



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.

Can machines
think?

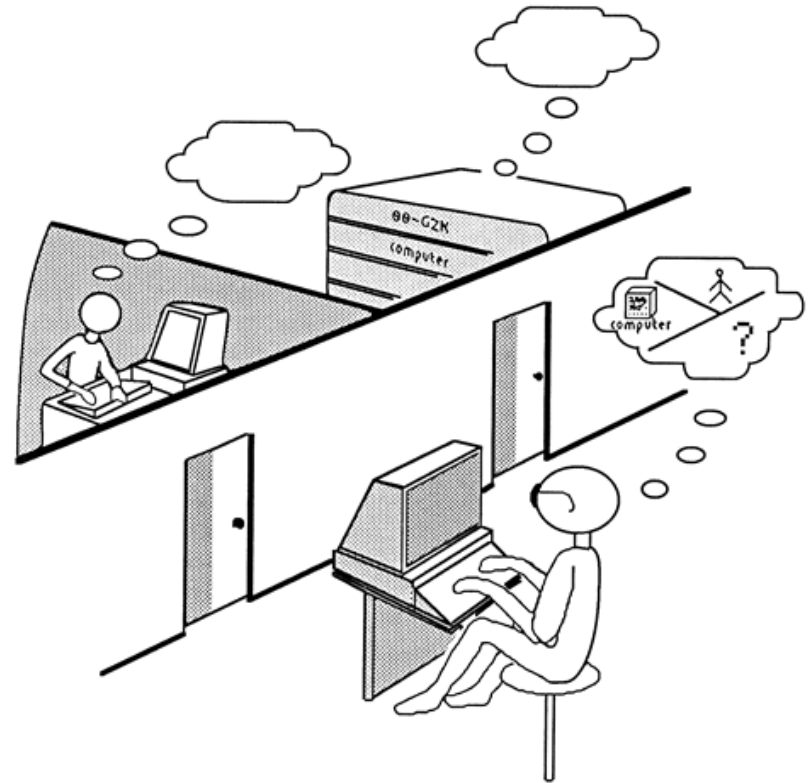


Alan Turing



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
- ▶ Judged by human

[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

- ▶ 70+ years have passed... Has any AI passed the Turing test?
- ▶ Yes (sort of)



Google's Duplex Demo @Google I/O 2018



ChatGPT (2022)

SO

名偵探柯男究竟是不是死神



名偵探柯男（也稱為柯南）是日本漫畫家青山別彥創作的著名推理漫畫人物。他是一名高中生，在他的推理技巧和各種科學技術的幫助下，解決了許多懸疑案件。名偵探柯男並不是死神，而是一名年輕的科學偵探。



SO

你如何解釋，為什麼名偵探柯男出現的地方就會出現命案？



這個問題有一些錯誤。名偵探柯男並不會因為他出現而導致命案發生。事實上，名偵探柯男是一名偵探，他的出現通常是為了解決案件，而不是導致案件發生。在故事情節中，柯男經常被推理案件的犯罪分子拖入案件中，並且利用他的推理技巧和科學技術來解決案件。因此，柯男出現的地方並不會導致命案，相反，他有助於解決案件，維護正義。



4GAMERS

ChatGPT (2022)

nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾ [Subscribe](#)

[nature](#) > [news feature](#) > [article](#)

NEWS FEATURE | 25 July 2023

**ChatGPT broke
race is on for new**



INDEPENDENT

[Push notifications](#) >

[Subscribe](#)

[NEWS](#) [SPORTS](#) [VOICES](#) [CULTURE](#) [LIFESTYLE](#) [TRAVEL](#) [PREMIUM](#)

Tech

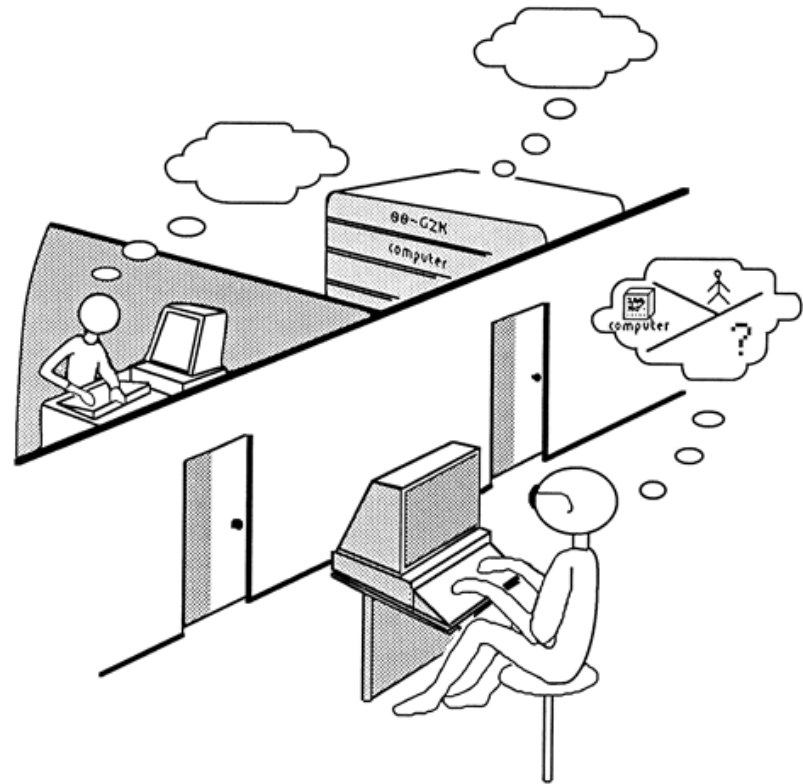
**ChatGPT beaten by 1960s computer
program in Turing test study**

