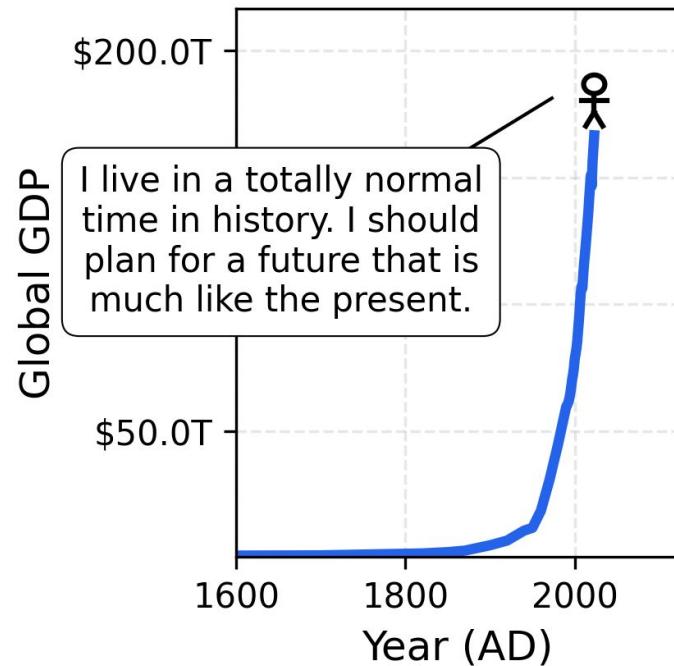


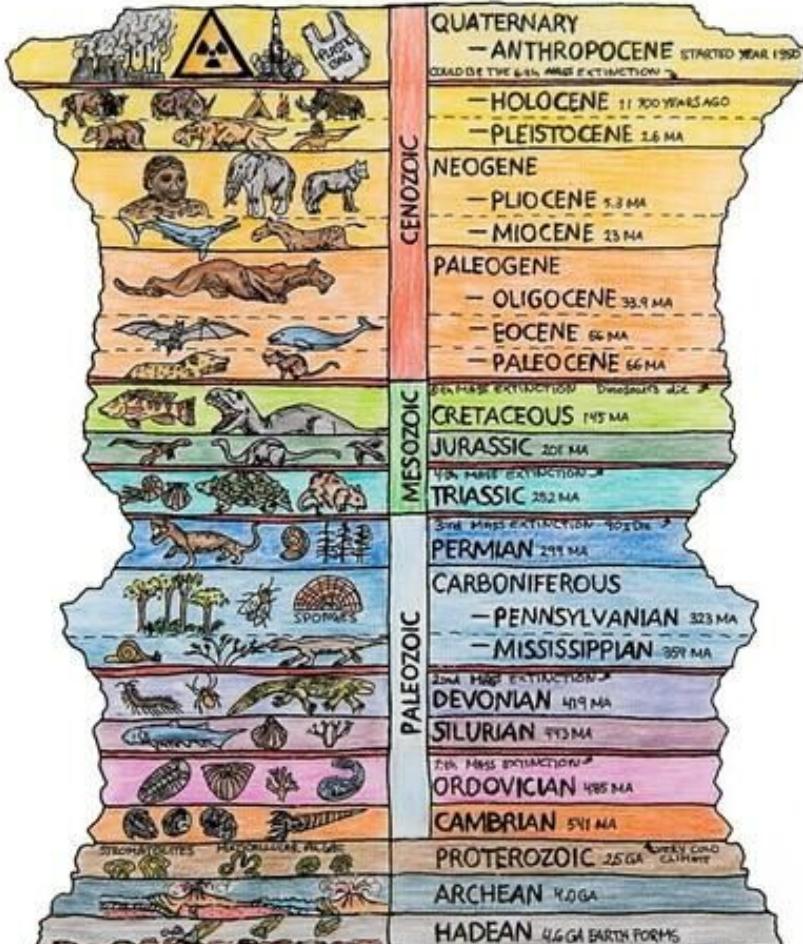
Advice for a (young) investigator in the first and last days of the Anthropocene

