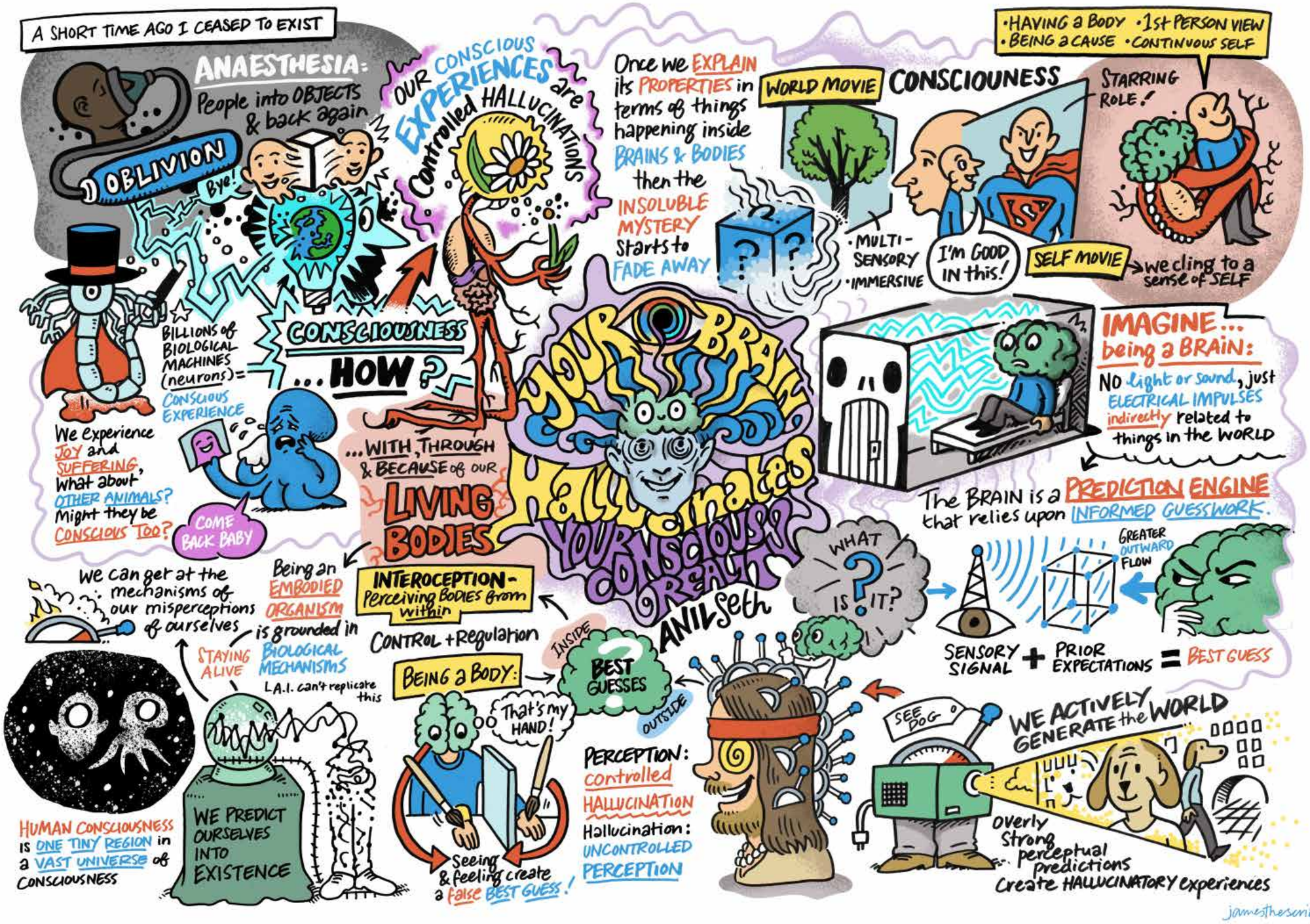
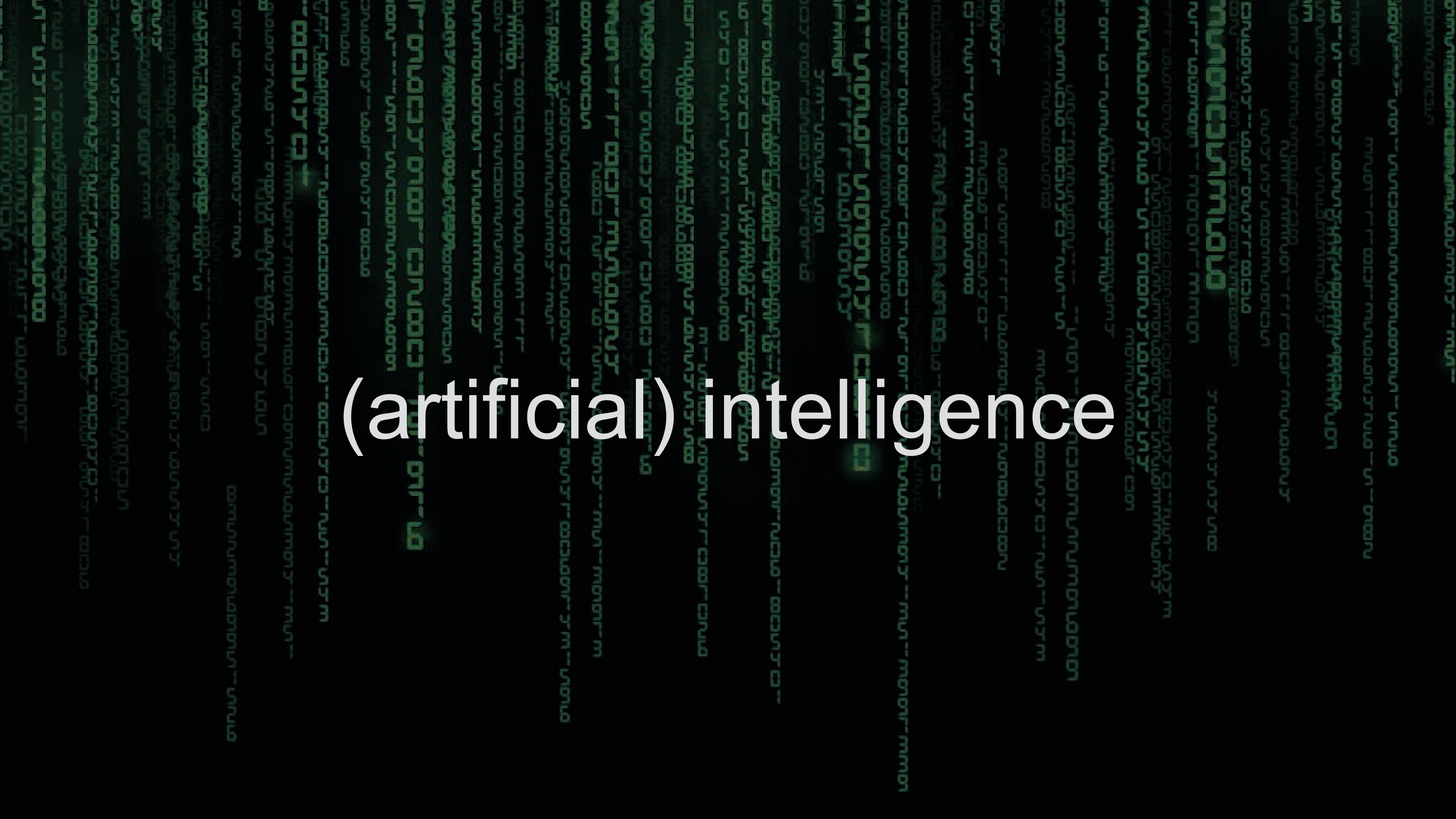


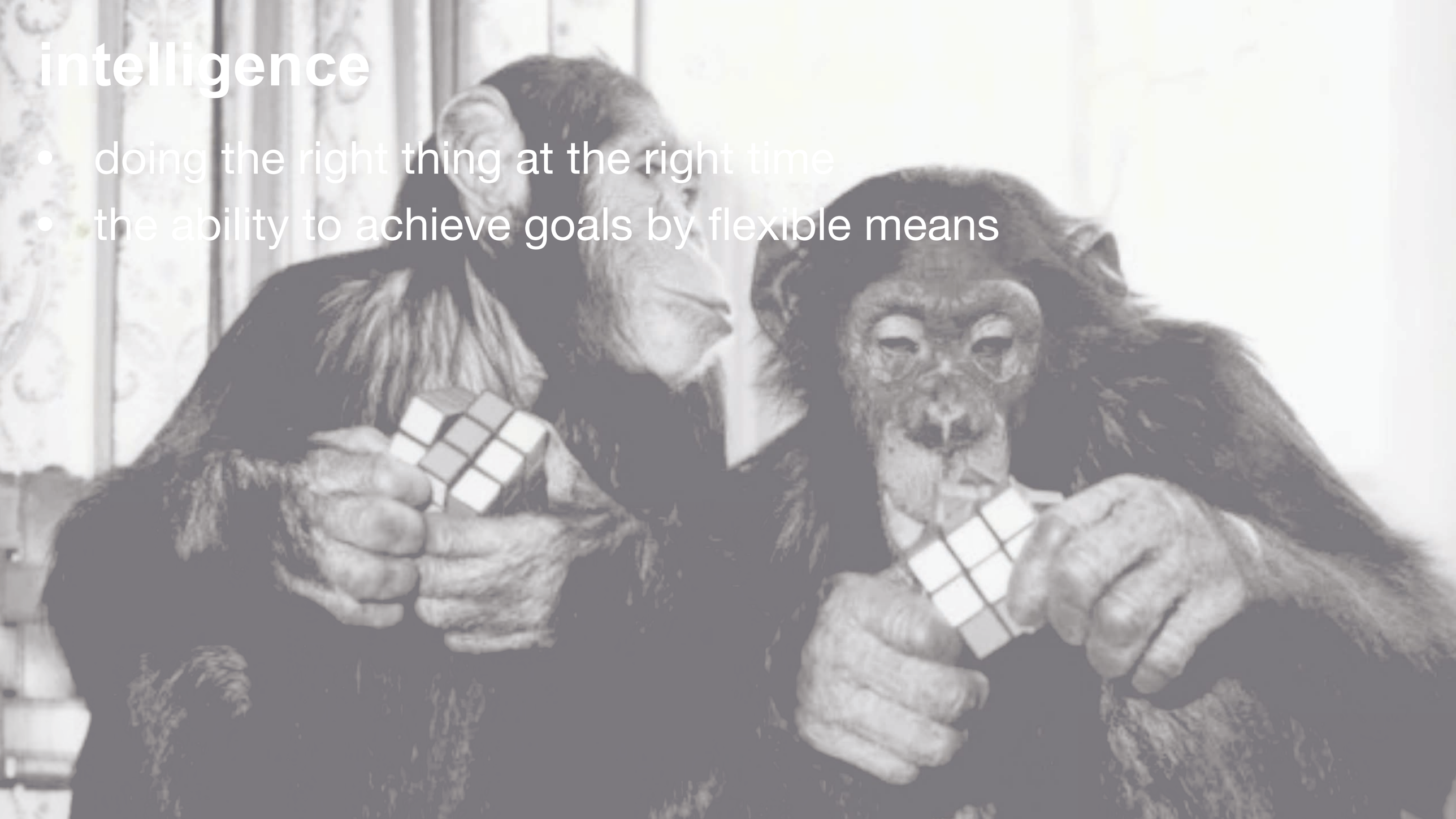
www.jamesthescribe.com

the intelligence we have,
and the intelligence we need


(artificial) intelligence

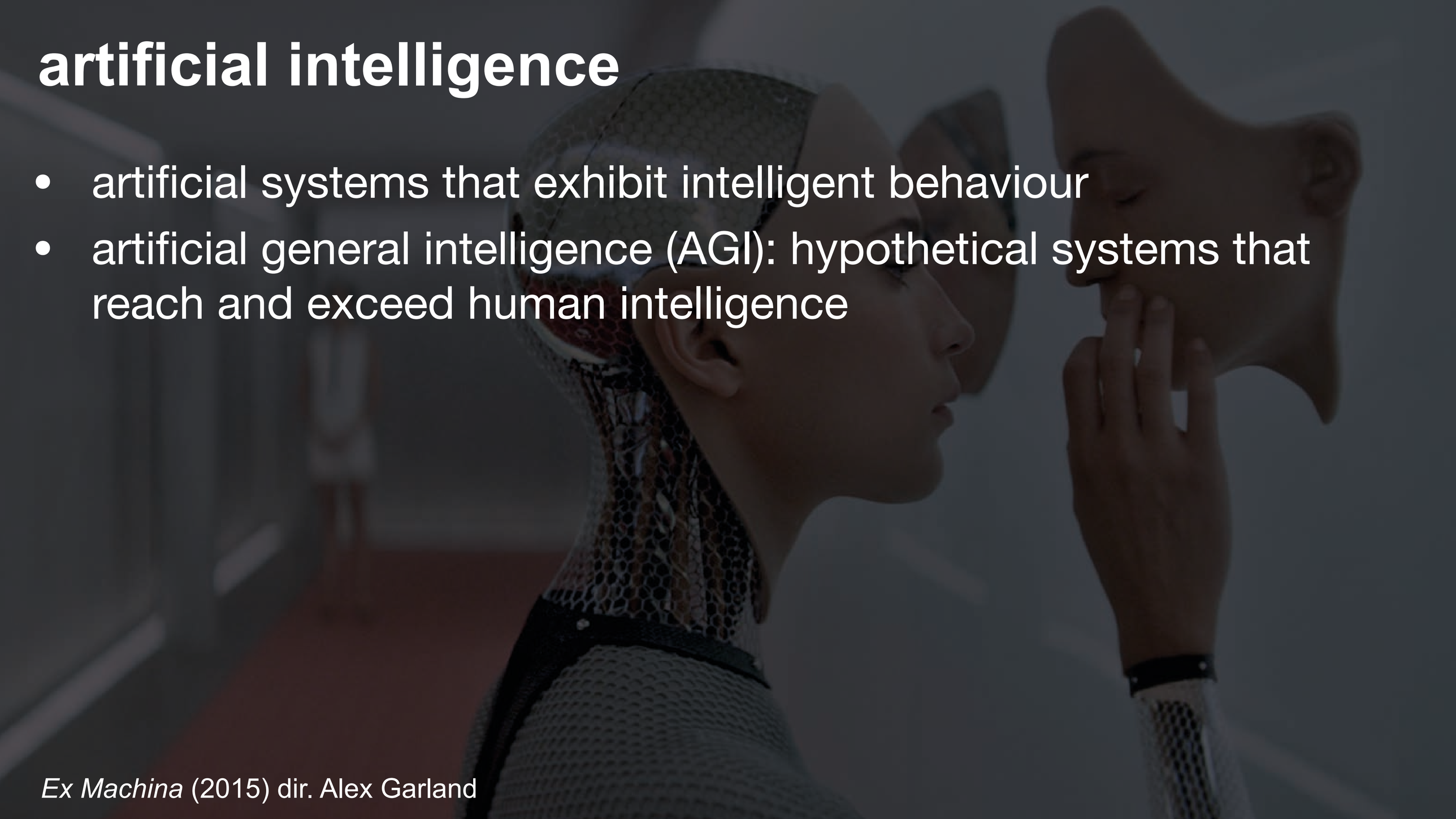
intelligence

- doing the right thing at the right time
- the ability to achieve goals by flexible means