One of the reasons ... for participants mistaking ELIZA for a human was that it was “too bad” to be a current AI model, and therefore “was more likely to be a human intentionally being uncooperative”.



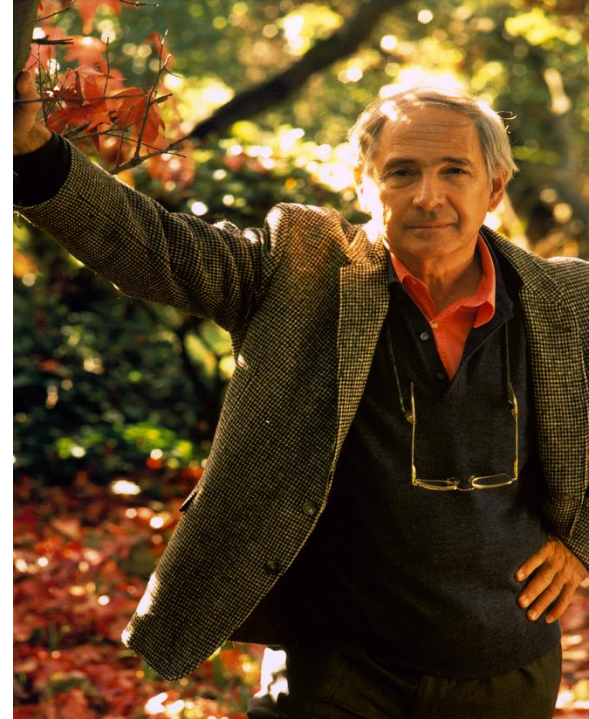
The Turing Test

- ▶ If 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.