Jascha Sohl-Dickstein
Anthropic



[h/t Tim Urban, Wait But Why]

- AI is transforming the world. Now.
 - You should take that seriously in your career and research.
 - You have **immense leverage** on how the whole thing turns out.
-
- ... and a few practical suggestions

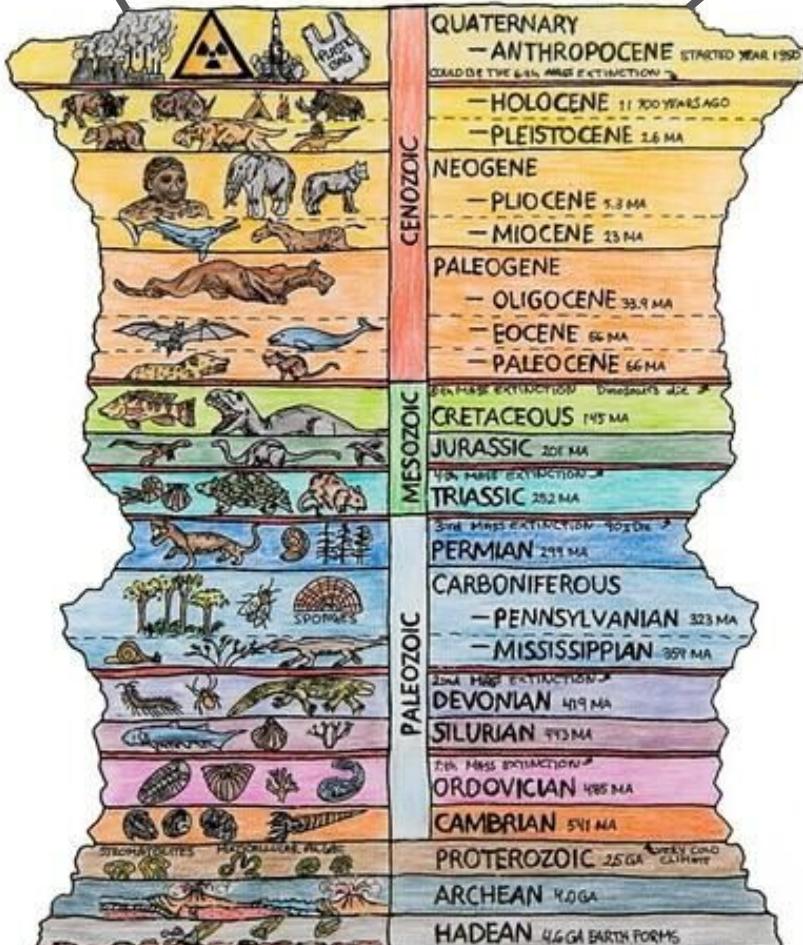


Anthropocene

Proposed geologic epoch characterized by impact of human activities on the Earth (cf. 6th mass extinction)

Started around 1950

?