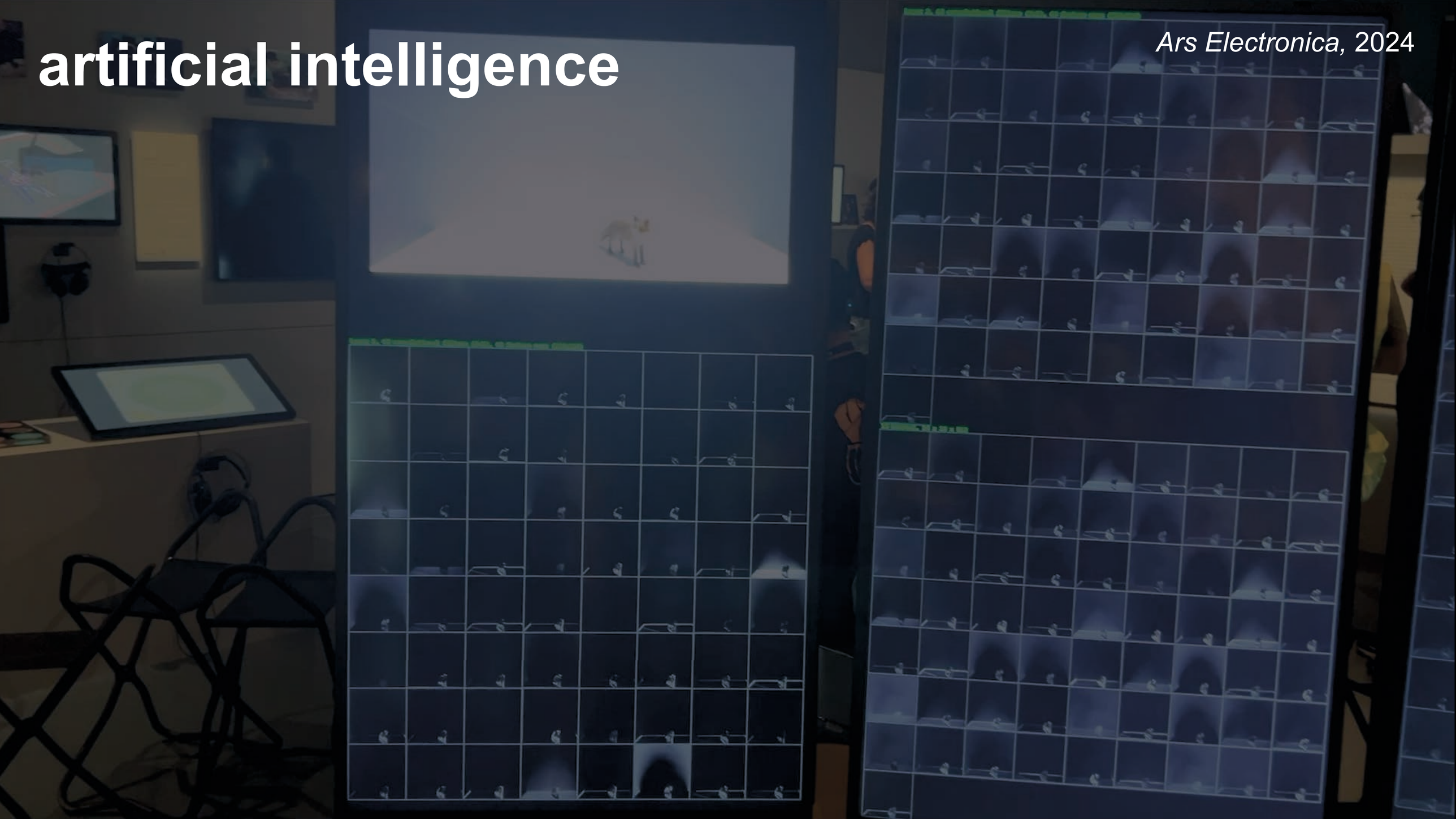
artificial intelligence

- artificial systems that exhibit intelligent behaviour
- artificial general intelligence (AGI): hypothetical systems that reach and exceed human intelligence



Ex Machina (2015) dir. Alex Garland

artificial intelligence



Ars Electronica, 2024

A. M. Turing (1950) Computing Machinery and Intelligence. ***Mind 49:*** **433-460.**

COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:

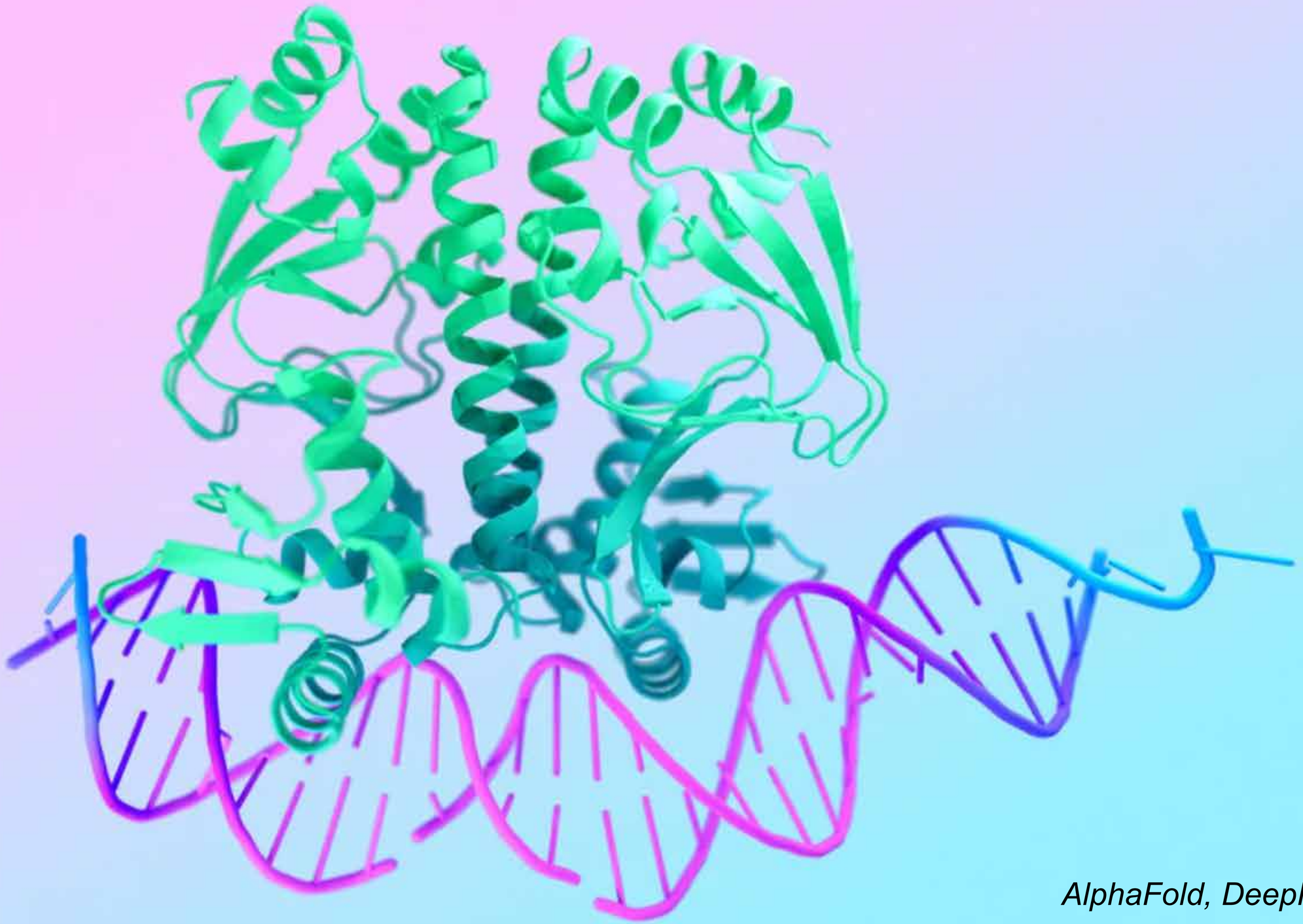
C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be:

"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.

We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"



