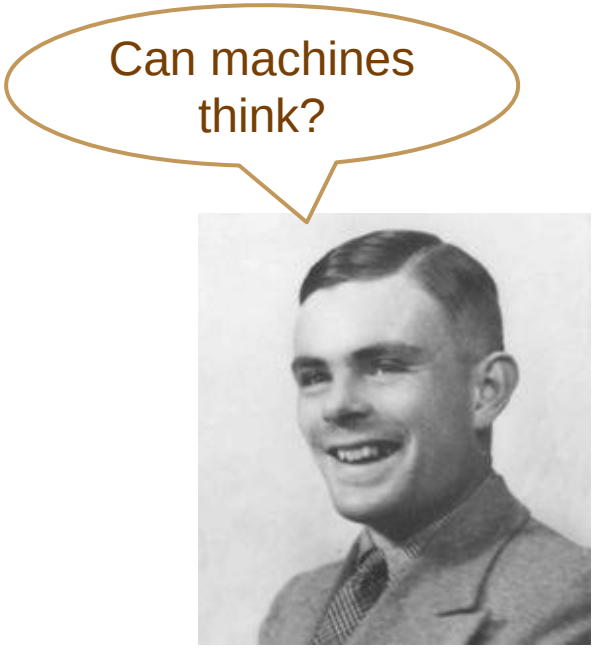
Philosophical Questions of AI

- Can machines act intelligently?
- If they can, do they act in the same way that human do?
- If they do, would they have conscious minds?
- ……

Test for Intelligence

- We know how to judge intelligence in humans, so how can we test for intelligence in machines?

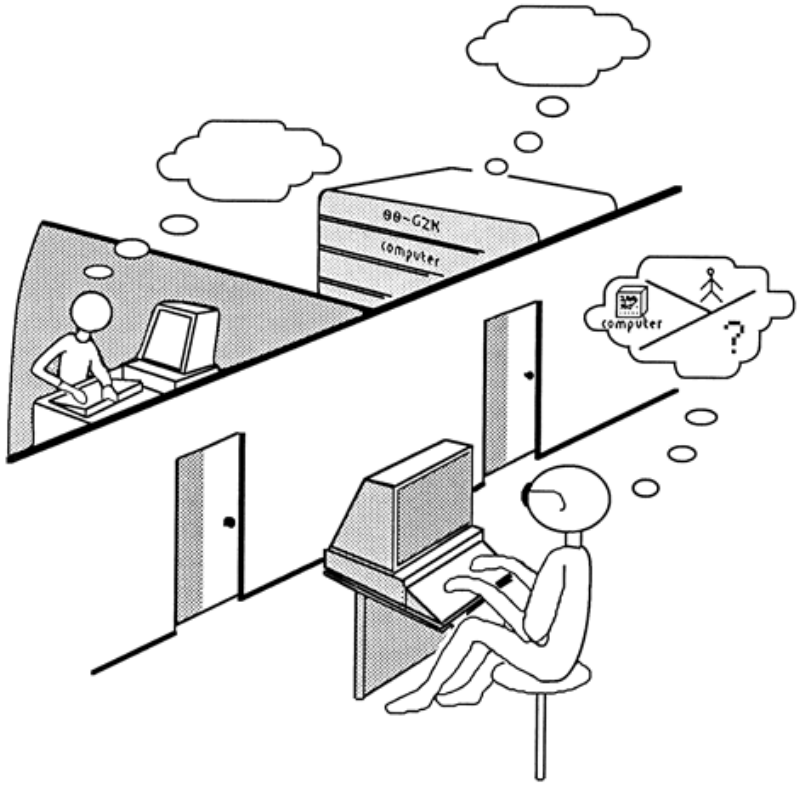
- The Turing Test
  - In 1950, Alan Turing thought up a test to see if a machine is as intelligent as a human.



