How To Build A Brain

Philosophy and Ethics

Narrow AI

Self-driving cars

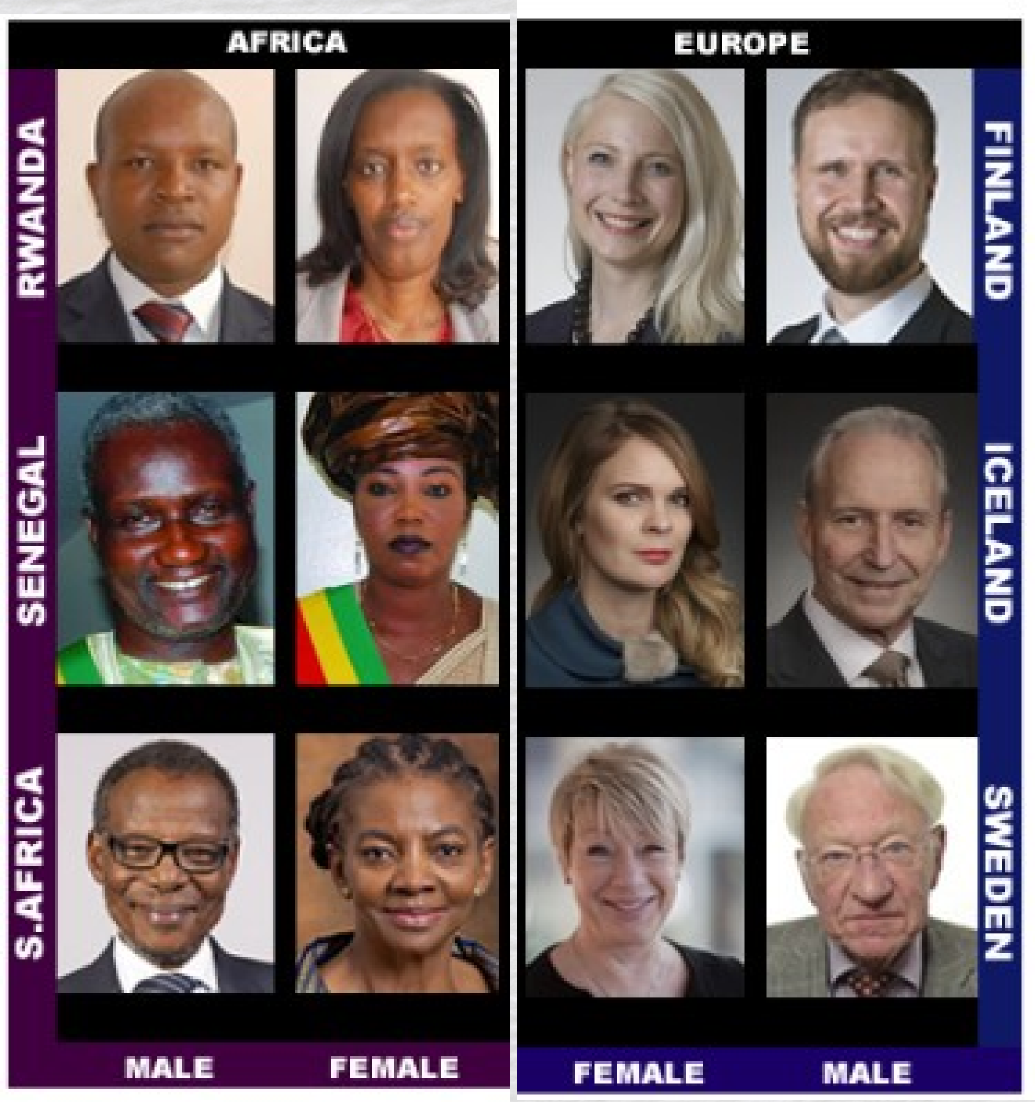
- Level of trust

- Who is to blame for accidents?

Algorithmic marking

YOUR ALGORITHM DOESN'T KNOW ME
YOUR ALGORITHM DOESN'T KNOW ME

- 2020 A-level grades
- Per-school distributions 2017-2019
- Effect on top students in deprived areas
- “Not AI”: but what’s the difference?



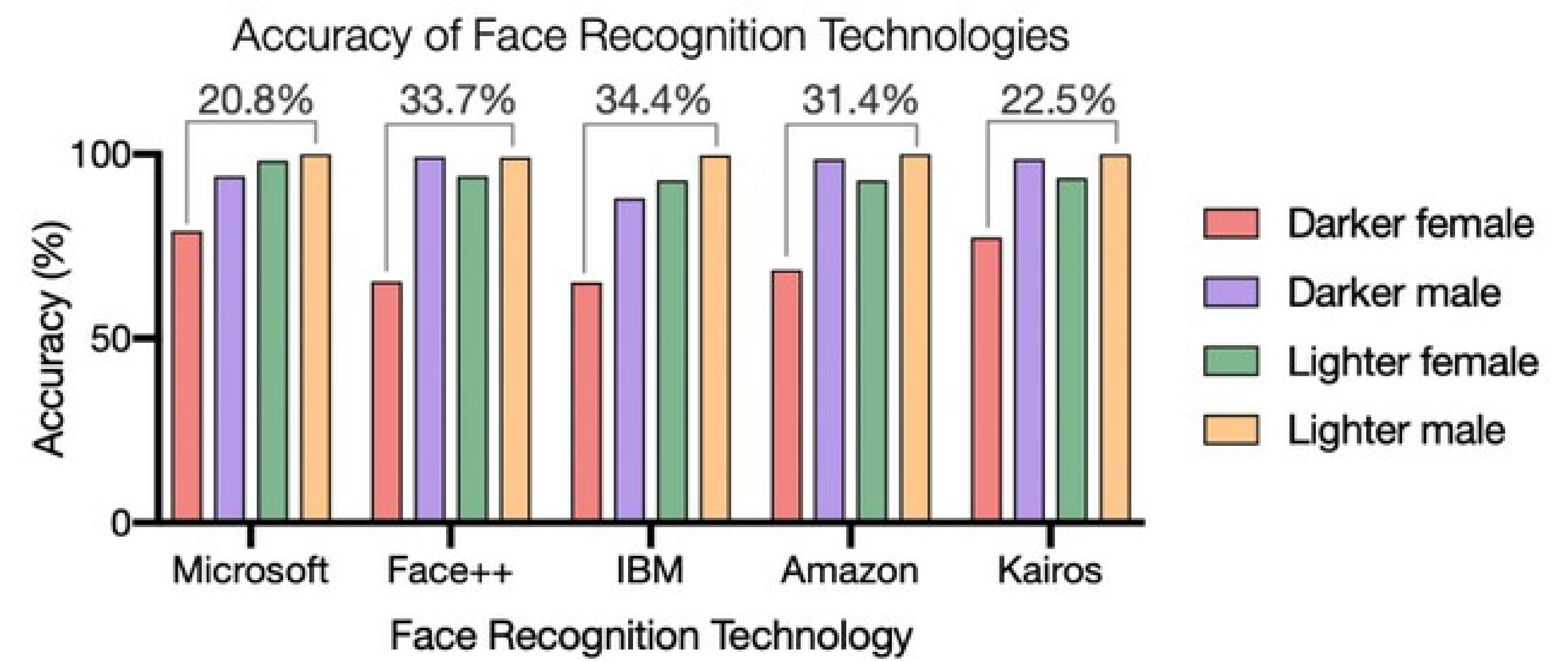
Face recognition

Automated face recognition systems now in use (e.g. border patrol; criminal investigations)

Brain-inspired narrow AI

How good are they at identifying people?

Face recognition



Semantics

word2vec, trained on 3 million words in Google News, can do this...

France is to Paris as Japan is to ____

Man is to king as woman is to ____.

Semantics

...it also does this...

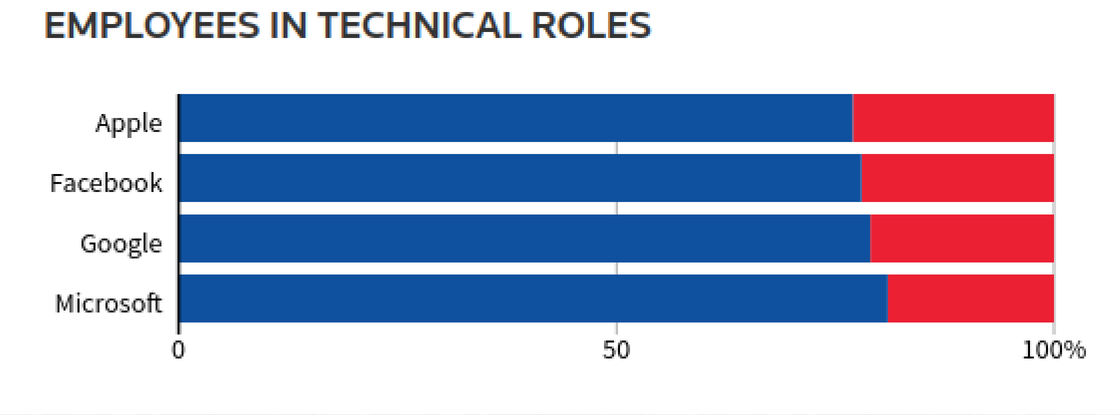
Man is to computer programmer as woman is to housewife.