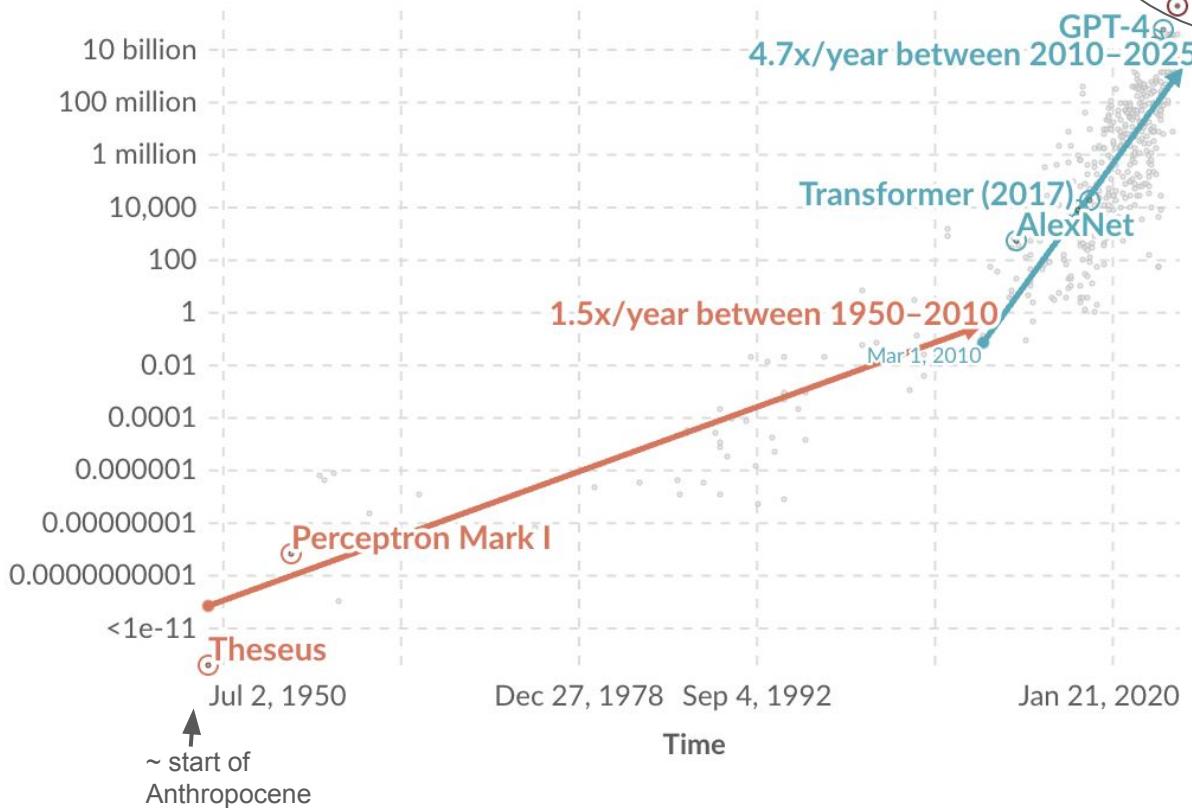
Anthropocene

Proposed geologic epoch characterized by impact of **human** activities on the Earth (cf. 6th mass extinction)

