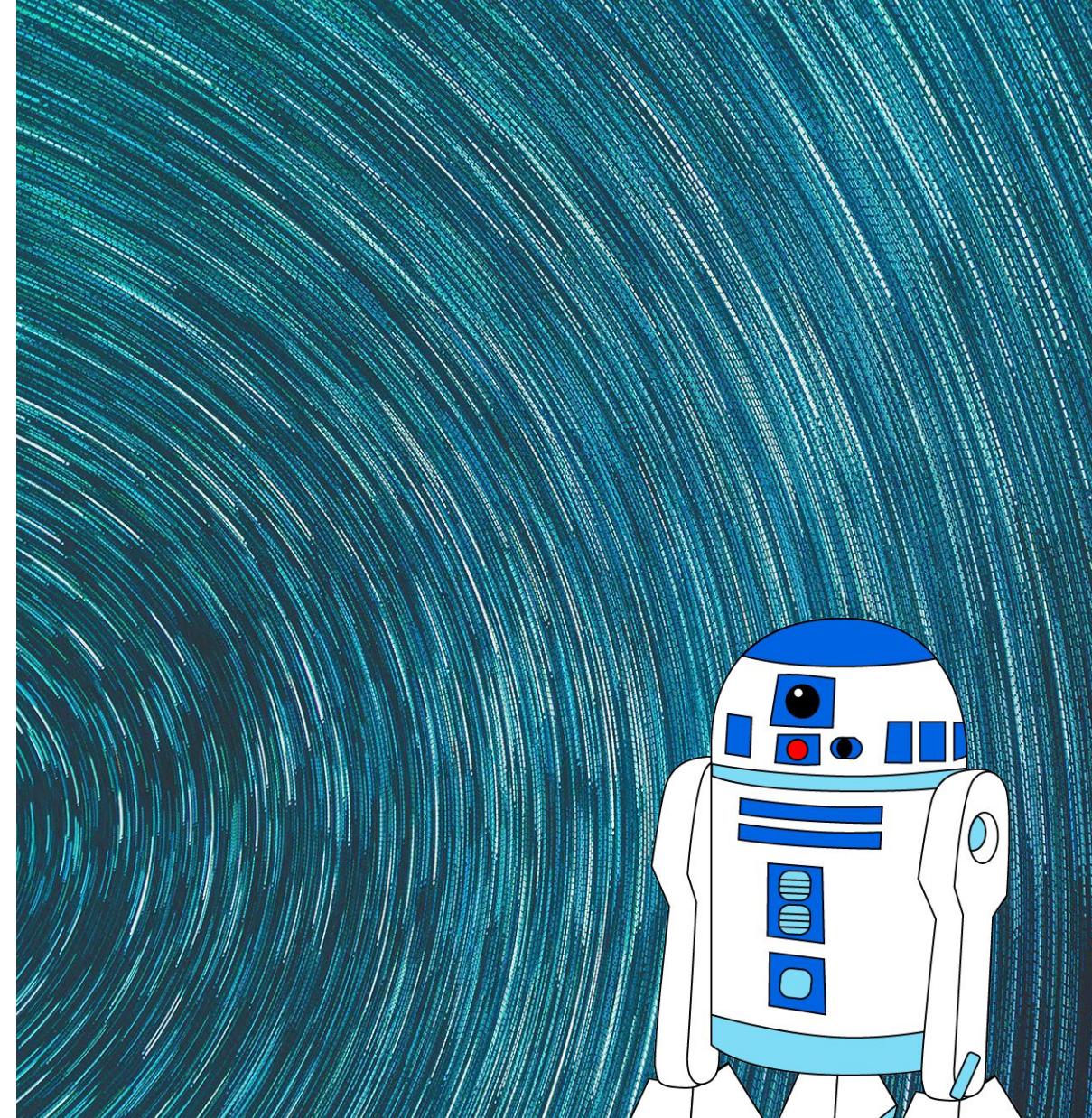


CIS 421/521:
ARTIFICIAL INTELLIGENCE

Artificial Intelligence in Philosophy



René Descartes (1596-1650)



cogito ergo sum
“I think, therefore I
am.”

Principle of *dualism* – that the mind or thinking self is essentially incorporeal or spiritual – that the mind exists separately from the body: "if a foot or arm or any other part of the body is cut off, nothing has thereby taken away from the mind."

René Descartes (1596-1650)

