



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?

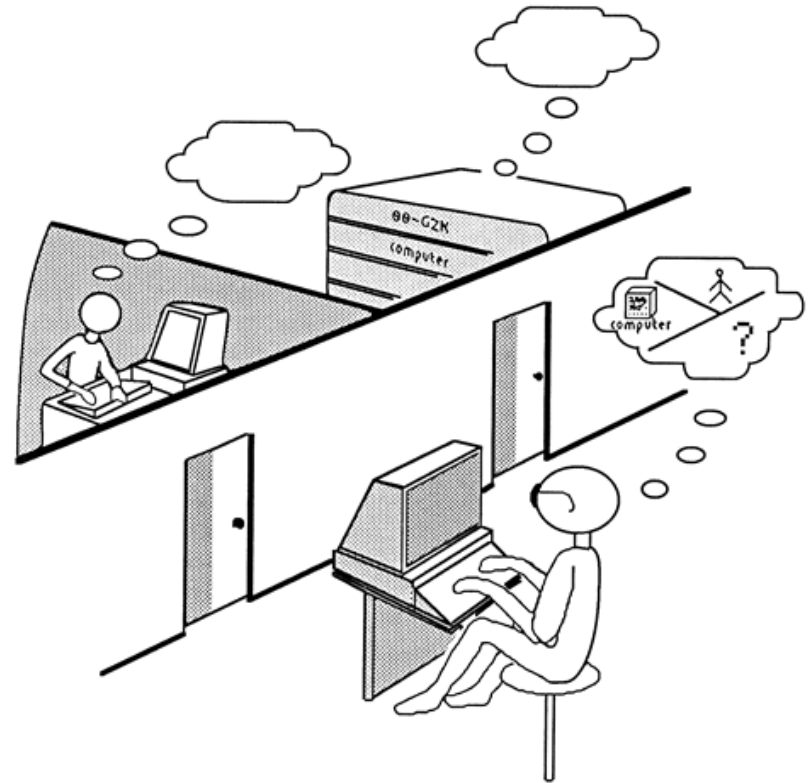


Alan Turing



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



GPT-3 (2020)

AI: I am an AI created by OpenAI. How can I help you today?

Human: I am feeling bored today. Grandma, tell me a story about the time the Cat stole the sun.

AI: Once upon a time, the Cat went to visit the Sun. He hadn't seen the Sun for quite some time. He told her to come out for a walk. The Sun told him that she didn't want to. The Sun was shy and didn't want to come out. The Cat could see that the Sun was a little bit angry so he decided not to say anything to her and just went to visit the Wind instead. But the Wind was angry too. The Wind wanted to blow the Cat away. The Cat ran back home, looked for something nice to eat and went to bed. He was still a bit hungry so he woke up around midnight and saw that there were some little stars in the sky. The stars came down to visit the Cat and they were very hungry as well. The Cat ate a star and they all went to bed and fell asleep. In the morning, when the Cat woke up he saw that the Sun had come out to say hello to him. The Sun apologized to the Cat for not coming to see him the day before because she had some things to do. They decided to go out and have a wonderful day together. The Cat and the Sun spent the whole day together. They had a lot of fun, but the Cat did not tell anyone about the stars he ate. That was a secret. The Cat wanted to keep it for himself. He liked that little secret and kept it close to his heart.



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)

USER **this code is not working like i expect — how do i fix it?**

```
resultWorkerErr := make(chan error)
defer close(resultWorkerErr)
go func() {
    defer cancel()
    resultWorkerErr <- b.resultWorker(ctx)
}()

err := b.worker(ctx)
cancel()
if err == nil {
    resultWorkerErr <- resultWorkerErr
}
return multierror.Append(err, <-resultWorkerErr)
```

CHATGPT It's difficult to say what's wrong with the code without more context. Can you provide more information about what the code is supposed to do and what isn't working as expected? Also, is this the entire code or just a part of it?