Started around 1950

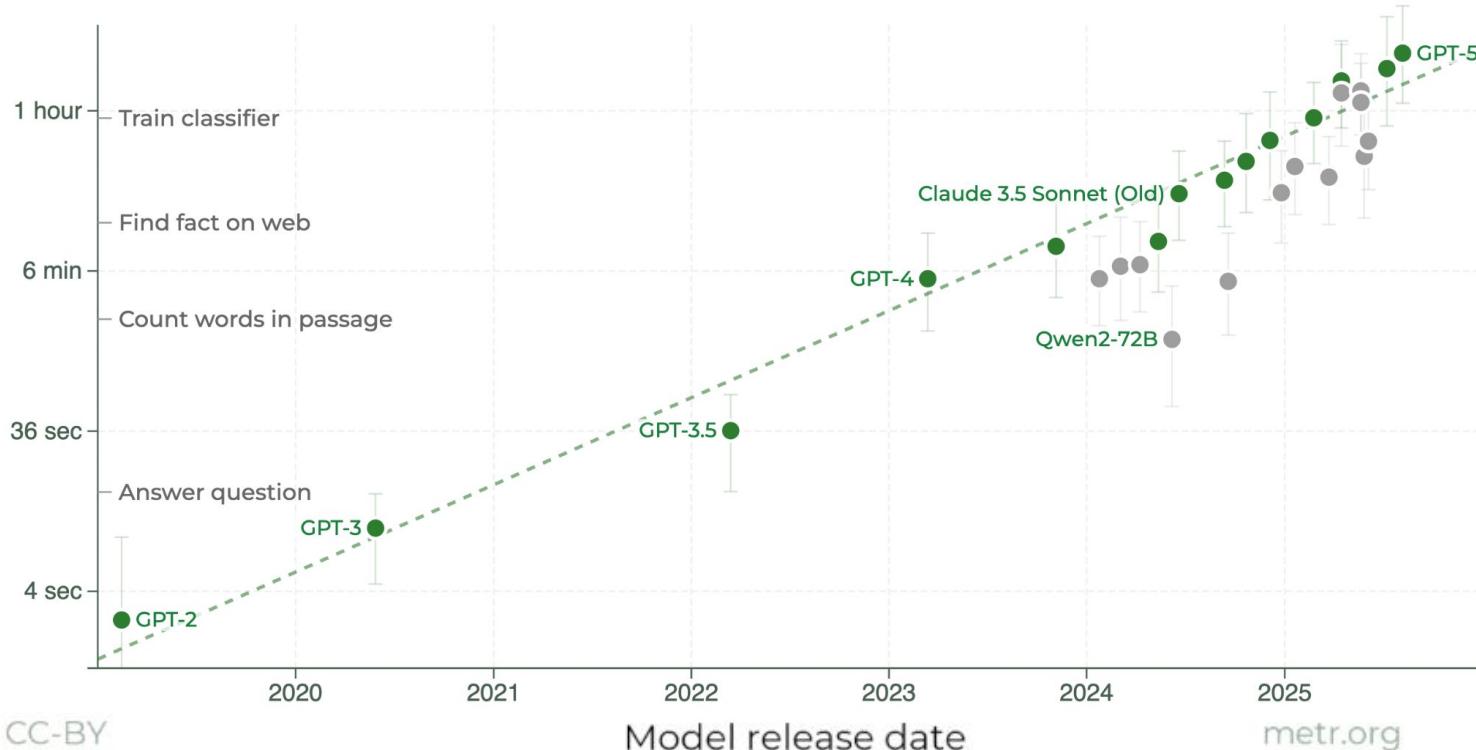
AI progress in the Anthropocene

Training computation (petaFLOP)



AI progress in the Anthropocene

The length of tasks AI can do is doubling every 7 months
Task length (at 50% success rate)

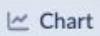


AI progress in the Anthropocene

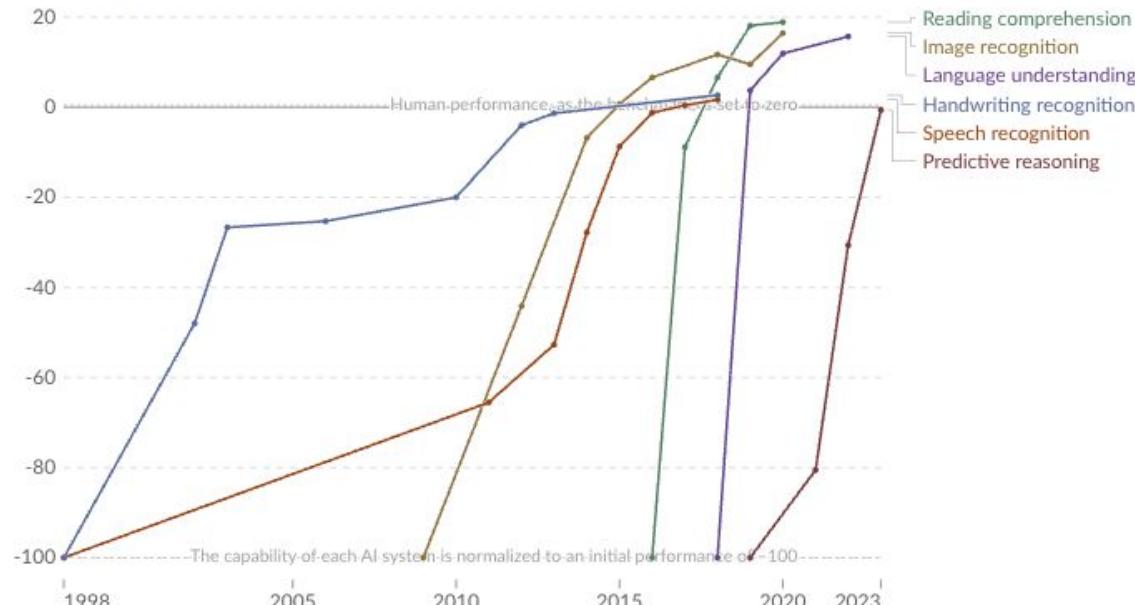
Test scores of AI systems on various capabilities relative to human performance

