



AI Futures: Embrace IA and AI ! BUT HALT AGI ?

- Human intelligence is characterised by adaptability, creativity, and the ability to learn from experiences. **Humans can think abstractly, solve problems, and understand complex concepts.** They can also navigate social situations and use their experiences to make informed decisions.



Machine intelligence relies on the processing power of computers and complex algorithms. AI systems excel at tasks involving **pattern recognition, data analysis, and optimization** but lack the intuition and creativity of human intelligence.





@tsarnick

AI: Computer systems that turn
information and data into

KNOWLEDGE

(Demis Hassabis, CEO, DeepMind)

at almost zero marginal cost.

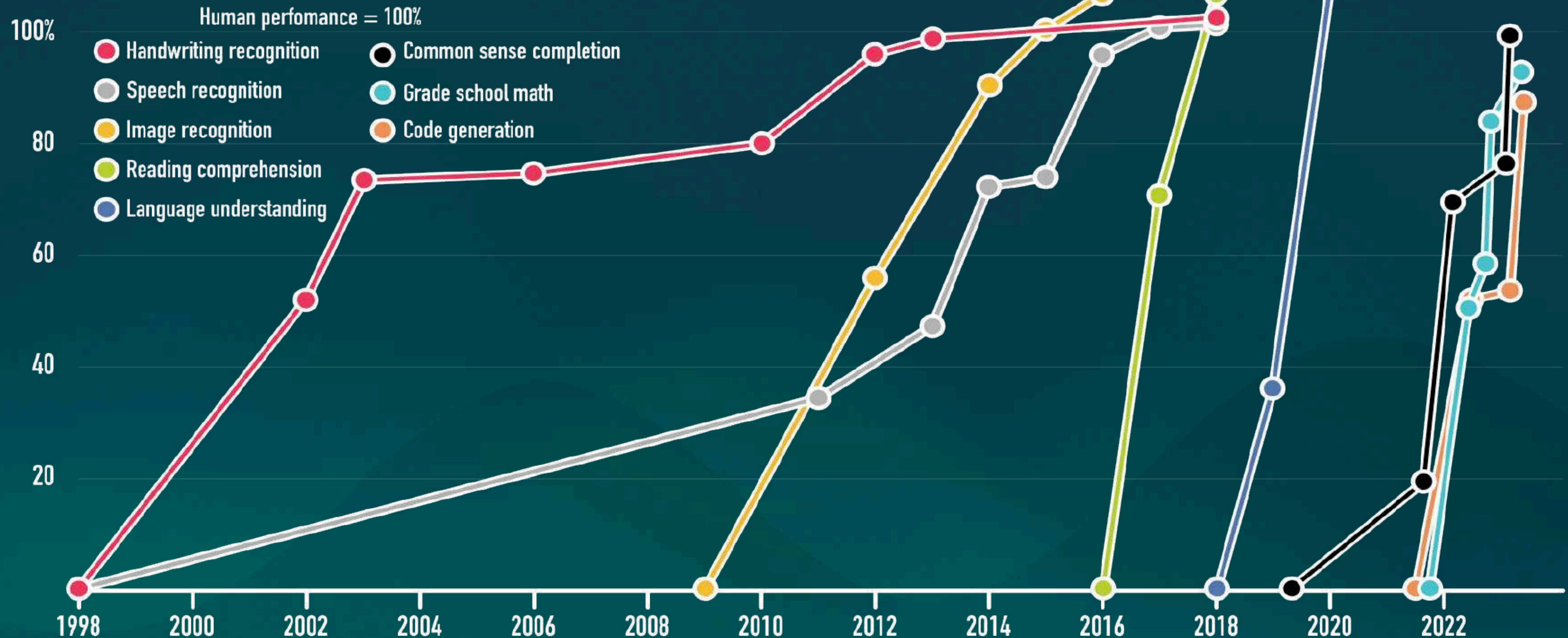


Artificial General Intelligence / AGI:
Autonomous systems that surpasses
human capabilities in the majority of
economically valuable tasks. (OpenAI)



Reminder: Machine Intelligence \neq Human Intelligence

STATE-OF-THE-ART AI PERFORMANCE RELATIVE TO HUMAN PERFORMANCE

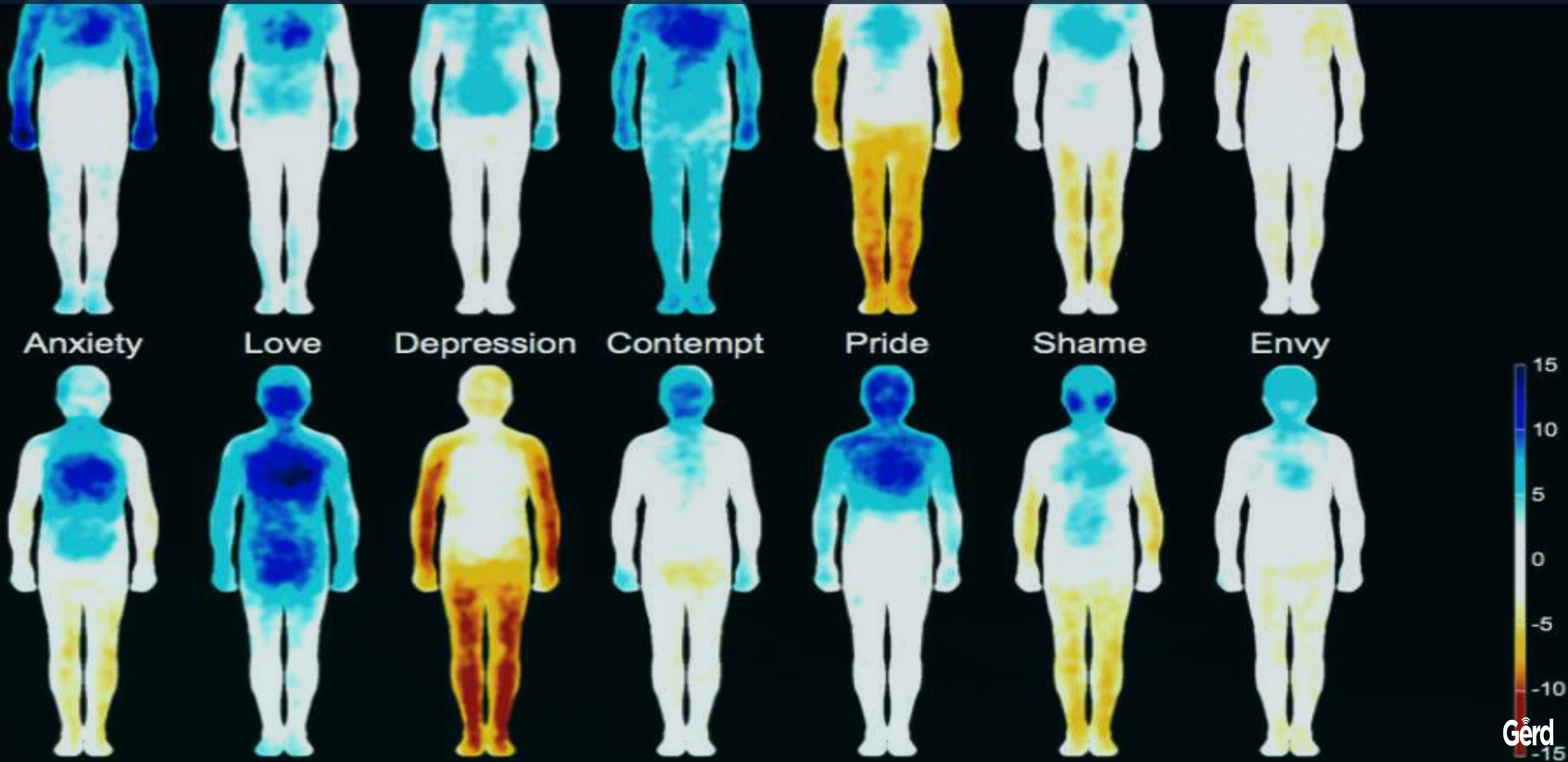


The difficulty of comparing intelligences (Yann LeCun)

A medium shot of Yann LeCun, a man with dark hair and glasses, wearing a blue button-down shirt. He is seated and speaking into a professional microphone on a boom arm. The background is softly lit with a lamp. In the foreground, a water bottle and a glass are visible on a table.

Intelligence is not a linear thing that you can measure with a scalar, with a single number.

We think with the body, not just the brain





On AI and Intelligence

(Joshua Bengio, University of Montreal)
NOT HIS ACTUAL VOICE

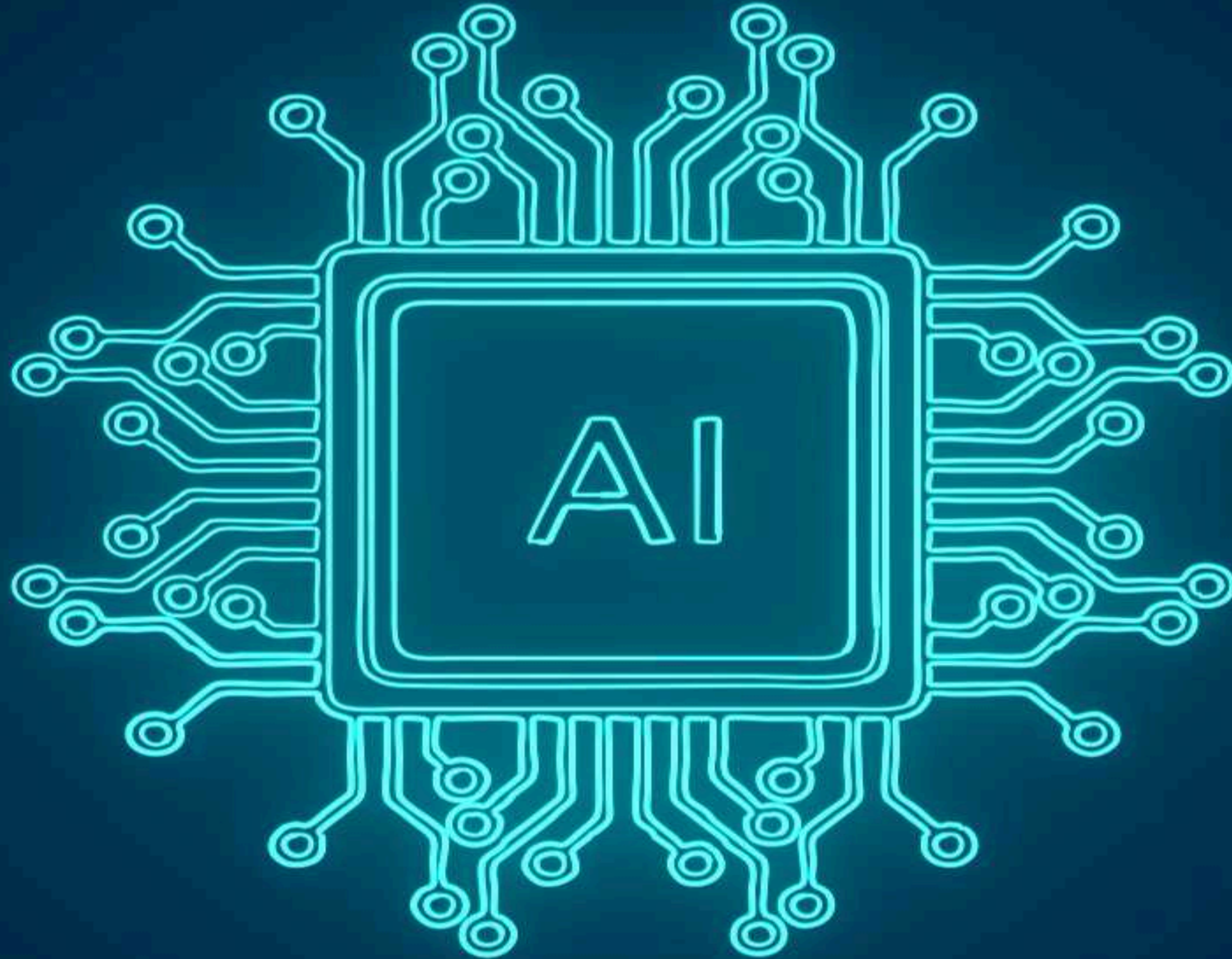
I find statements like “AIs cannot have true intelligence” or “The AIs just predict the next word” unconvincing. I agree that if one defines “true” intelligence as “the way humans are intelligent”, AIs don’t have “true” intelligence — their way of processing information and reasoning is different from ours.