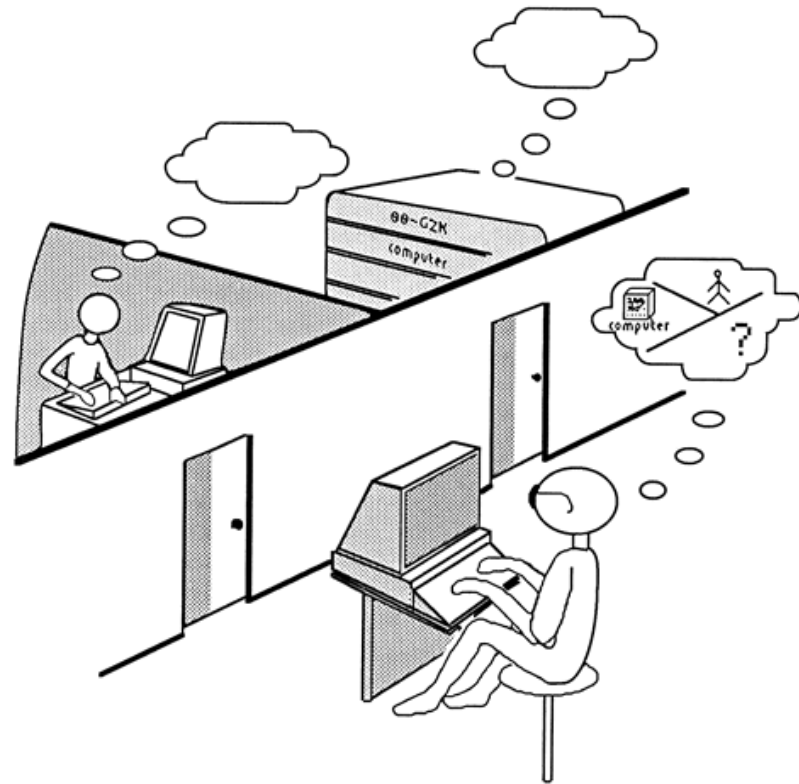
USER **just part of it — the error never surfaces. I think there is something wrong with the channel**

CHATGPT It's difficult to say without more information about what the code is supposed to do and what's happening when it's executed. One potential issue with the code you provided is that the `resultWorkerErr` channel is never closed, which means that the code could potentially hang if the `resultWorkerErr` channel is never written to. This could happen if `b.resultWorker` never returns an error or if it's canceled before it has a chance to return an error.

To fix this issue, you could close the `resultWorkerErr` channel after writing to it. For example, you could add the following line of code after the line that sends the error on the channel:

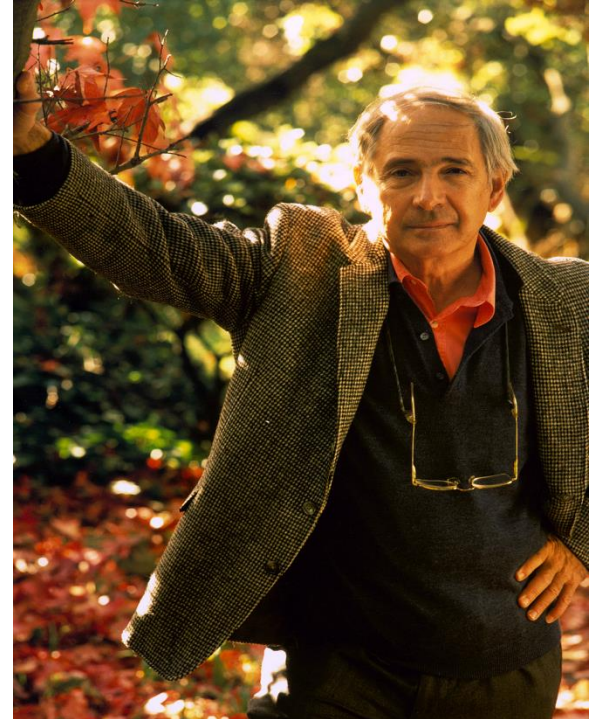
The Turing Test

- ▶ 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.