OurWorld
in Data

Within each domain, the initial performance of the AI is set to -100. Human performance is used as a baseline, set to zero. When the AI's performance crosses the zero line, it scored more points than humans.

Table  Chart

 Edit capabilities



Data source: Kiela et al. (2023) – [Learn more about this data](#)

OurWorldinData.org/artificial-intelligence | CC BY

Note: For each capability, the first year always shows a baseline of -100, even if better performance was recorded later that year.

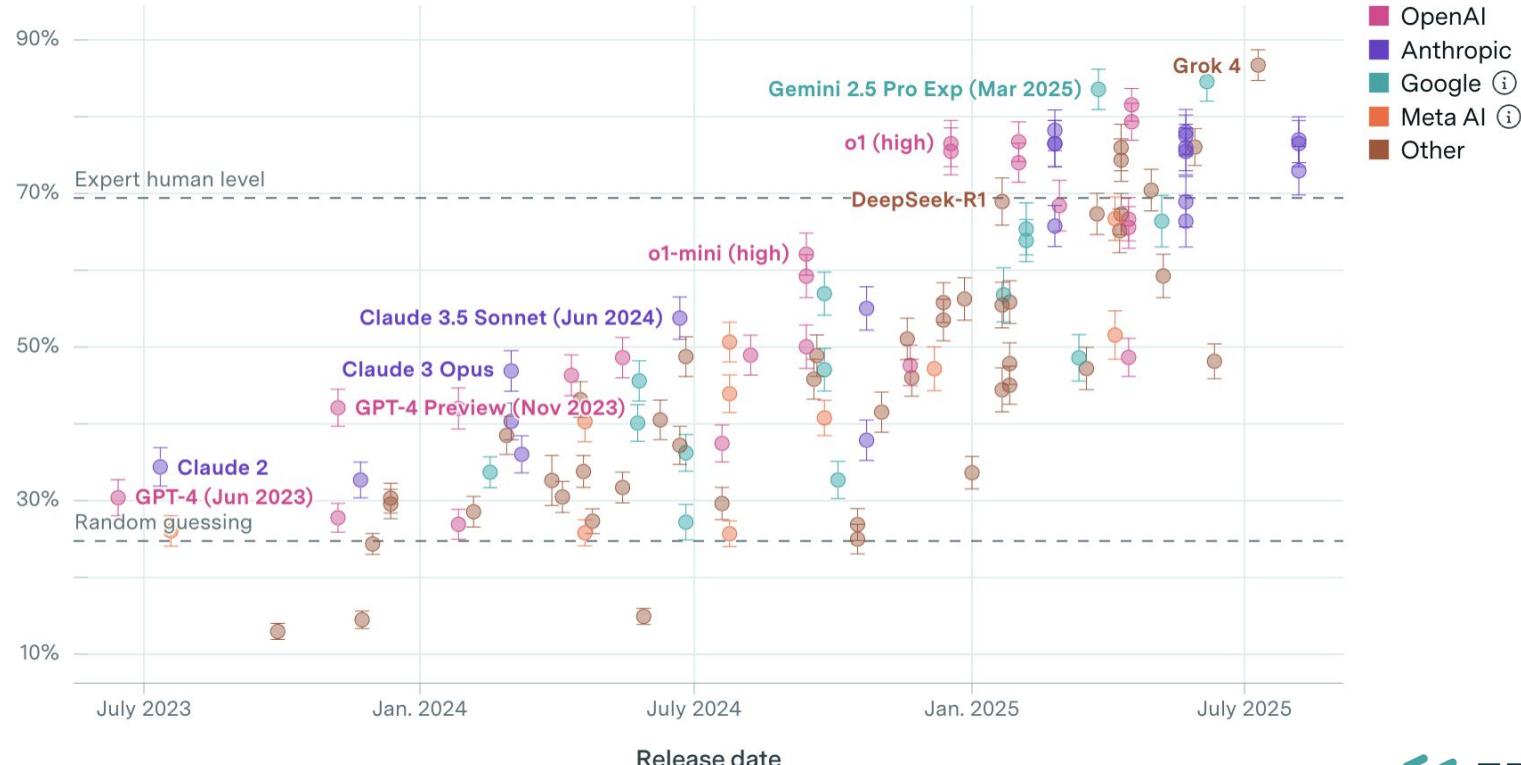