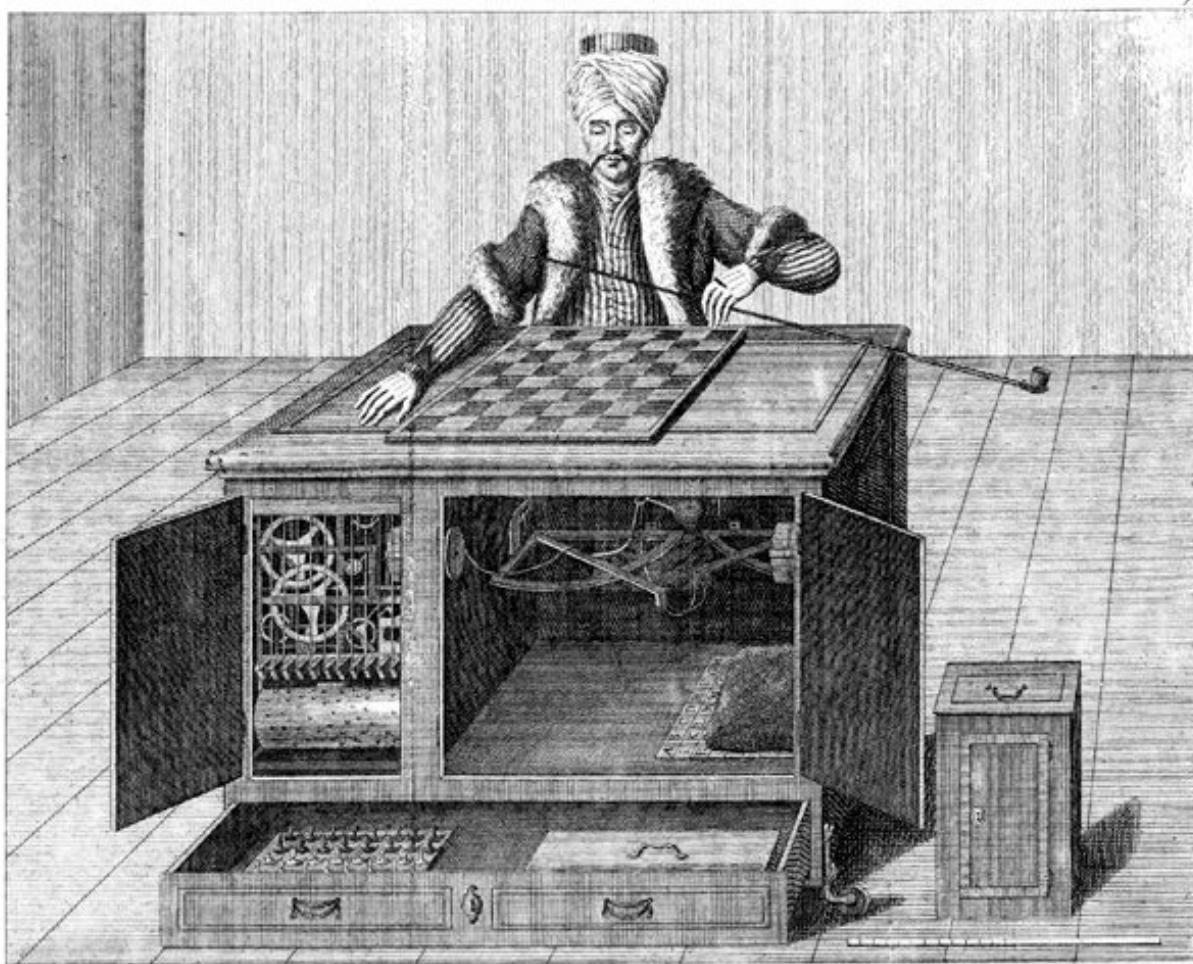


**How can I know that
you are not an
automaton?**

Rene Descartes wondered if he could know for sure that others who looked and behaved like him weren't in fact automata.

Bodies of people and animals are nothing more than complex machines - the bones, muscles and organs could be replaced with cogs, pistons and cams.

17th and 18th century automatons



W. de Kempelen sel.
Chez A. Moehl excud: Basilea
Per Schachspieler wicervordem Spiele gezeigt wird von mir Le Jouer d'echec, tel qu'on le montre avant le jeu, par devant.



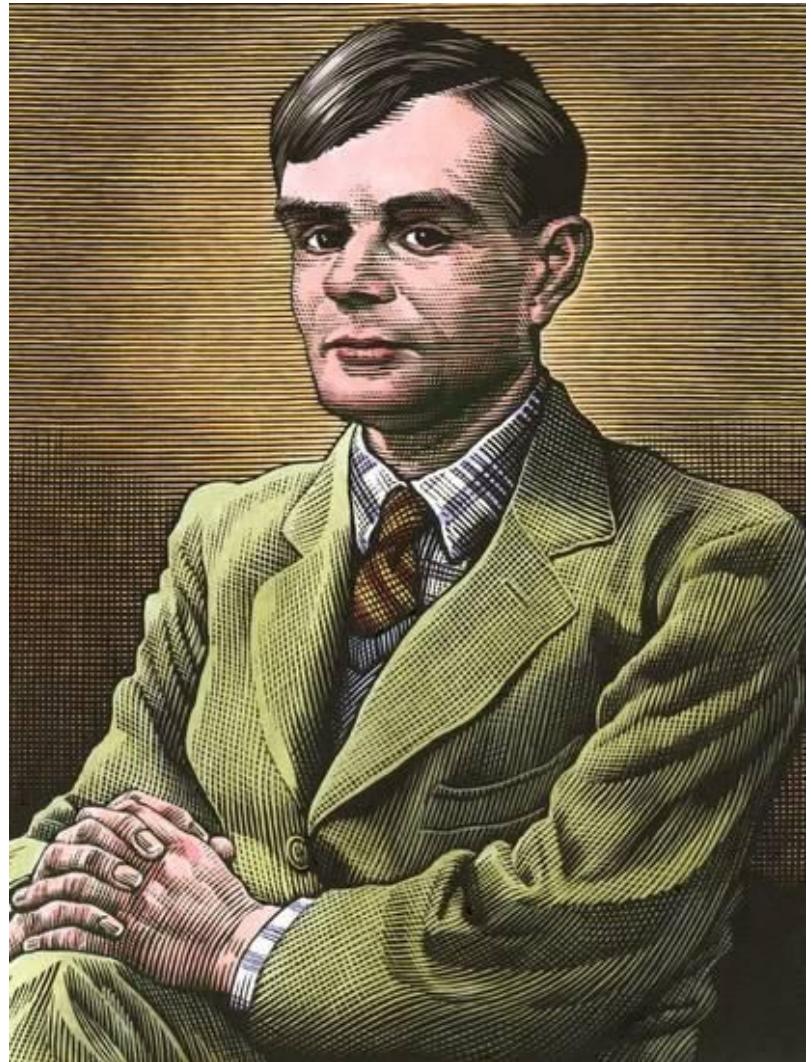
René Descartes (1596-1650)



**Let's design a test that
only a person could
pass.**

"if there were machines bearing the image of our bodies, and capable of imitating our actions. For example, if touched in a particular place it may demand what we wish to say to it; if in another it may cry out that it is hurt. However there would be two tests to know that they were not really men."... They could never use properly use language.