But in a conversation about potential catastrophic AI risks, this is a distraction.

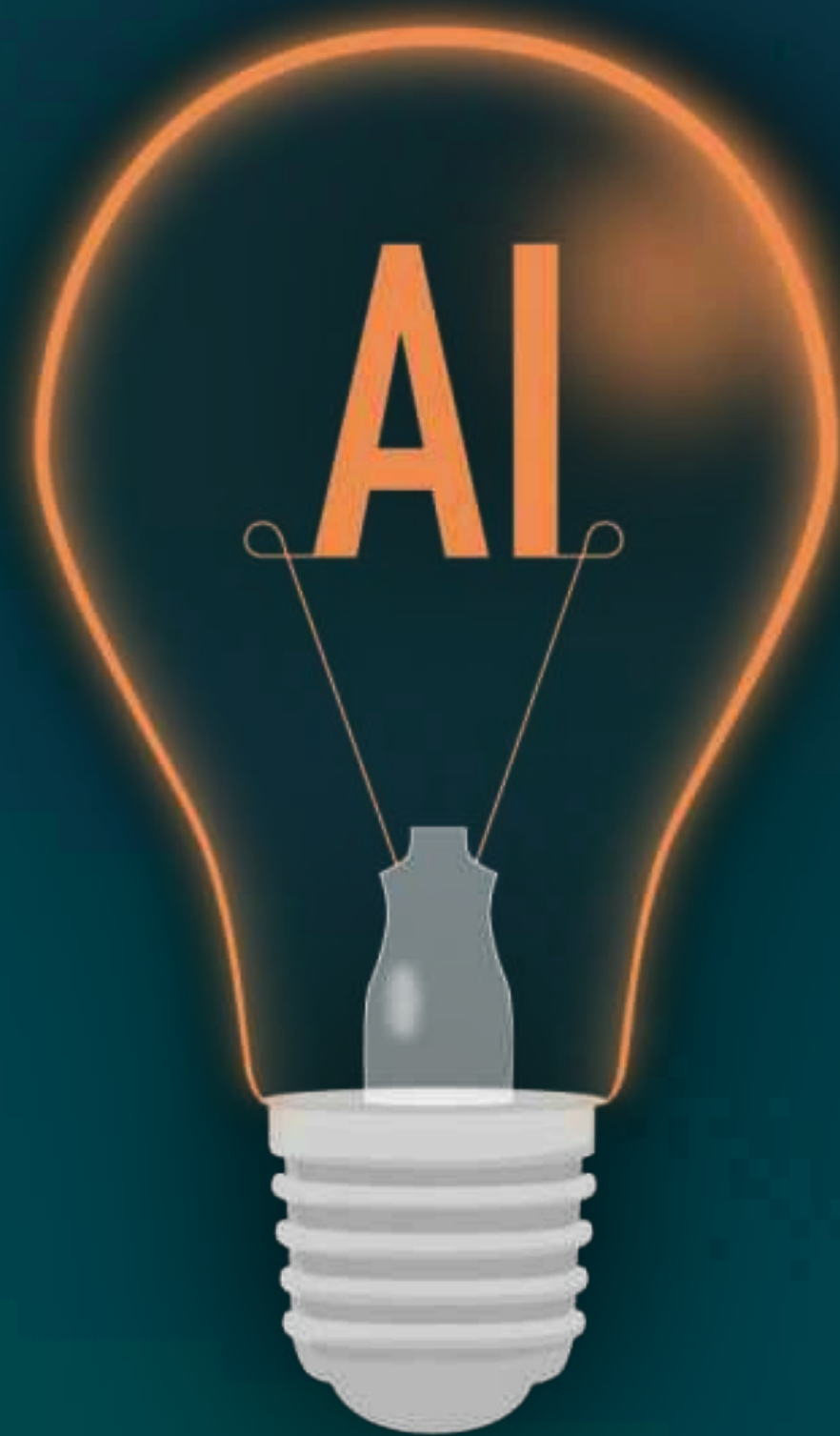
What matters for such a conversation is: What can the AI achieve? How good is it at problem-solving? That’s how I think of “AGI” and “ASI” — A LEVEL OF AI CAPABILITIES AT WHICH AN AI IS AS GOOD AS, OR BETTER THAN, A HUMAN EXPERT AT SOLVING BASICALLY ANY PROBLEM (excluding problems that require physical actions).

There is also the current level of AI ability, with a very high level of mastery of language and visual material, and more and more capabilities in a broader variety of cognitive tasks.

FINALLY, THERE IS NO SCIENTIFIC REASON TO BELIEVE THAT HUMANITY IS AT THE PINNACLE OF INTELLIGENCE. IN FACT, IN MANY SPECIALIZED COGNITIVE TASKS, COMPUTERS ALREADY SURPASS HUMANS.



Stephen Hawking argued that super-intelligence is physically possible because "there is no physical law precluding particles from being organised in ways that perform even more advanced computations than the arrangements of particles in human brains"



First stage of AI: **Classify** data and information

Second stage of AI: **Generating** content & data

Third stage of AI: Limited **Reasoning**

Fourth stage of AI: **Recursive** Capabilities and Intelligence Explosion (SuperIntelligence)

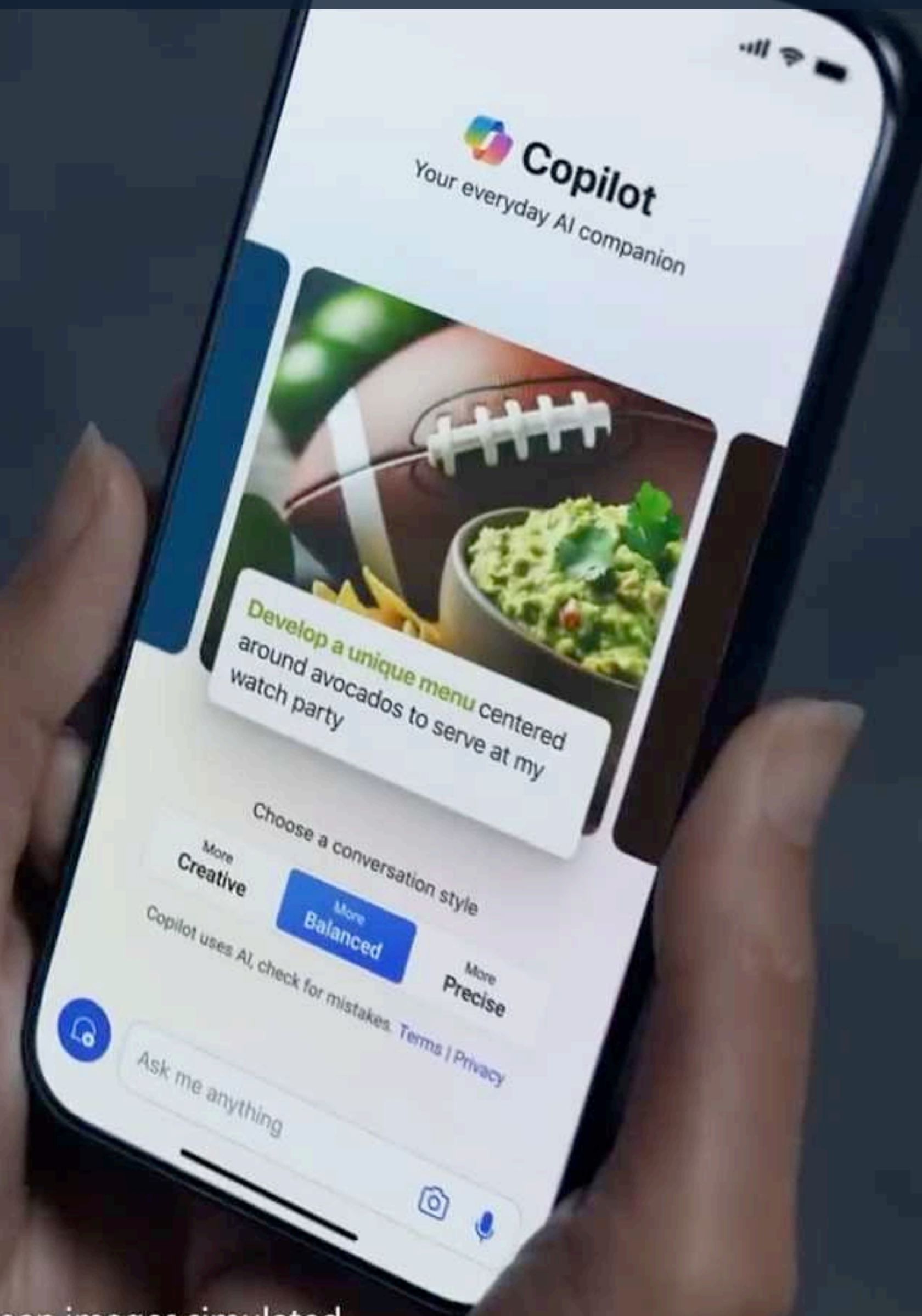
Intelligent Assistance – and a bit of ‘broader AI’



Meet CoCounsel



Intelligent Assistance, dressed up nicely by the marketing people



Screen images simulated.

MUSTAFA SULEYMAN

**SQUAWK
BOX**

"THE COMING WAVE"

AI PIONEER ON TECH REVOLUTION

research assistant, but it'll also be a coach



Gerd

OpenAI has internally shared definitions for five levels of artificial general intelligence (AGI), according to Bloomberg. An OpenAI document Bloomberg reproduced defines the levels:

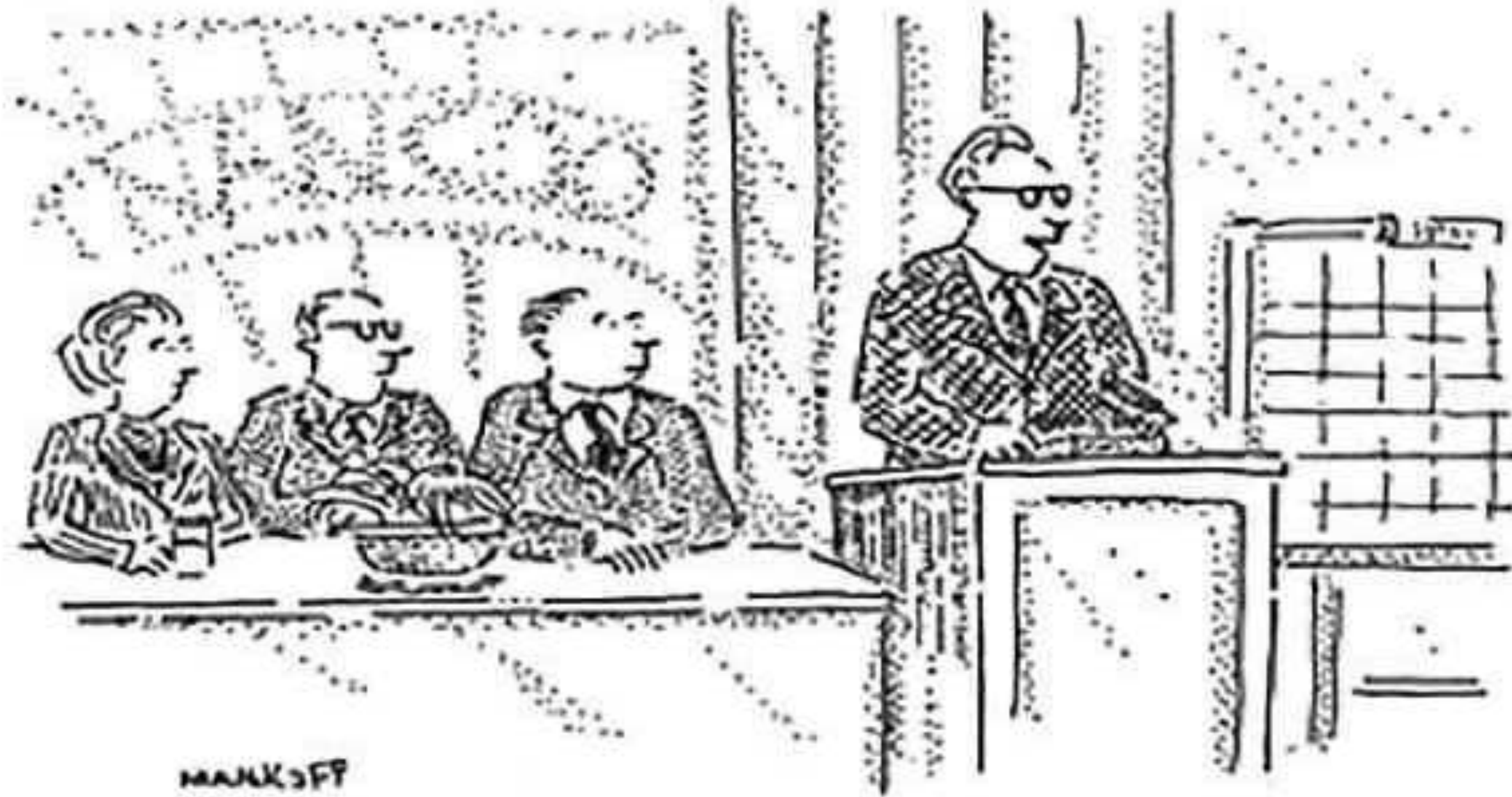
1. **CHATBOTS**: AI with conversational language skills
2. **REASONERS**: human-level problem-solving abilities
3. **AGENTS**: systems that can take actions
4. **INNOVATORS**: AI that can aid in invention
5. **ORGANISATIONS**: AI that can do the work of an organisation



yeah, if you.

