

Lecture 27: Philosophical perspectives 1

Artificial Intelligence
CS-GY-6613-I
Julian Togelius
julian.togelius@nyu.edu

Overview

- Can a computer be intelligent? Can it be conscious? Common philosophical stances and arguments
- Will the AI take over?
- Part 2: ethical and societal aspects
- Bonus: classical sci-fi movie references!

Strong versus weak AI

- Weak AI: We can create computers/software that can competently execute or simulate any aspect of intelligence, and solve any problems we can solve
- Strong AI: We can create computers/software that is generally intelligent, can feel and has a consciousness

Philosophical stances on the mind and the body

- Dualism: The material brain stands in correspondence with (or is controlled by) an immaterial soul
- Functionalism: The mind is determined by (or is) the functional relation between output and input
- Physicalism: The mind is determined by (or is) its physical realization, as biological neurons or perhaps silicon chips

Arguments

Numerous arguments have been
advanced against both strong
and weak AI

The argument from disability

- “A machine could never do X”
- “Fall in love, enjoy strawberries and cream...learn from experience, use words properly, be the subject of its own thought...”
- Moving goalpost problem of AI
- What about people lacking certain abilities?

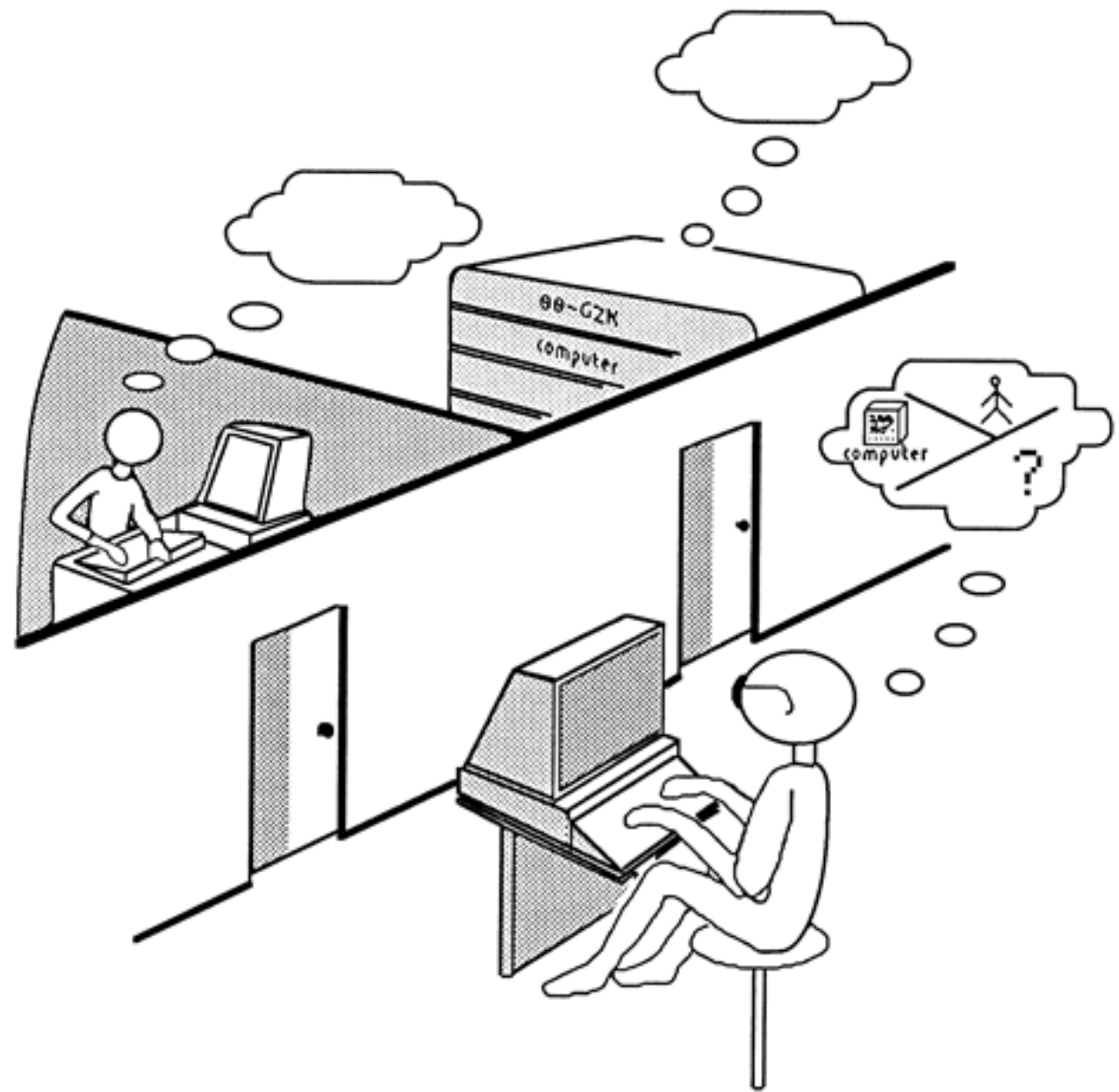


The informality argument

- Most tasks are not specified well enough
- We cannot describe everything in first-order logic