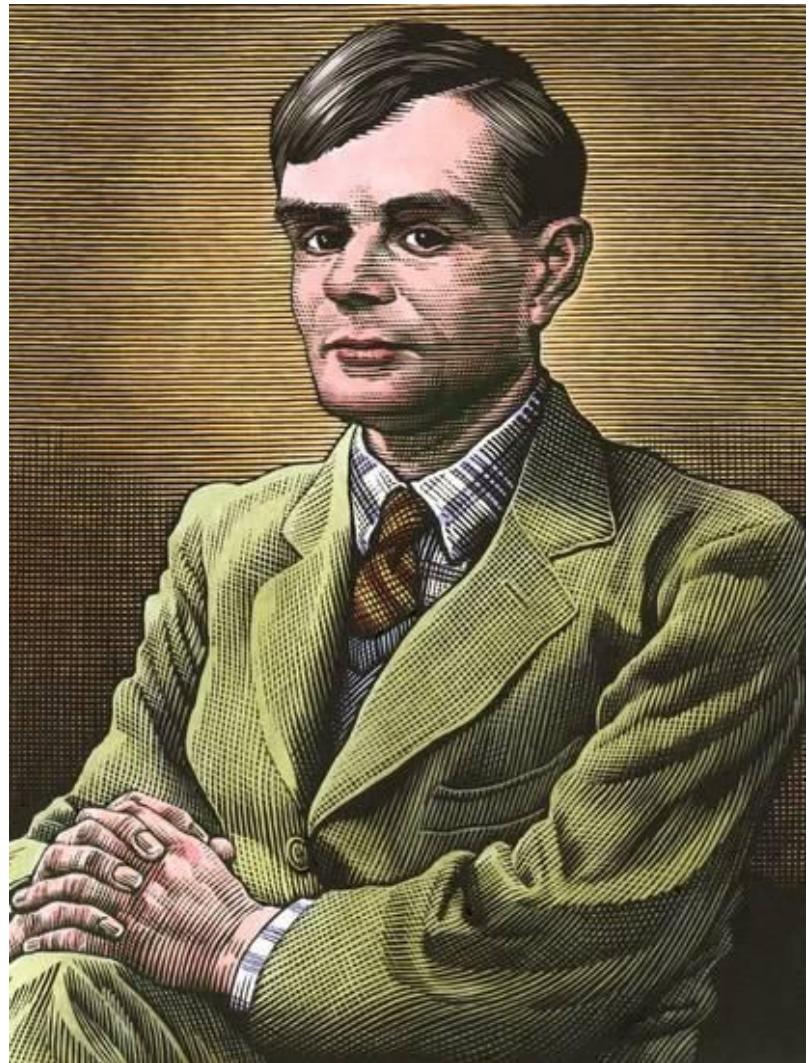
Alan Turing (1912-1954)



Can machines think? Let's operationalize the question.

"I propose to consider the question, 'Can machines think?'" Because "thinking" is difficult to define, Turing chooses to "replace the question by another, which is closely related to it and is expressed in relatively unambiguous words." Turing's new question is: "Are there imaginable digital computers which would do well in the *imitation game*?"

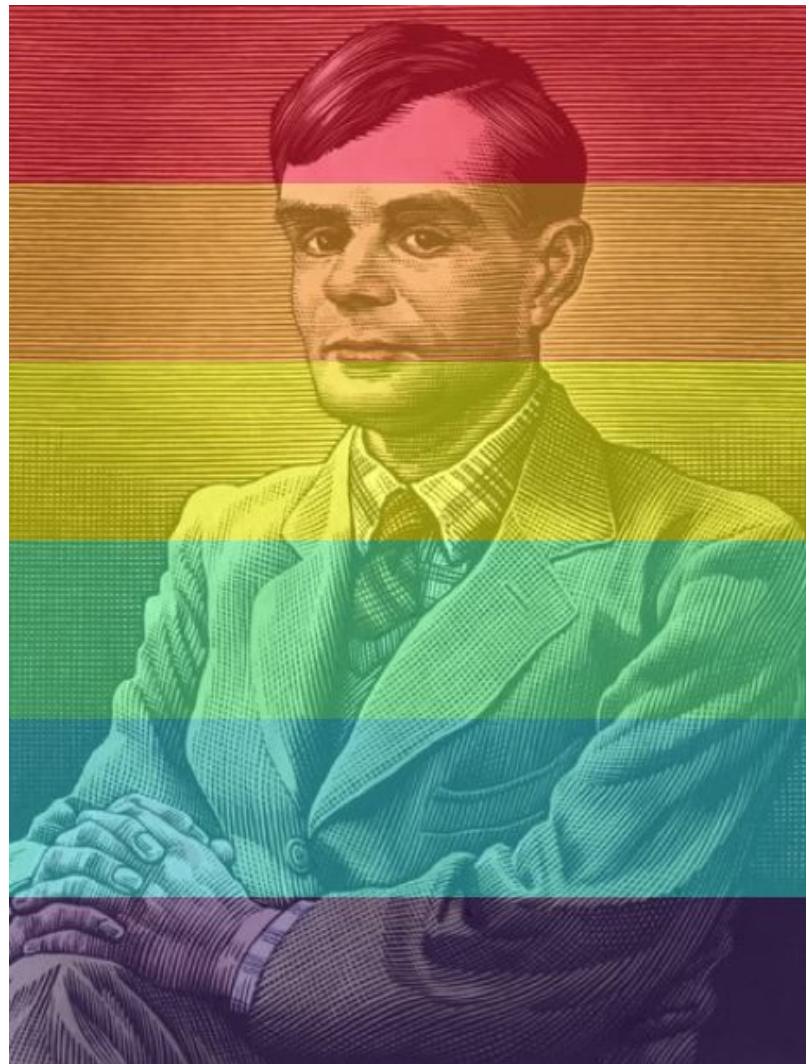
Alan Turing (1912-1954)