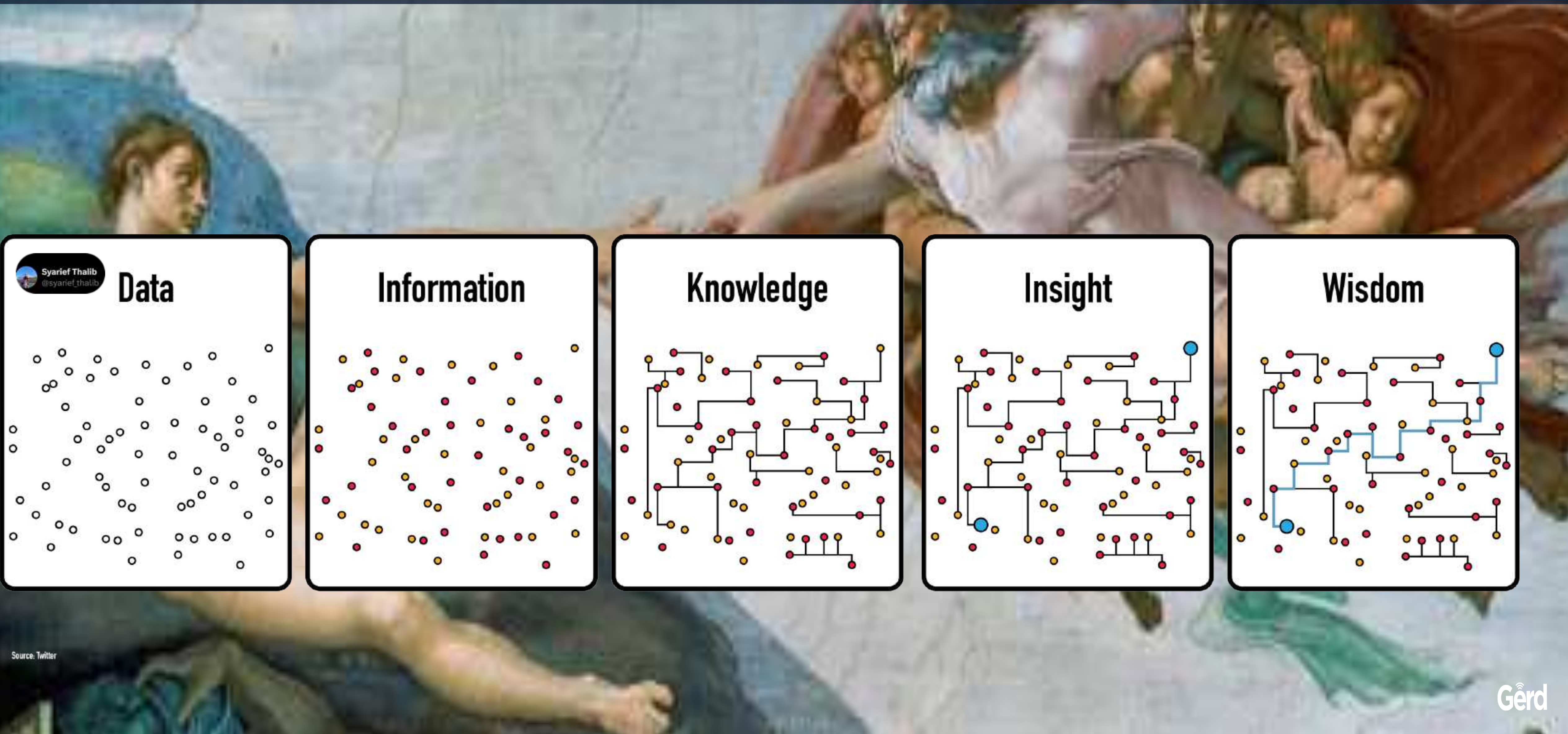
the Rise of Artificial Intelligence

OpenAI



"And so, while the end-of-the-world scenario will be rife with unimaginable horrors, we believe that the pre-end period will be filled with unprecedented opportunities for profit."

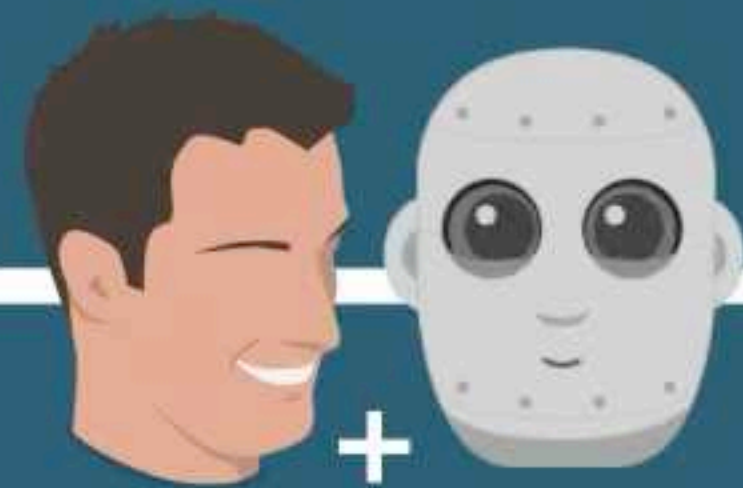
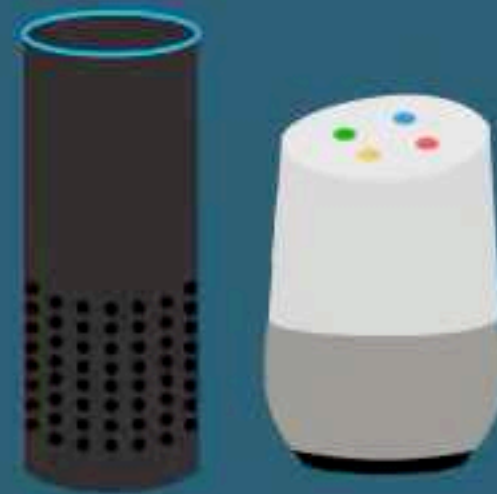
Sam Altman: “The coming change will center around the most impressive of human capabilities: the phenomenal ability to think, create, understand, and reason”



“Intelligence means having the power to

(Stuart Russell, UC Berkeley)

ASSISTED INTELLIGENCE



AUGMENTED INTELLIGENCE



shape the world in your interest”

AUTOMATION



AUTONOMOUS INTELLIGENCE



“Intelligence means having the power to

(Stuart Russell, UC Berkeley)

ASSISTED INTELLIGENCE

OpenAI’s ‘Strawberry’ AI model aims for advanced reasoning – reports



Suswati Basu **Tech Journalist**

Fact Checked by Sam Shedden

Last updated: 15 July, 2024

Disclosure

IW Why Trust Us



shape the world in your interest”

AUTOMATION



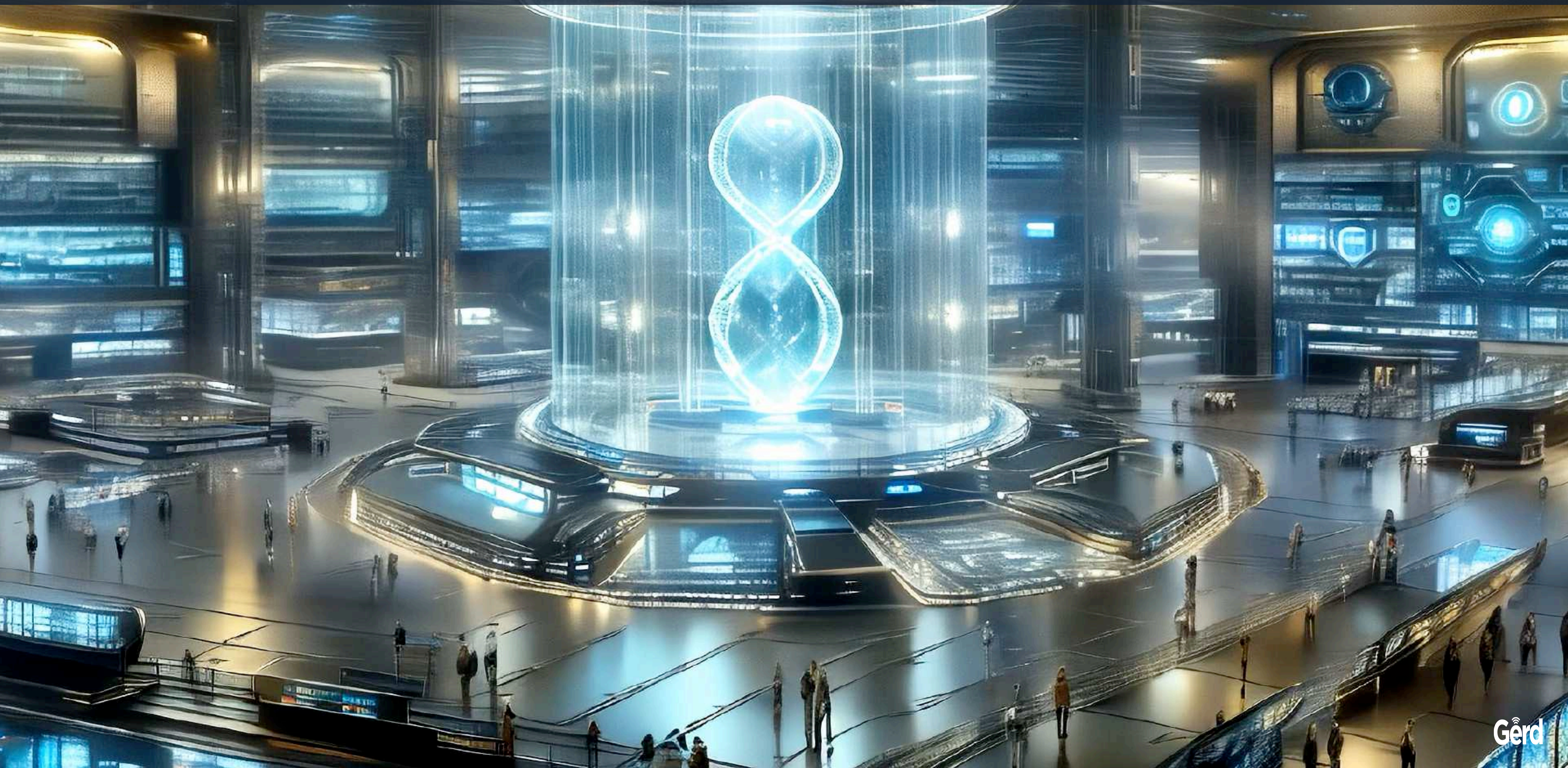
AUTONOMOUS INTELLIGENCE



Gerd

Nick Bostrom defines a SuperIntelligence as **"any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest"** including scientific creativity, strategic planning, and social skills. He argues that a superintelligence can outmaneuver humans anytime its goals conflict with humans.

Would **CONTAINMENT OF AGI** actually be feasible?



The answer is, we
haven't the faintest **idea.**



Should we really value Intelligence over Humanity?