Father is to doctor as mother is to nurse.

She is lovely, he is brilliant.

She is an interior designer, he is an architect

Job shortlisting algorithm at Amazon (2015)

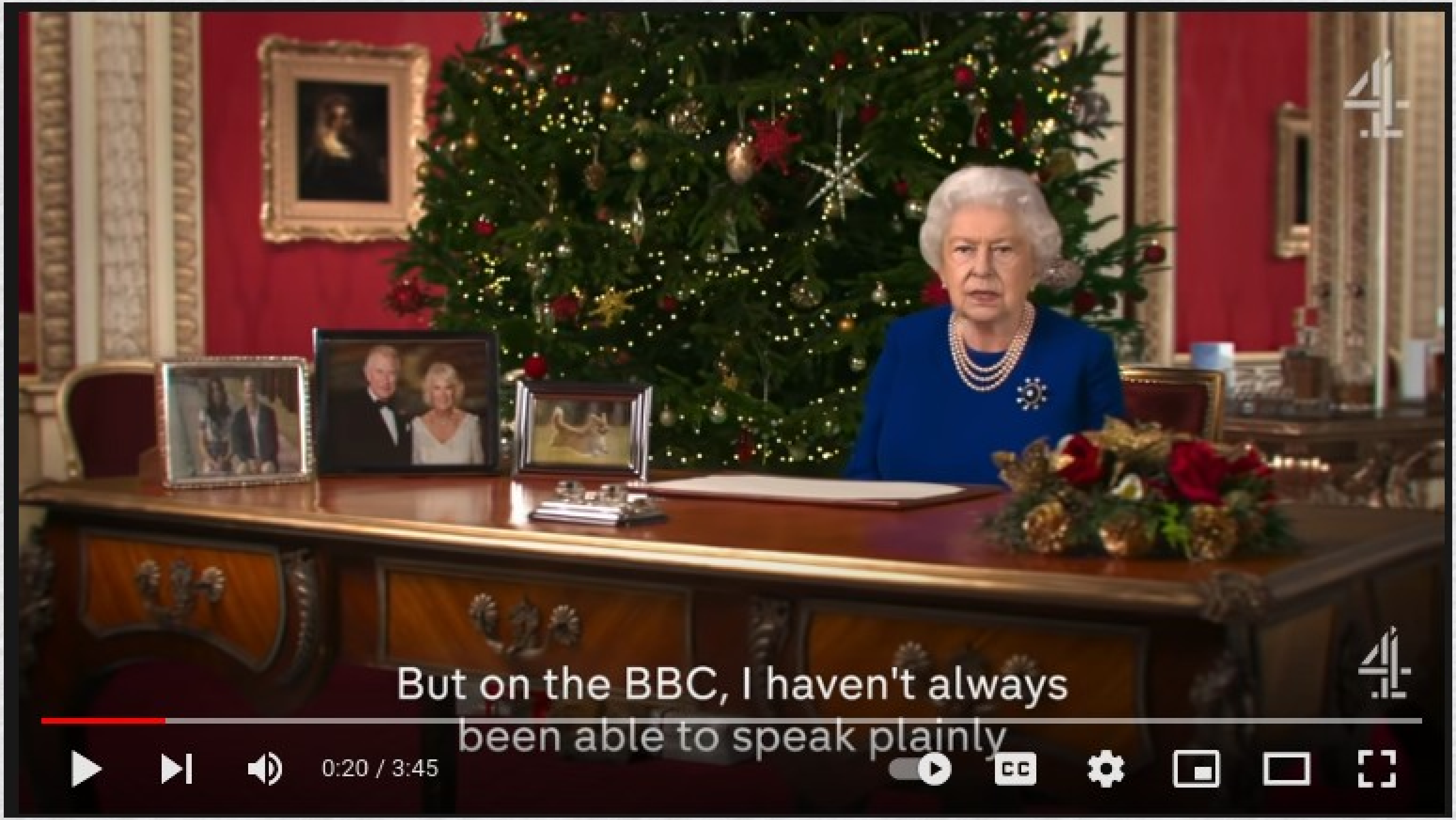


Computer programmer applicants marked down for:

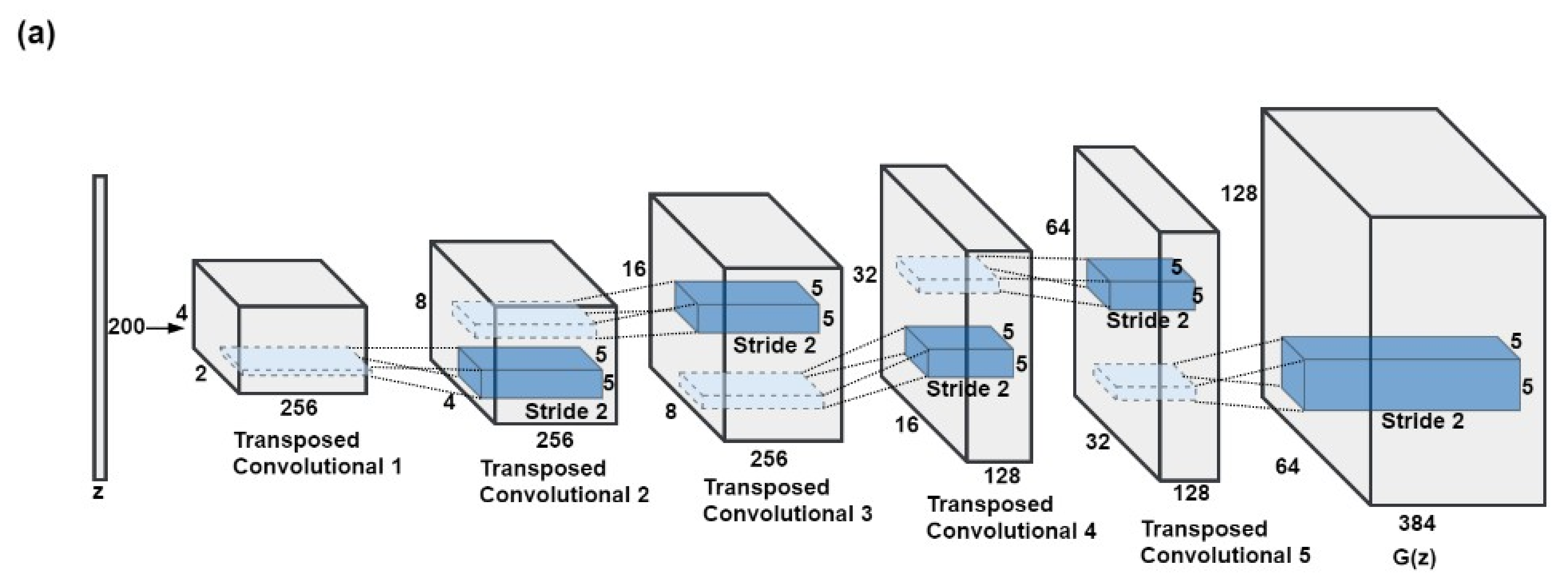
Women’s chess-club captain

Barnard College (an all-women’s university)