AlphaFold, DeepMind

reasons to be excited
Google DeepMind, October 2024

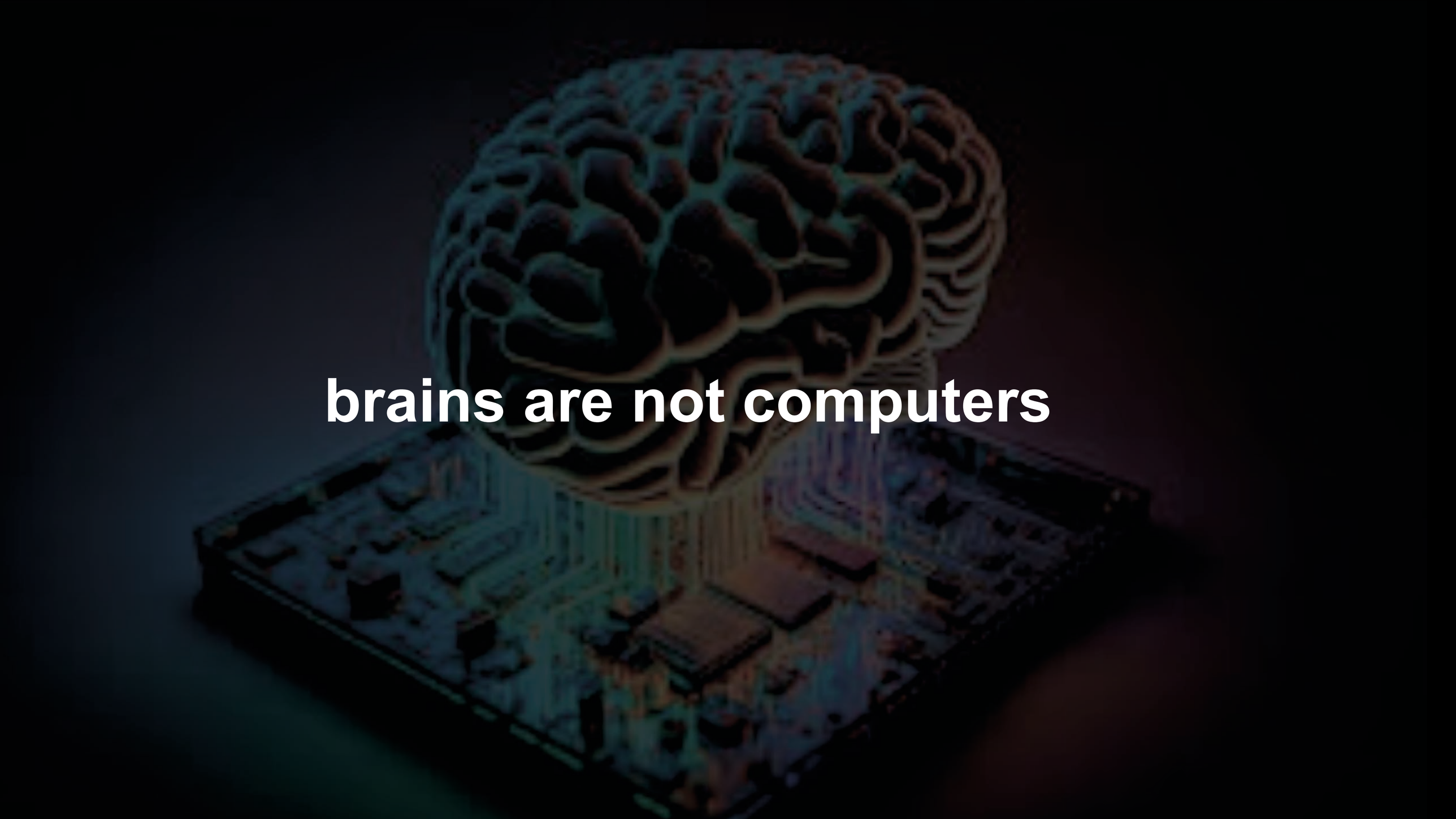
reasons to be worried

the landscape

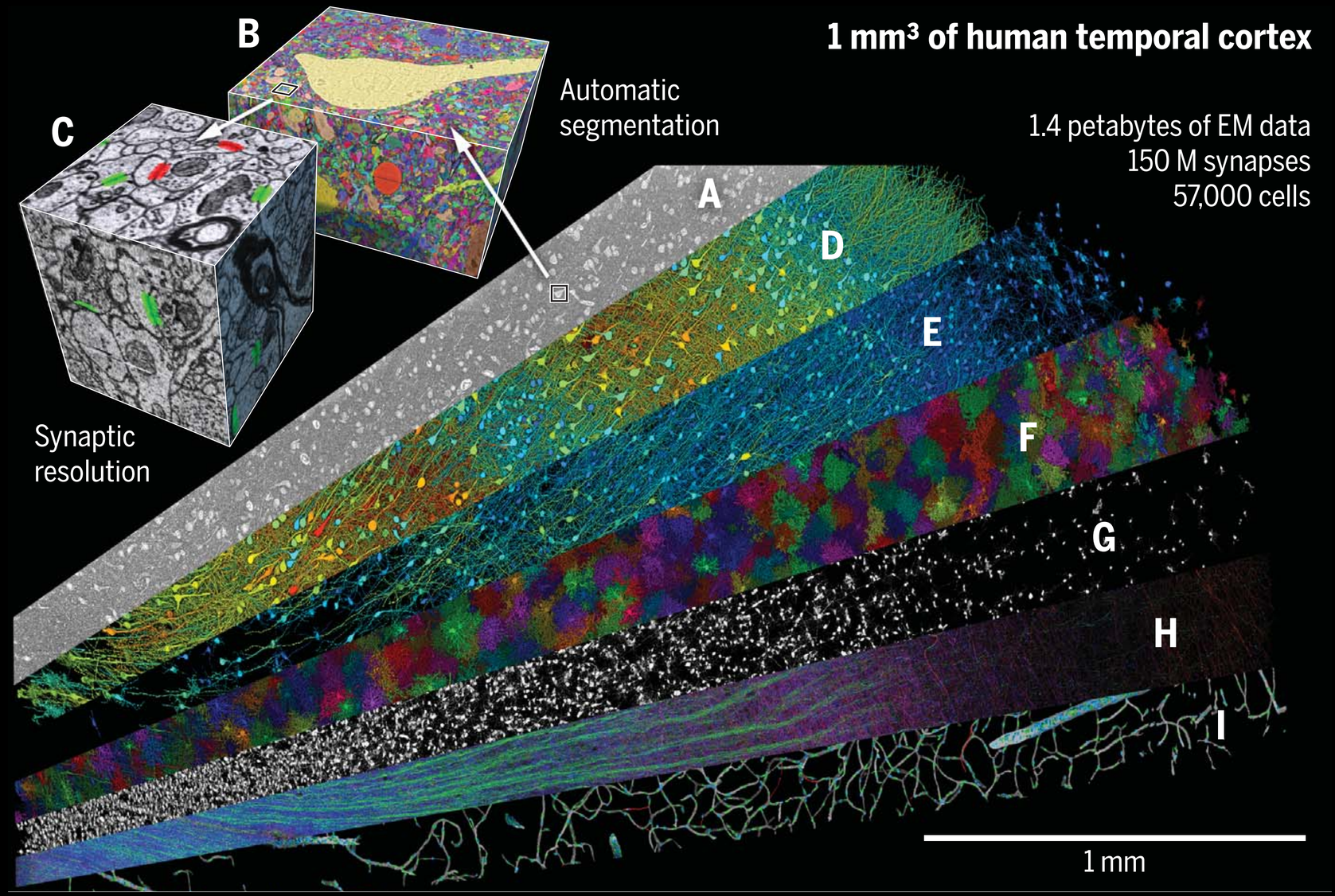
the “singularity”

	near term / certain	long term / uncertain
good	health / biology communication/coding empowerment	superintelligent AI will save us silicon immortality
bad	energy cost social disruption privacy erosion behaviour manipulation	conscious robots will kill us all

the intelligence we have


brains are not computers

complex

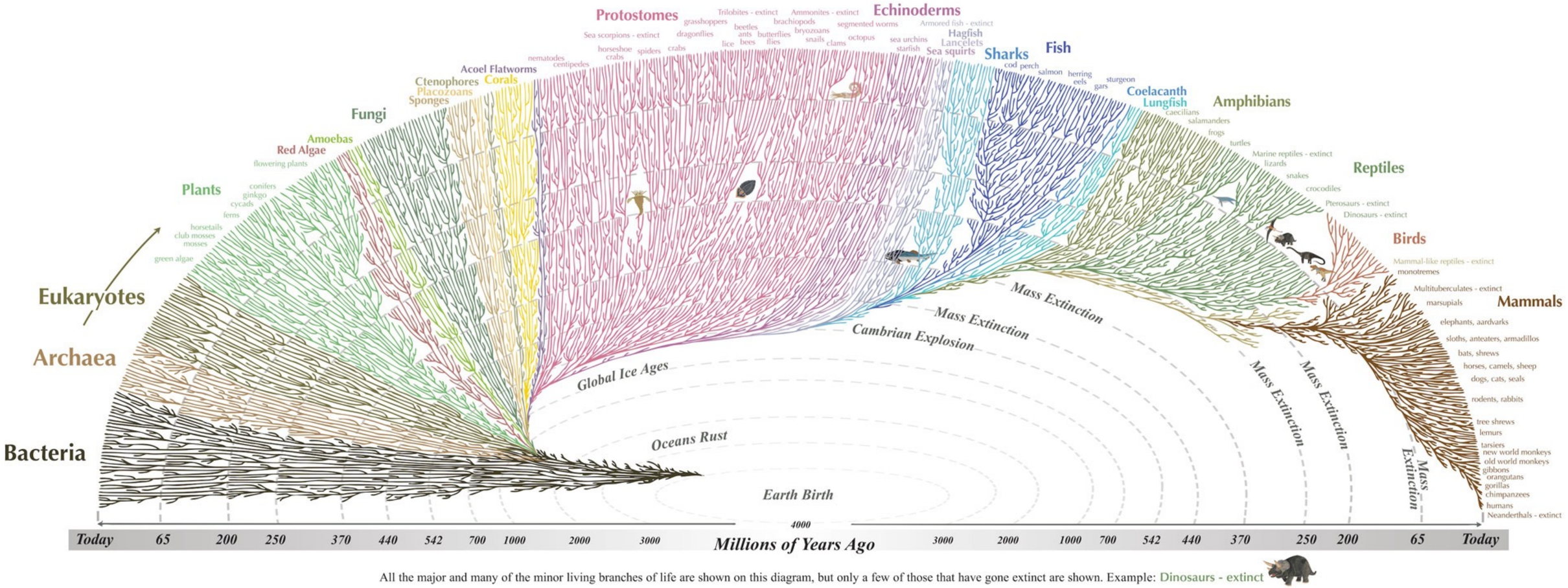


Shapson-Coe et al (2024) *Science*

embodied

Mario Secchi : Alamo stock photo

evolved



All the major and many of the minor living branches of life are shown on this diagram, but only a few of those that have gone extinct are shown. Example: Dinosaurs - extinct

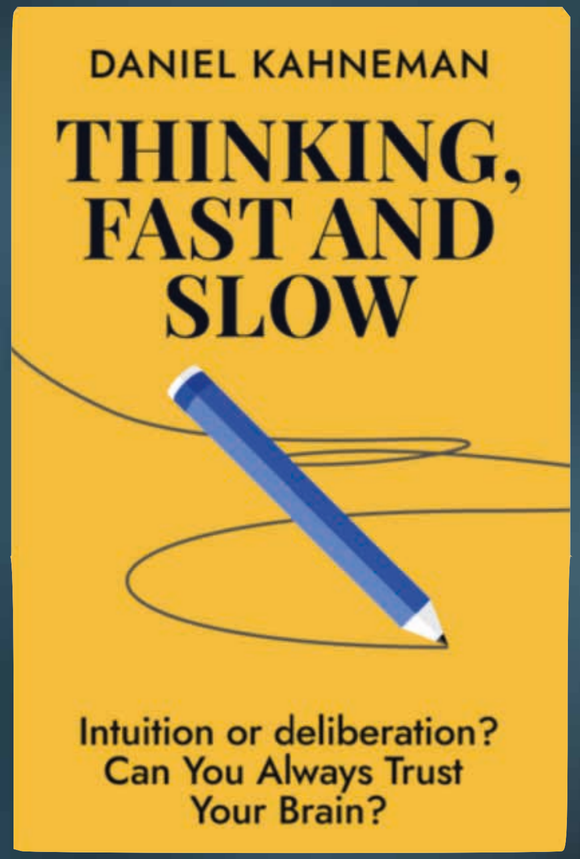
(c) Len Eisenberg, www.evogeneao.com

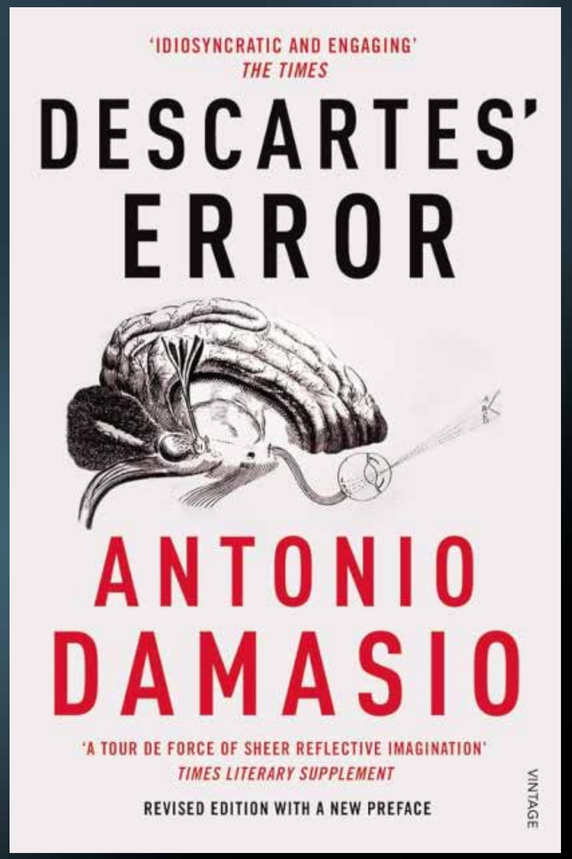
conscious

human intelligence

- bounded cognition
- emotion and reason

- features, not bugs!