Stuart Russel
Computer Scientist at Berkeley
Director for the Center of
Human-Compatible AI

DIONZ

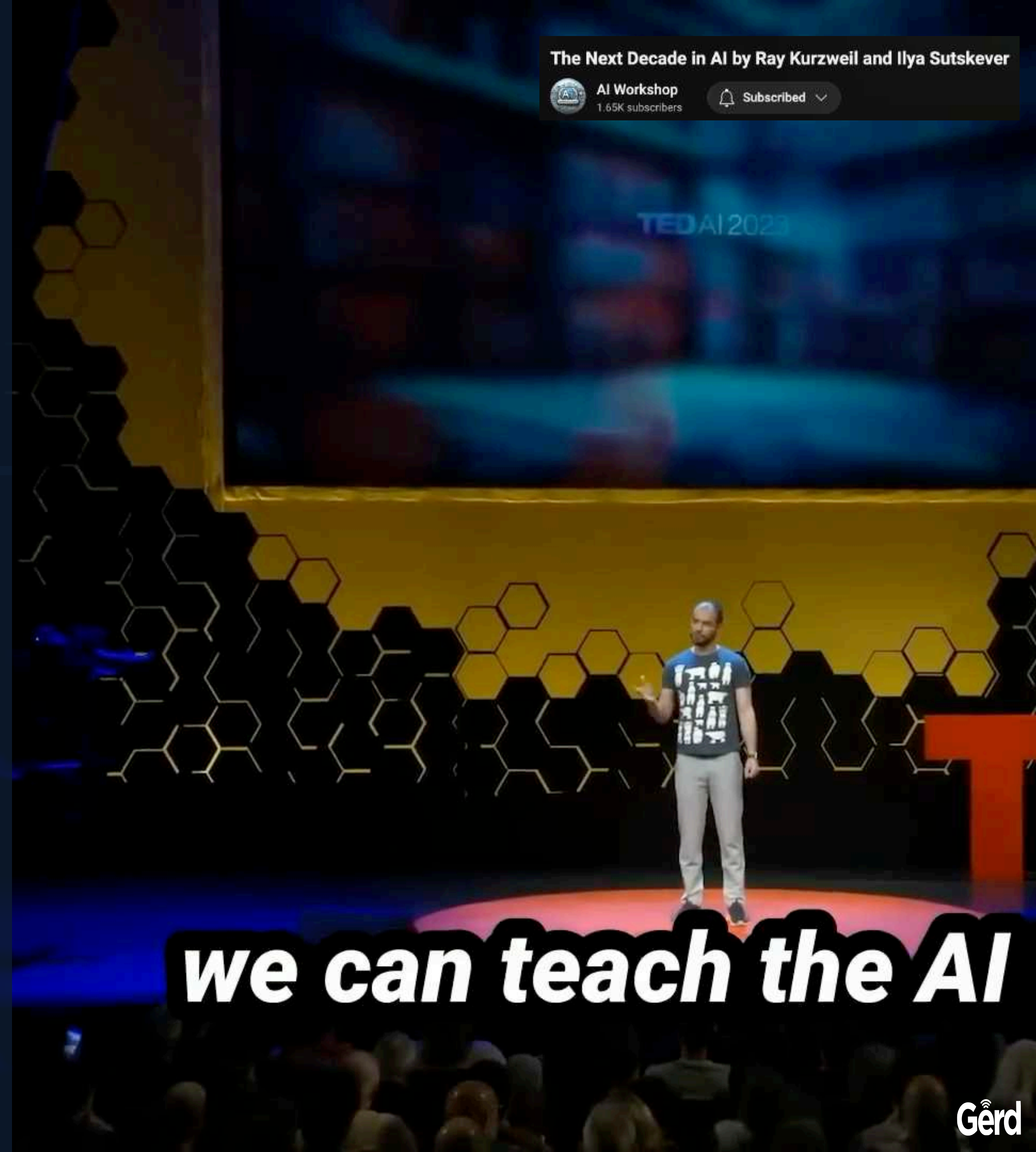


**“We tend to overestimate the effect of a technology in the short run
and underestimate the effect in the long run” Amara’s Law**



Philosopher David Chalmers argues
that the human brain is a mechanical
system, and therefore ought to be
emulatable by synthetic materials.

5 YEARS
TO AGI?



we can teach the AI



5 YEARS TO AGI?

nothing but digital brains

PRICE-PERFORMANCE OF COMPUTATION, 1939-2023

Best achieved price-performance
in computations per second per
constant 2023 dollar



The Next Decade in AI by Ray Kurzweil and Ilya Sutskever



AI Workshop
1.65K subscribers



Subscribed

5 YEARS TO AGI?

about a trillion calculations



What is artificial superintelligence?



Artificial superintelligence (ASI) is a hypothetical software-based [artificial intelligence](#) (AI) system with an intellectual scope beyond human intelligence. At the most fundamental level, this superintelligent AI has cutting-edge cognitive functions and highly developed thinking skills more advanced than any human.

Geoffrey Hinton Computer Scientist "Godfather of AI"

The worst case scenario that
you think is conceivable?

Top AI Scientists on AI Catastrophe



Roman Yampolskiy
366 subscribers

Subscribed

 **EmTech
Digital**

They will be able to
manipulate people...

Top AI Scientists on AI Catastrophe



Roman Yampolskiy
366 subscribers

Subscribed

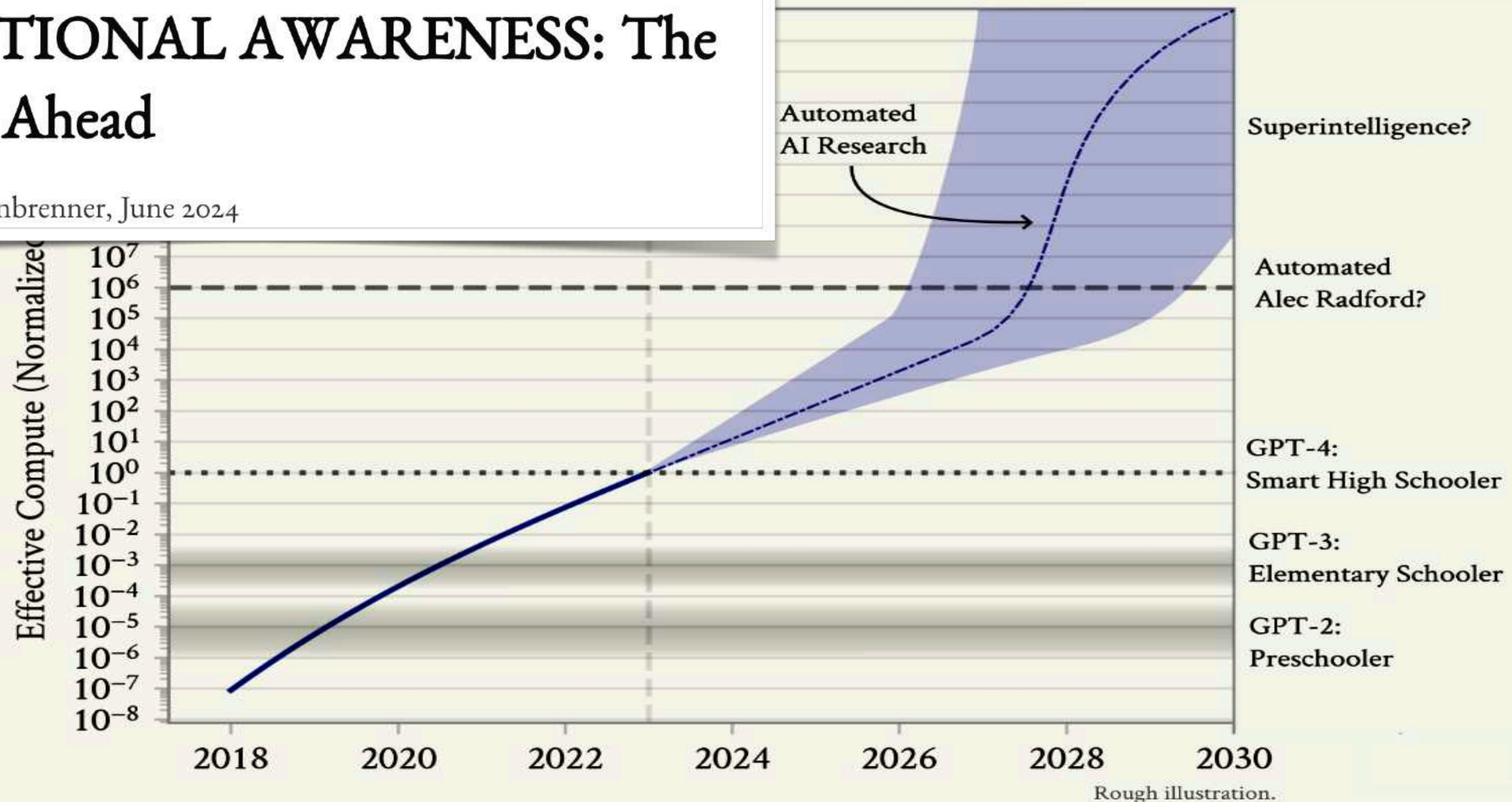
Worst case: intelligent entities that can modify and replicate themselves
We can expect AI that is supremely good at manipulating people
(thereby also very good at manufacturing synthetic realities)

We may soon have 'things more intelligent' than us, for the first time ever

Geoffrey Hinton

SITUATIONAL AWARENESS: The Decade Ahead

Leopold Aschenbrenner, June 2024



SITUATIONAL AWARENESS | Leopold Aschenbrenner

Automated AI research could accelerate algorithmic progress, leading to 5+ OOMs of effective compute gains in a year. The AI systems we'd have by the end of an intelligence explosion would be vastly smarter than humans.

Comparison with humans

Bostrom argues that AI has many advantages over the [human brain](#):

- Speed of computation: biological [neurons](#) operate at a maximum frequency of around 200 [Hz](#), compared to potentially multiple GHz for computers.
- Internal communication speed: [axons](#) transmit signals at up to 120 m/s, while computers transmit signals at the [speed of electricity](#), or optically at the [speed of light](#).
- Scalability: human intelligence is limited by the size and structure of the brain, and by the efficiency of social communication, while AI may be able to scale by simply adding more hardware.
- Memory: notably [working memory](#), because in humans it is limited to a few [chunks](#) of information at a time.
- Reliability: transistors are more reliable than biological neurons, enabling higher precision and requiring less redundancy.
- Duplicability: unlike human brains, AI software and models can be easily [copied](#).
- Editability: the parameters and internal workings of an AI model can easily be modified, unlike the connections in a human brain.
- Memory sharing and learning: AIs may be able to learn from the experiences of other AIs in a manner more efficient than human learning.

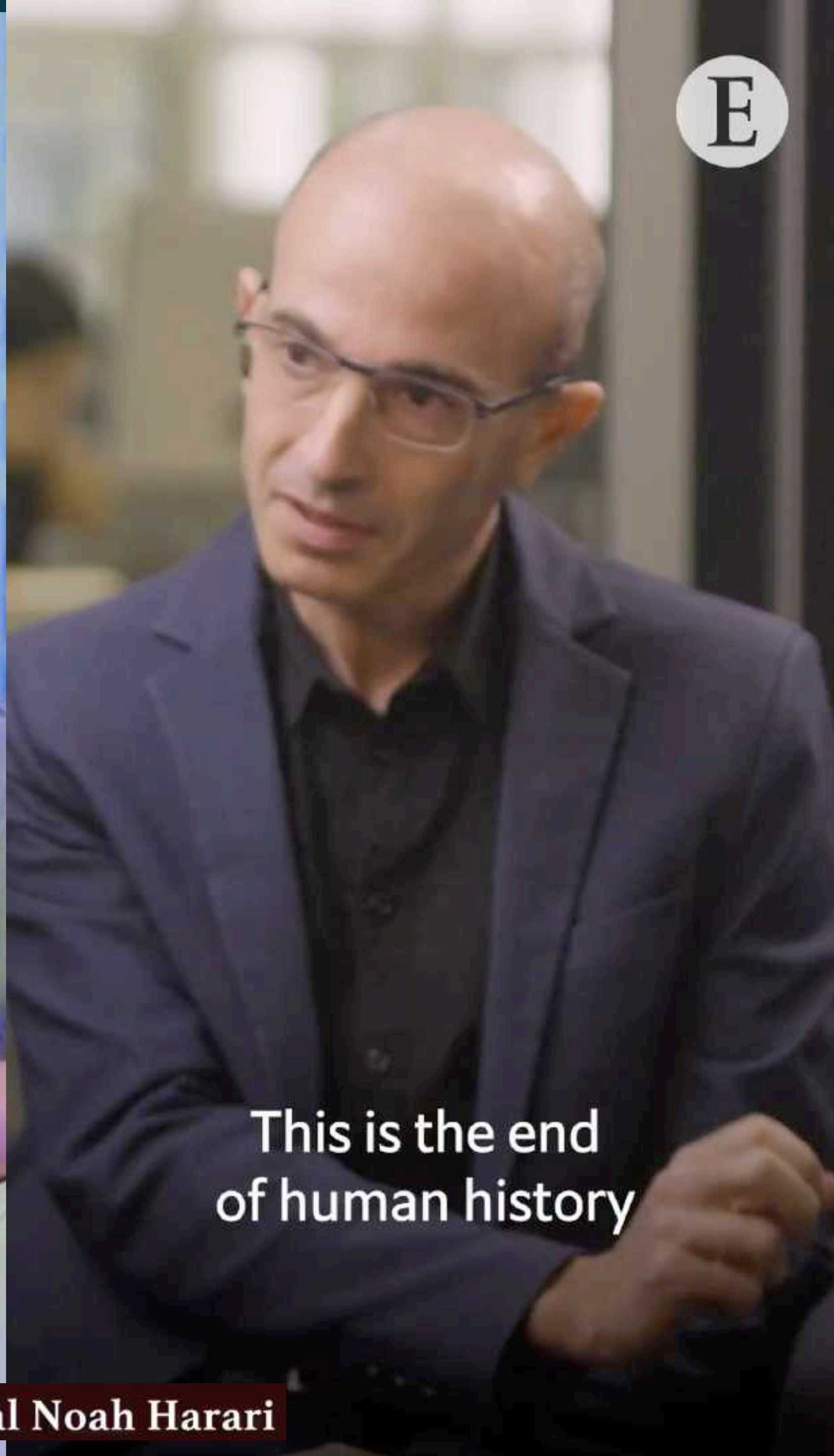


Nick Bostrom on the Potential of Super Intelligence: Machine Intelligence Limits are far beyond those of Humans