Alan Turing

The Turing Test

- Put a person at a computer running a chat program connected to two other computers:
  - one has a person on the other end
  - the other has a computer
- If the 1st person can't tell which is the machine and which is human, then it's true Artificial Intelligence



Criticism of the Turing Test

- Human vs. general intelligence
  - [AIMA] Analogy with aeronautical engineering: “making machines that fly so exactly like pigeons that they can fool even other pigeons.”
- Behaviorism
  - More about this later

State-of-the-art

- Many attempts to produce computer programs (chatbots) that pass the Turing Test.
- Loebner Prize: annual competition (since 1991) of computer programs for passing the Turing Test.
  - https://aisb.org.uk/sample-page/
  - The prize is reported as defunct since 2020.

State-of-the-art

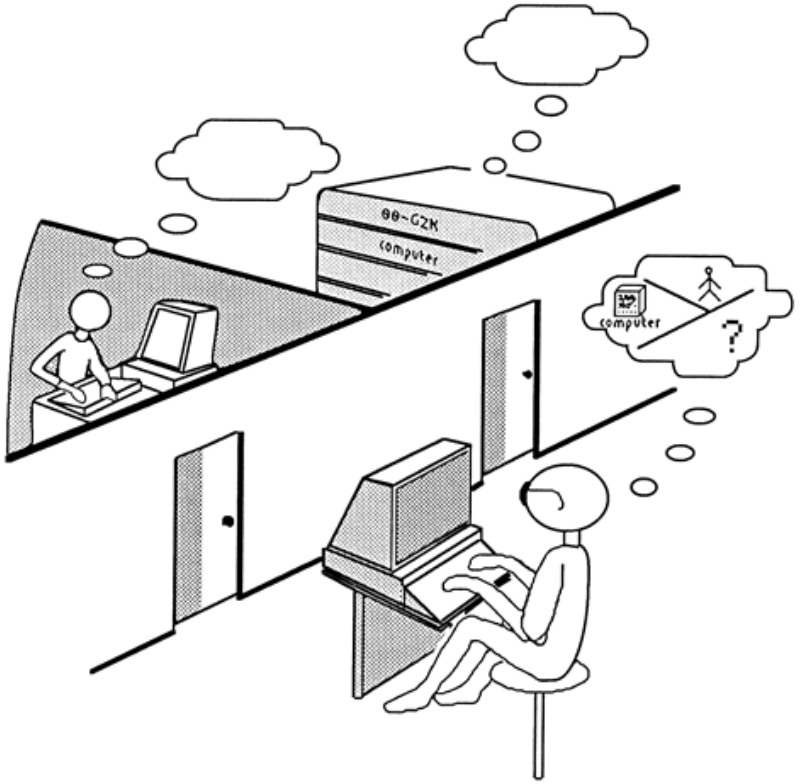
amazon alexa

Google ASSISTANT

Google's Duplex Demo @Google I/O 2018

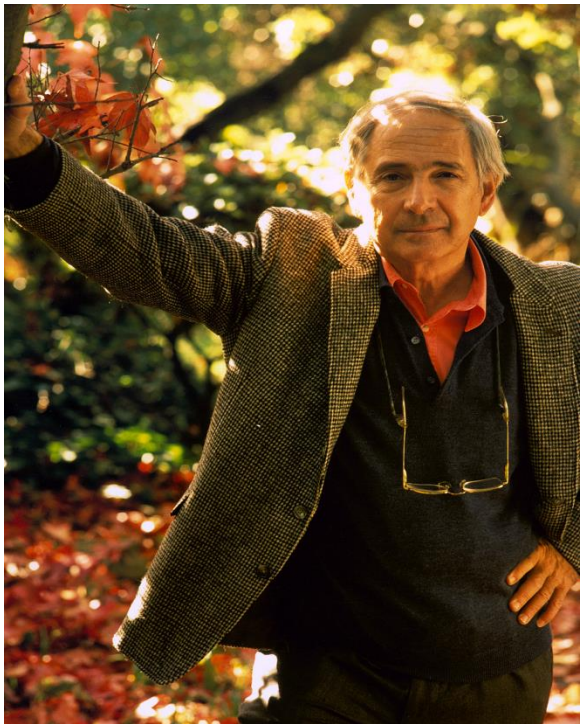
The Turing Test

- If eventually a program passes the Turing test, then does it really have intelligence like a human?