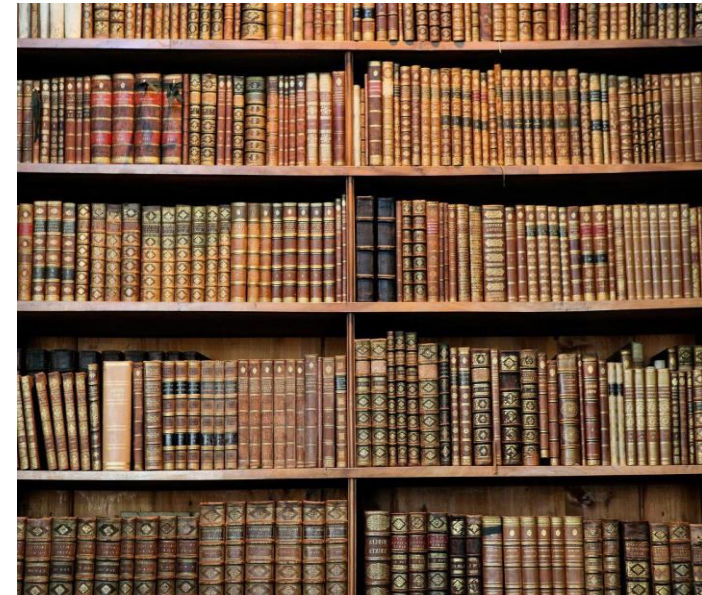


Searle's Chinese Room

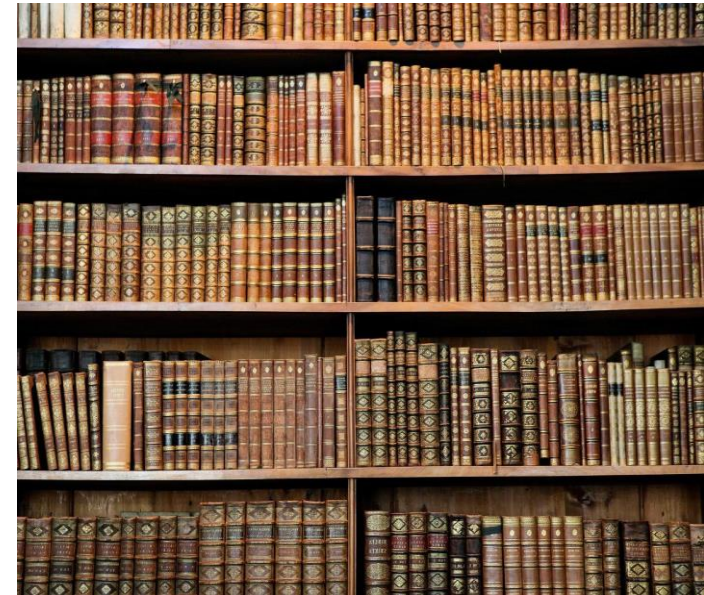
- ▶ Searle's New Job
 - ▶ He's told that he works for a company that makes decoration drawings
 - ▶ Currently, the company needs to update their drawings
 - ▶ Searle's job is to receive "input" drawings and update them to the new drawings.



Searle's Chinese Room



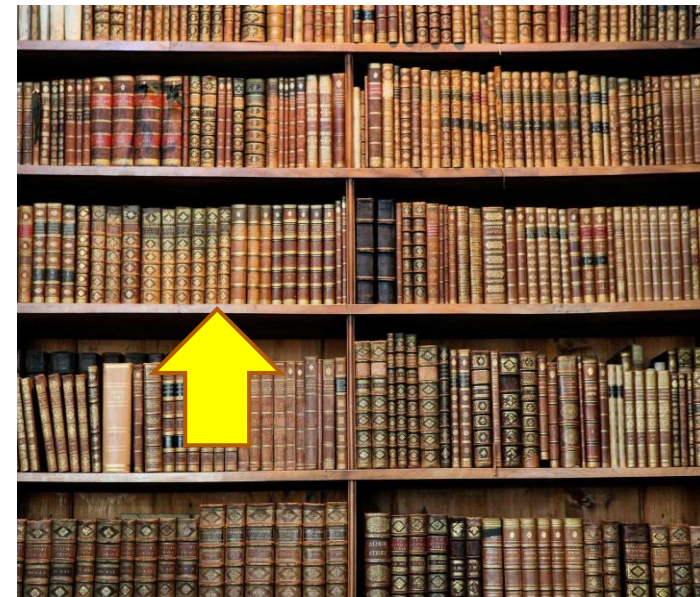
Searle's Chinese Room



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".