**Can a computer pass
as human in a
conversation?**

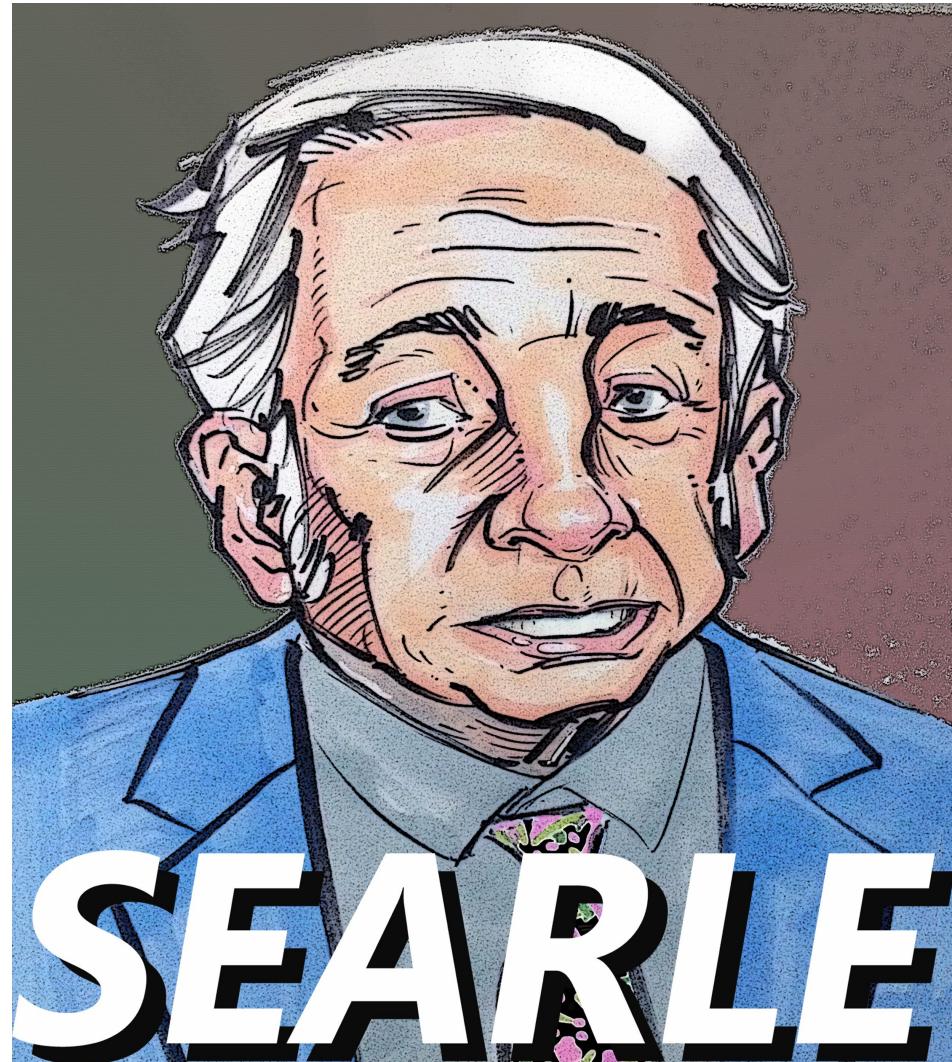
A human evaluator would judge text-based conversations between a human and a machine designed to generate human-like responses. If the evaluator cannot reliably tell the machine from the human, the machine passes the test. The test results do not depend on the ability to give correct answers to questions, only how closely one's answers resemble those a human would give.

Alan Turing (1912-1954)



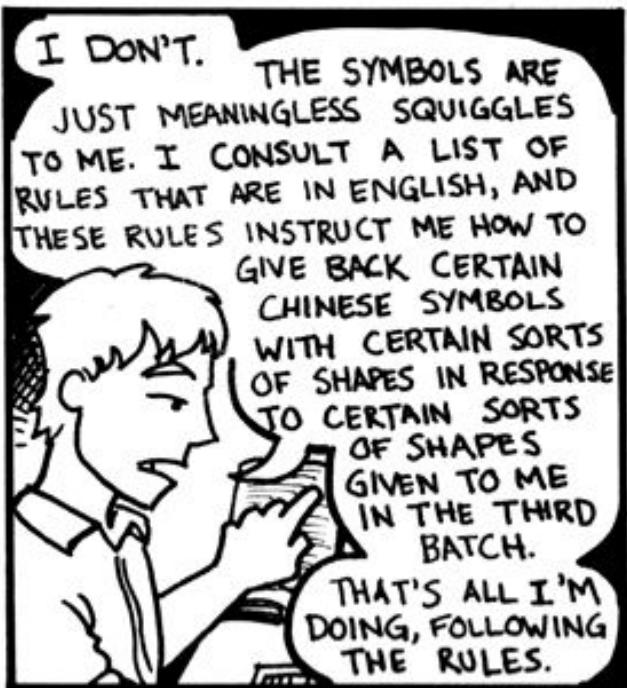
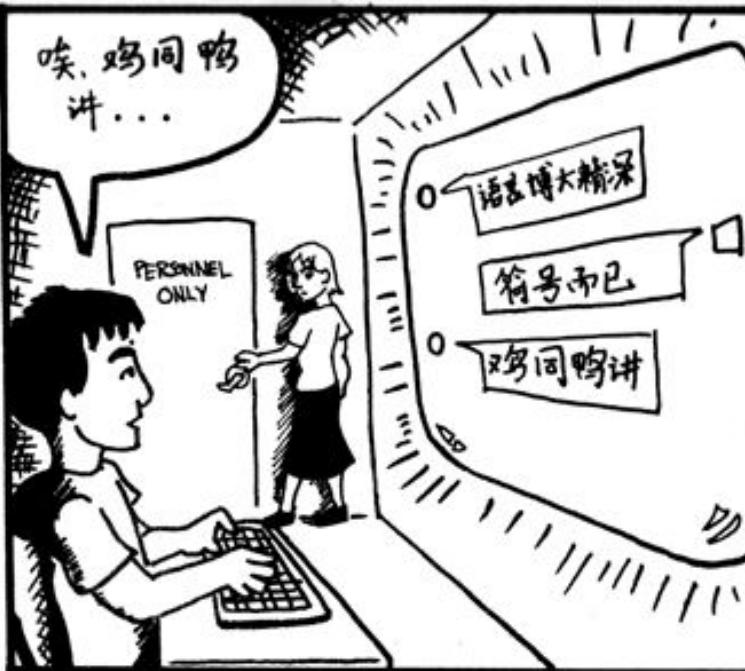
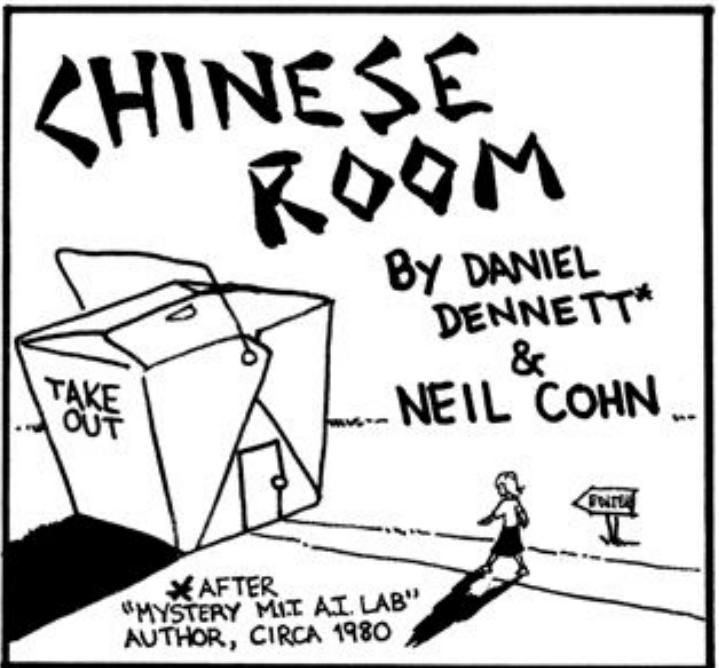
Turing was prosecuted in 1952 for being homosexual. He received chemical castration as an alternative to prison. Turing killed himself in 1954. It wasn't until the 2000s that Britain finally realized its gross injustice and its complicity in killing someone who should have been treated as a national hero.