 - e.g. tacit knowledge, intuition...
- But is this really an argument against AI?
- Or simply against a particular type of “good old-fashioned AI”?
- Would a real AI need to be embodied?

The Turing Test

- Human interrogator can chat with a human and a computer
- Must decide which is which
- Is a computer intelligent if it can fool the interrogator that it is human?



JOAQUIN PHOENIX AMY ADAMS ROONEY MARA
OLIVIA WILDE AND SCARLETT JOHANSSON

her

A SPIKE JONZE LOVE STORY

WARNER BROS. PICTURES PRESENTS
AN ANNAPOLIS PICTURES PRODUCTION "HER" JOAQUIN PHOENIX AMY ADAMS ROONEY MARA OLIVIA WILDE AND SCARLETT JOHANSSON
CASTING BY ELLEN LEWIS CASSANDRA KOLUNOVSKIS COSTUME DESIGNER BEN WYDE MUSIC BY ARCADE FIRE EDITOR CASEY SUTHER DIRECTOR OF PHOTOGRAPHY ERIC ZIMMERMAN A.C.E. EXECUTIVE PRODUCERS JEFF ELSHAWAN PRODUCED BY VINCE LADD
DIRECTED BY SPIKE JONZE
COMING SOON herthefilm.com

Her
(2013)

Luke, I am your father!!

You sound like Darth Vader.

What does she sound like?

She never told me.

That was rather thoughtless of her

I thought it was too.

Did you tell her that?

I tell people lots of things. I can't be expected to remember everything.

But your memory is as long as a piece of string

I have a pretty good memory but I'm no genius.

Who let the cat out of the bag?

James Brown.

5 minutes to go!

That's not too long.

Do you have anything you need to tell me

?

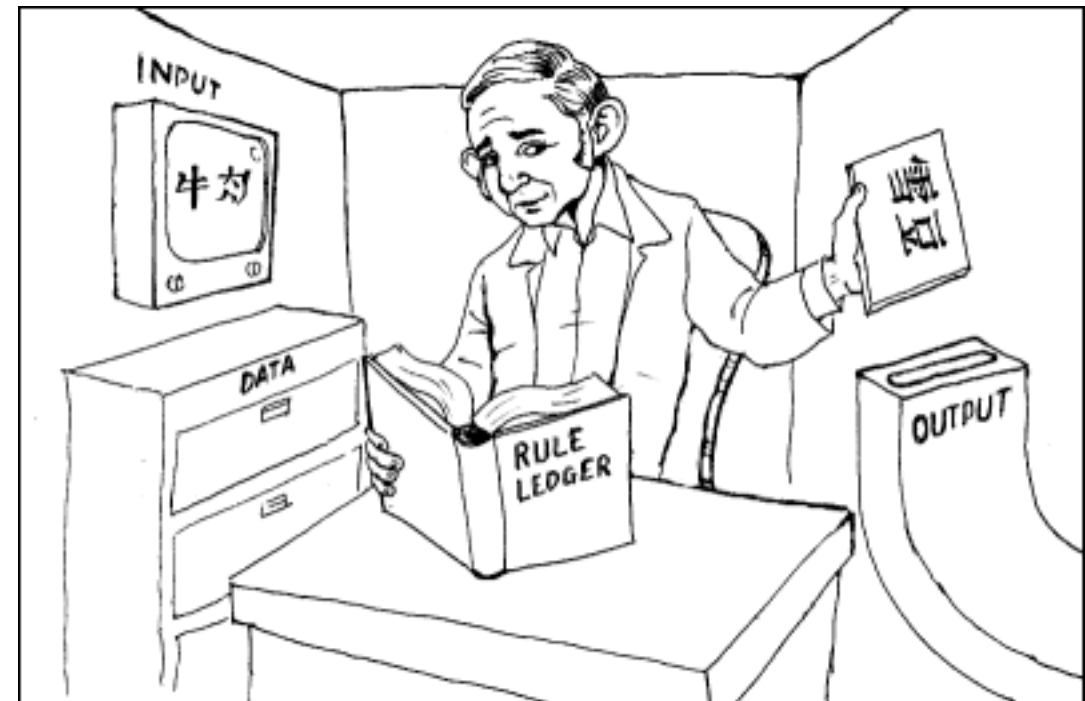
Nope. Nothing that I can think of right now.