Searle's Chinese Room Argument

- John Searle
  - Famous philosopher at the University of California, Berkeley
  - The inventor of the term "strong AI"!
  - Wrote "Minds, Brains and Programs" in 1980, which described the "Chinese Room Argument"



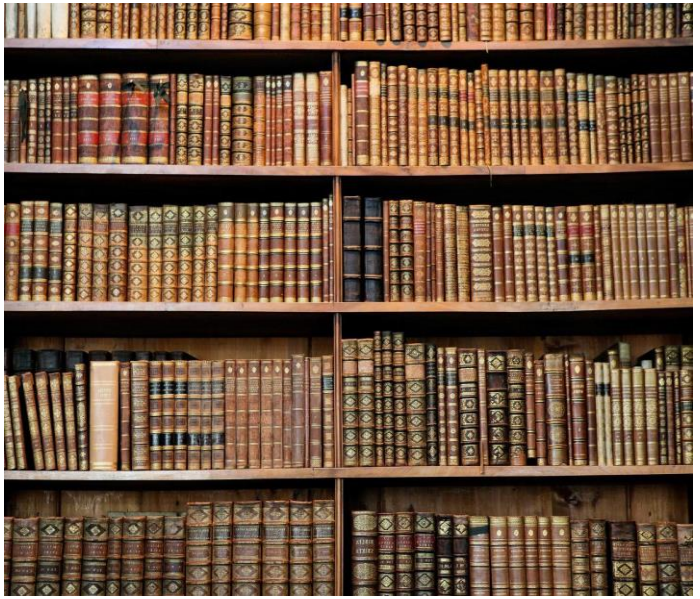
Searle's Chinese Room

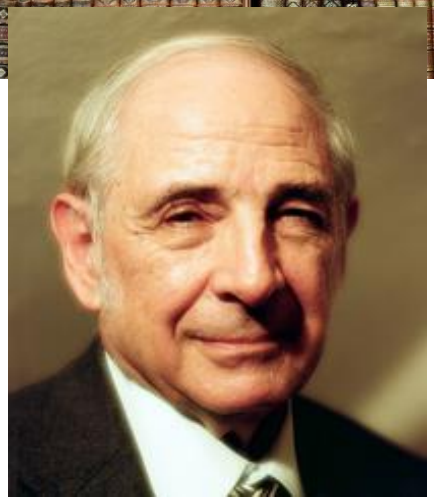
- John Searle
  - Doesn't know any Chinese language.
  - Never seen a Chinese character.
  - Doesn't even know that there are languages other than English.

Searle's Chinese Room

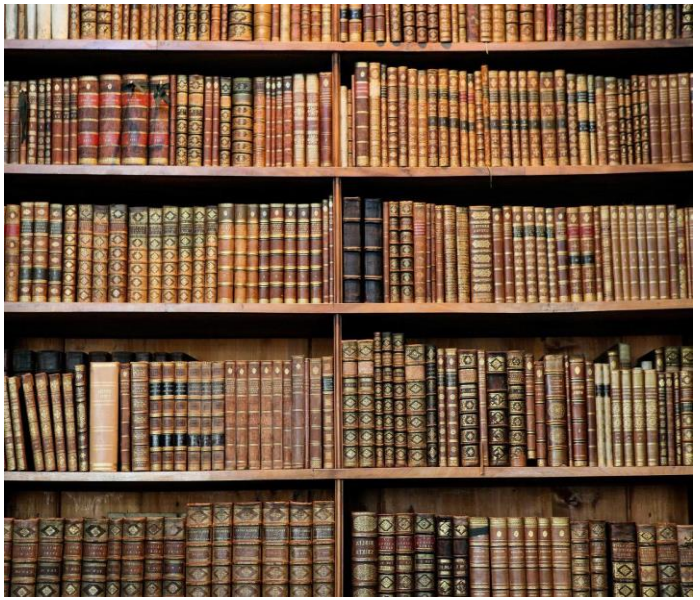
- Searle's New Job
  - He's told that he works for a company that makes decoration drawings
  - Currently, the company needs to update their drawings
  - Searle's job is to receive "input" drawings and update them to the new drawings.