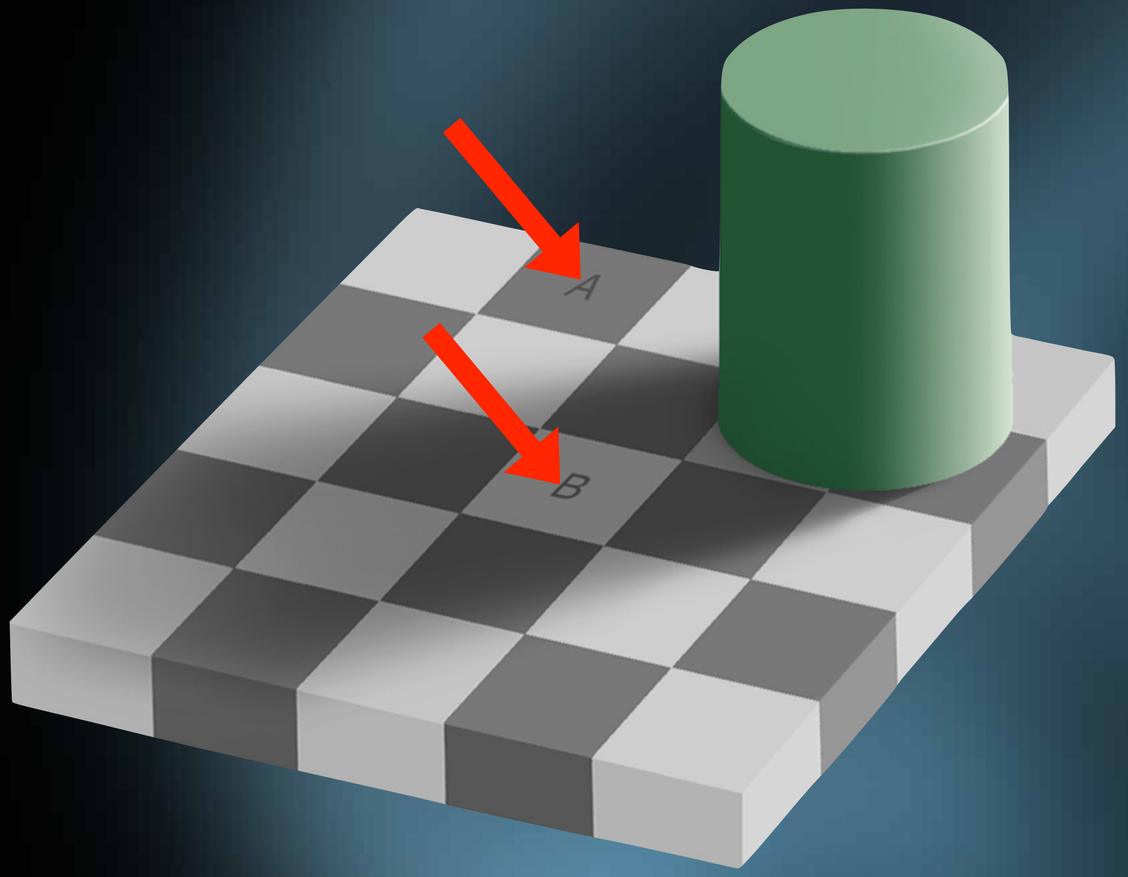


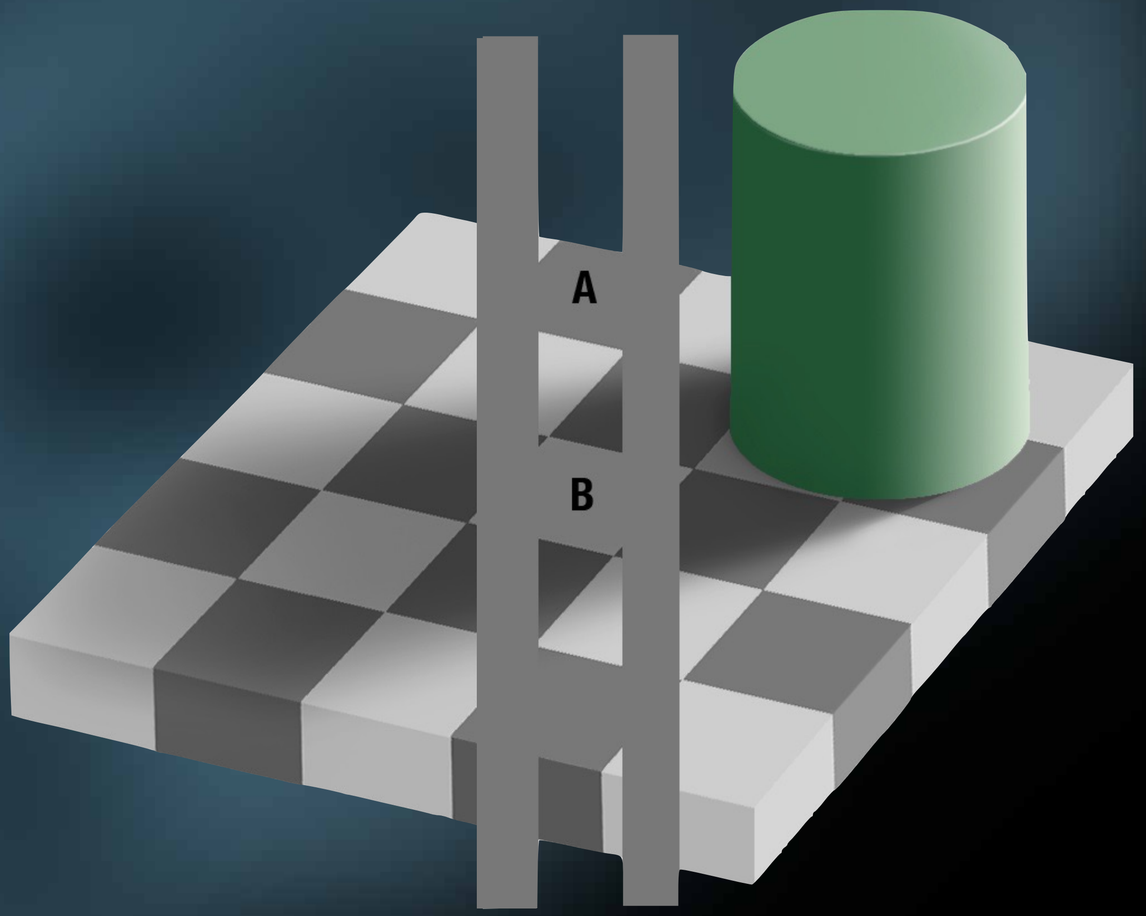


perception

Brighton, Jan 2024

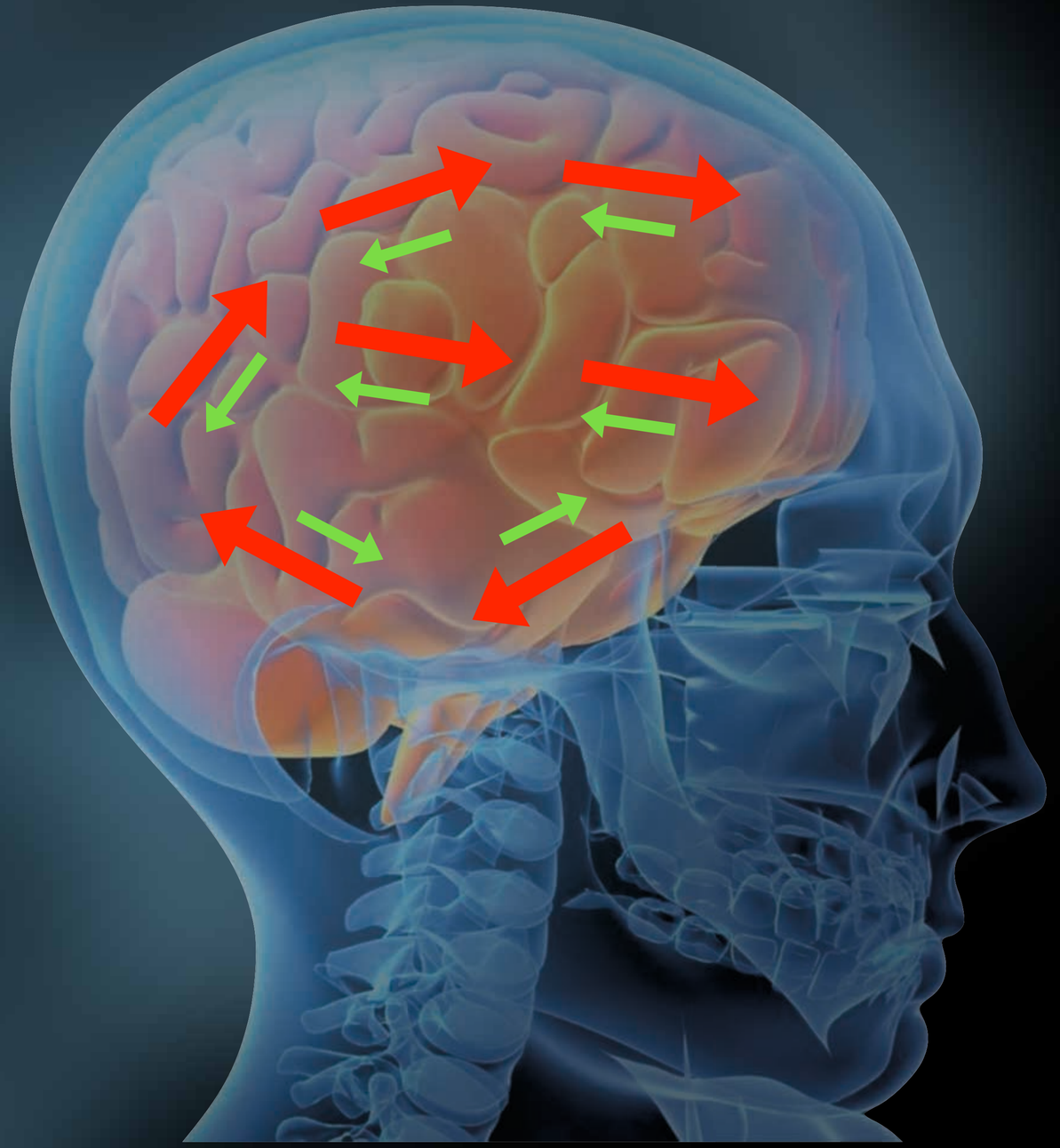
perception as inference





Edward H. Adelson (1995)

controlled
hallucination



prediction error

prediction

Seth (2013) *Trends in Cognitive Sciences*

interoception

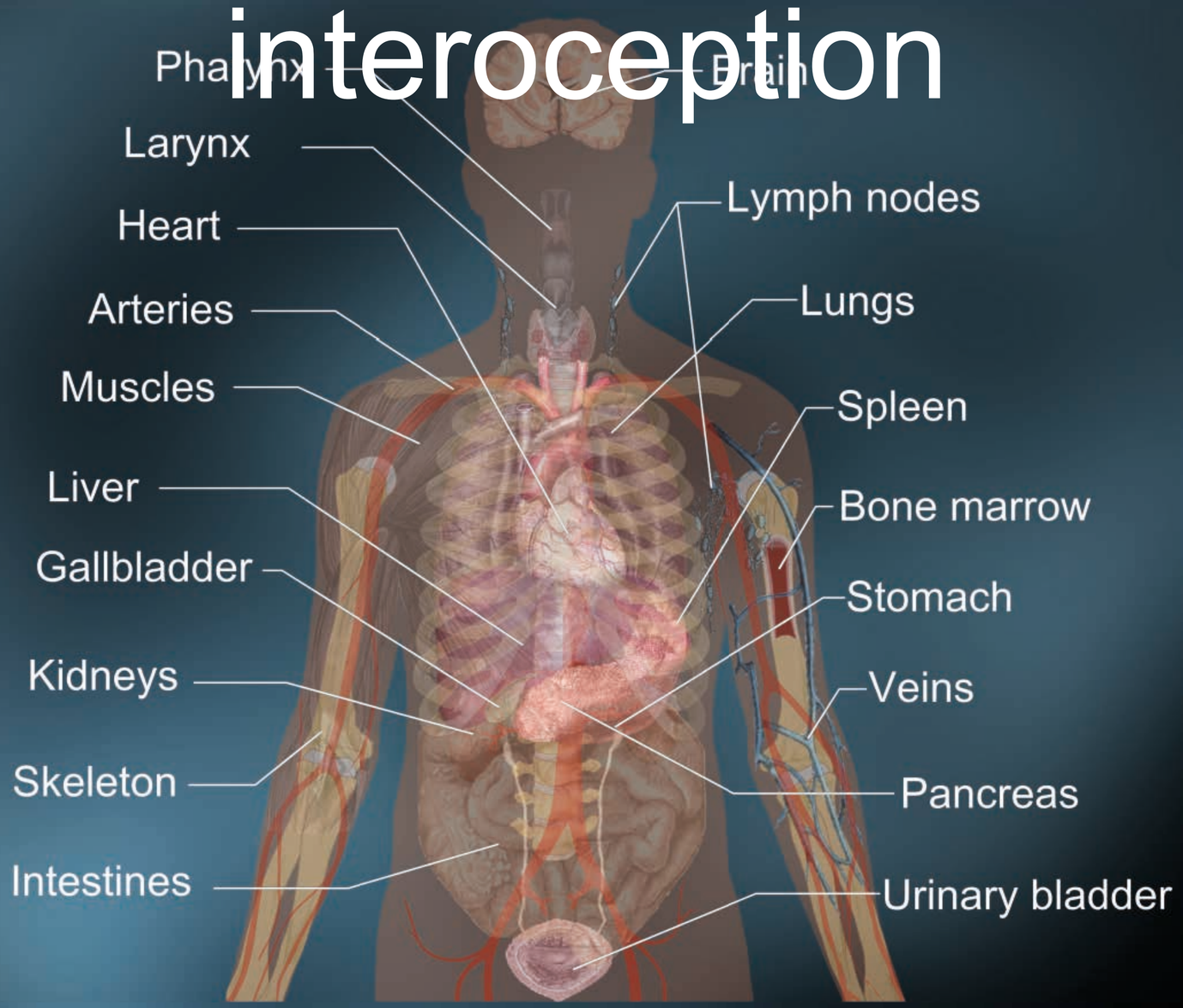
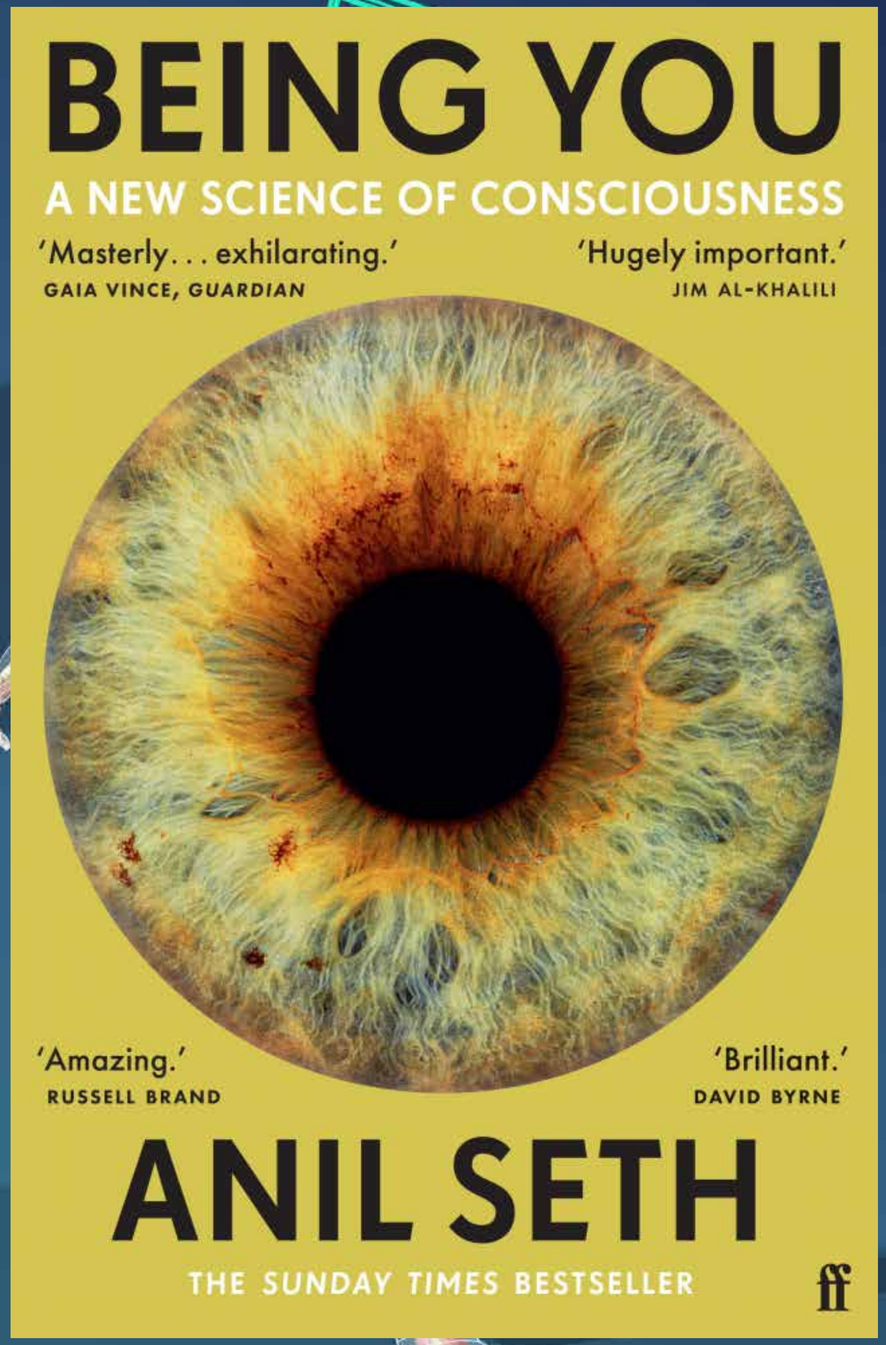


image credit: Mikael Häggström, with permission.