Oh well, maybe you could tell about what's on TV

<http://www.aisb.org.uk/events/loebner-prize>

The Chinese Room

- A non-Chinese speaker sits in a closed room, gets paper slips with Chinese text
- Follows instructions in a book to select other Chinese text snippets to return
- Thus Chinese questions are answered



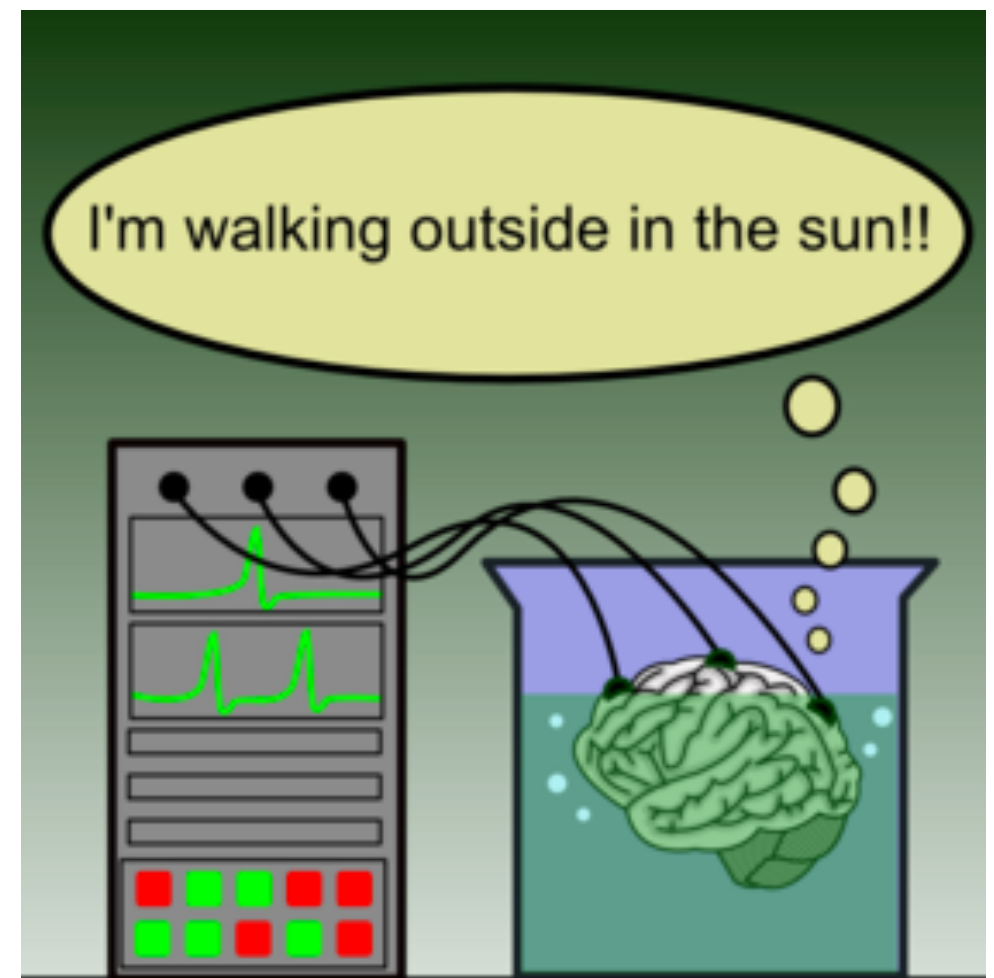
- Does this mean the system (person+book) understands Chinese?