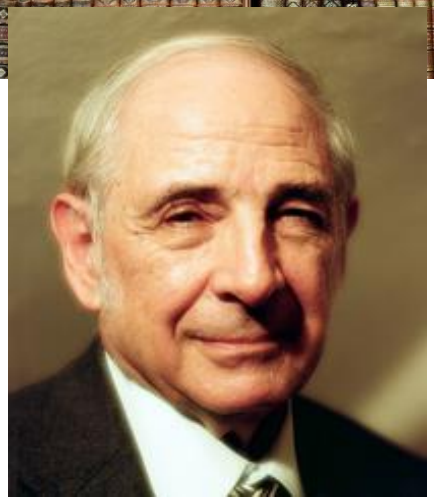
Searle’s Chinese Room





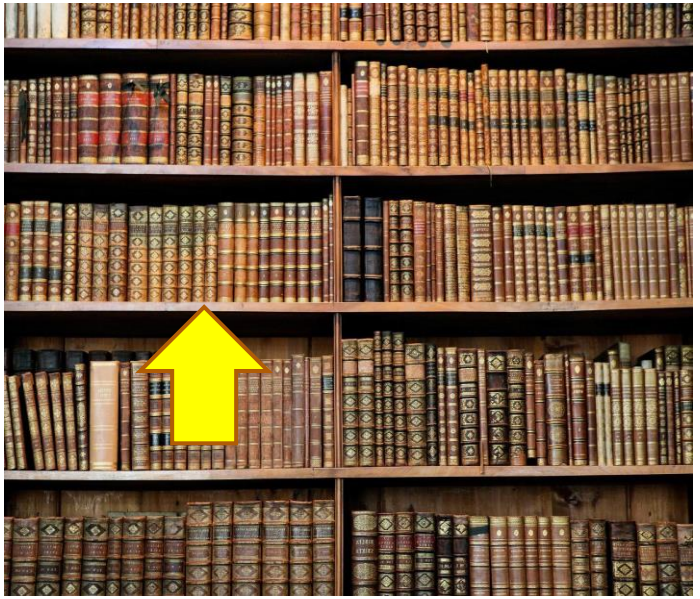
Searle’s Chinese Room

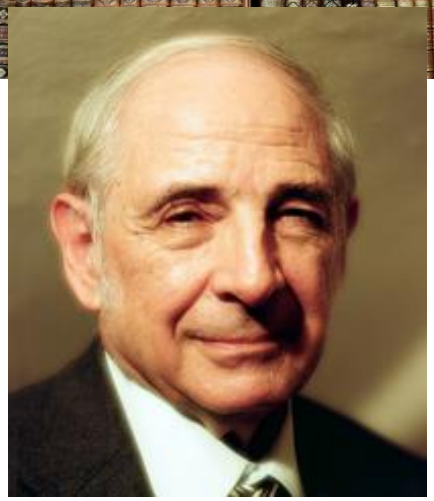




什麼 帶來
快樂