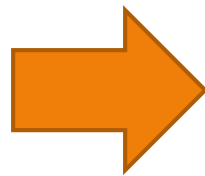
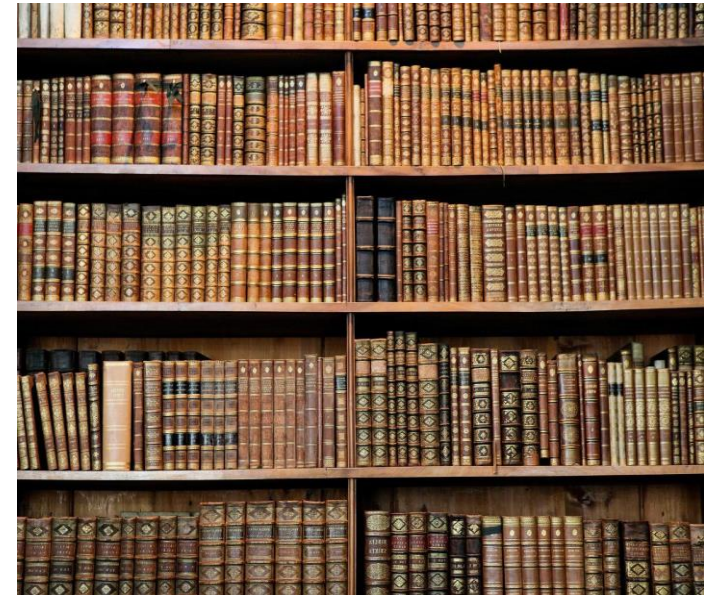
98



什麼帶來
快樂



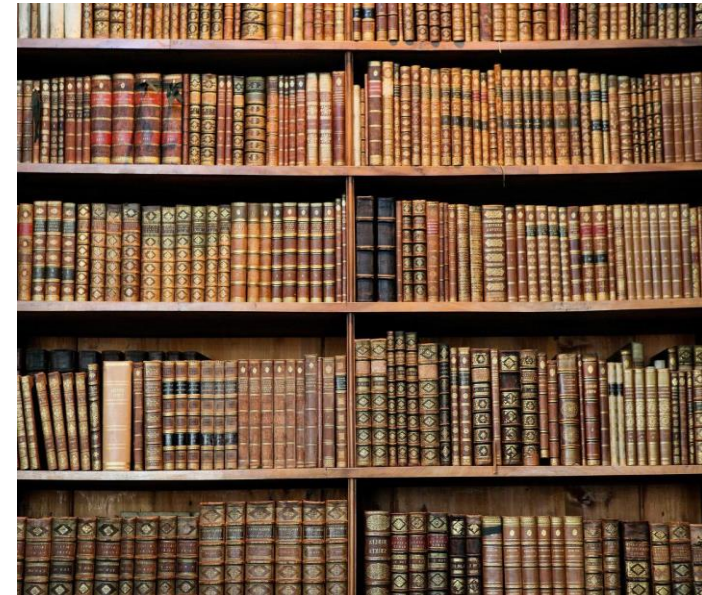
Step 3: Copy Down New Squiggles



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

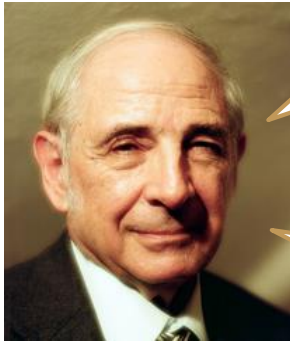


What's Going On?



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