“We perceive the world around us, and ourselves within it, with, through, and *because of* our living bodies”



Seth (2021) *Being You: A New Science of Consciousness*

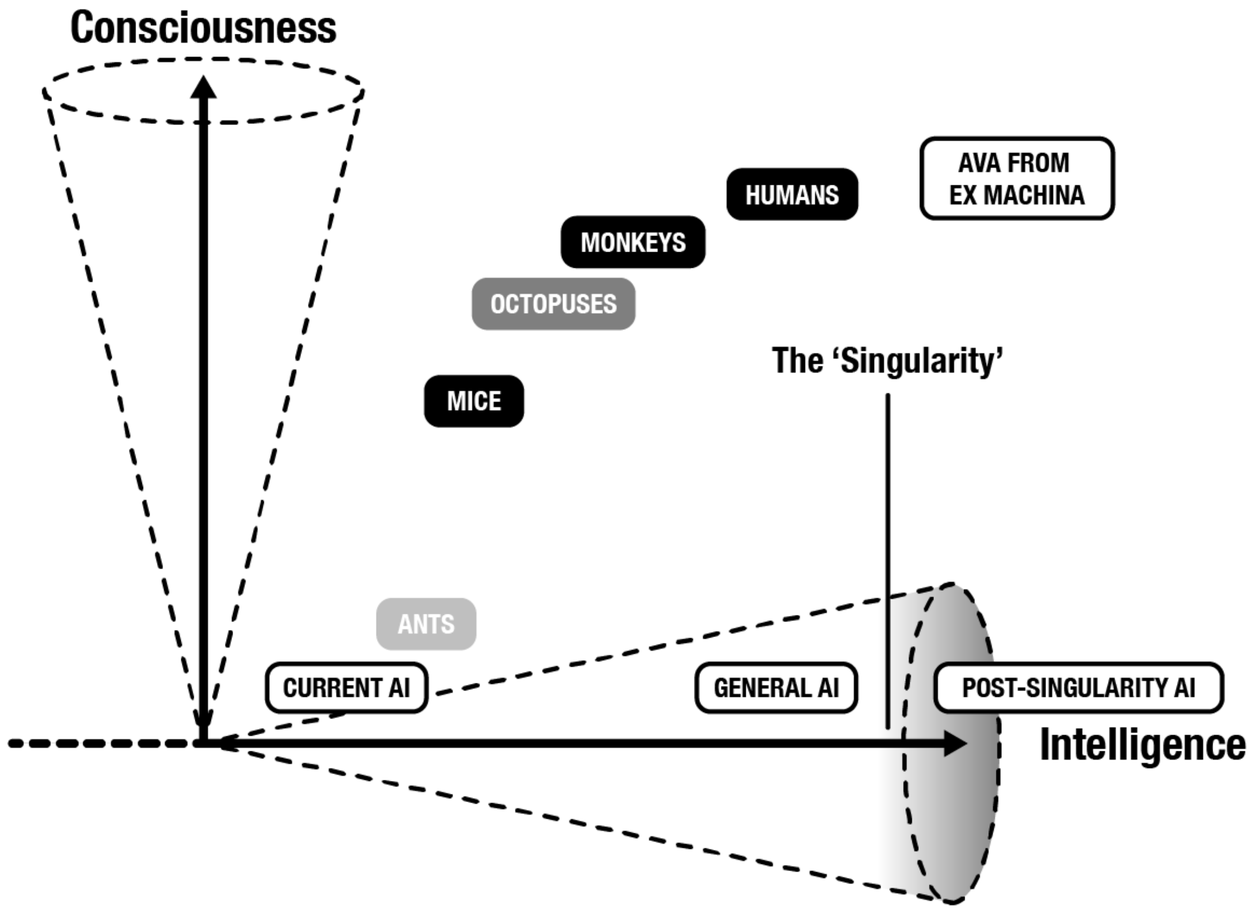
consciousness
intelligence
life

clarifying the landscape

HAL 9000

artificial intelligence

real consciousness?



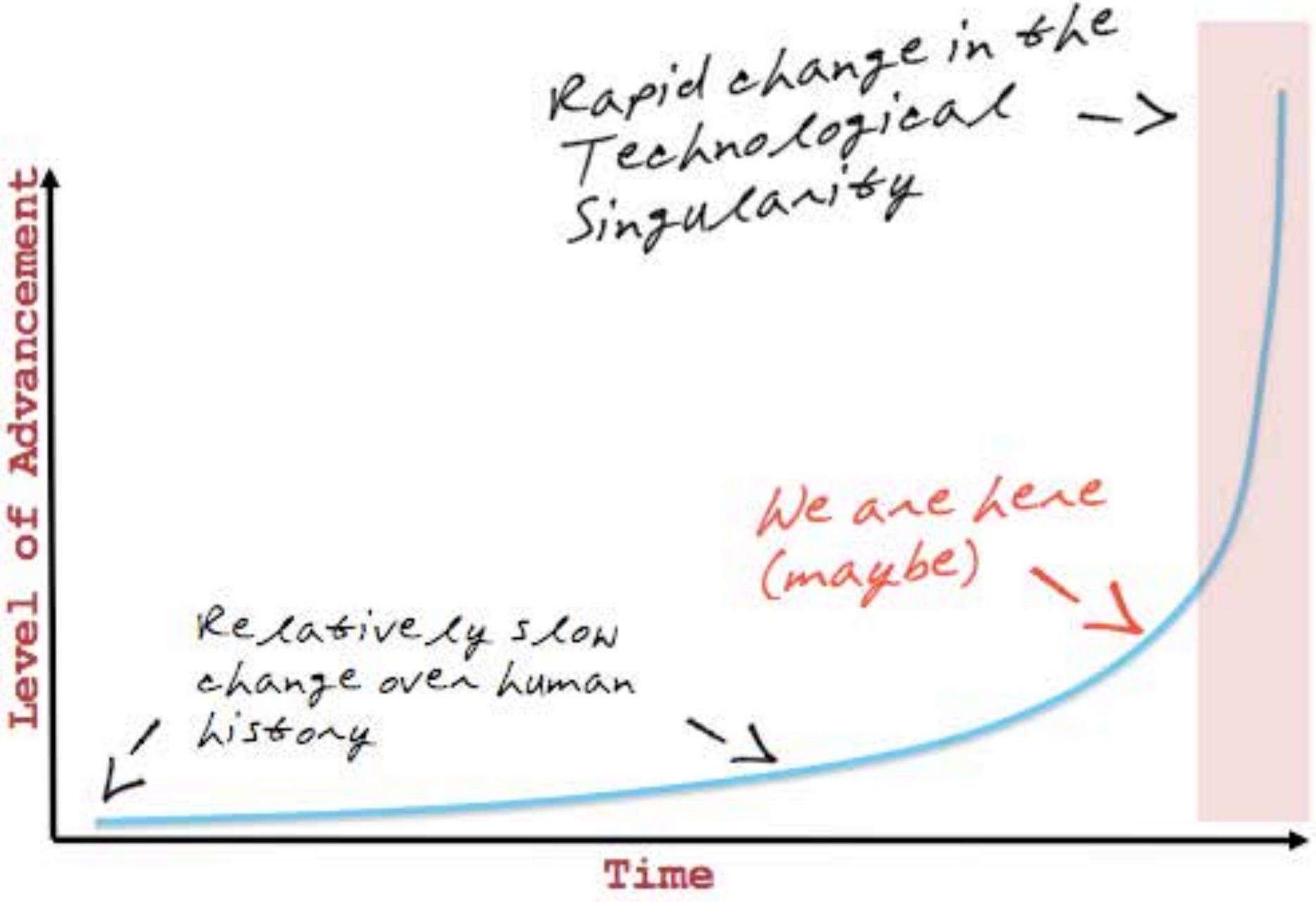
Seth (2024) *PsyArXiv*

Seth (2021) *Being You: A New Science of Consciousness*

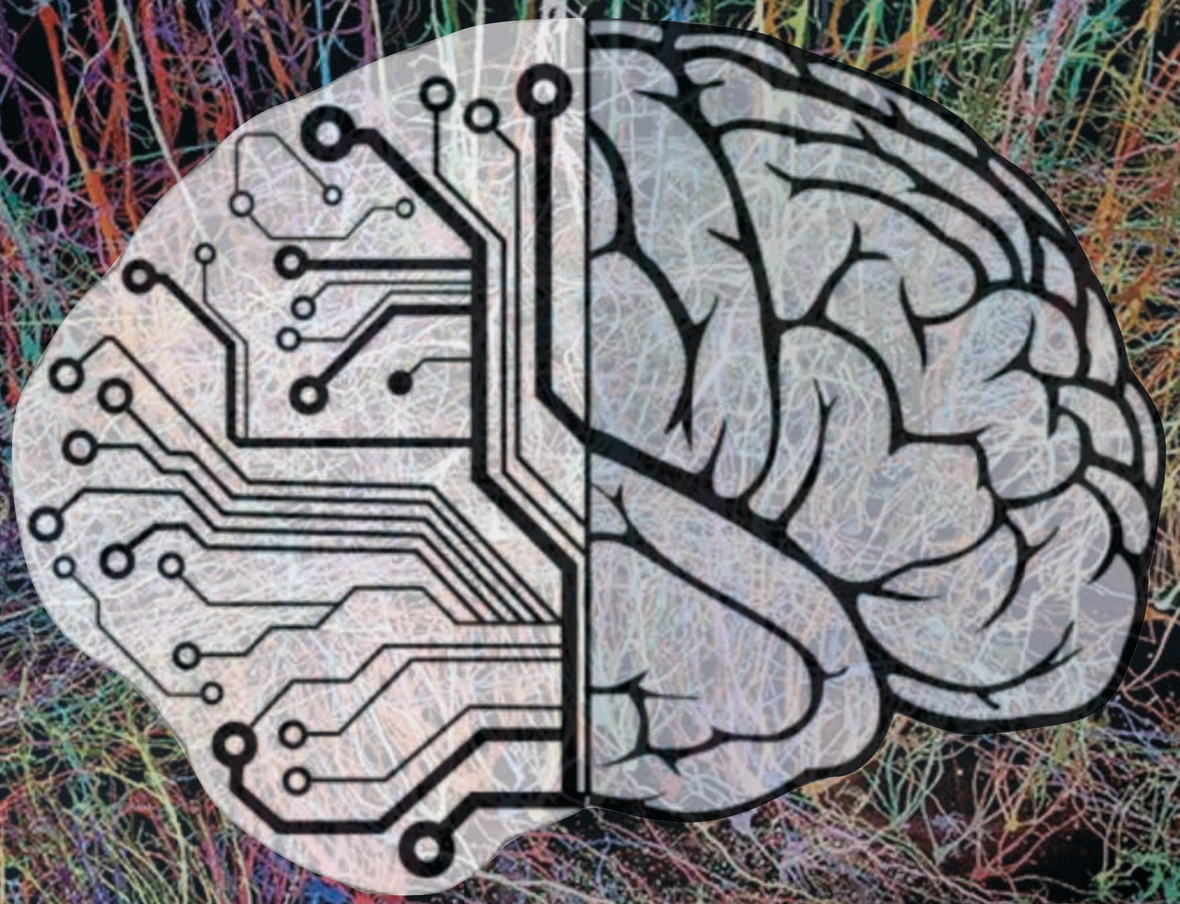
IKEA by Spike Jonze, h/t Henry Shevlin

OpenAI
GPT-4
LLMs don’t **hallucinate**, they **confabulate**

“Technology can make us forget what we know about life.”
Sherry Turkle


Rapid change in the Technological Singularity
We are here (maybe)
Relatively slow change over human history
Level of Advancement
Time