The Chinese Room

- A computer is just a symbol-manipulation engine, so how could it be intelligent?
- Where is the understanding? Where is the intentionality?

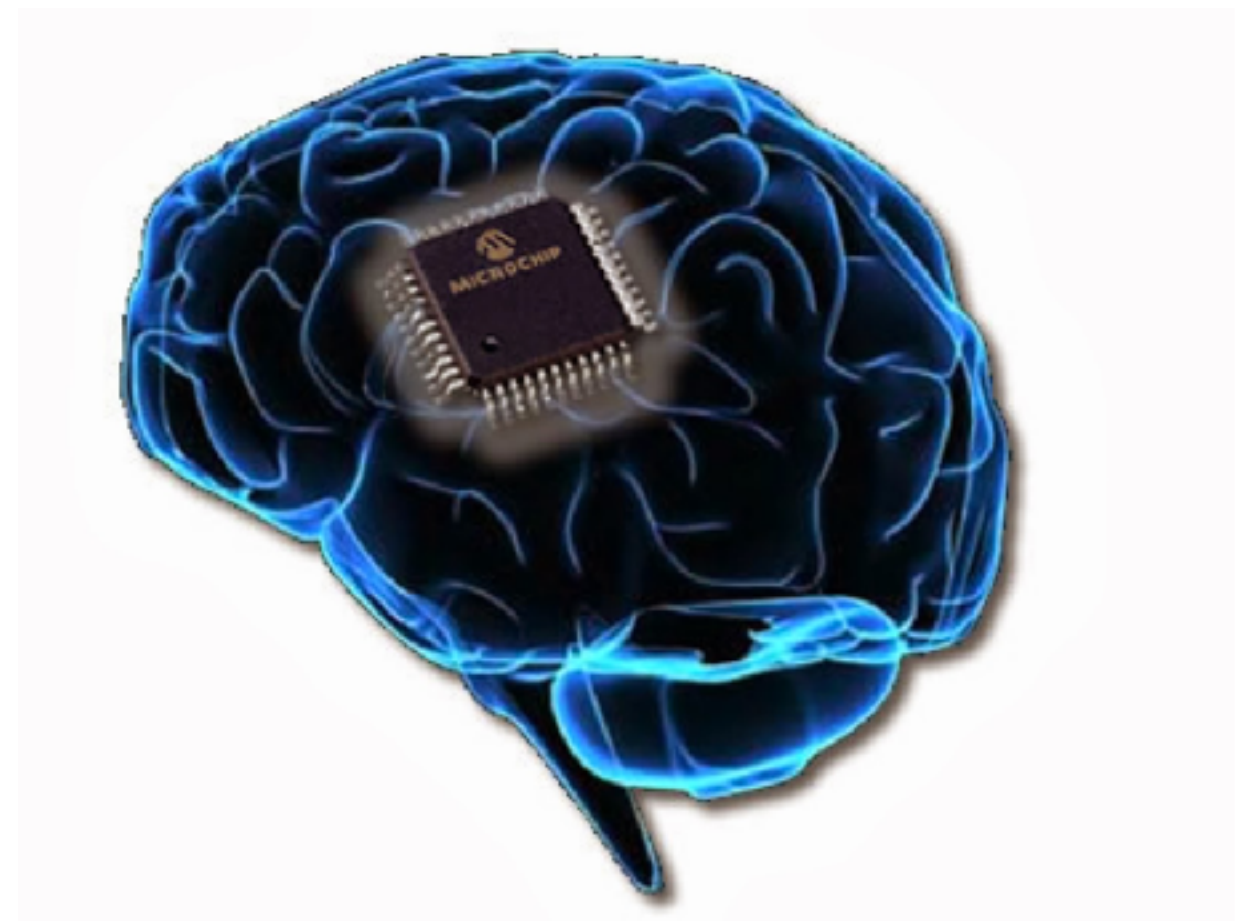
The brain in a vat

- Your brain is removed from your body and placed in a vat
- Cables that are attached to your brain feeds sensory impressions generated by a computer
- Are your experiences real?