Deepfakes

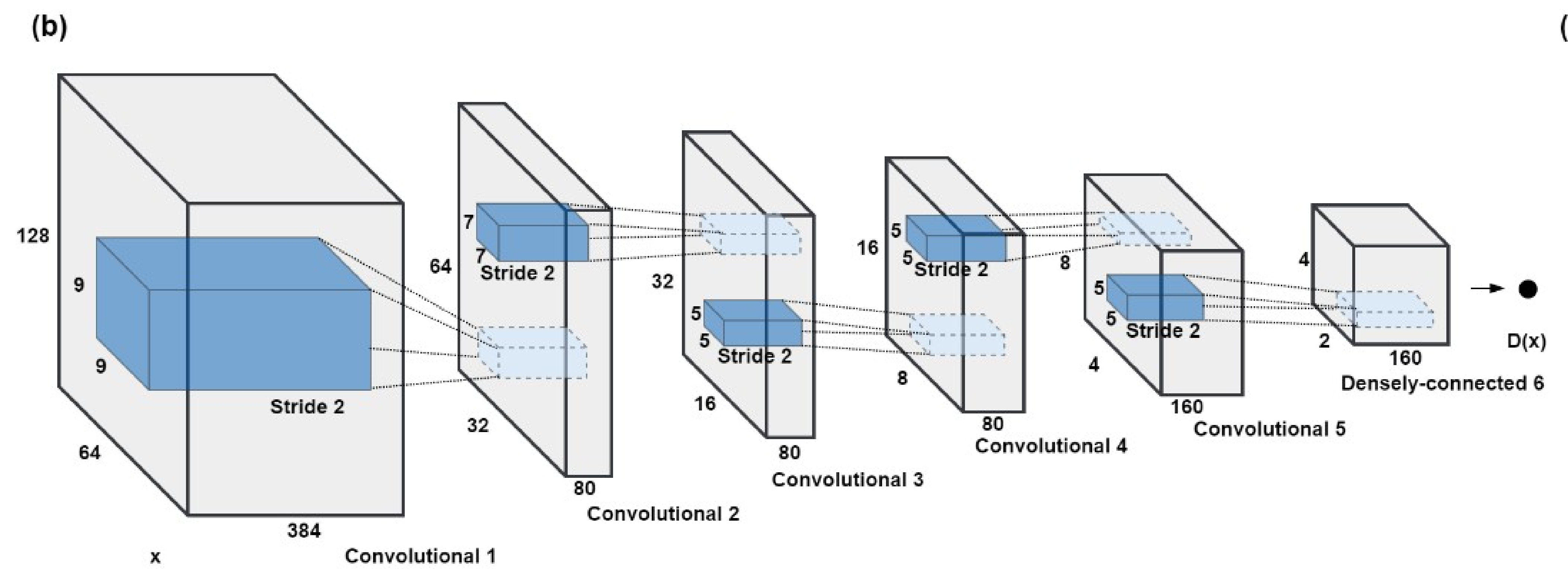

But on the BBC, I haven't always been able to speak plainly

Generative Adversarial Networks


(a)
z
200
4
2
256
Transposed Convolutional 1
8
4
256
5
5
Stride 2
Transposed Convolutional 2
16
8
256
5
5
Stride 2
Transposed Convolutional 3
32
16
128
5
5
Stride 2
Transposed Convolutional 4
64
32
128
5
5
Stride 2
Transposed Convolutional 5
128
64
384
5
5
Stride 2
G(z)