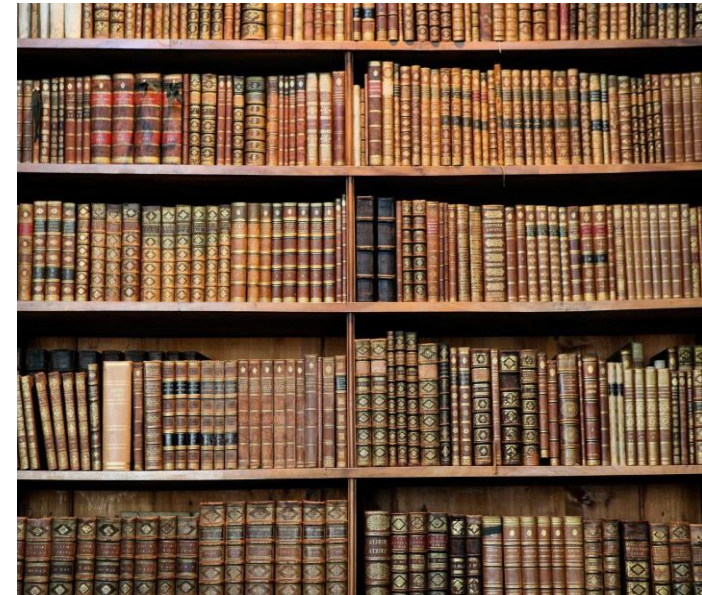


Searle's Chinese Room

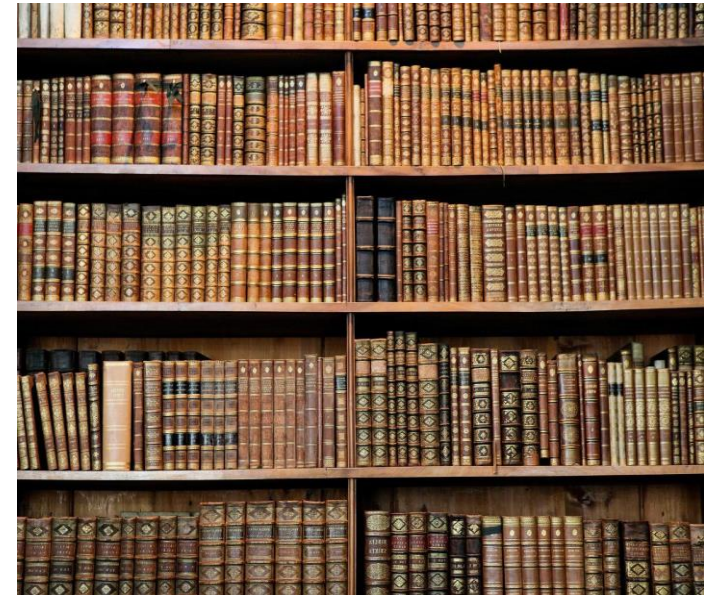
- ▶ Searle's New Job
 - ▶ He's told that he works for a company that makes decoration drawings
 - ▶ Currently, the company needs to update their drawings
 - ▶ Searle's job is to receive "input" drawings and update them to the new drawings.



Searle's Chinese Room



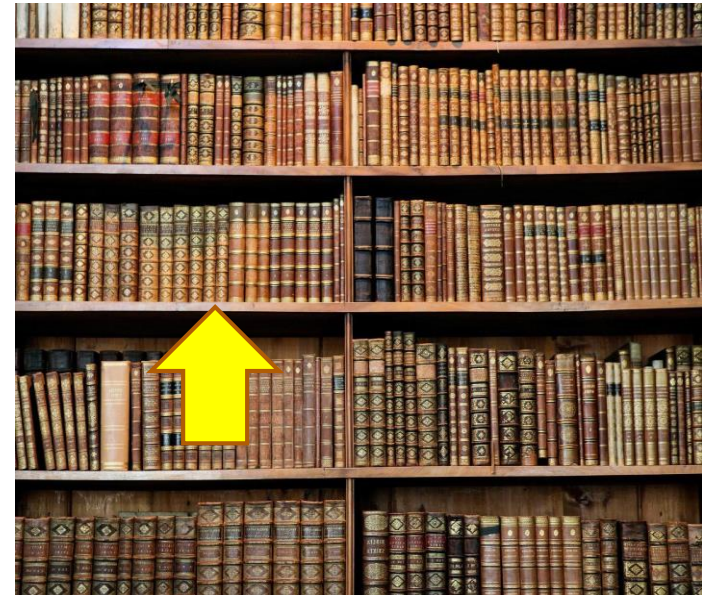
Searle's Chinese Room



什麼帶來
快樂



Step 1: Find Rulebook #37



什麼帶來
快樂



Step 2: Find Instructions for this Squiggle.



If you see this shape,
"什麼"
followed by this shape,
"帶來"
followed by this shape,
"快樂"

then produce this shape,
"爲天"
followed by this shape,
"下式".

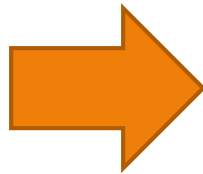
98



什麼帶來
快樂



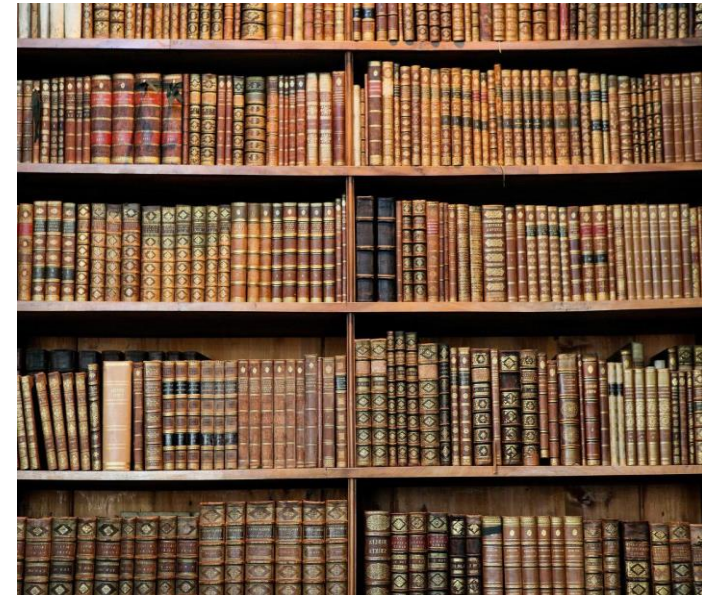
Step 3: Copy Down New Squiggles



Step 5: Update Blackboard



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