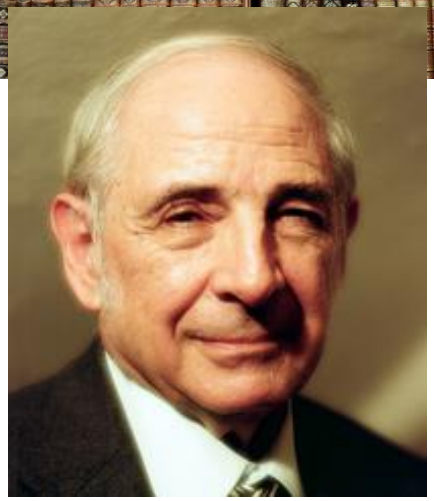
Step 1: Find Rulebook #37

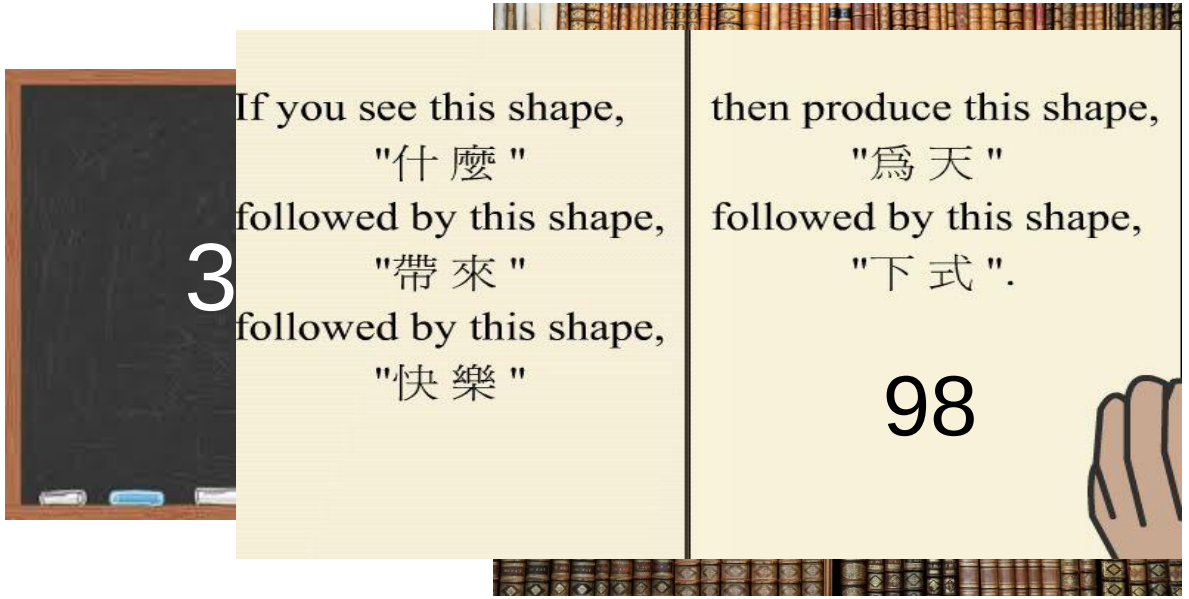




什麼 帶來
快樂

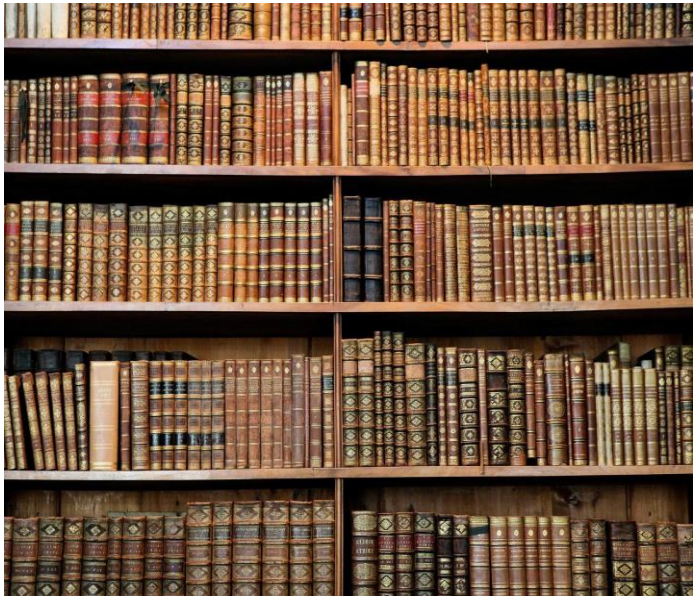
Step 2: Find Instructions for this Squiggle.