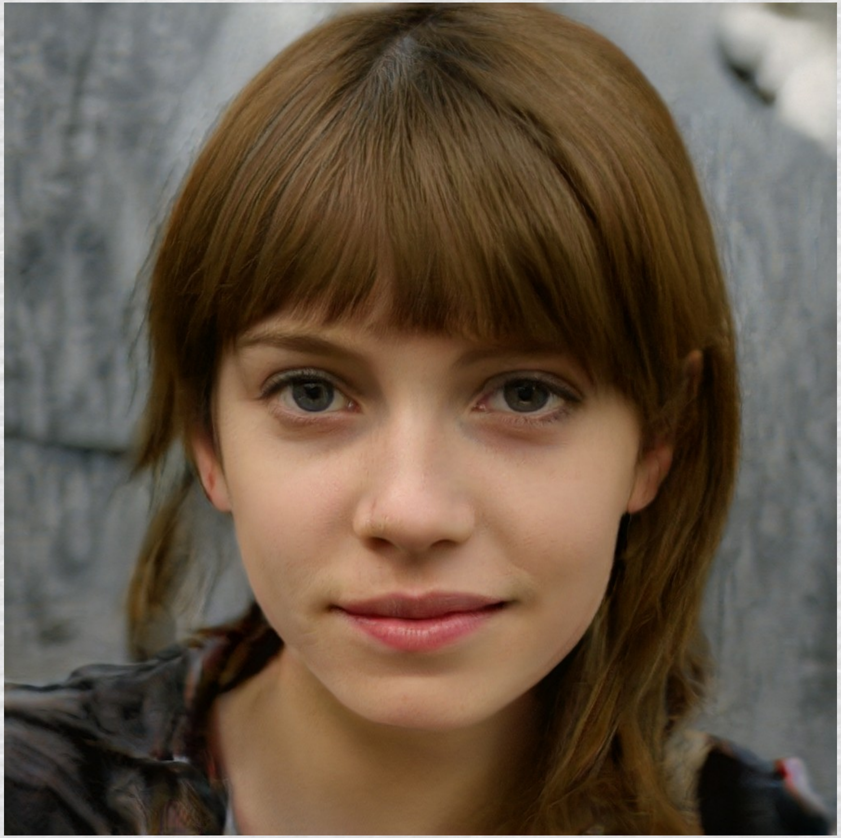
Generator

Generative Adversarial Networks



Detector

Not a real person

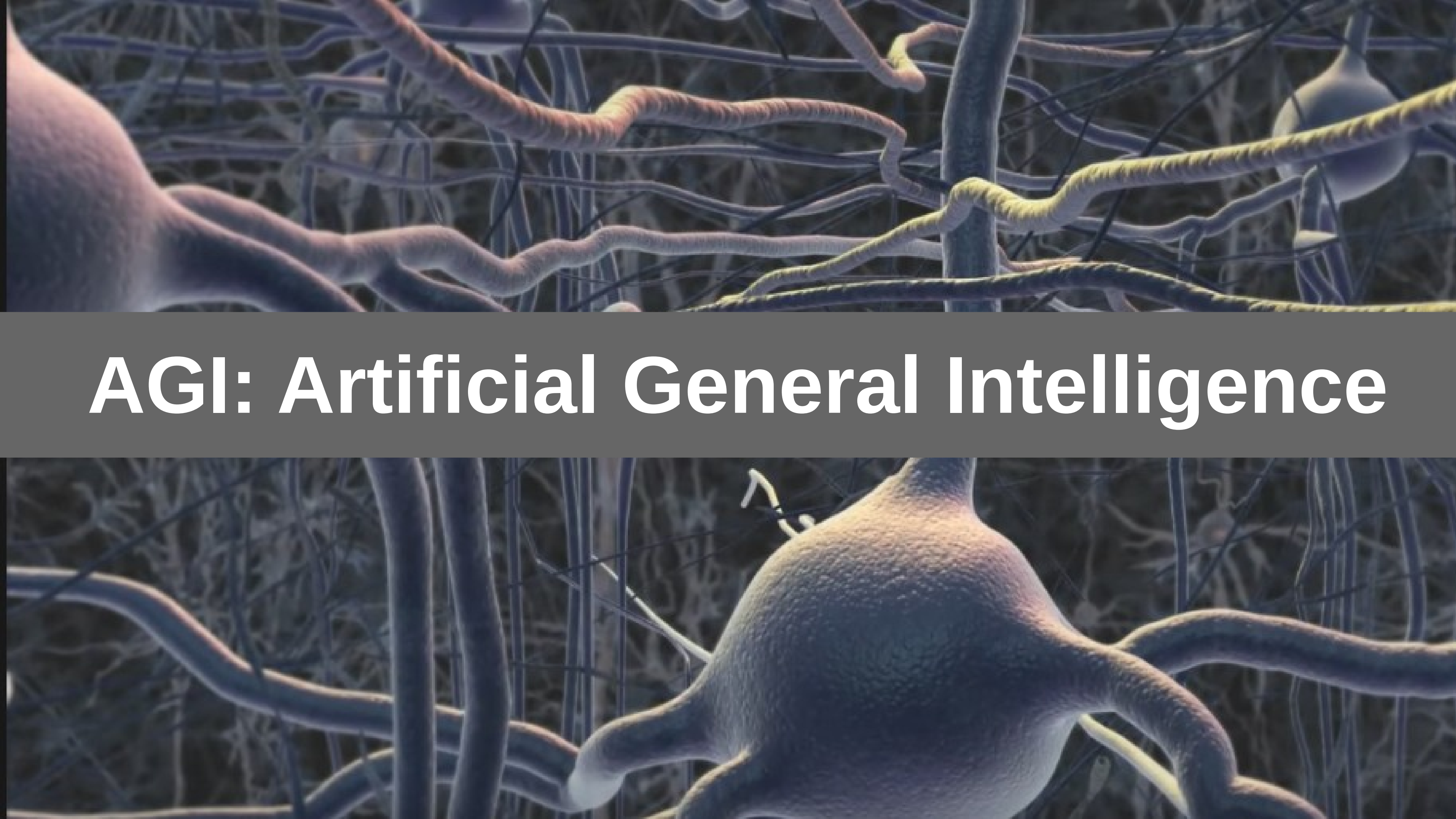


Generated by StyleGAN

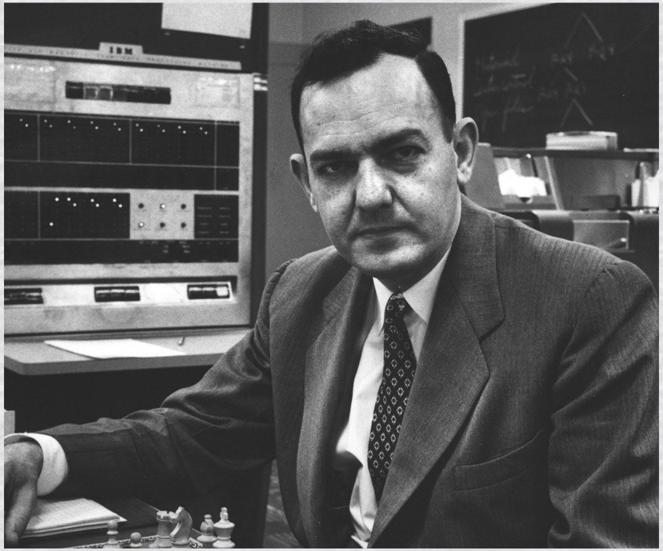
Art by AI

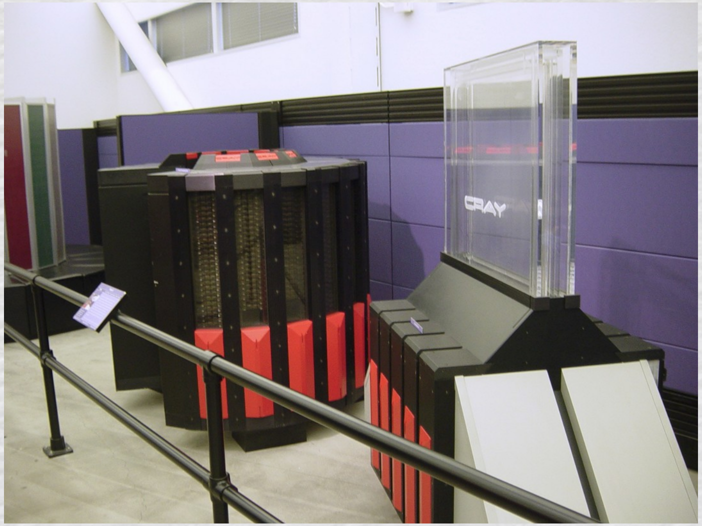
Generated by StyleGAN

AGI: Artificial General Intelligence



When AGI? Optimism



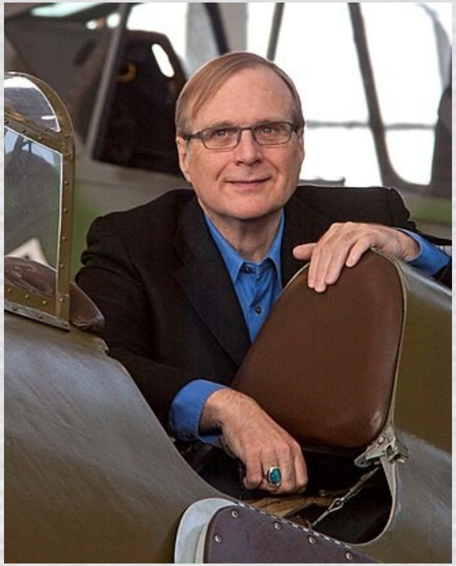


"machines will be capable, within twenty years, of doing any work a man can do." (Herb Simon, 1965)

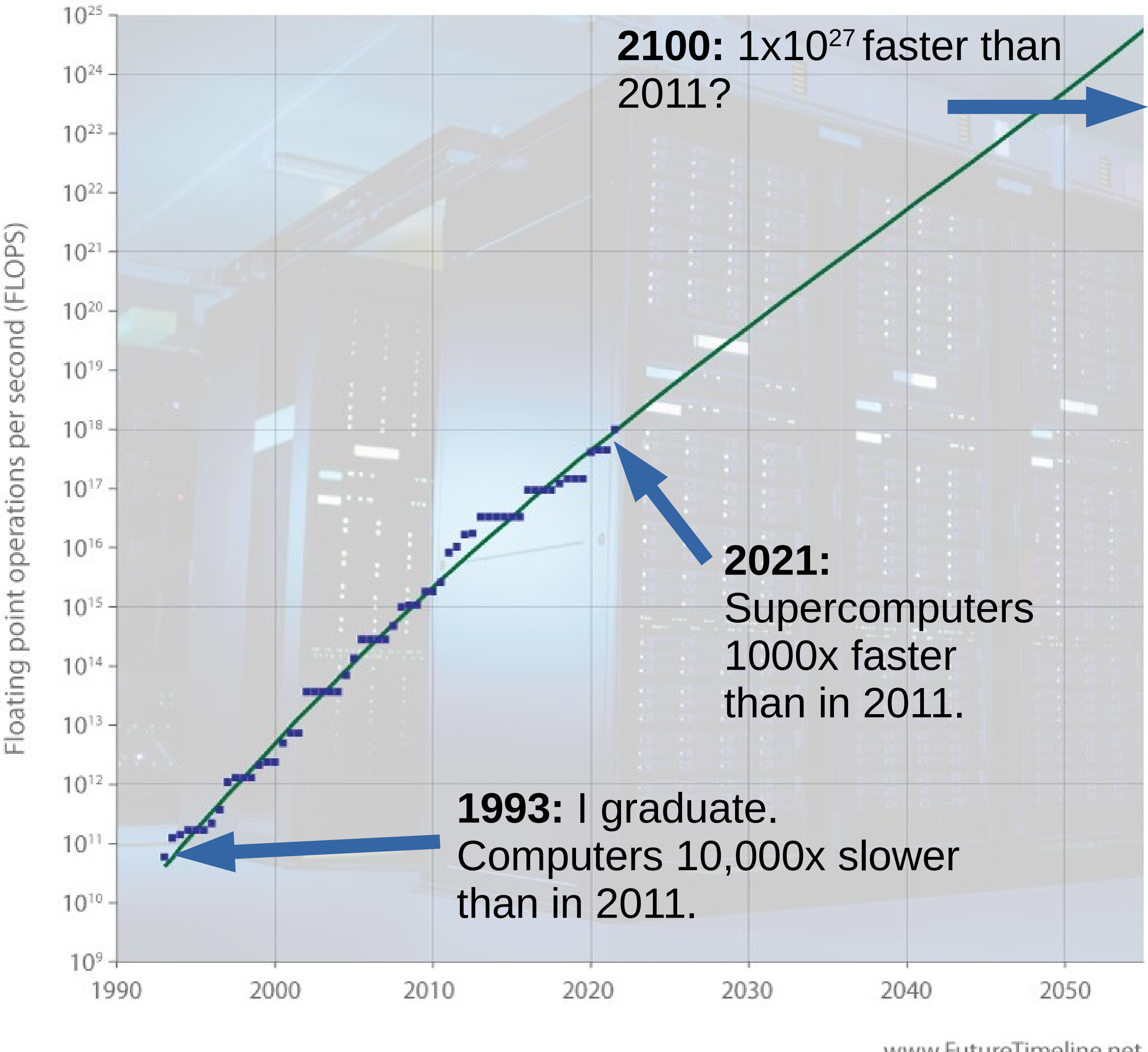
1985: Cray-2 is world’s fastest computer

2020: Xbox is 500 times faster than the Cray-2. Has not yet managed to get a job at McDonalds

A 21st century perspective



“AGI is not likely in the 21st century” (Paul Allen, 2011)