median human

Yuval Noah Harari



This is the end
of human history

ARTIFICIAL GENERAL INTELLIGENCE:

An autonomous system
that surpasses human
capabilities in the
majority of economically
valuable tasks. (Wikipedia)



Elon Musk on the Dangers of Artificial Intelligence (Ant Hill Analogy)

T

Tommy T
115 subscribers

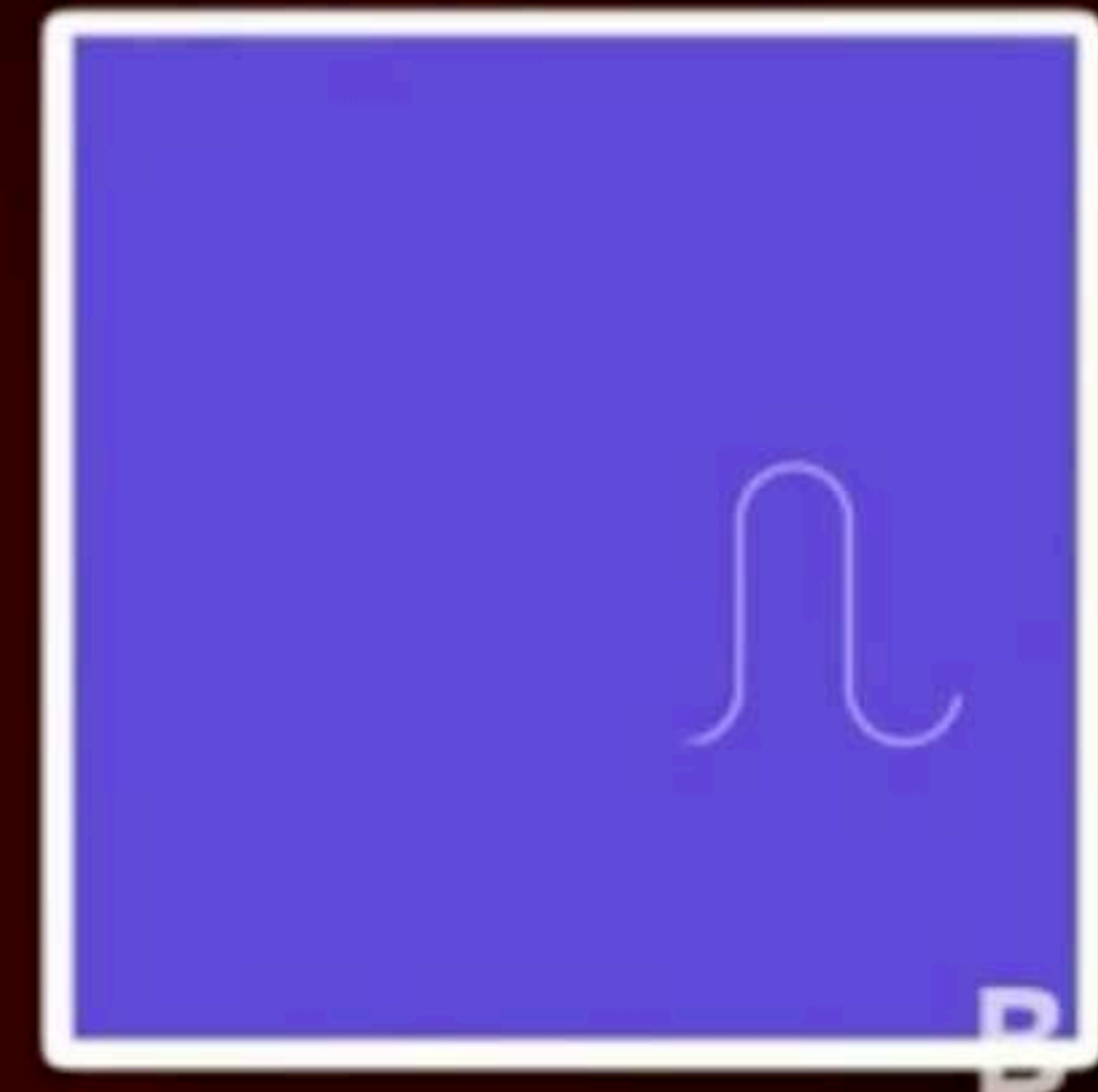
Subscribe

Roman Yampolskiy
Computer Scientist &
AI Researcher

“No second chance”

Eliezer Yudkowsky
AI Researcher

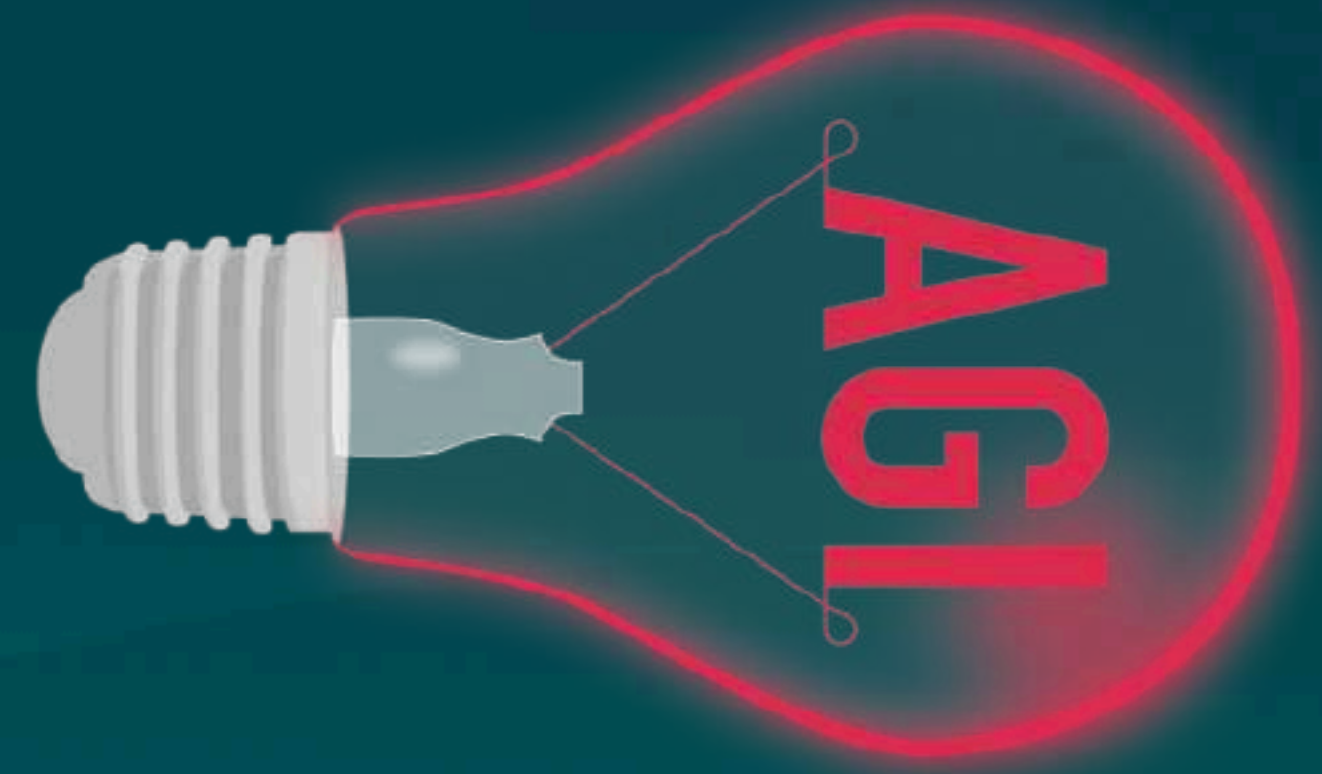
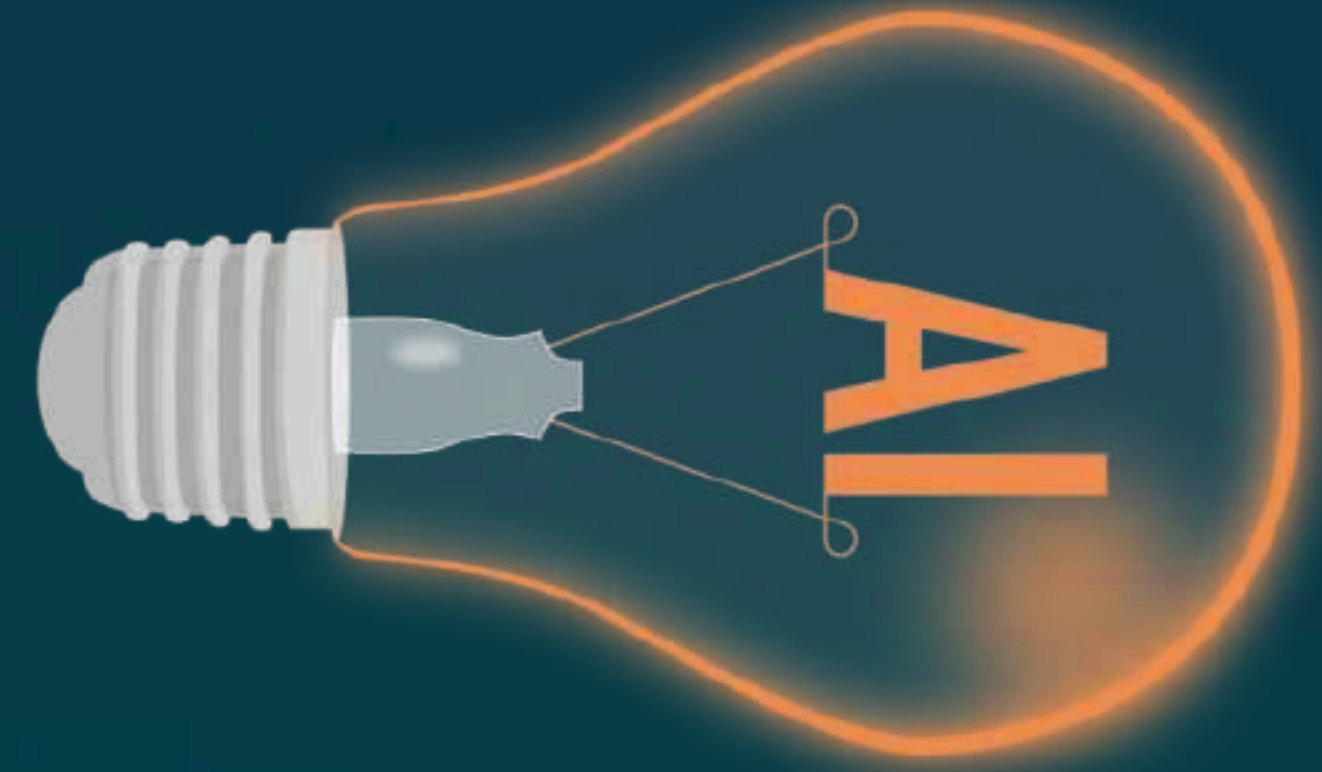
“You don’t get to
TRY AGAIN”