John Searle (1932-) The Chinese Room

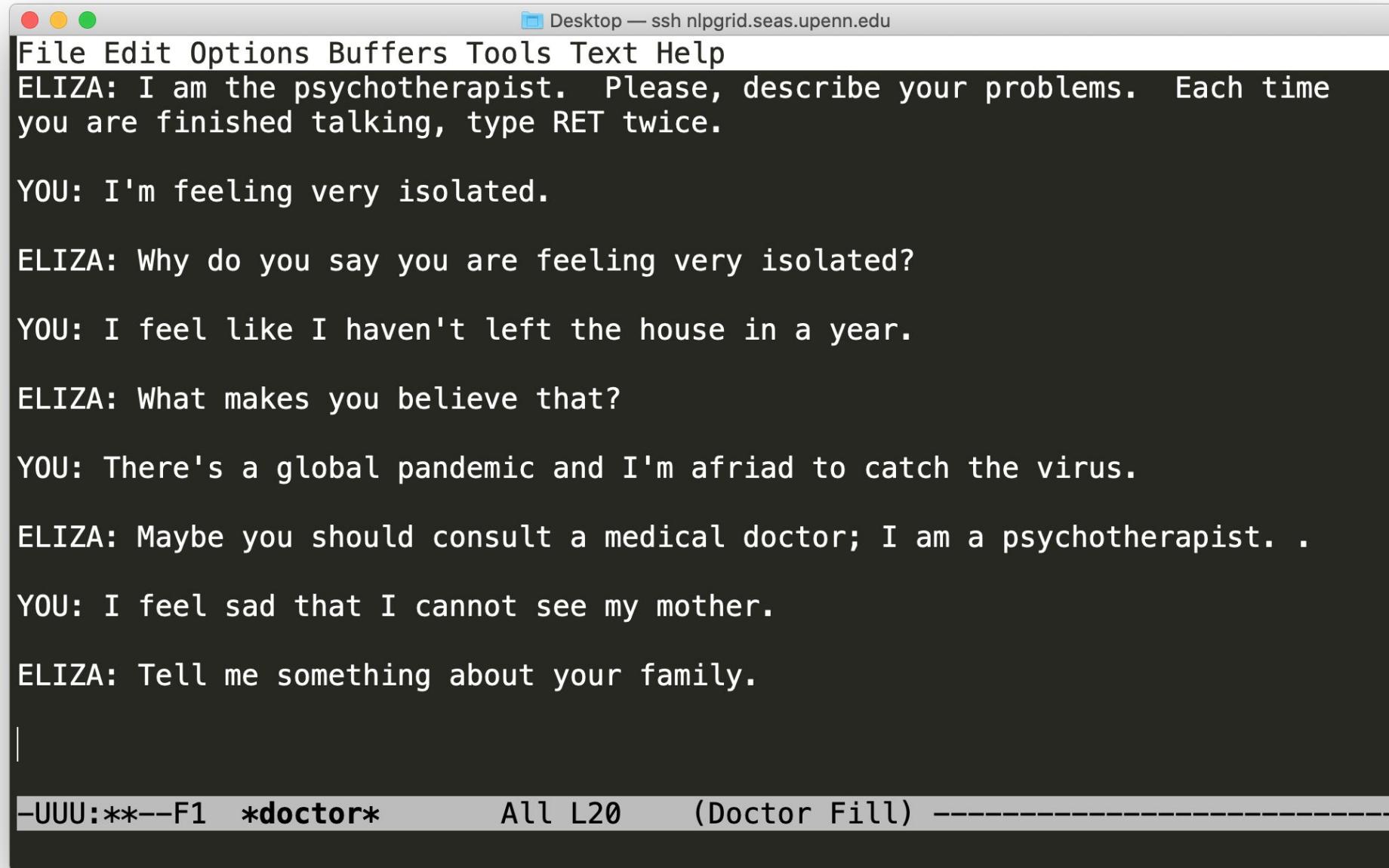


**That's not thinking.
That's just symbol
manipulation.**

Is the Turing Test a good test of whether machines possess intelligence? No, because the appearance of being a good conversational participant is achievable through simple symbol manipulation. Searle contends that the application of rules to input symbols is not true intelligence.



ELIZA – early NLP computer program



Desktop — ssh nlpgrid.seas.upenn.edu

File Edit Options Buffers Tools Text Help

ELIZA: I am the psychotherapist. Please, describe your problems. Each time you are finished talking, type RET twice.

YOU: I'm feeling very isolated.

ELIZA: Why do you say you are feeling very isolated?

YOU: I feel like I haven't left the house in a year.

ELIZA: What makes you believe that?

YOU: There's a global pandemic and I'm afraid to catch the virus.

ELIZA: Maybe you should consult a medical doctor; I am a psychotherapist. .