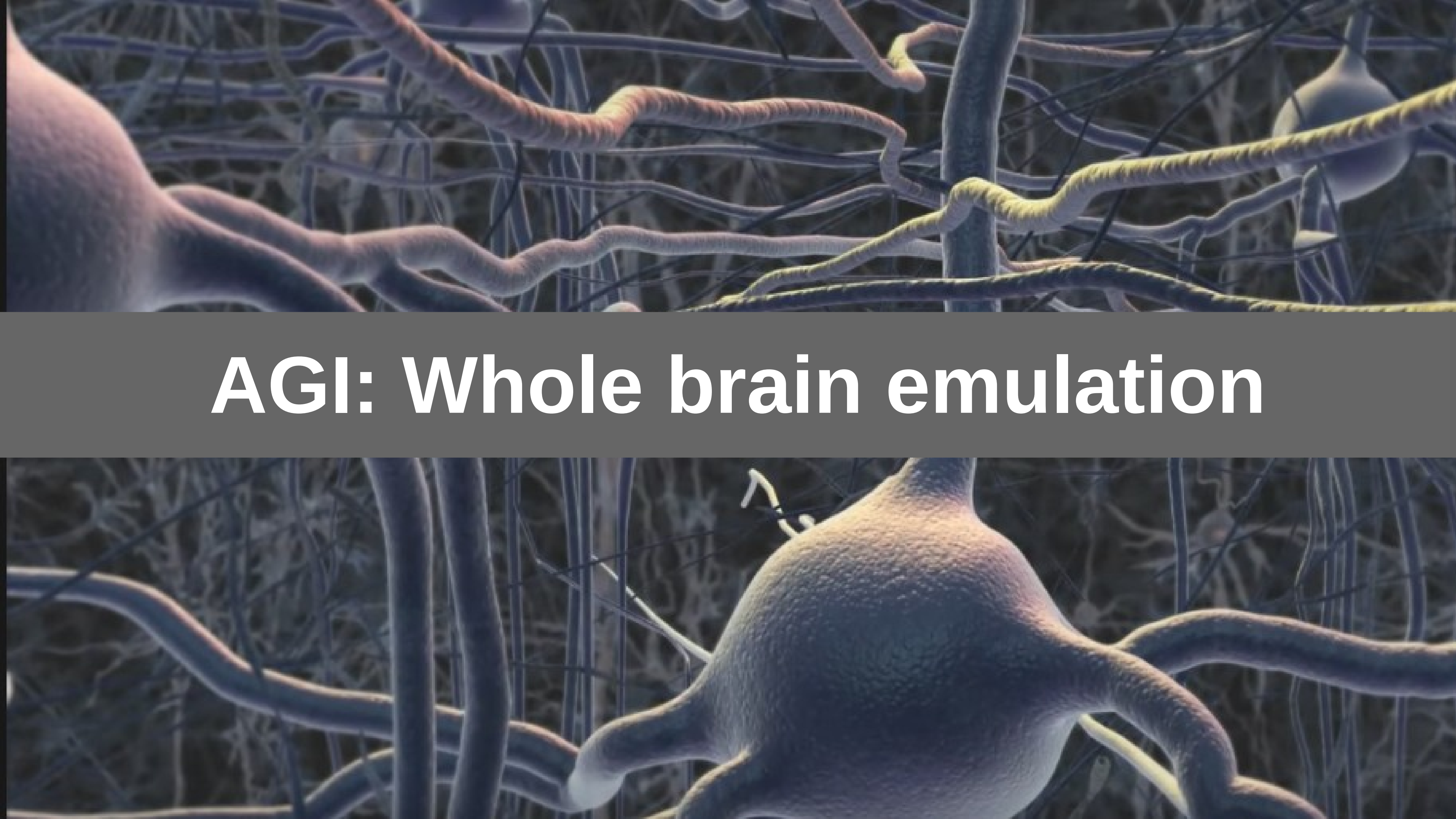
2100: 1x10^{27} faster than 2011?

2021: Supercomputers 1000x faster than in 2011.

1993: I graduate. Computers 10,000x slower than in 2011.

www.FutureTimeline.net

AGI: Whole brain emulation

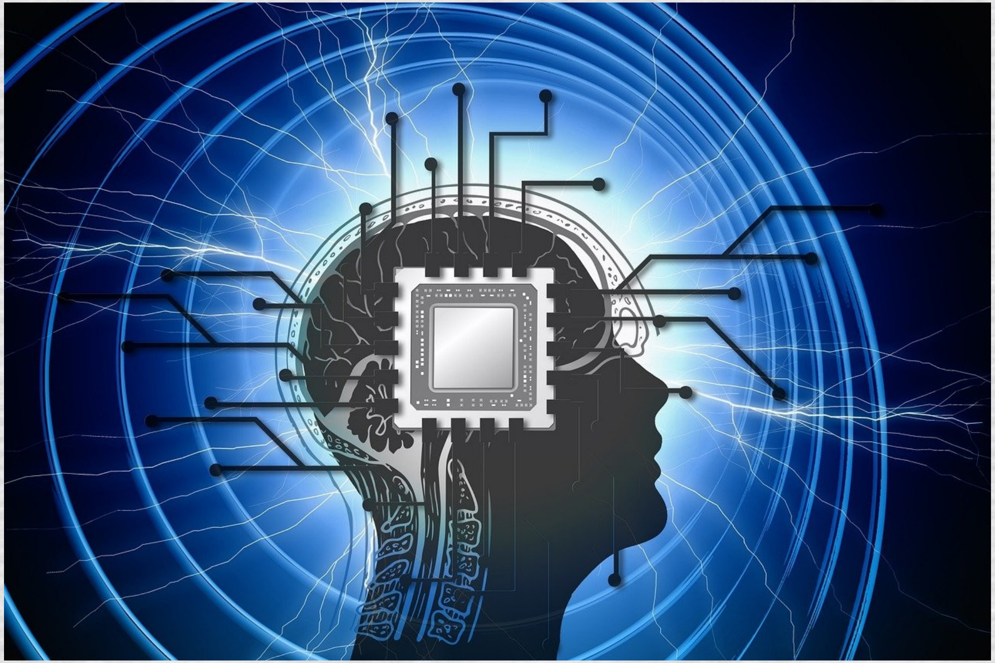


Harder than we thought

Ray Kurzweil

1997: Brain is 10 petaFLOPs

Ray Kurzweil



2006: Blue Brain

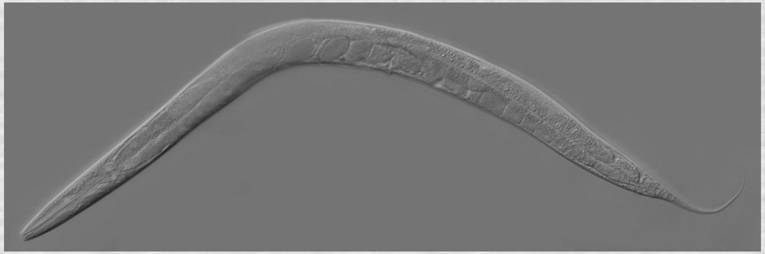
IBM Blue Gene

10,000 neuron simulation in real time
(85 billion neurons in human brain)

Henry Makram

"It is not impossible to build a human brain and we can do it in 10 years" (Makram, 2009)

OpenWorm

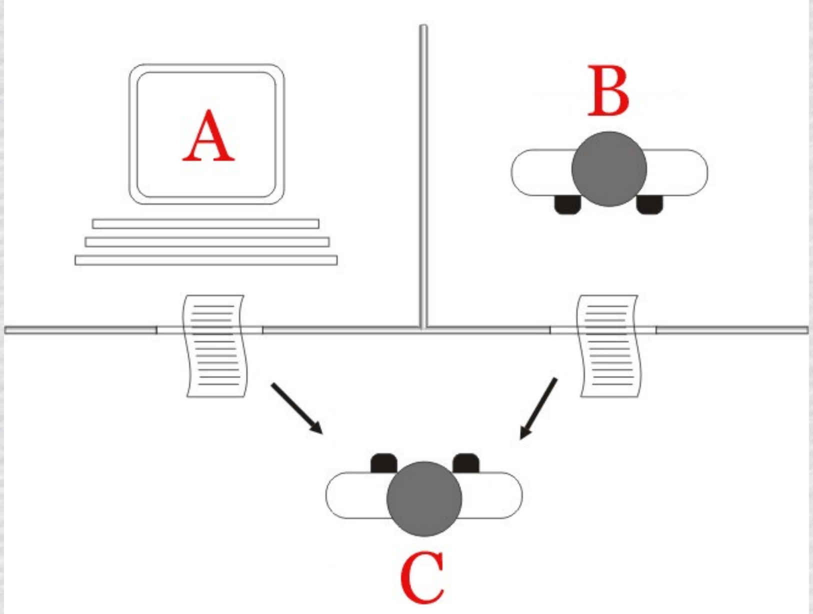


C. Elegans – A 1mm worm with 302 neurons.

OpenWorm is a project to do whole-brain emulation of this worm. It has not yet succeeded.