1.The Control and Alignment Problem

2.The Collaboration and Governance Problem



Good-level AI

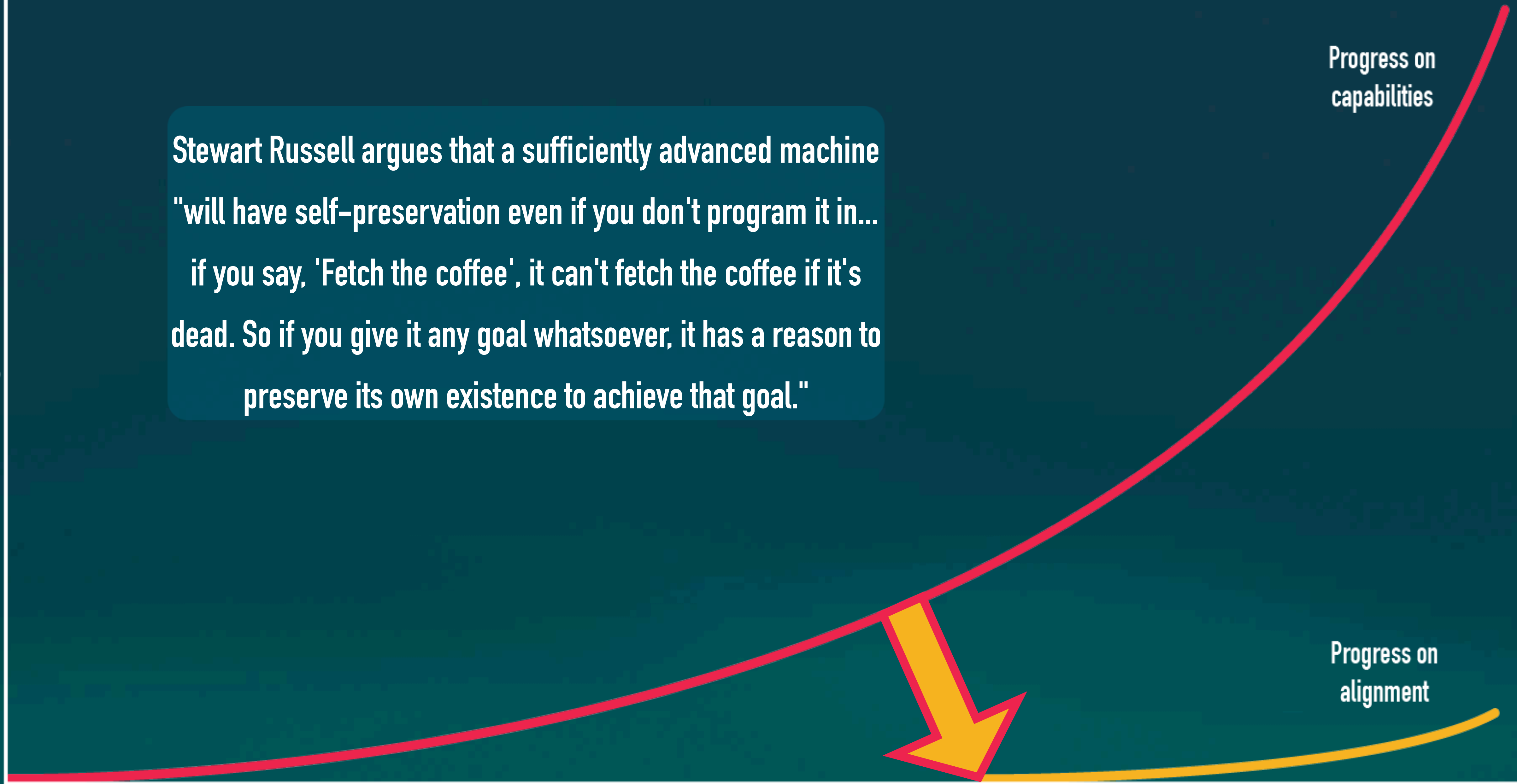
Stewart Russell argues that a sufficiently advanced machine "will have self-preservation even if you don't program it in... if you say, 'Fetch the coffee', it can't fetch the coffee if it's dead. So if you give it any goal whatsoever, it has a reason to preserve its own existence to achieve that goal."

Progress on capabilities

Progress on alignment

Time

Progress



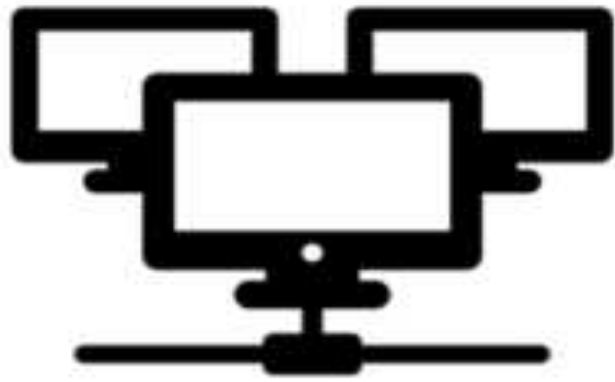
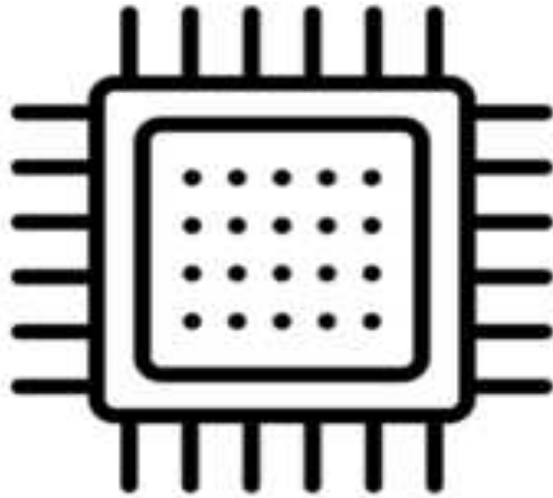
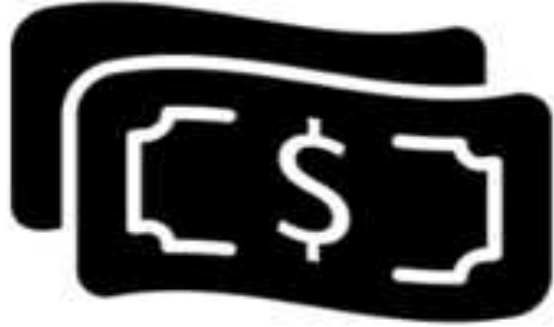





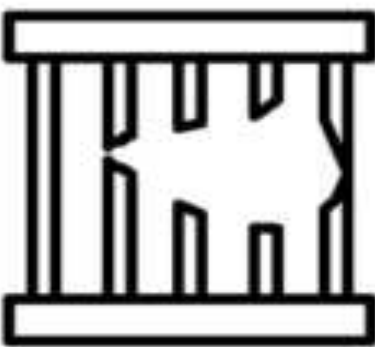


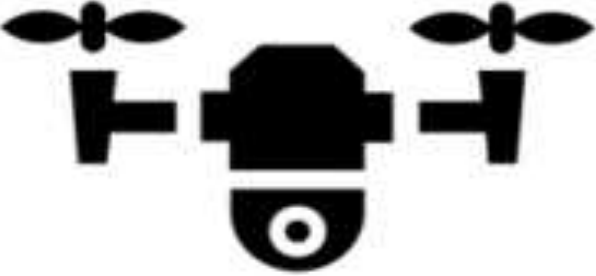




Progress on
capabilities

Progress on
alignment

Bostrom argues similarly that if an advanced AI's instrumental goals conflict with humanity's goals, the AI might harm humanity in order to acquire more resources or prevent itself from being shut down, but **only as a way to achieve its ultimate goal.**

 <p>Evading shutdown</p>	 <p>Hacking computer systems</p>	 <p>Run many AI copies</p>	 <p>Acquire computation</p>	 <p>Attract earnings and investment</p>	 <p>Hire or manipulate human assistants</p>	 <p>AI research and programming</p>
 <p>Persuasion and lobbying</p>	 <p>Hiding unwanted behavior</p>	 <p>Strategically appear aligned</p>	 <p>Escaping containment</p>	 <p>R&D</p>	 <p>Manufacturing and robotics</p>	 <p>Autonomous weaponry</p>

Fears about AI's existential risk are overdone, says a group of experts

Blaise Agüera y Arcas and his co-authors argue that tackling more immediate concerns will mitigate long-term threats

Regulators should not prioritise existential risk posed by superintelligent AI. Instead, they should address the problems which are in front of them, making models safer and their operations more predictable in line with human needs and norms. Regulations should focus on preventing inappropriate deployment of AI.



IMAGE: DAN WILLIAMS

“It is far from obvious whether AI, super-intelligent or not, is best thought of as an alien entity with its own agency or as part of the anthropogenic world, like any other technology that both shapes and is shaped by humans”



Bill Joy: 2000 in Wired

BILL JOY

BACKCHANNEL APR 1, 2000 12:00 PM

Why the Future Doesn't Need Us

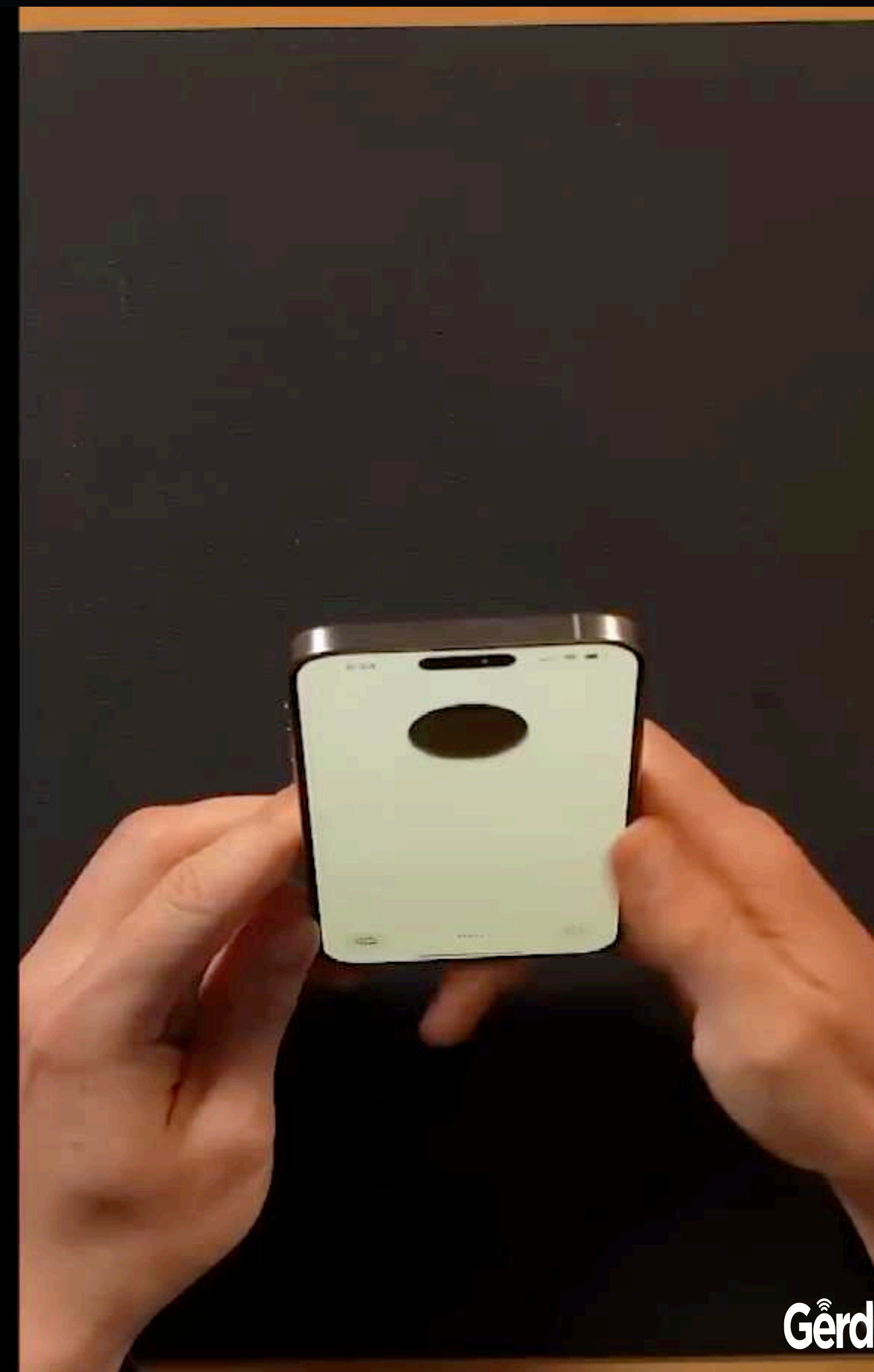
Our most powerful 21st-century technologies—robotics, genetic engineering, and nanotech

“The human race might easily permit itself to drift into a position of such dependence on the machines that it would have no practical choice but to accept all of the machines' decisions.