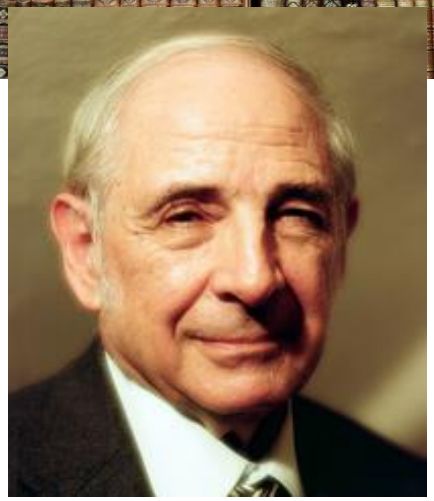


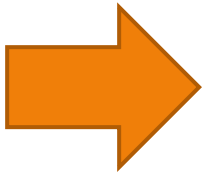


什 麼 帶 來
快 樂

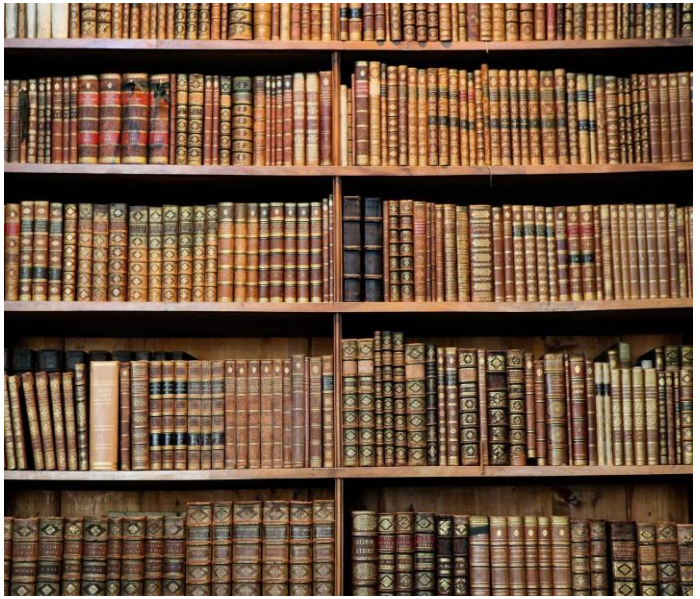
Step 3: Copy Down New Squiggles







What's Going On?



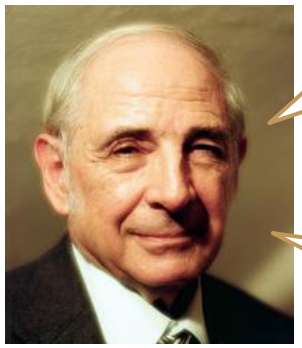
- Searle is running a Chinese AI program
  - The number on the blackboard correspond to different states: tired, hungry, in a hurry, bored…
  - Each book contains rules of what one would say, given the state, in response to the question

What's Going On?

What's Going On?

- Searle is running a Chinese AI program

- Searle passes the Chinese version of the Turing test!
- Yet he has no understanding of what's going on



Similarly, a program cannot give a computer "understanding".

No intelligence in a computer!

The Systems Reply

- Searle is part of a larger system. Searle doesn't understand Chinese, but the whole system (Searle + room + rules) does understand Chinese.

Searle's Response

- Searle argues that in theory, he could just memorize all the rules, and get rid of the rest of the system. Now the entire system = Searle, but Searle still does not understand Chinese.

- Counter-response
  - If Searle could internalize the rules, part of his brain would understand Chinese. Searle's brain would house two personalities: English-speaking Searle and Chinese-speaking system.

More replies to Searle

- The systems reply
- The robot reply
- The brain simulator reply
- The complexity reply
- The other minds reply

- See: http://en.wikipedia.org/wiki/Chinese_room#Replies