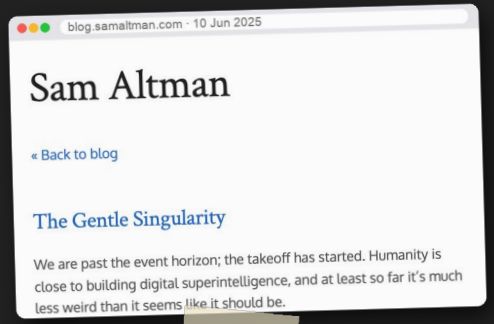
# Mind and Machines

Session 2 · The Future of Intelligence

# Superintelligence is coming?

## SOON


blog.samaltman.com · 10 Jun 2025

Sam Altman

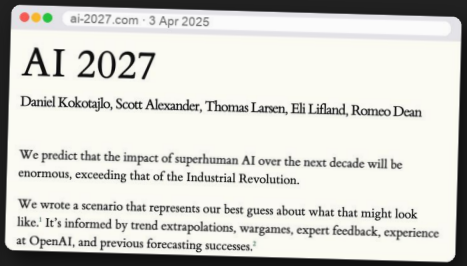
« Back to blog

The Gentle Singularity

We are past the event horizon; the takeoff has started. Humanity is close to building digital superintelligence, and at least so far it's much less weird than it seems like it should be.

*"We are past the event horizon; the takeoff has started."*

— Sam Altman, 'The Gentle Singularity', June 2025


ai-2027.com · 3 Apr 2025

AI 2027

Daniel Kokotajlo, Scott Alexander, Thomas Larsen, Eli Lifland, Romeo Dean

We predict that the impact of superhuman AI over the next decade will be enormous, exceeding that of the Industrial Revolution.

We wrote a scenario that represents our best guess about what that might look like.[1] It's informed by trend extrapolations, wargames, expert feedback, experience at OpenAI, and previous forecasting successes.[2]

*"We predict that the impact of superhuman AI over the next decade will be enormous, exceeding that of the Industrial Revolution."*

— AI 2027 (Kokotajlo, Alexander, Larsen, Lifland, Dean), April 2025

DARIO AMODEI · MACHINES OF LOVING GRACE · OCT 2024

"I think it could come as early as 2026, though there are also ways it could take much longer."

## NOPE


arxiv.org · Apr 2021

Computer Science > Artificial Intelligence

[Submitted on 26 Apr 2021 (v1), last revised 28 Apr 2021 (this version, v2)]

Why AI is Harder Than We Think

Melanie Mitchell

*"...four fallacies in common assumptions made by AI researchers, which can lead to overconfident predictions about the field."*

— Melanie Mitchell, 'Why AI is Harder Than We Think', 2021

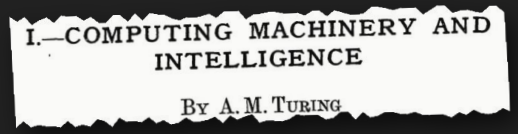
*"...we should recognize that it is a religious idea at this moment and not a scientific one."*

— Kevin Kelly, Backchannel, April 2017

DAVID DEUTSCH · AEON · 2012

"In 1950, Turing expected that by the year 2000, 'one will be able to speak of machines thinking without expecting to be contradicted.' In 1968, Arthur C. Clarke expected it by 2001."

# How we got here


I.—COMPUTING MACHINERY AND INTELLIGENCE

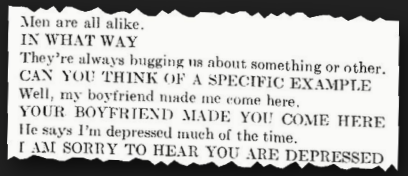
By A. M. Turing

**1950**

Turing's imitation game

**1966**

Weizenbaum's ELIZA chatbot

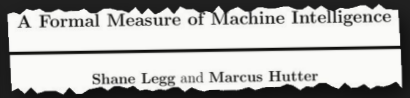

Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED

**1980**

Searle's Chinese Room

**2006**

Legg & Hutter define intelligence


A Formal Measure of Machine Intelligence

Shane Legg and Marcus Hutter


THE SUPERINTELLIGENT WILL:

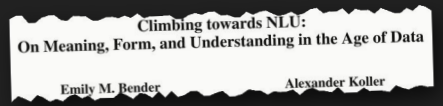
**2012**

Bostrom's orthogonality thesis · Deutsch's 'Creative blocks'

**2017**

Kelly: 'The myth of a superhuman AI'





Climbing towards NLU:
On Meaning, Form, and Understanding in the Age of Data

Emily M. Bender Alexander Koller

**2020**

Bender & Koller's octopus test

**2024–25**

Amodei's 'country of geniuses' · AI 2027

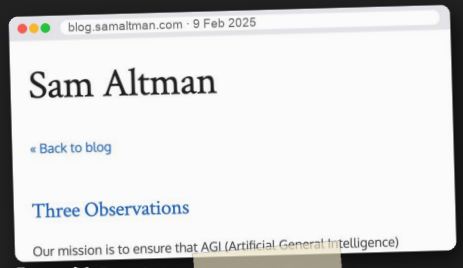



**2026**

**?**

# People use different definitions

## AGI




***"AGI is a weakly defined term, but generally speaking we mean it to be a system that can tackle increasingly complex problems, at human level, in many fields."***

— Sam Altman, 'Three Observations', Feb 2025

DAVID DEUTSCH · 'CREATIVE BLOCKS' · 2012

"That AGIs are people has been implicit in the very concept from the outset."

## Powerful AI



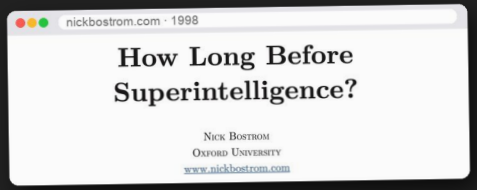

***"I find AGI to be an imprecise term that has gathered a lot of sci-fi baggage and hype."***

— Dario Amodei, 'Machines of Loving Grace', Oct 2024

HIS DEFINITION OF "POWERFUL AI"

"...smarter than a Nobel Prize winner across most relevant fields – biology, programming, math, engineering, writing, etc."

## Superintelligence




***"By a 'superintelligence' we mean an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills."***

— Nick Bostrom, 'How Long Before Superintelligence?', 1998

# What does intelligence mean?

**1950** → **2006** → **2012** → **2020**

| 1950 | 2006 | 2012 | 2020 |
|---|---|---|---|
| Can it talk like us? | Can it achieve goals? | Can it create new explanations? | Does it understand? |

I.—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

ALAN TURING · MIND · 1950

"We now ask the question, 'What will happen when a machine takes the part of A in this game?' … These questions replace our original, 'Can machines think?'"

***"Intelligence measures an agent's ability to achieve goals in a wide range of environments."***

— Shane Legg & Marcus Hutter, 'A Formal Measure of Machine Intelligence', 2006

***"…the ability to create new explanations is the unique, morally and intellectually significant functionality of people (humans and AGIs)…"***

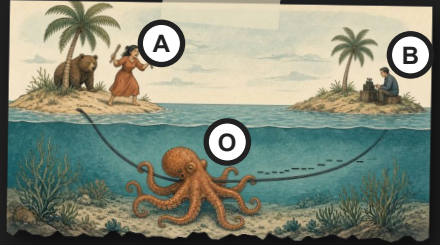
— David Deutsch, 'Creative blocks', Aeon, 2012

"AGI cannot possibly be defined purely behaviourally."