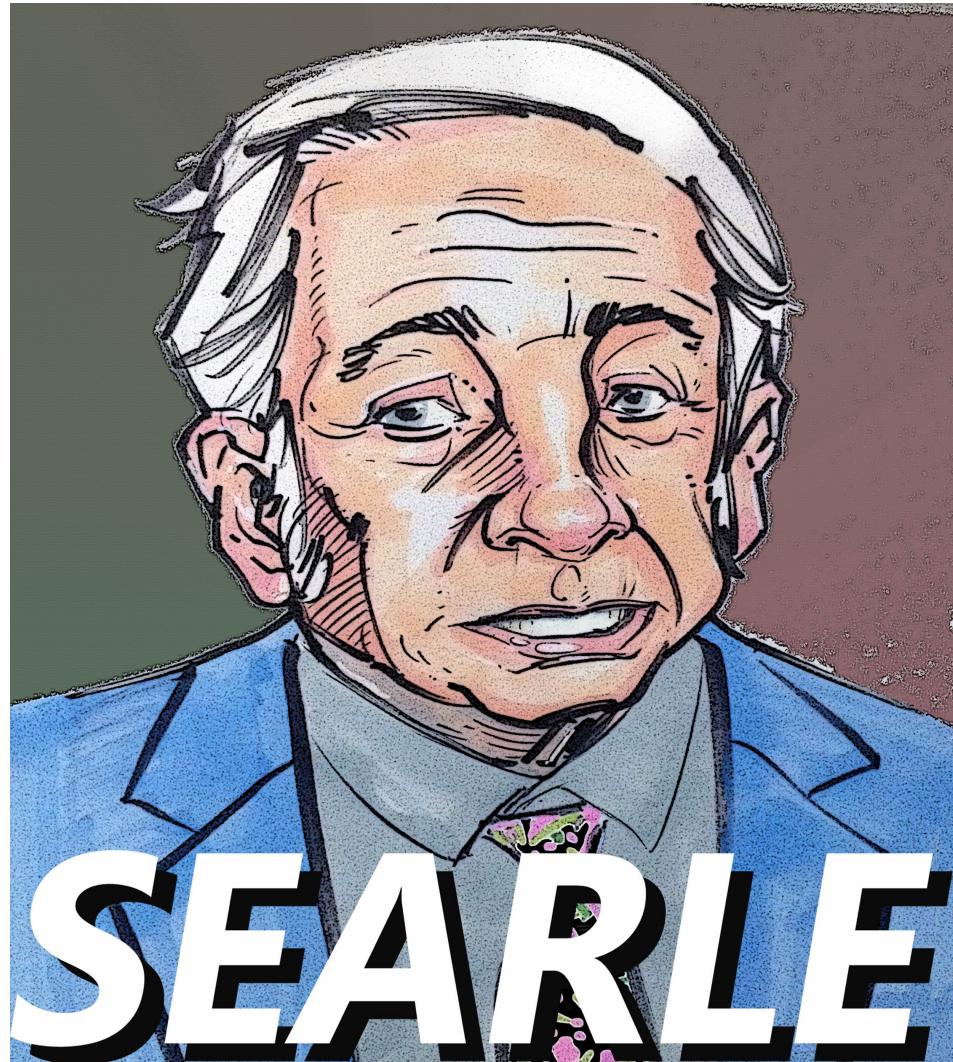
YOU: I feel sad that I cannot see my mother.

ELIZA: Tell me something about your family.

|

-UUU:**--F1 *doctor* All L20 (Doctor Fill) -----

John Searle (1932-) The Chinese Room

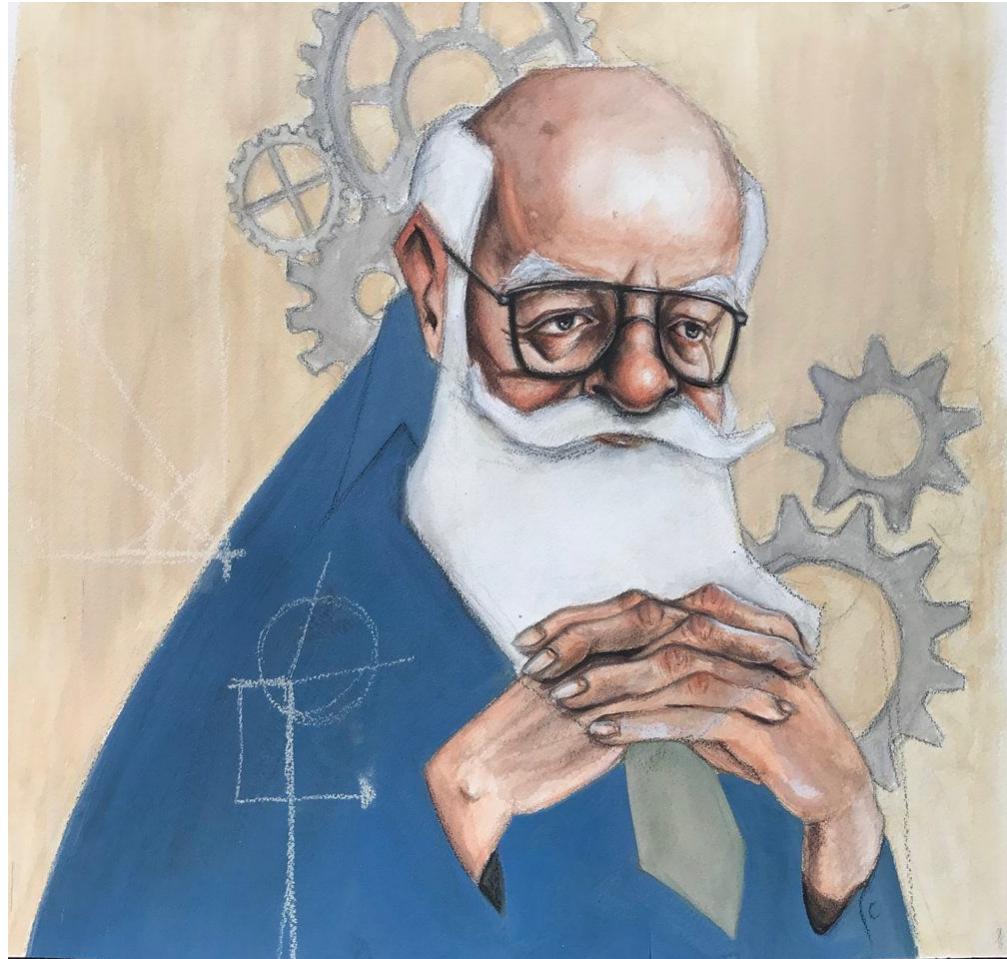


Brains cause minds, so computers can't have minds.

Searle contrasts strong AI with weak AI. In strong AI, the computer really is a mind in the sense that it can be literally said to understand and have other cognitive states. In weak AI, computers just simulate thought, their seeming understanding isn't real understanding.

He argues that (biological) brains cause minds.