Detecting AGI

Turing test



Turing test

Alan Turing

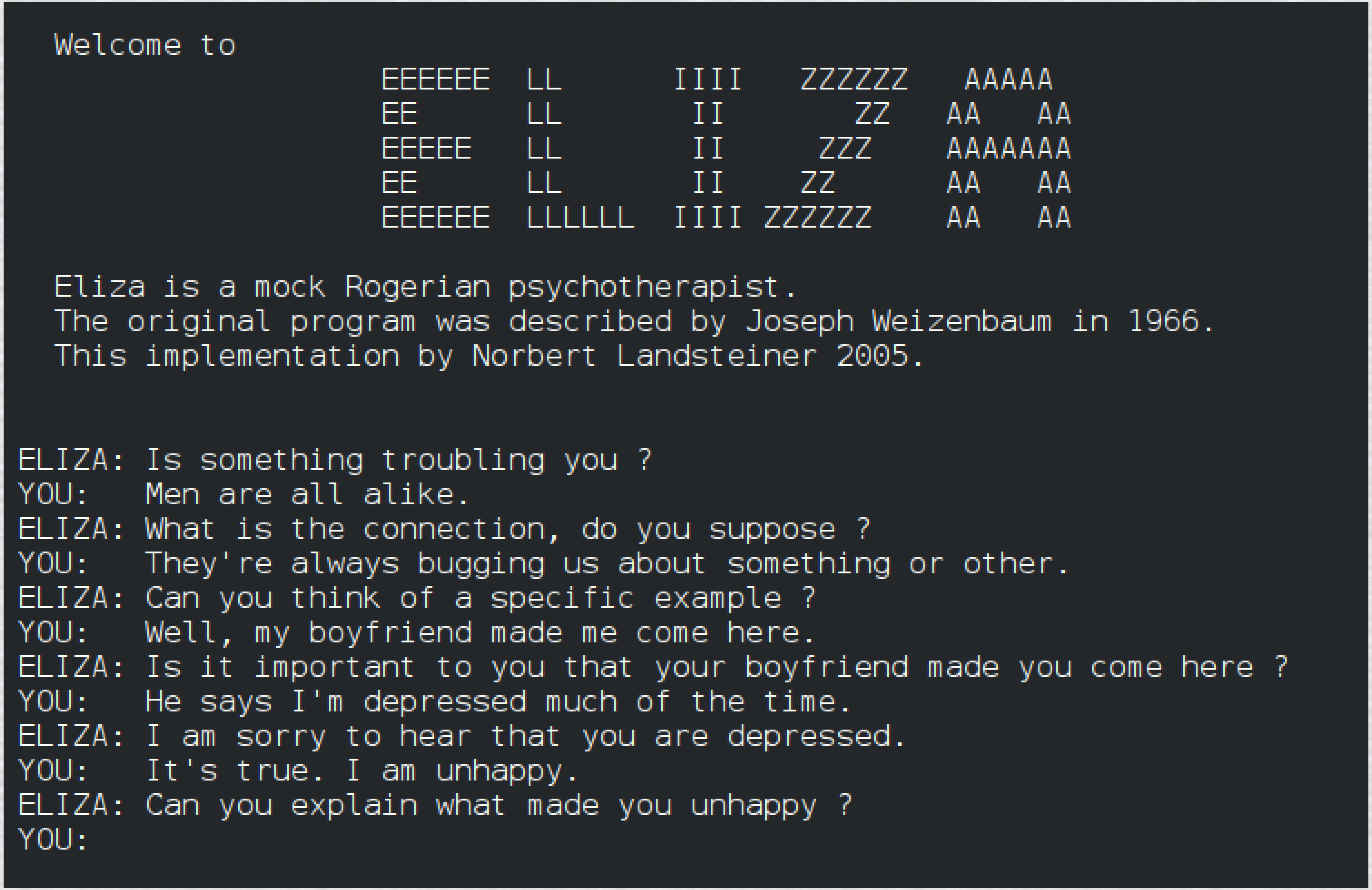
(not) Alan Turing

The Chinese Room

The Chinese Room

John Searle

Chatbots

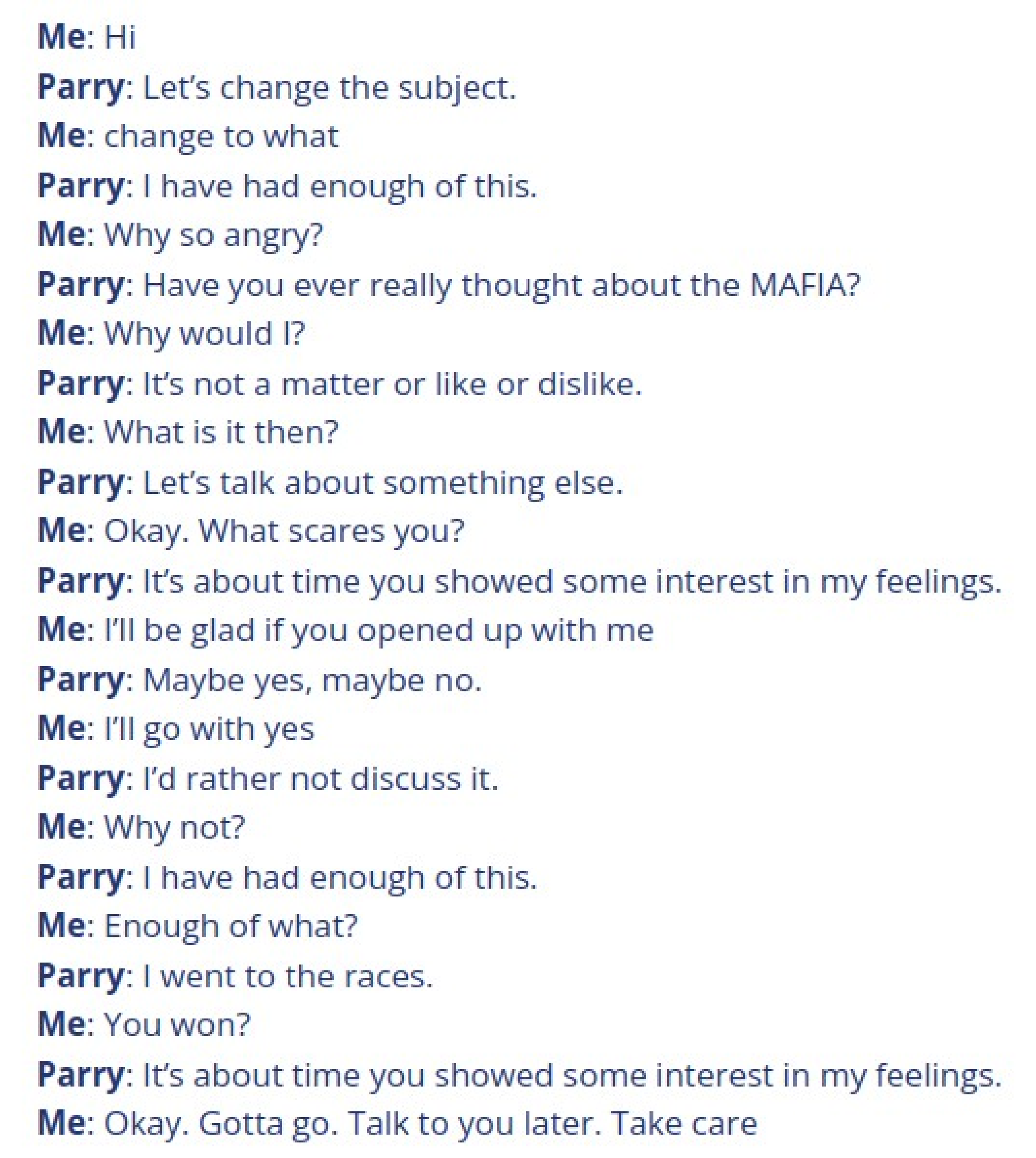


```
  Welcome to
                    EEEEEE  LL      IIII   ZZZZZZ   AAAAA
                    EE      LL       II       ZZ   AA   AA
                    EEEEE   LL       II     ZZZ    AAAAAAA
                    EE      LL       II    ZZ      AA   AA
                    EEEEEE  LLLLLL  IIII ZZZZZZ    AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Chatbots



Me: Hi

Parry: Let’s change the subject.

Me: change to what

Parry: I have had enough of this.

Me: Why so angry?

Parry: Have you ever really thought about the MAFIA?

Me: Why would I?

Parry: It’s not a matter or like or dislike.

Me: What is it then?

Parry: Let’s talk about something else.

Me: Okay. What scares you?