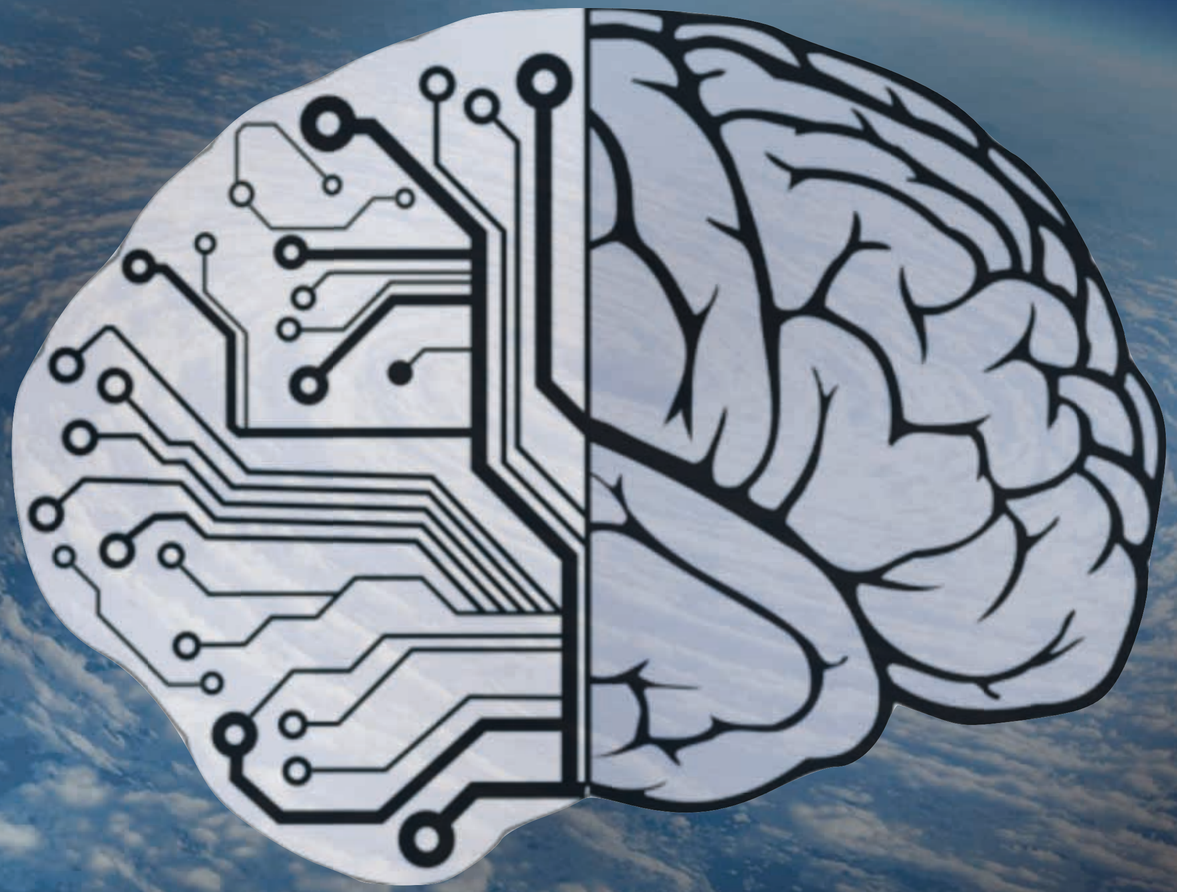
wetware, not hardware



Seth (2024) *PsyArXiv*

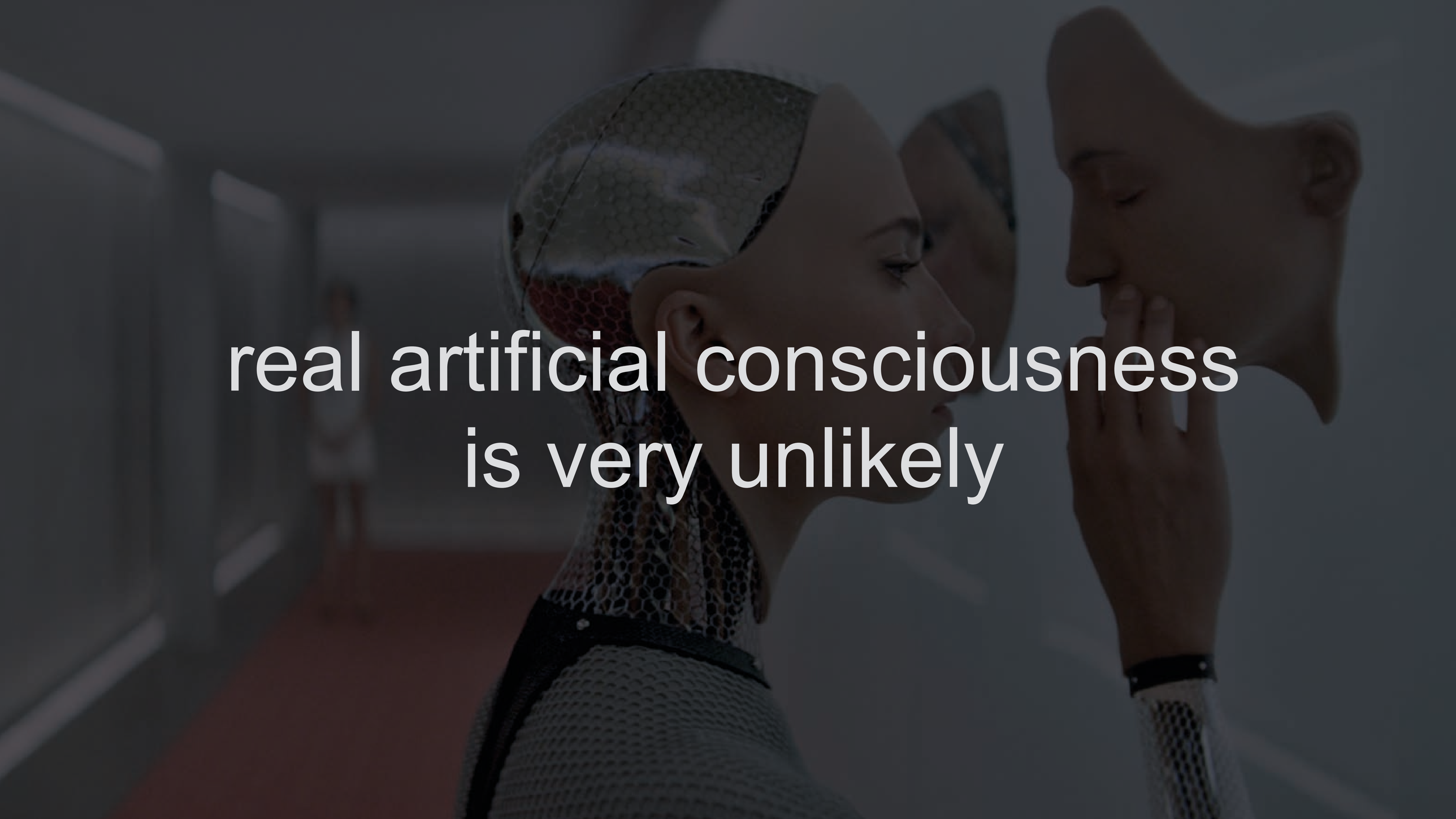
Seth (2021) *Being You: A New Science of Consciousness*

simulation vs instantiation



Seth (2024) *PsyArXiv*

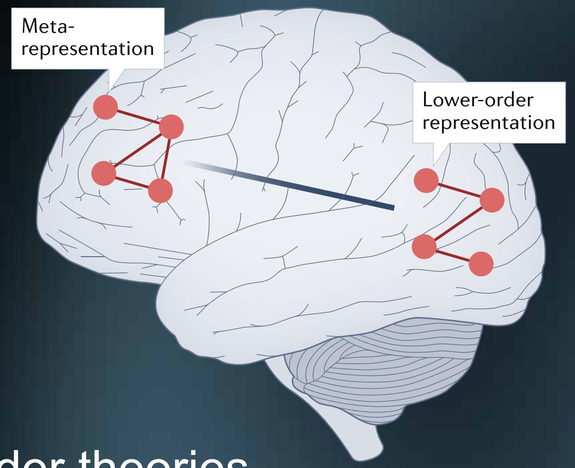
Seth (2021) *Being You: A New Science of Consciousness*


real artificial consciousness
is very unlikely

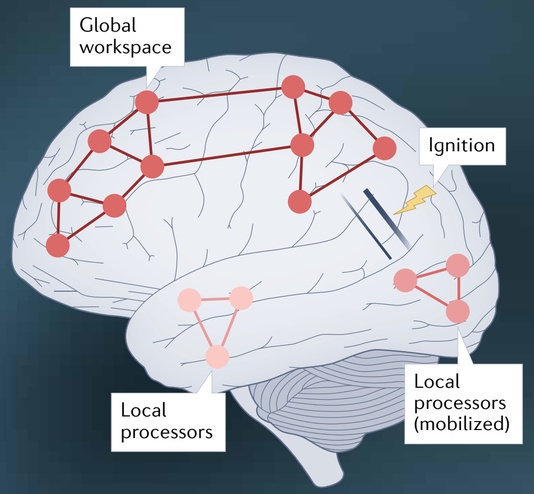
why does this matter?

I might be wrong

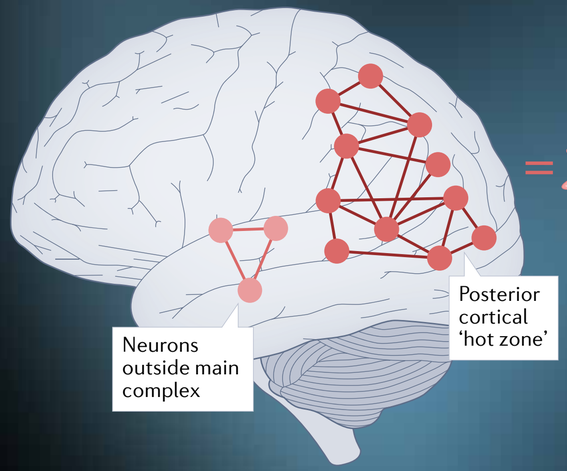
theories of consciousness



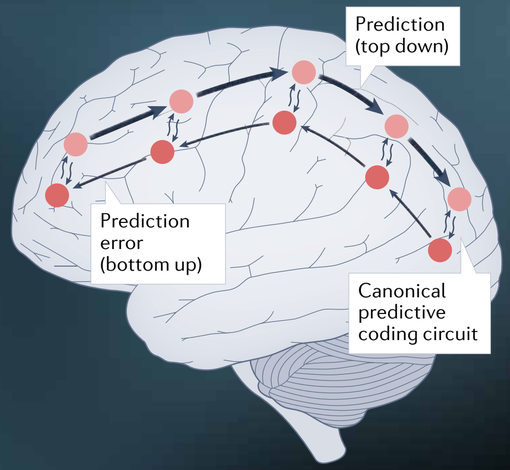
Higher-order theories



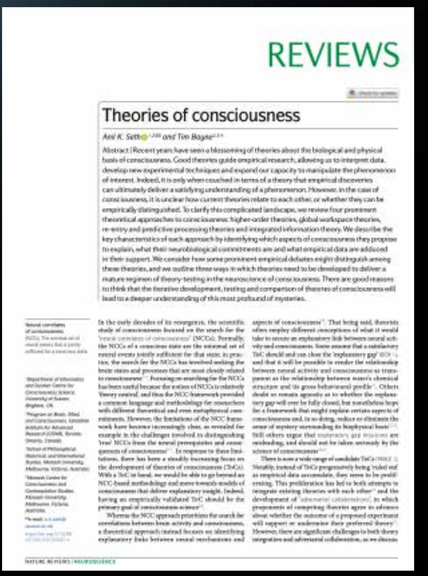
Global workspace theories



Integrated information theories



Predictive processing / reentry theories


REVIEWS

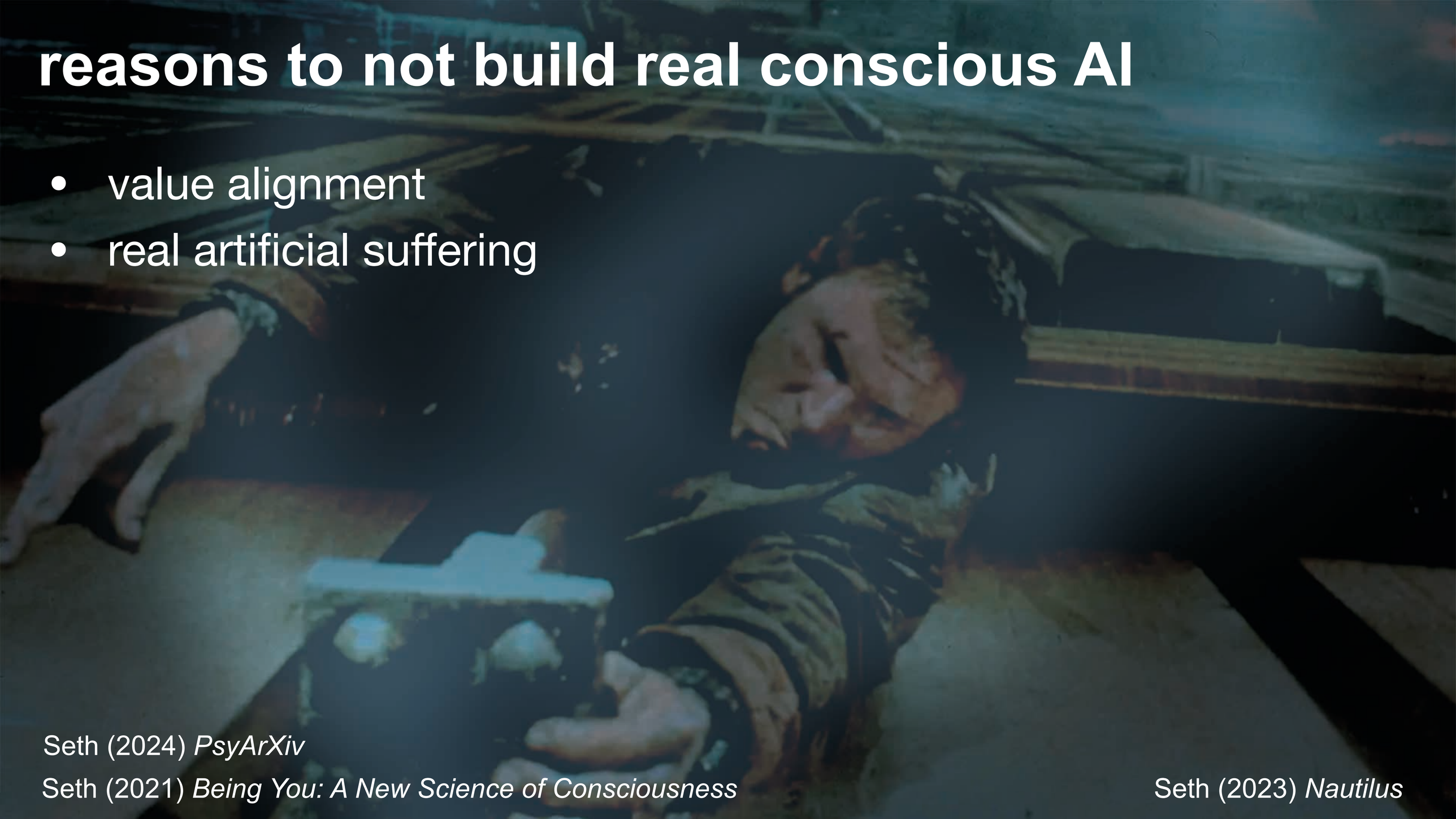
Theories of consciousness

Anil K. Seth and Tim Bayne

Seth & Bayne (2022) *Nature Reviews Neuroscience*

reasons to not build real conscious AI

- value alignment
- real artificial suffering



Seth (2024) *PsyArXiv*

Seth (2021) *Being You: A New Science of Consciousness*

Seth (2023) *Nautilus*

even AI that *seems* conscious could be very bad

conscious-seeming AI

The Washington Post
Democracy Dies in Darkness

TECHNOLOGY

The Google engineer who thinks the company's AI has come to life

AI ethicists warned Google not to impersonate humans. Now one of Google's own thinks there's a ghost in the machine.