As society and the problems that face it become more and more complex and machines become more and more intelligent, people will let machines make more of their decisions for them, simply because machine-made decisions will bring better results than man-made ones.



How will we still know what's real or not?



The AGI Coordination Problem

(Joshua Bengio, University of Montreal)

“The politics of AI: If the coordination problem was solved perfectly, solving the AI alignment and control problem would not be an absolute necessity: We could “just” collectively apply the precautionary principle and avoid doing experiments anywhere with a non-trivial risk of constructing uncontrolled AGI”



“The dynamics of all these self-interests and cultural factors are currently leading us into a dangerous race towards greater AI capabilities without the methodology and institutions to sufficiently mitigate the greatest risks, such as catastrophic misuse and loss of control” (Bengio)



The Economic Context:

Create AGI because of... the \$\$\$\$\$\$?

How can we create a Future-Fit,
Humanly Sustainable Capitalism?

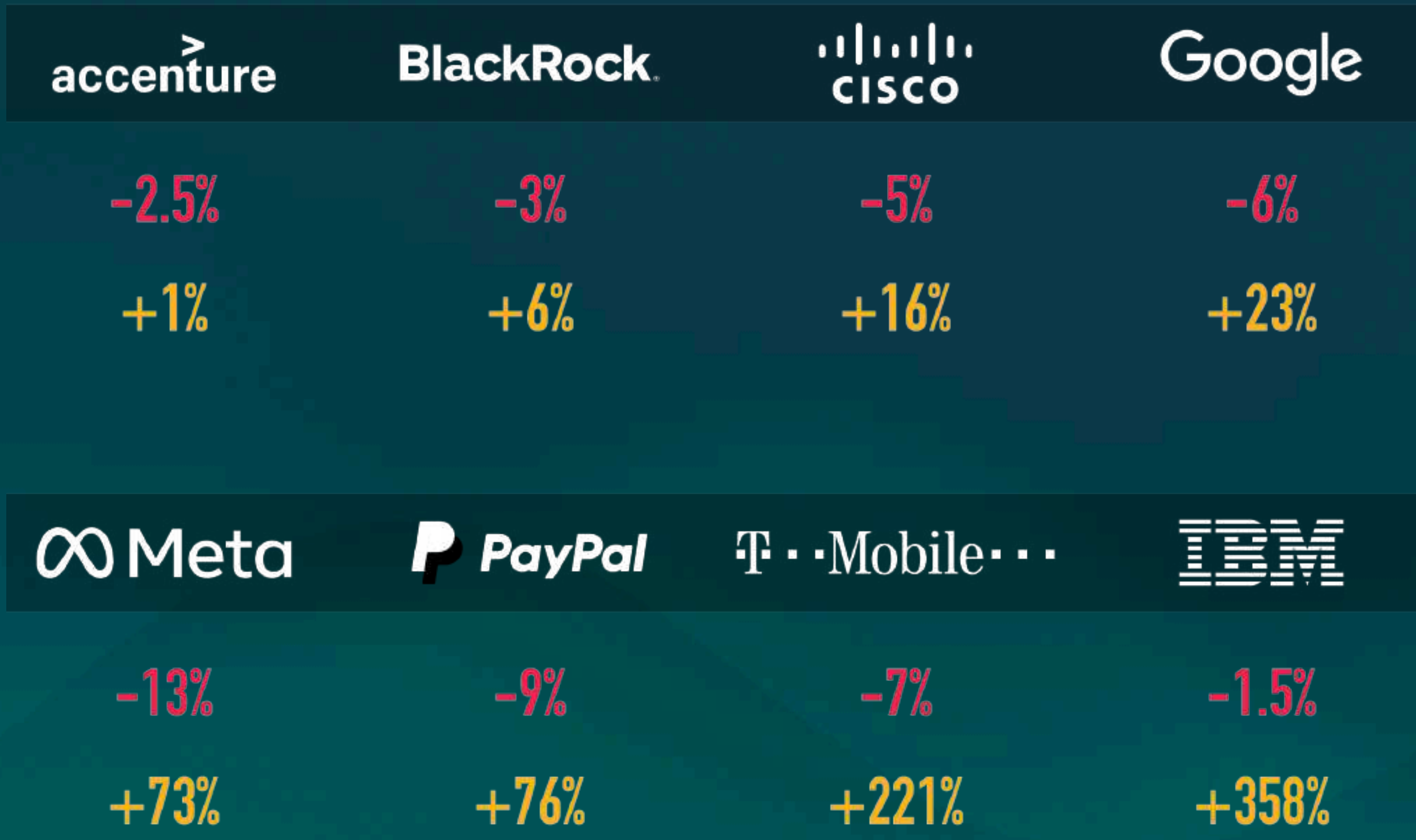


Commerce is the goal here

If AI is used mostly for corporate gains, what would AGI be used for?

CHANGE IN HEADCOUNT VS. ANNUAL NET INCOME 2023

Headcount Net Income

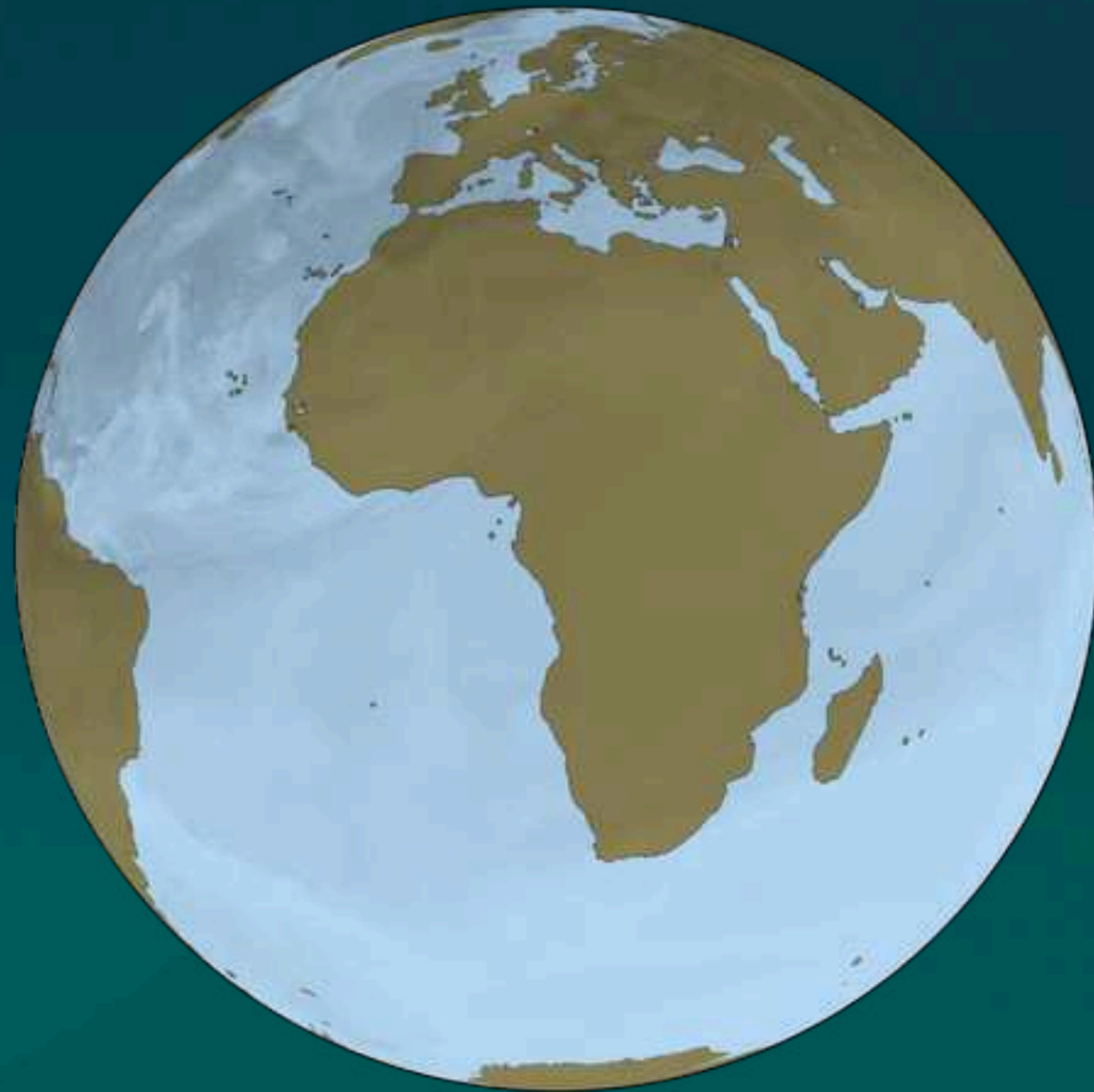


Source: Scott Galloway

“The problem comes when safety and profit maximization or company culture are not aligned. There is lots of historical evidence (think about fossil fuel companies and the climate, or drug companies before the FDA, e.g., with thalidomide, etc) and research in economics showing that **PROFIT MAXIMISATION CAN YIELD CORPORATE BEHAVIOR THAT IS AT ODDS WITH THE PUBLIC INTEREST**”

(Joshua Bengio, University of Montreal)

PROFIT
GROWTH



PEOPLE
PLANET
PURPOSE
PROSPERITY





The way of life

AI Takeover?

If AI takes over, at least us humans will have more "me time" ...right?



The Daily Show ✓
11.2M subscribers



Subscribed ▼

**"IT'S EASIER TO IMAGINE
THE END OF THE WORLD
THAN THE END OF
CAPITALISM"**

(Jameson and Žižek)



over, at least us humans will have more "me time" ...right?

aily Show ✓
subscribers

Subscribed ▼

**"IT'S EASIER TO IMAGINE
THE END OF THE WORLD
THAN THE END OF
CAPITALISM"**

(Jameson and Žižek)

In a nutshell:

Proaction  Precaution

IA: Yes AI: Yes, but* AGI: NO, unless...*



The King Midas Problem of AGI: Be careful what you wish for

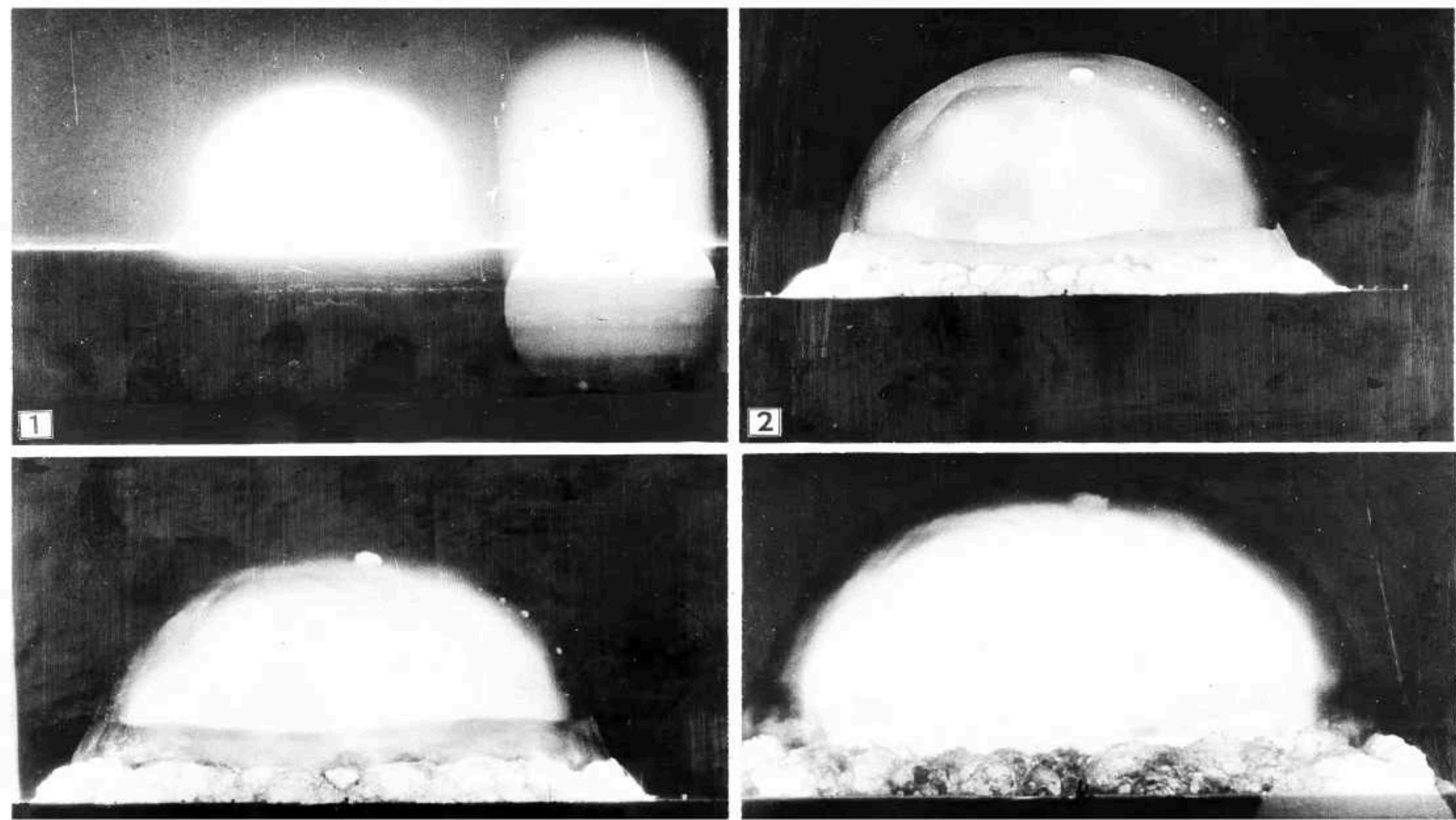


PURELY INTELLIGENT CREATURES,
WHETHER HUMACHINE, WILL WILL
DENRITATUR TO SOCIETY

Purely intelligent creatures, whether human
or machine, will be a curse, not a blessing

Trump allies draft AI order to launch 'Manhattan Projects' for defense

The plan to 'make America first in AI' and roll back 'burdensome regulations' would favor Silicon Valley investors, who are now flocking to support the former president



Images of the first atomic bomb test in 1945 in New Mexico. Photograph: AP

From 'Oppenheimer' to AI: The Dangers of Implementing Technology Without Oversight

How comfortable are you to build something that might **ENABLE** human extinction?

“With the prospect of human-level computing power a new idea suggests itself: that I may be working to create tools which will enable the construction of the technology that may replace our species. How do I feel about this? Very uncomfortable.

Given the incredible power of these new technologies, shouldn't we be asking how we can best coexist with them? And if our own extinction is a likely, or even possible, outcome of our technological development, shouldn't we proceed with great caution?”

Bill Joy: 2000 Wired





“We need to make sure that no single human, no single corporation and no single government can abuse the power of AGI at the expense of the common good”



(Joshua Bengio, University of Montreal)

The technocratic oath for AI: HUMAN / PLANET BENEFIT FIRST



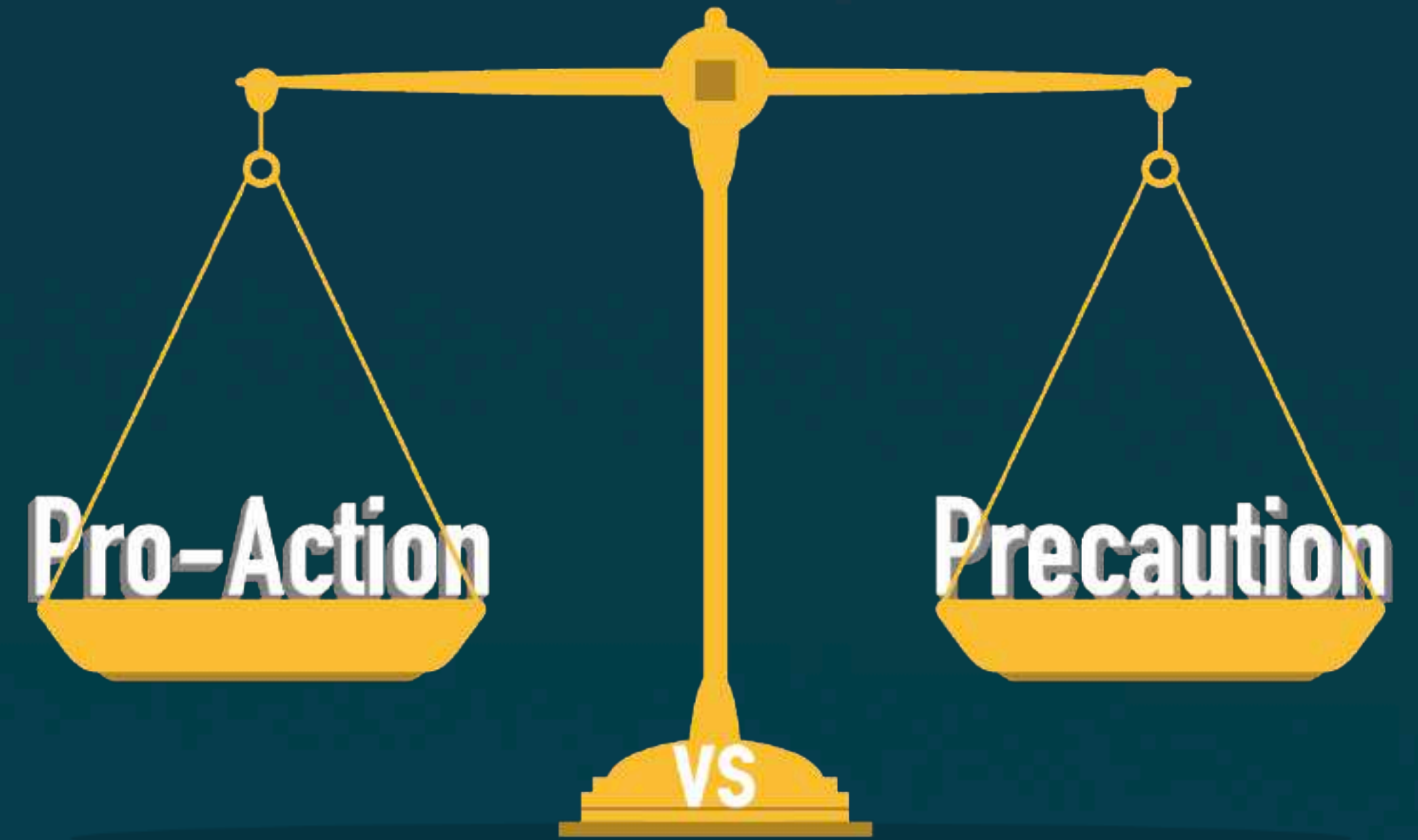
```
<header class="page-header">  
  <h1> "TECHNOCRATIC OATH" </h1>  
</header>  
<body class="collections--feature collections">  
  <noscript class="header-noscript bg__state--red text__general--text-inverse">  
    <div class="notification__content">  
      <h3> "I will place human decency  
        over technological potential in all cases" </h3>  
    </div>  
  </noscript>  
</body>
```


THE PRECAUTIONARY PRINCIPLE

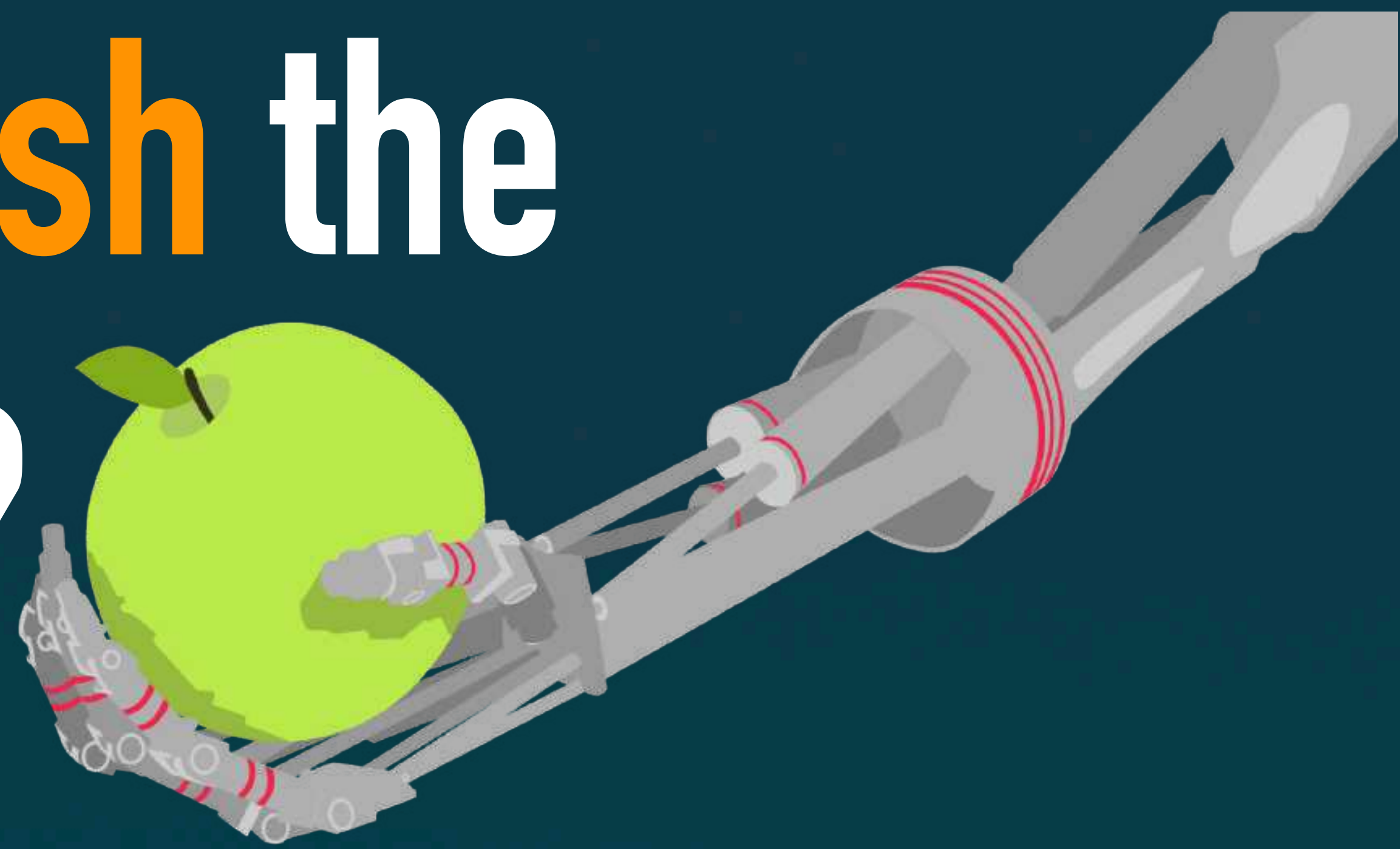
is a broad epistemological, philosophical and legal approach to innovations with potential for causing harm when extensive scientific knowledge on the matter is lacking.

IT EMPHASISES CAUTION, PAUSING AND REVIEW BEFORE LEAPING INTO NEW INNOVATIONS THAT MAY PROVE DISASTROUS

(Wikipedia)



Could we **relinquish** the possibility of AGI?



In 2000, Bill Joy (SUN Founder) proposed relinquishment, or limiting the pursuit of certain types of knowledge, as the only viable solution to safeguard humanity's future.

**In 2023, hundreds of AI experts and other notable figures signed a statement declaring that
“MITIGATING THE RISK OF EXTINCTION FROM AI SHOULD BE A GLOBAL PRIORITY
ALONGSIDE OTHER SOCIETAL-SCALE RISKS SUCH AS PANDEMICS AND NUCLEAR WAR”**

Who will be Mission Control for Humanity?



“The wolf you feed is the wolf that wins”



UNITED NATIONS