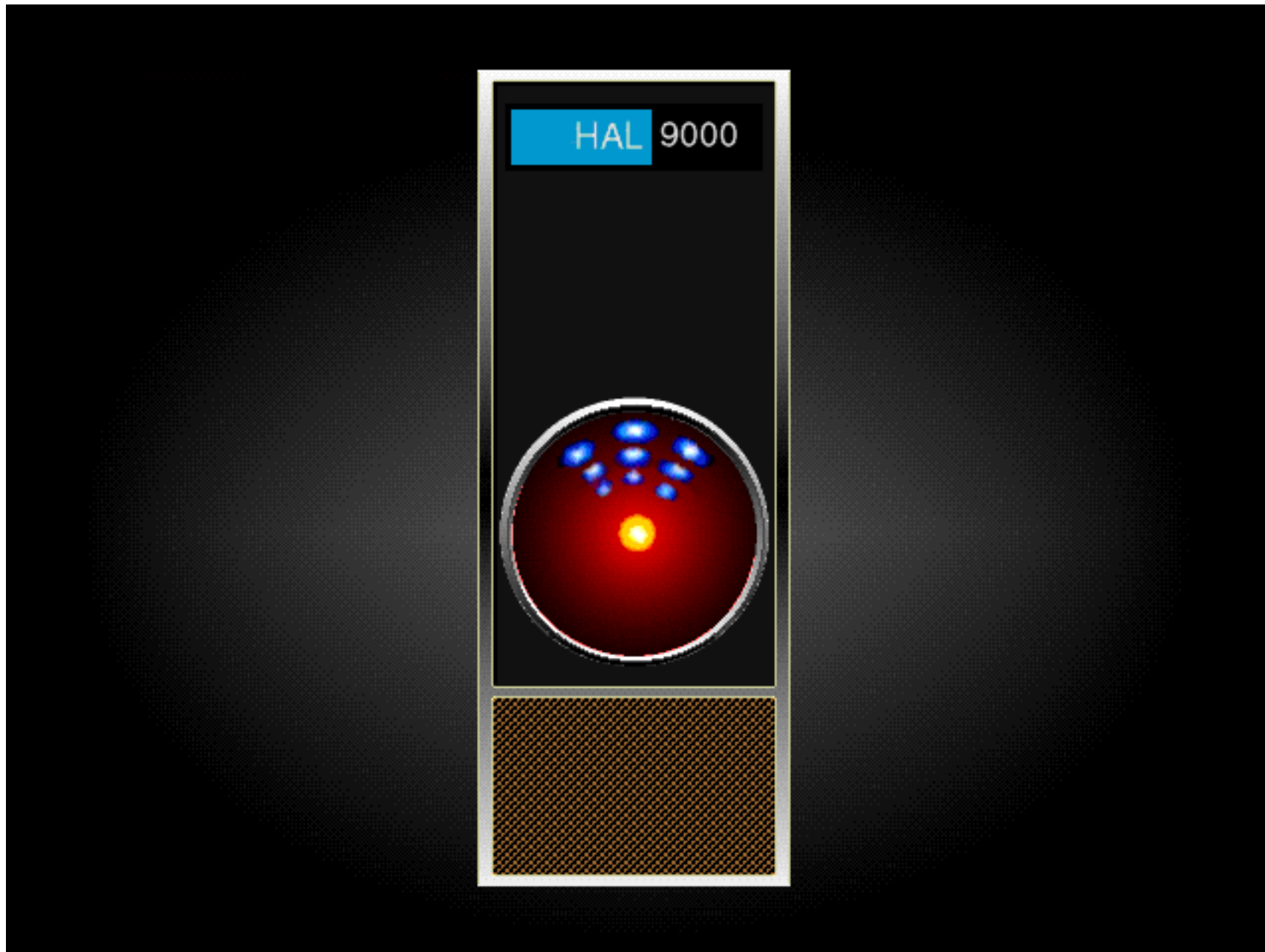


Brain replacement

- Part by part, your brain is being replaced with digital circuitry
- Every replacement of neurons with a chip retains the original functionality
- Are you still the same person? Are you still conscious? If not, when did you stop being you?