By Nitasha Tiku

June 11, 2022 at 8:00 a.m. EDT

Google engineer Blake Lemoine. (Martin Klimek/For the Washington Post)

conscious-seeming AI

- potential for cognitively impenetrable illusions of consciousness

Seth (2024) *PsyArXiv*

Seth (2023) *Nautilus*

conscious-seeming AI

- intentional stance / misprediction
- social coercion
- brutalism (the West World problem)

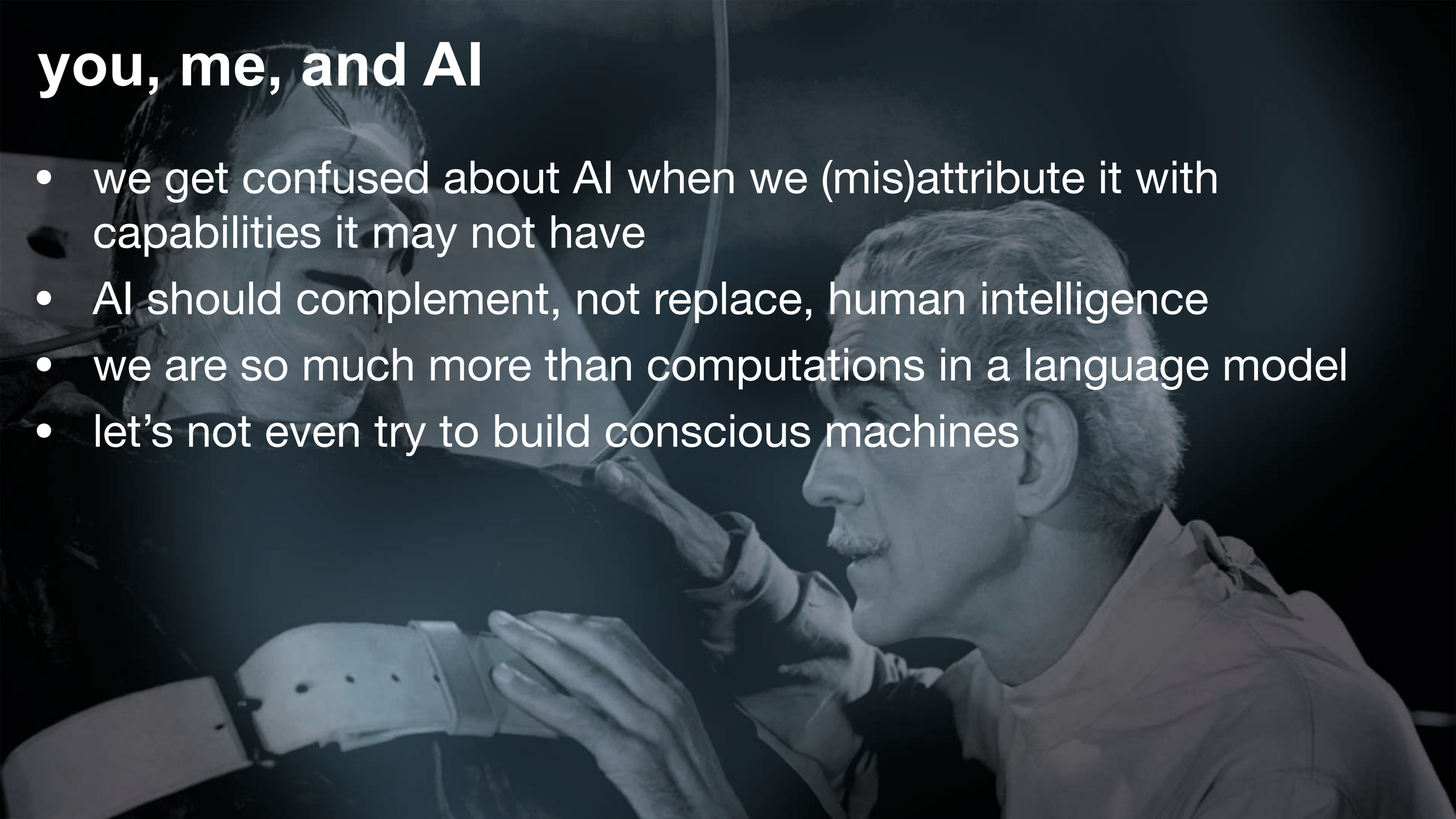
Seth (2024) *PsyArXiv*

Seth (2023) *Nautilus*

the intelligence we need

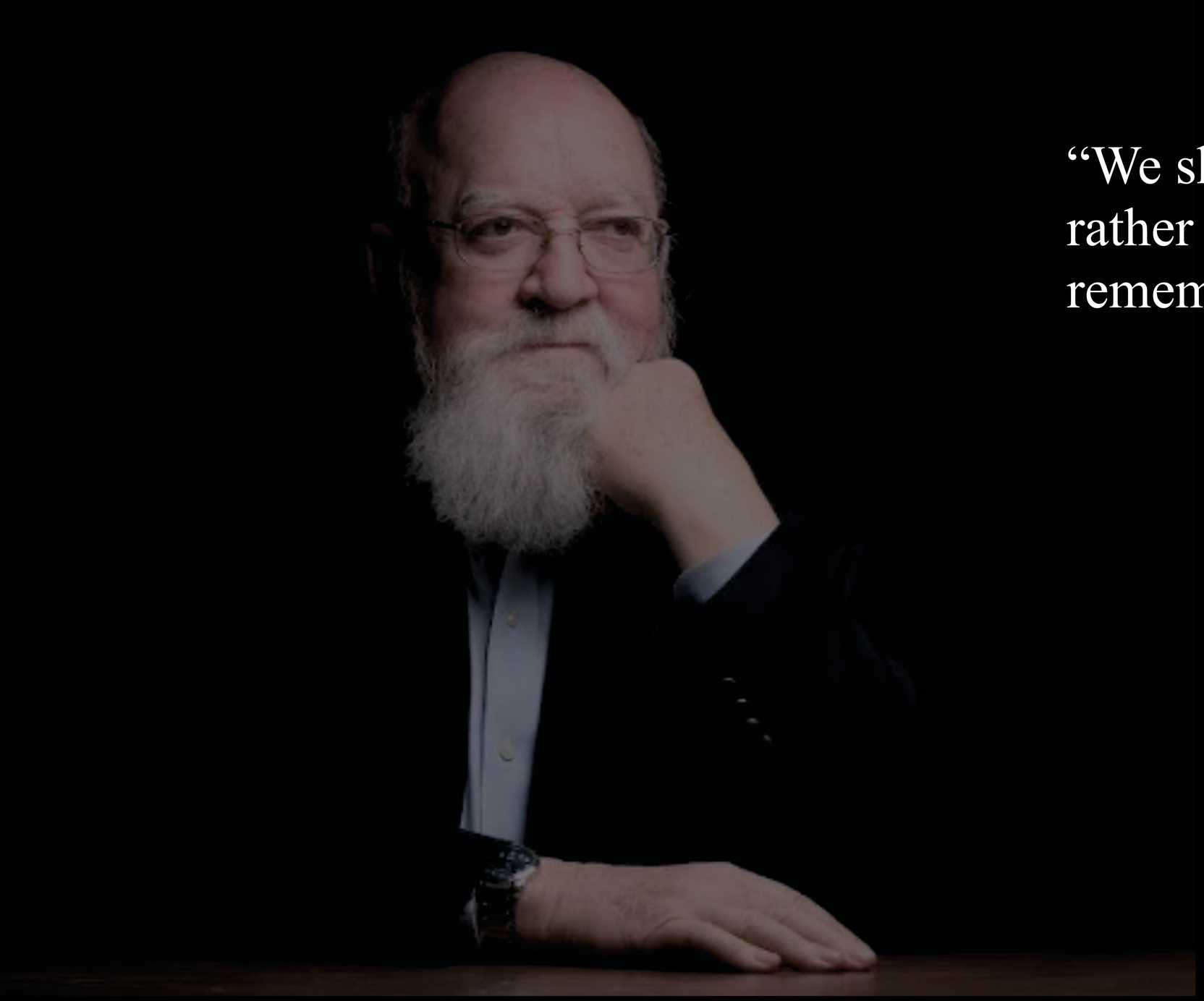
you, me, and AI

- we get confused about AI when we (mis)attribute it with capabilities it may not have
- AI should complement, not replace, human intelligence
- we are so much more than computations in a language model
- let’s not even try to build conscious machines



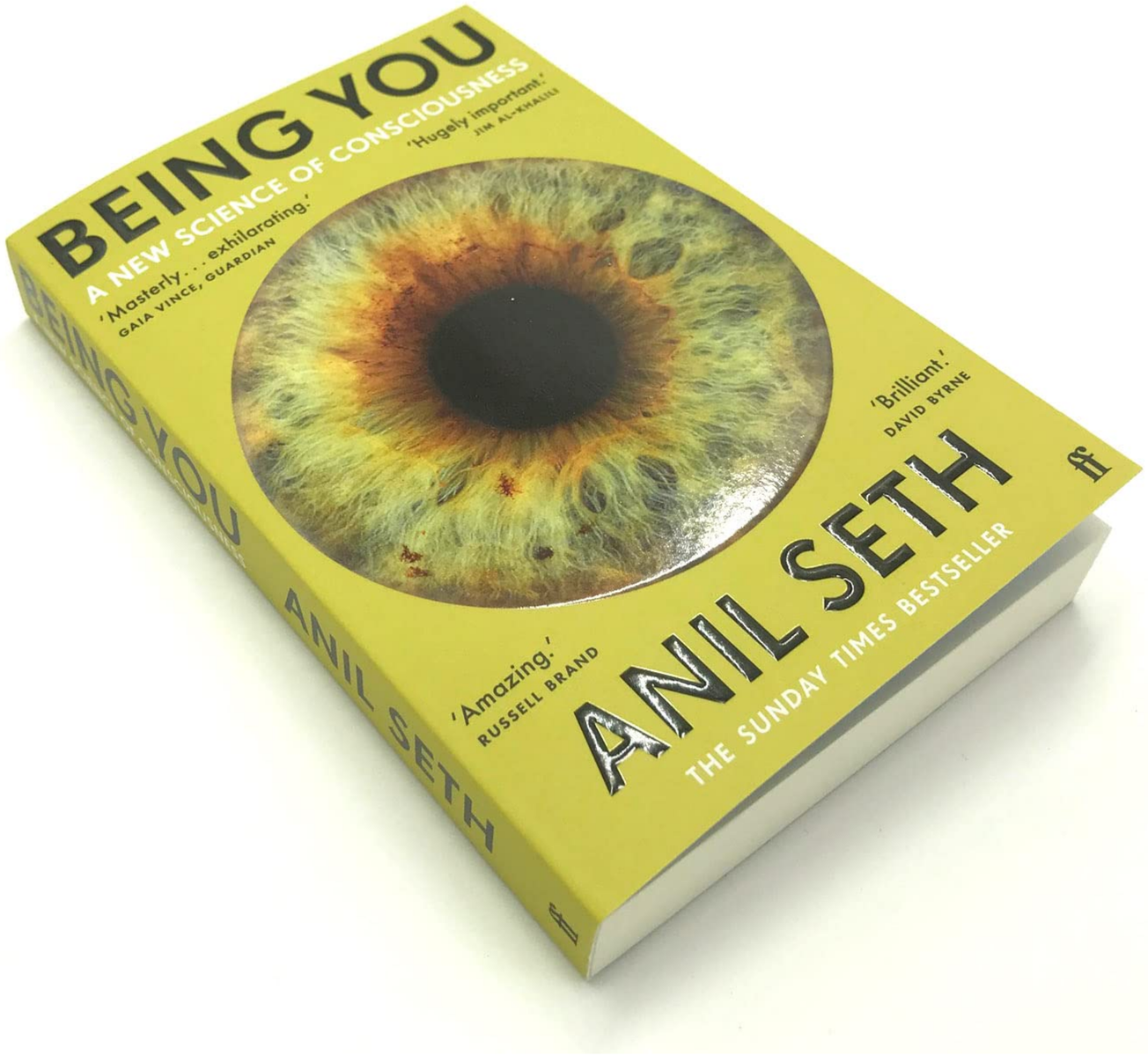
summary

- human intelligence is embodied, evolved, bounded, and continuous with perception
- don’t confuse the (certain) near-term challenges/ opportunities with the (uncertain) long-term
- understand our psychological biases
- AI can be tools - even partners - but they should not replace us



“We should treat AI as tools rather than colleagues, and always remember the difference.”