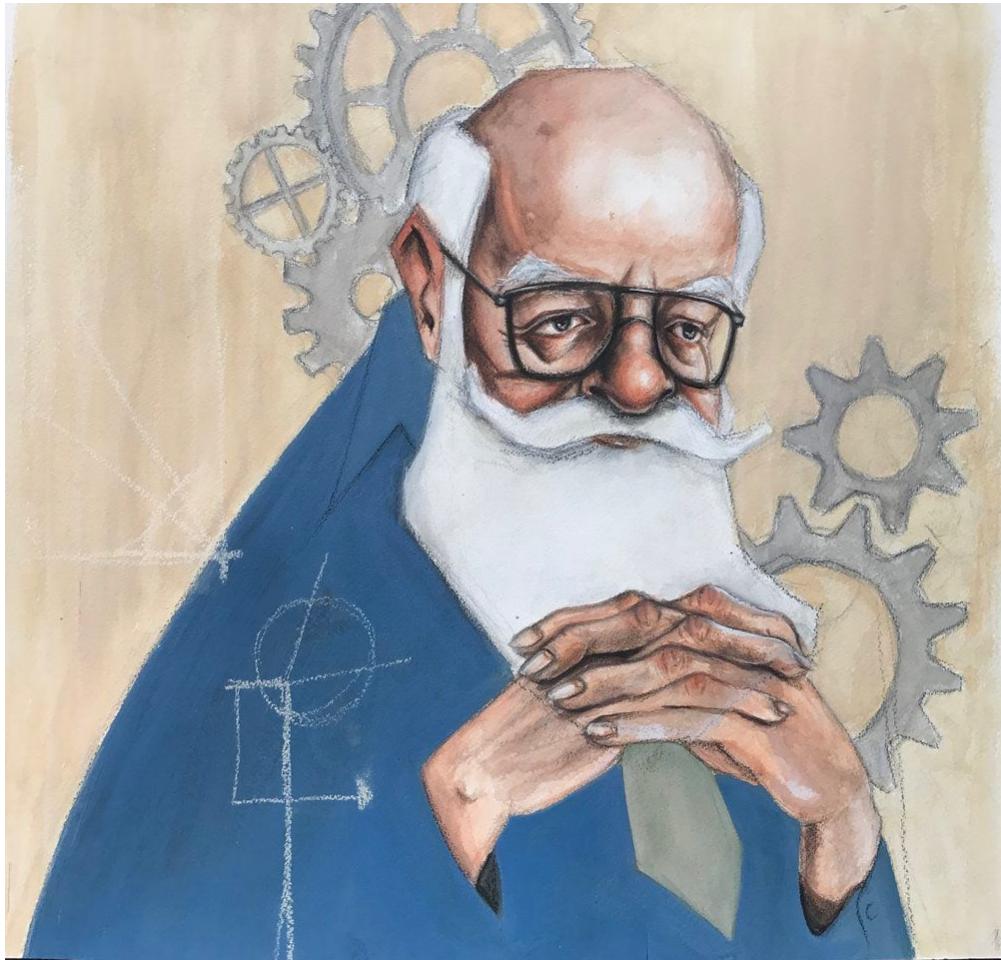
Daniel Dennett (1942-) Brain in a Vat



**Where am I if my brain is in
a vat instead of my body?**

Daniel Dennett wrote a short story called "Where Am I?" where he describes being recruited by the Pentagon to have his brain removed from his body and connected via radio links attaching his severed nerves. Body is sent on a secret mission to diffuse bomb without radiation harming brain.