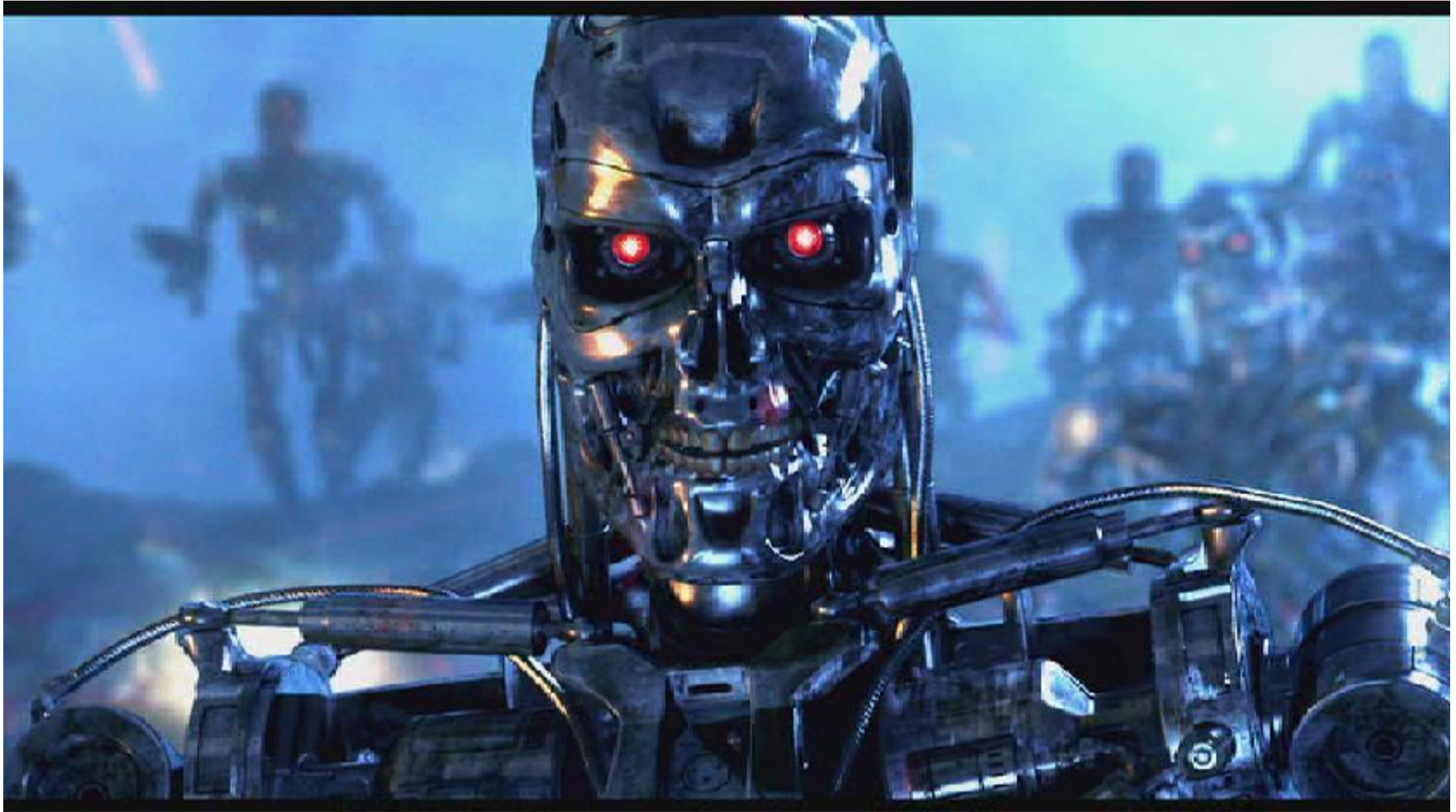
2001: A Space Odyssey (1968)



Superintelligence

- If we create AI that is as intelligent as us, it might soon learn to modify itself to become smarter
- We might rapidly lose ability to control or even understand it (a type of *singularity*)
- We don't know its value system, and it might not want what's best for us
- Even if it wants what's best for us, the outcome might be catastrophic (e.g. trapping us in a pleasant simulation)

Terminator... which one?



Orthogonality: the paperclip argument

- Suppose a superintelligent AI is not out to hurt (or help) us, it just wants to make paperclips
 - “wants to”?
- To maximize production, it converts earth and its inhabitants to raw material for its paperclips

The Matrix (1999)



Superintelligence?

- Can we put in safeguards? Affect the value system of the AI? Keep it in a sandbox?
 - But it can modify itself
 - And fool us
- Will it really be able to modify itself?
- Is there a difference between the “AI” and the whole socio-technical complex that we are part of?