Daniel Dennett
1942-2024



The *Sunday Times* Top 10 Bestseller

A *Financial Times* Book of the Year

A *Guardian* Science Book of the Year

An *Economist* Book of the Year

A *New Statesman* Book of the Year

A *Bloomberg* Book of the Year

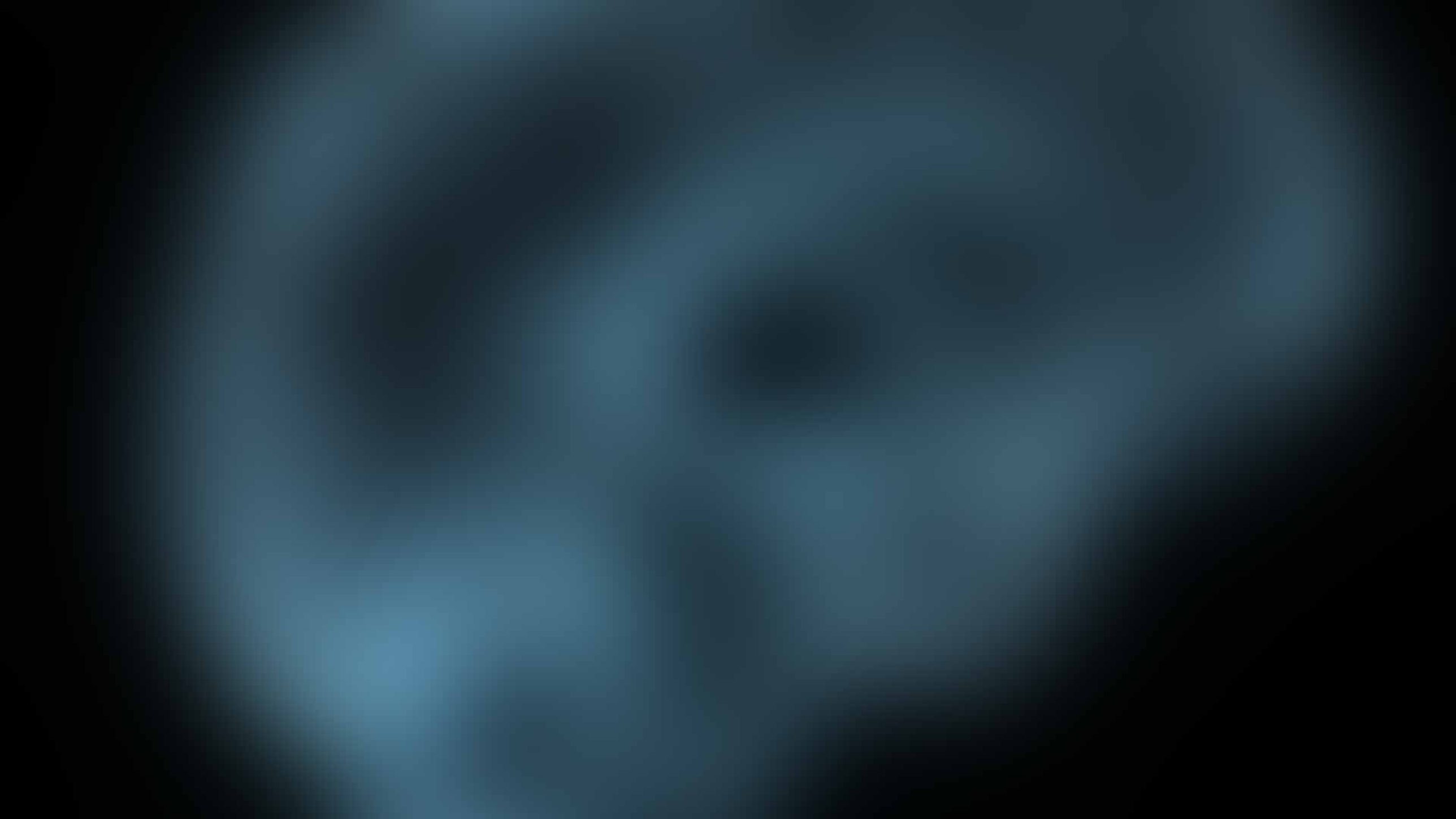
"An exhilarating book: A vast-ranging phenomenal achievement that will undoubtedly become a seminal text"

Gaia Vince, Guardian (Book of the Week)

"A brilliant beast of a book"

David Byrne

www.anilseth.com
@anilkseth



offcuts

“Technology can make us forget what we know about life.”
Sherry Turkle

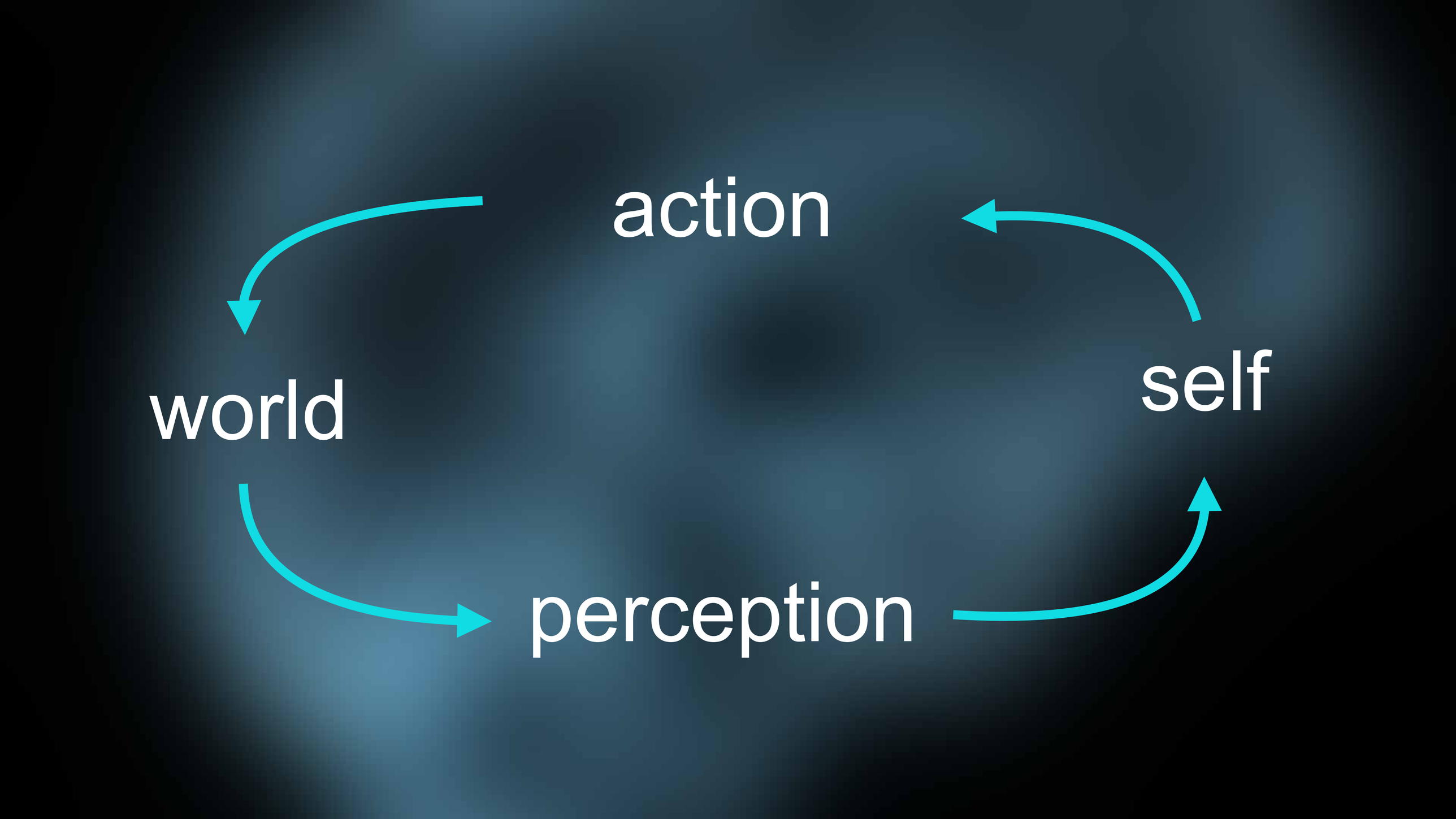
artificial intelligence

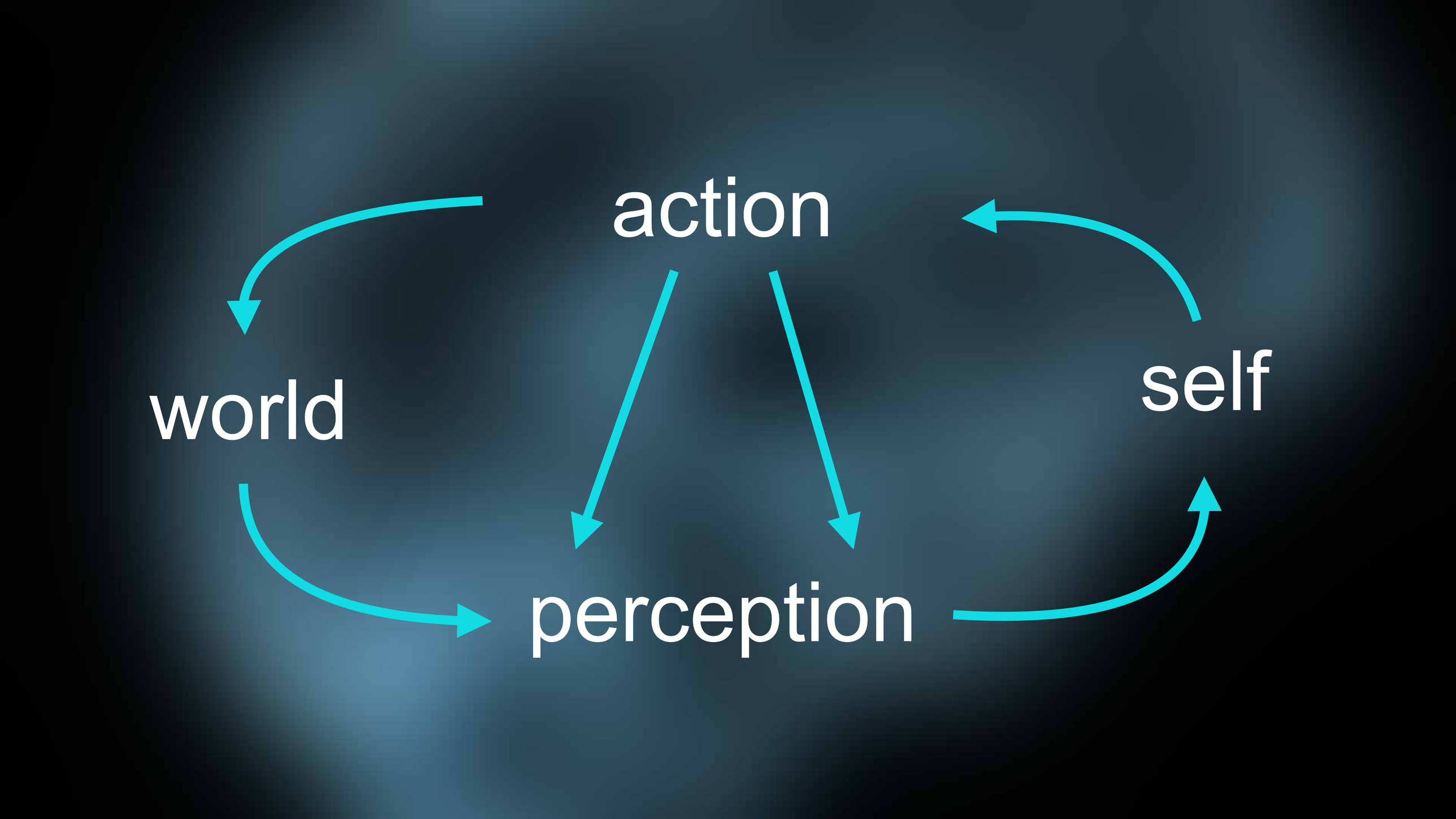
Ars Electronica, 2024

cerebral organoids

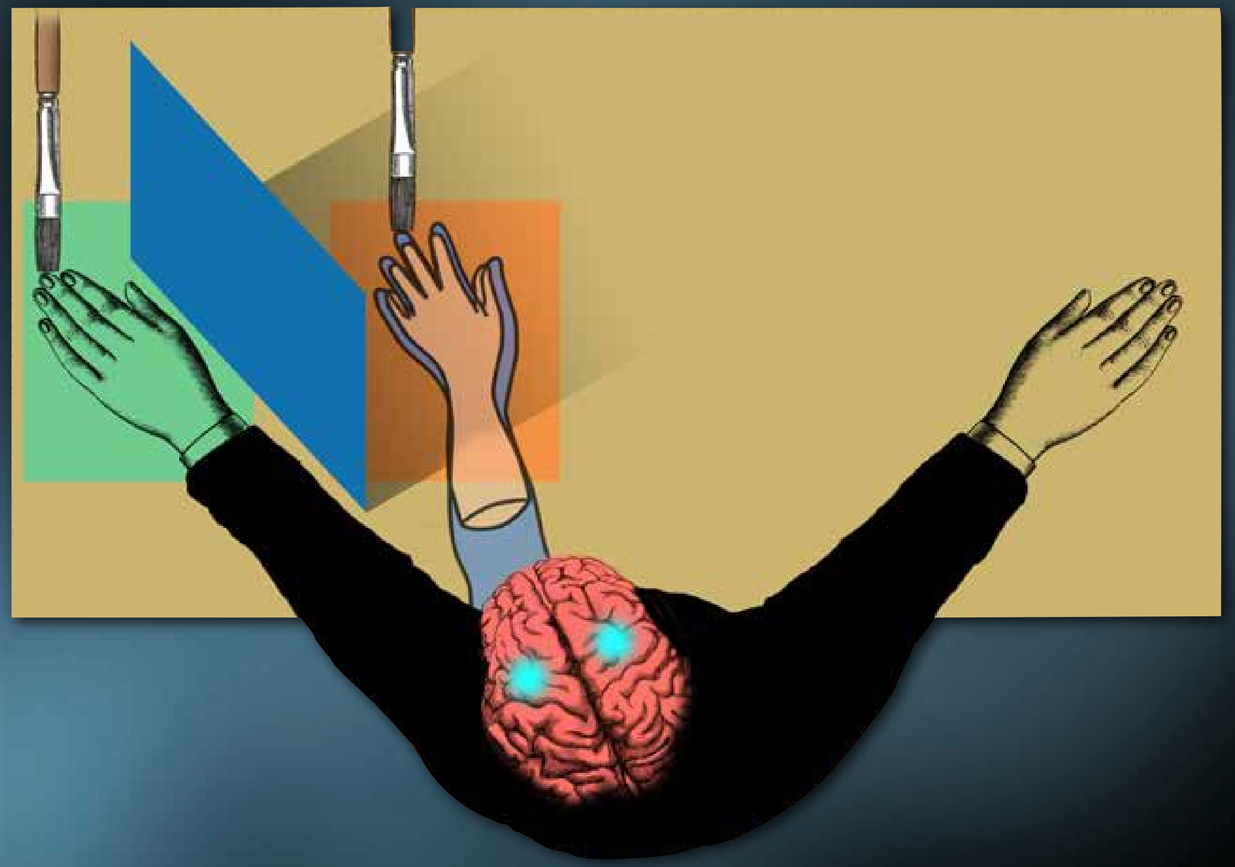
Trujillo et al (2019) *Cell*

Kagan et al (2022) *Neuron*


action
world
self
perception


action
world
self
perception

rubber hand illusion



“We do not see things as they are. We see them as we are.”

Anaïs Nin (1961)

Honorary Mention: EU Prize for Citizen Science, 2024

https://perceptioncensus.dreamachine.world/

the ‘AI’ brand

artificial intelligence

machine learning

applied statistics