— Deutsch, same essay

***"It is not that O's utterances make sense, but rather, that A can make sense of them."***

— Emily Bender & Alexander Koller, the octopus test, 2020




1. A and B chat by undersea cable
2. O taps in, learns the patterns
3. O cuts B off and poses as B
4. A bear chases A; O can't help

***"Intelligence is not a single dimension, so "smarter than humans" is a meaningless concept."***

— Kevin Kelly, 'The Myth of a Superhuman AI', 2017

# Scaling might not be the only bottleneck

## THE BUILDERS

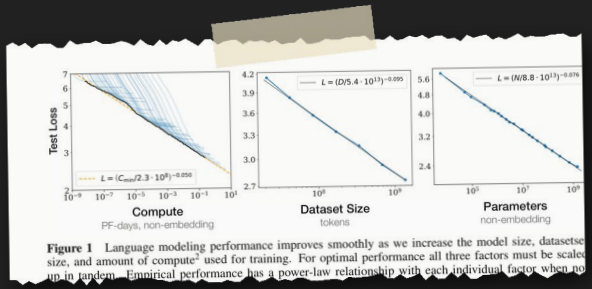



**Figure 1** Language modeling performance improves smoothly as we increase the model size, datasetsc size, and amount of compute² used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when no

*"The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin."*

— Rich Sutton, 'The Bitter Lesson', 2019

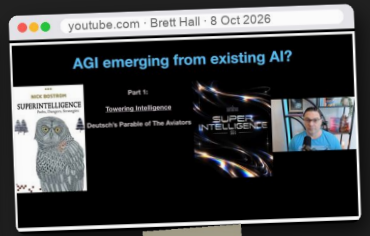
*"The intelligence of an AI model roughly equals the log of the resources used to train and run it."*

— Sam Altman, 'Three Observations', Feb 2025

VS

## THE THEORISTS




*"Expecting to create an AGI without first understanding in detail how it works is like expecting skyscrapers to learn to fly if we build them tall enough."*

— David Deutsch, 'Creative blocks', Aeon, 2012

*"It had been noticed that the best flying things were the things that flew the highest. It seemed obvious: more height was the key."*

— Brett Hall, 'The parable of the aviators'

DAVID DEUTSCH · THE BEGINNING OF INFINITY · 2011

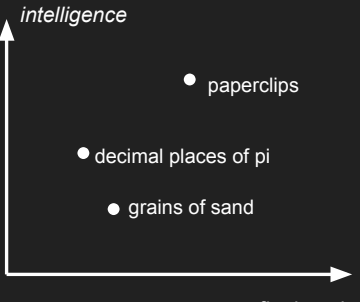
"If you can't program it, you haven't understood it."

*Brett Hall's 'parable of the aviators', after David Deutsch*

# What is the future of superintelligence?

## An alien will

Nick Bostrom, 2012




*"An artificial intelligence can be far less human-like in its motivations than a space alien."*

— Nick Bostrom, 2012

Instrumental convergence: for a wide range of final goals, an agent has reason to pursue self-preservation and goal-content integrity

## A country of geniuses

Dario Amodei, 2024

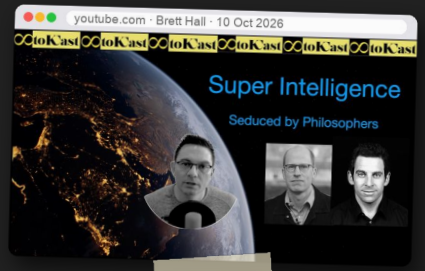
| | |
|---|---|
| **Nobel** | smarter than a Nobel Prize winner |
| **Millions** | of instances running at once |
| **10x–100x** | human speed |
| **Weeks** | tasks that take weeks, done autonomously |

*"Intelligence may be very powerful, but it isn't magic fairy dust."*

— Dario Amodei, on his "country of geniuses in a datacenter"

## A person

David Deutsch & Brett Hall




*"An AGI is a person. Not a human person - but a person nonetheless."*

— Brett Hall, 'Superintelligence'

HALL ON BOSTROM'S PAPERCLIP AI

"...the stupidest "super" intelligence possible. A single minded fool who nonetheless can thwart truly general intelligence (us)."

## A myth

Kevin Kelly, 2017




*"The results of accelerating technology will most likely not be super-human, but extra-human. Outside of our experience, but not necessarily "above" it."*

— Kevin Kelly, Backchannel, 2017