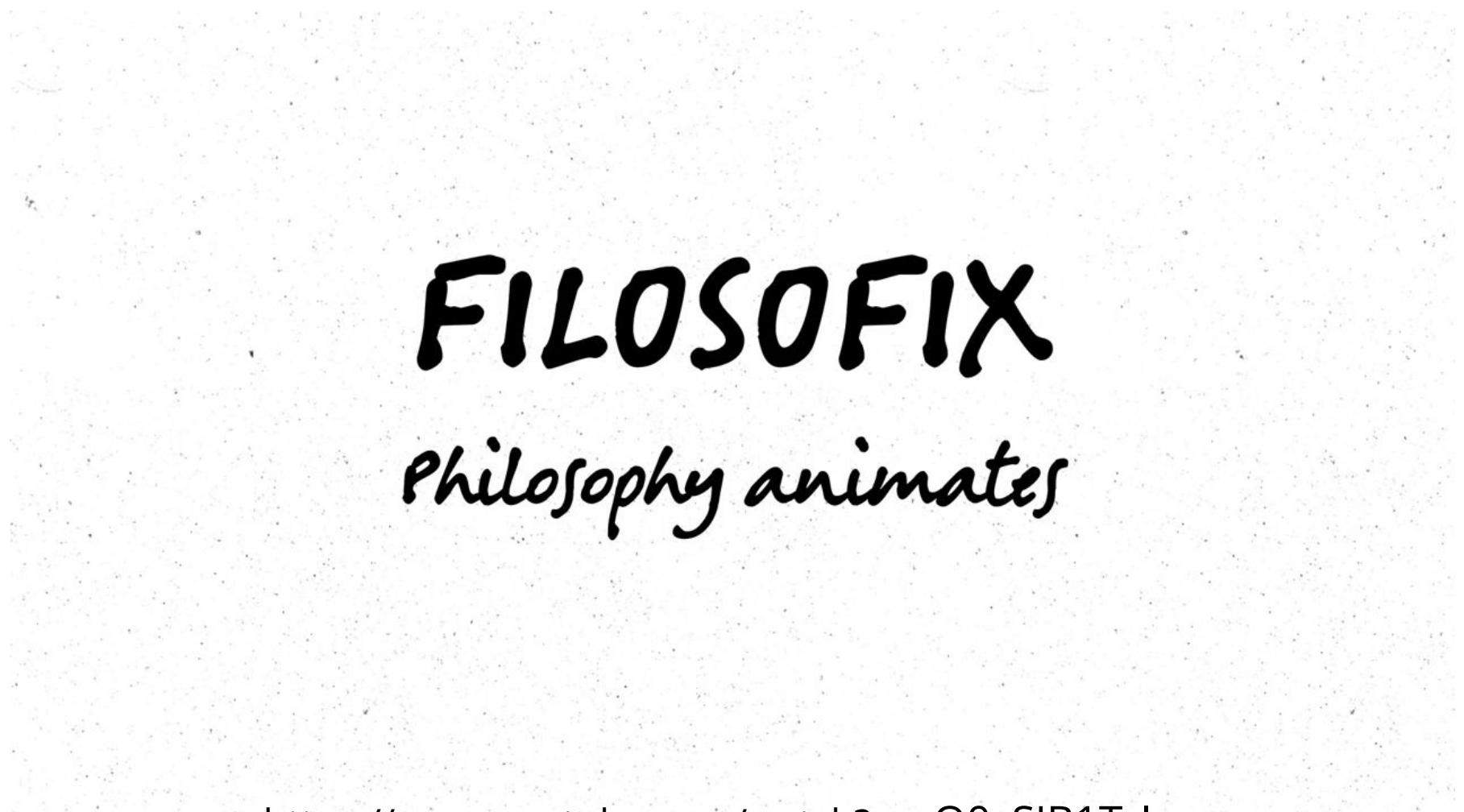
Daniel Dennett (1942-) Brain in a Vat



His body is destroyed by the radiation, but his consciousness continues in the vat. The scientists restore him in a new body. Then it is revealed that constructed a computer duplicate of my brain, reproducing both the complete information-processing structure and the computational speed of my brain in a giant computer program.

His brain in a vat is processing symbols. What makes this different than a computer processing symbols?

Brain in a vat



FILOSOFIX

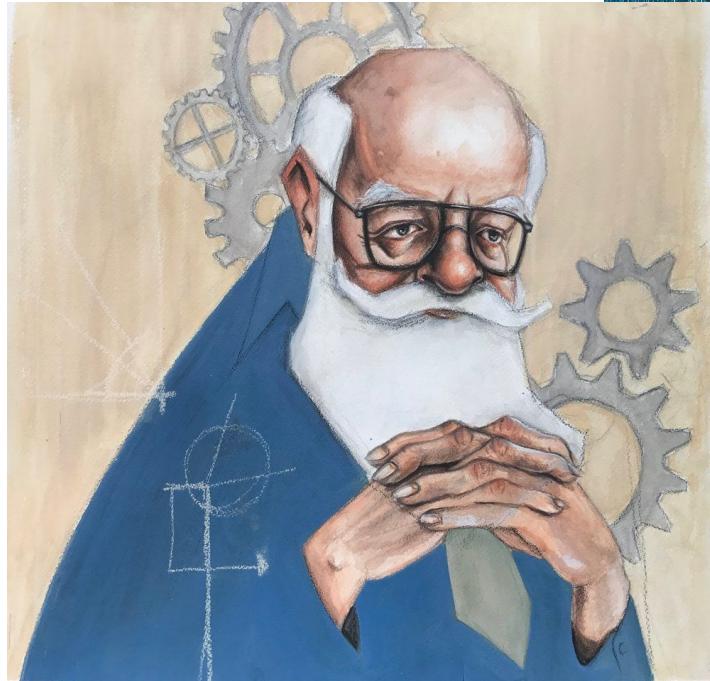
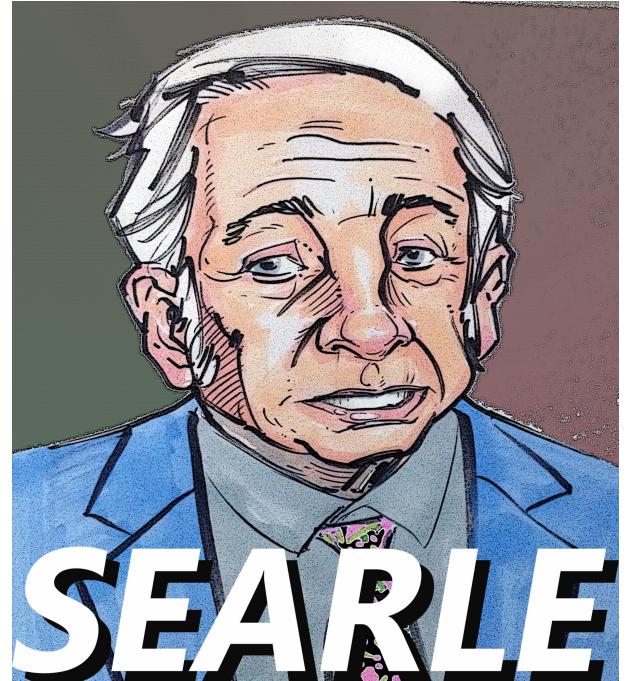
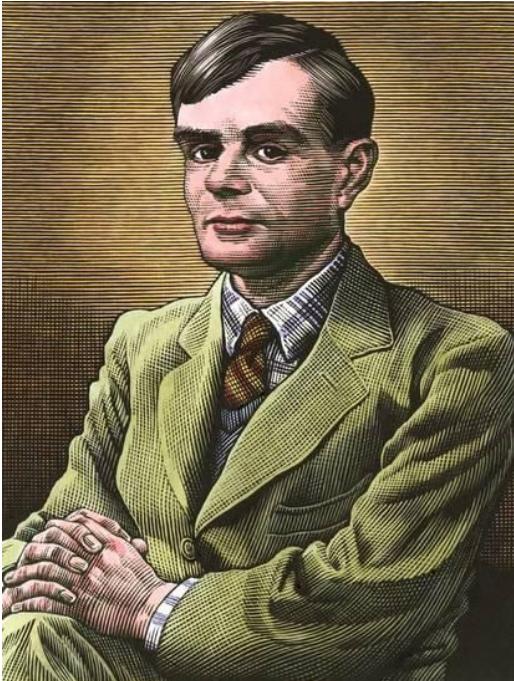
Philosophy animates

<https://www.youtube.com/watch?v=zO0sSJB1TrI>

Strong AI versus Weak AI

- *Actually thinking* versus merely *simulating thinking*
- Are you a brain in a vat?
- Would an AI program be equivalent?
- We will leave this to the philosophers and instead focus on practical AI programs that work.

But I will give you extra credit if you invent a sentient AI.



Machine Learning versus Artificial General Intelligence

Within the modern academic AI community, there's a focus on solving sub-problems that can be benchmarked on leaderboards.

There's a lack of attention to the question of whether it is possible to build systems that are truly intelligent, as we commonly understand intelligence.

If you're interested in philosophical questions about AI, I recommend listening to Lex Fridman's podcast.