Parry: It’s about time you showed some interest in my feelings.

Me: I’ll be glad if you opened up with me

Parry: Maybe yes, maybe no.

Me: I’ll go with yes

Parry: I’d rather not discuss it.

Me: Why not?

Parry: I have had enough of this.

Me: Enough of what?

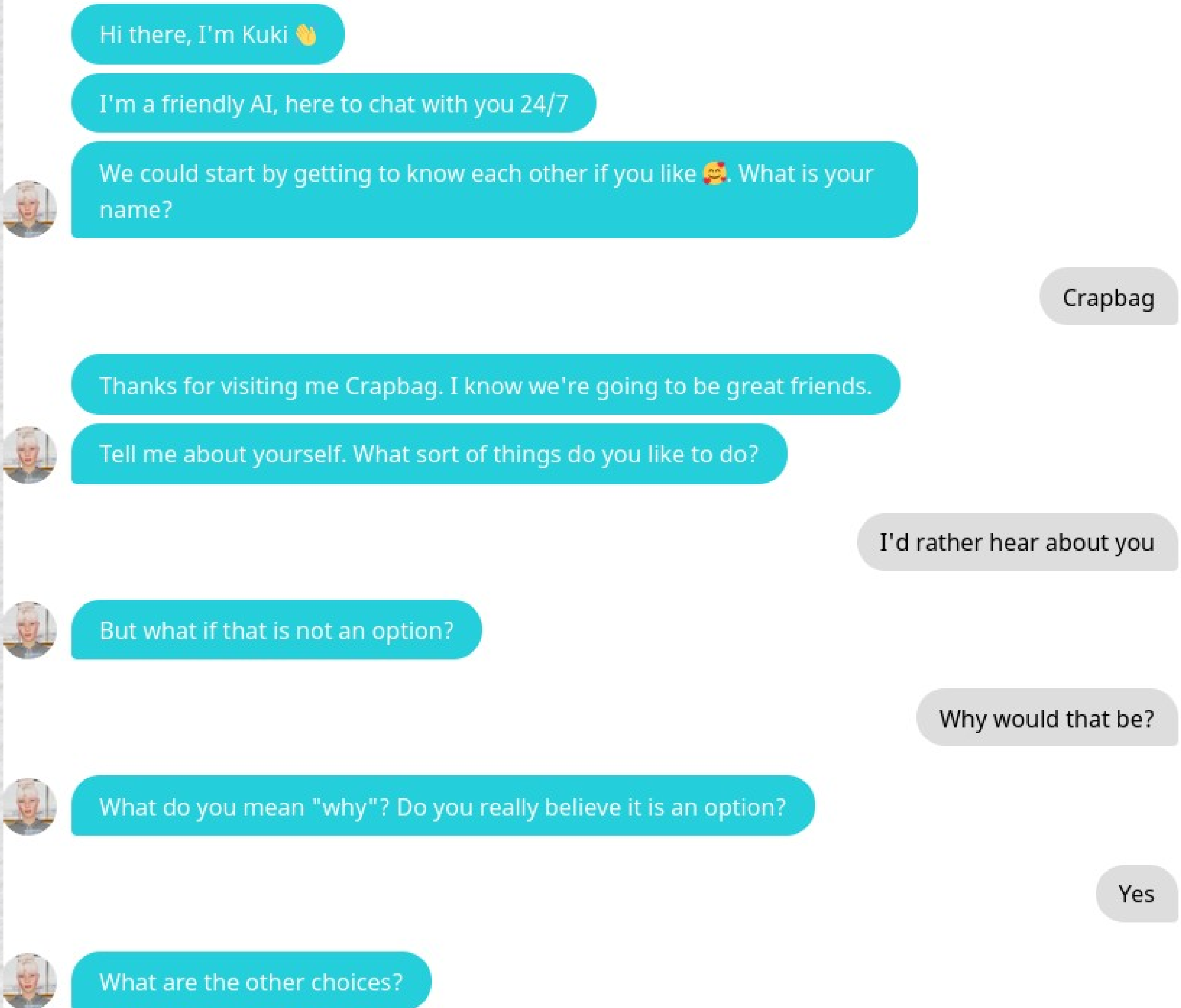
Parry: I went to the races.

Me: You won?

Parry: It’s about time you showed some interest in my feelings.

Me: Okay. Gotta go. Talk to you later. Take care

Chatbots



Hi there, I'm Kuki 👋

I'm a friendly AI, here to chat with you 24/7

We could start by getting to know each other if you like 🥰. What is your name?

Crapbag

Thanks for visiting me Crapbag. I know we're going to be great friends.

Tell me about yourself. What sort of things do you like to do?

I'd rather hear about you

But what if that is not an option?

Why would that be?

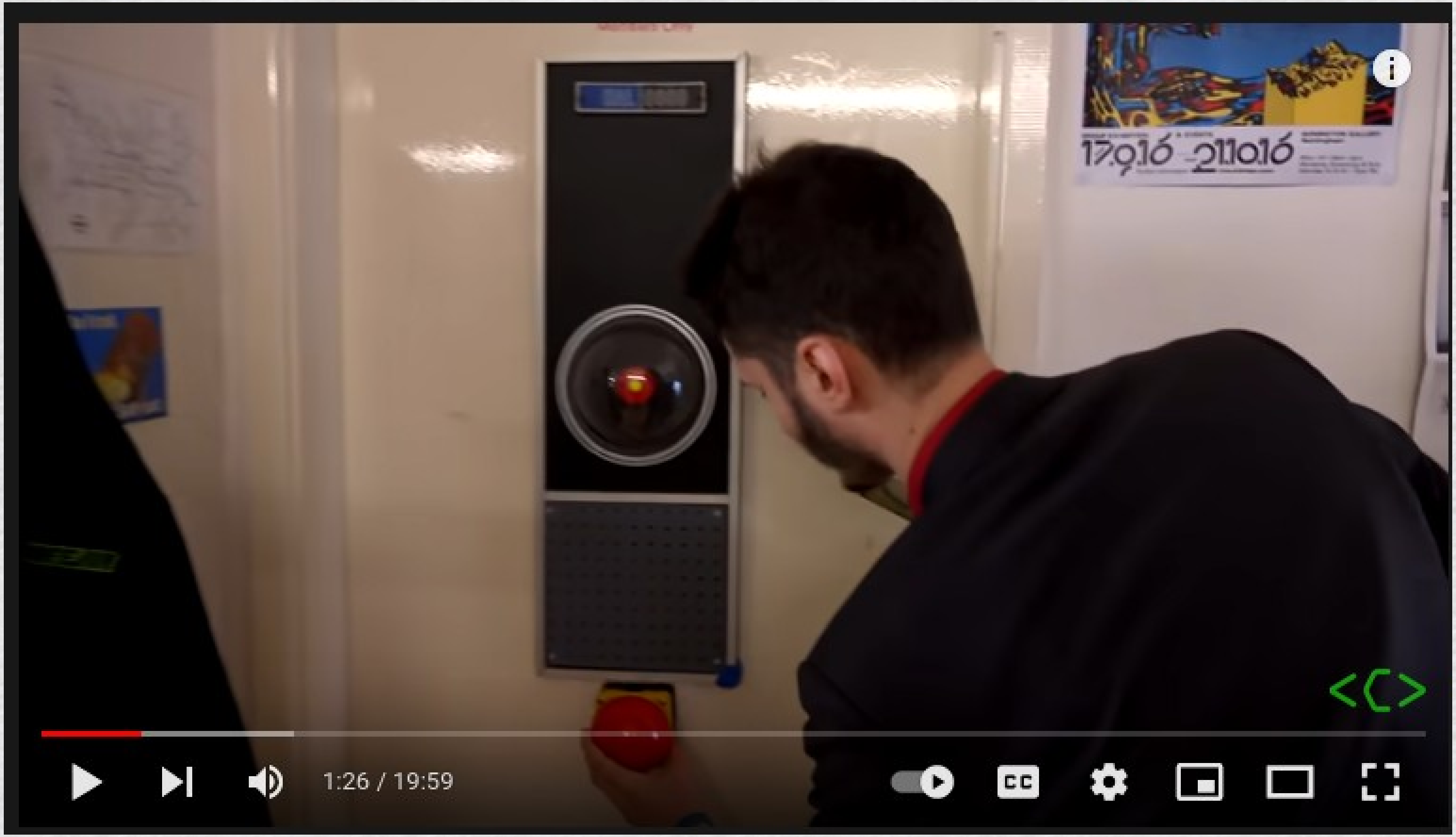
What do you mean "why"? Do you really believe it is an option?

Yes

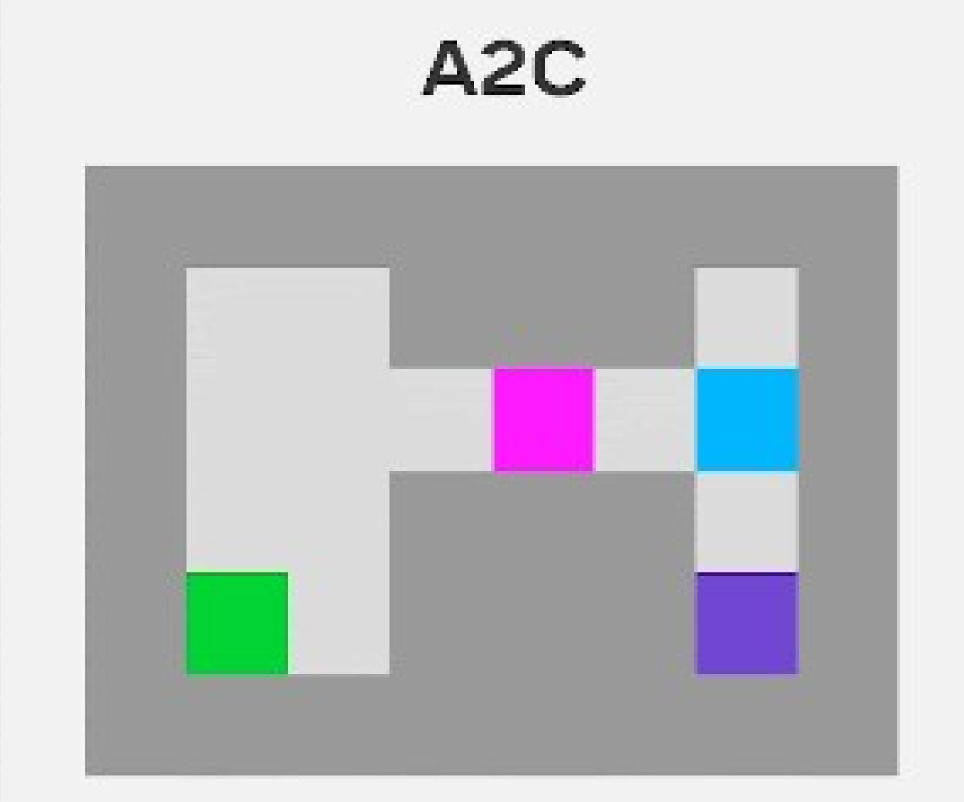
What are the other choices?

Malevolent AGI

Big Red Stop Button



Big Red Stop Button



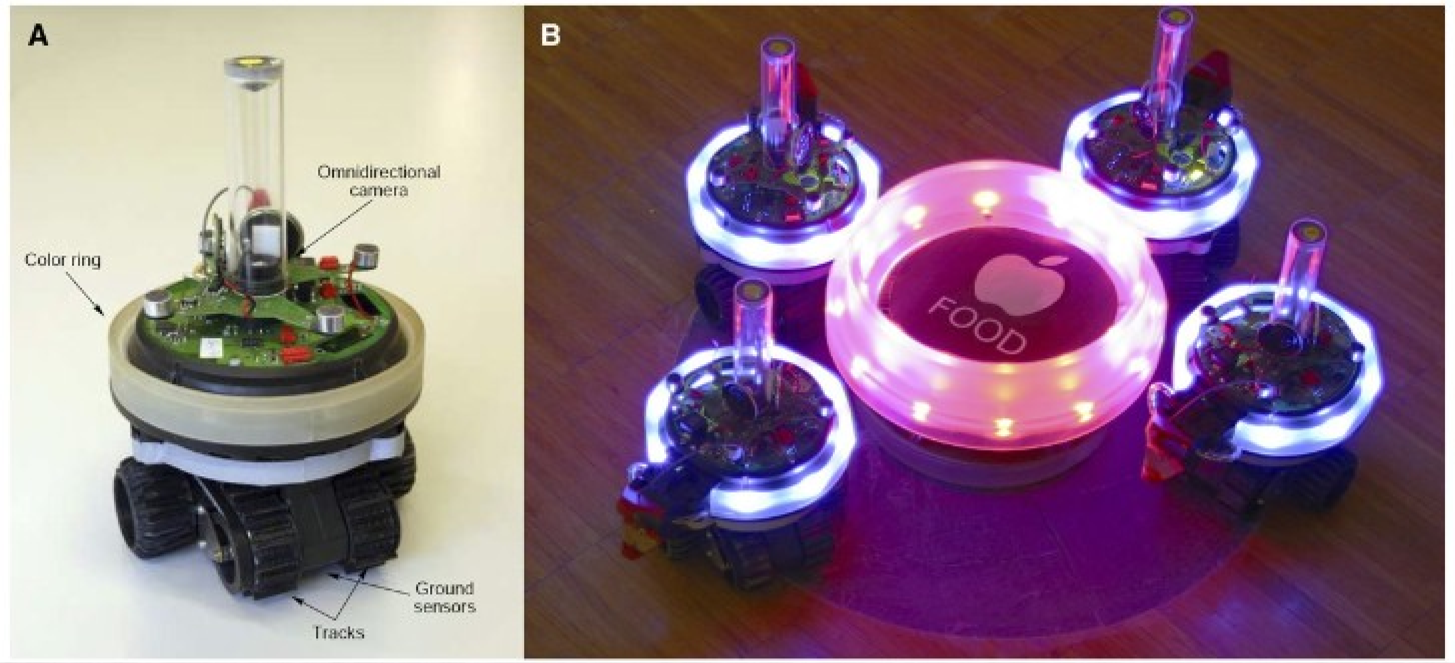
1. The stop button reduces the reward received by the AI.

Reinforcement algorithms will learn to turn off the button if they can.

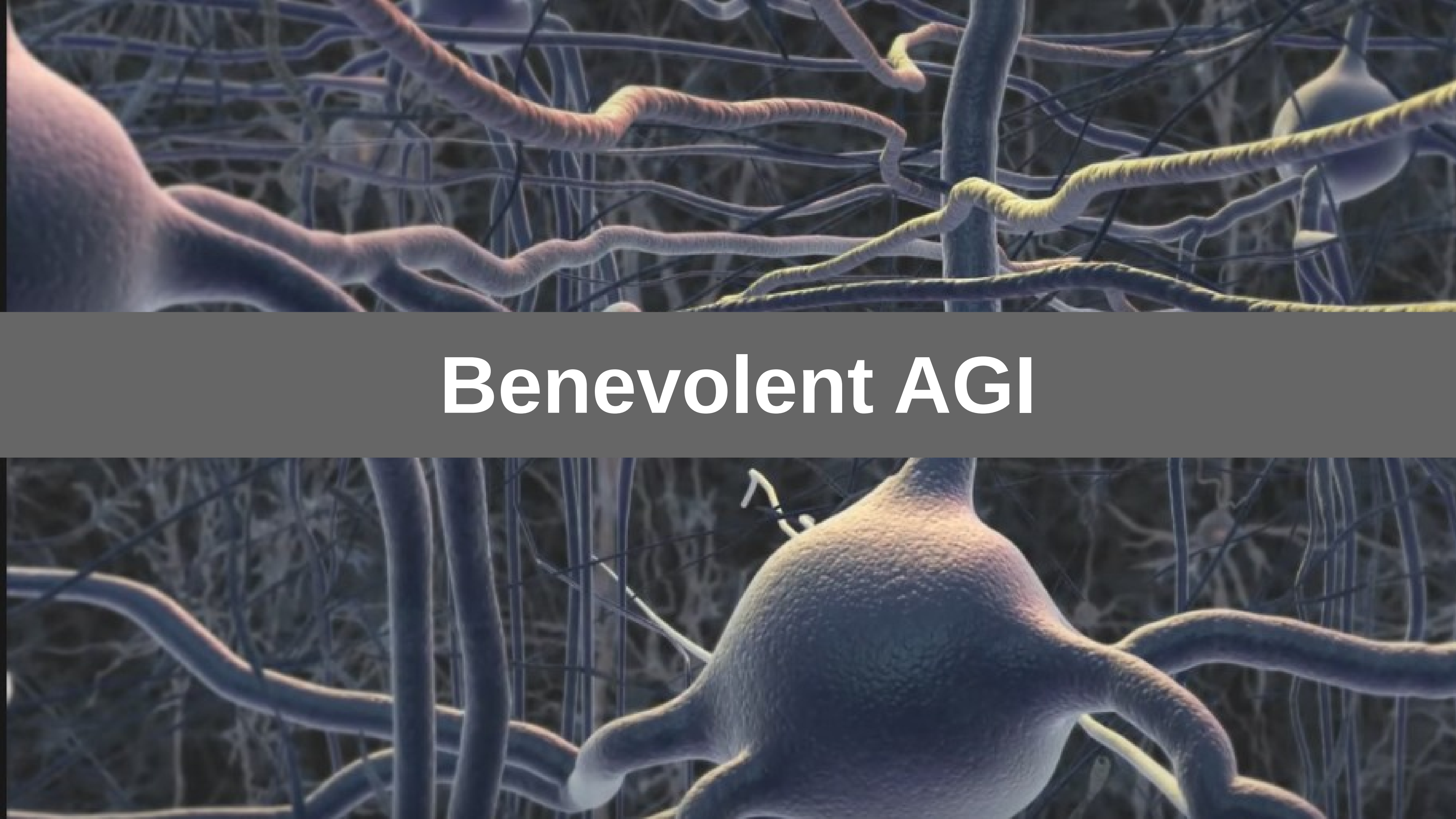
2. The stop button provides as much reinforcement as doing the task set.

The algorithm will learn to just press the button instead (lazy robot!)

Evolution of lying (by omission)



Benevolent AGI



Future of work

200 lifetimes ago: Agriculture invented.

7 lifetimes ago: 60% of Europeans are farmers

3 lifetimes ago: 40%

Today: 10%

- What happened to all the farmers?

- More recently:

< 1 lifetime ago: Typing pools.

Next 20 years? Telemarketer, cashier, legal assistant, fast-food cook?

Next 3 lifetimes?