AI progress in the Anthropocene

AI performance on a set of Ph.D.-level science questions

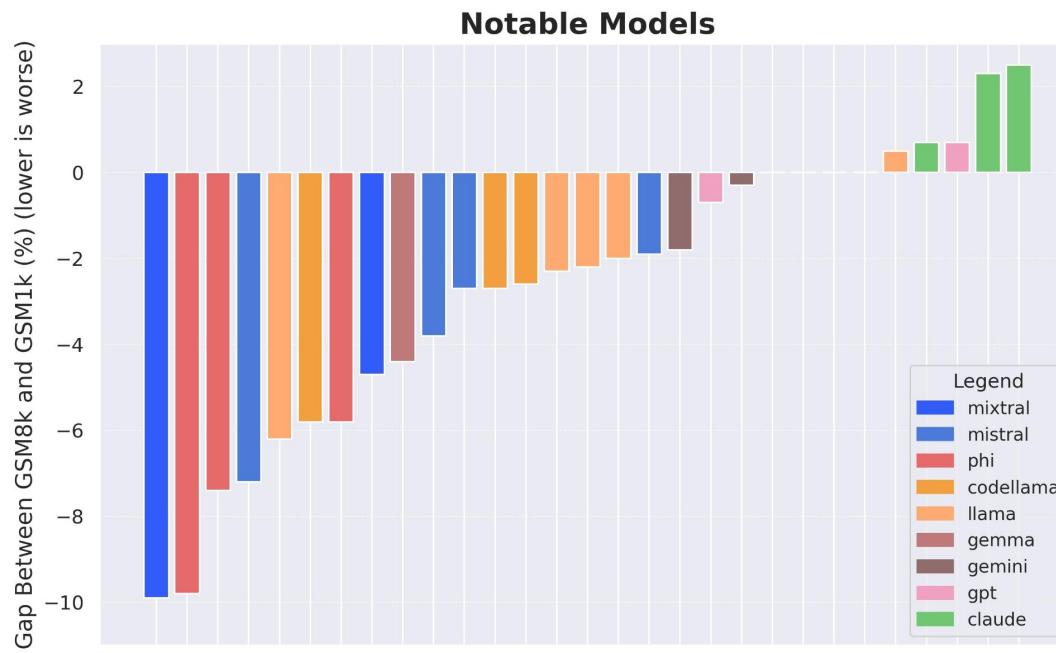
GPQA Diamond accuracy [\(i\)](#)

115 Results [\(i\)](#) Organization



AI progress in the Anthropocene

- Benchmark performance not explained by overfitting
- Relative performance on new questions, mimicking those in GSM8K benchmark



AI progress in the Anthropocene

- Benchmark performance not explained by overfitting

The New York Times

Google A.I. System Wins Gold Medal in International Math Olympiad

OpenAI said it, too, had built a system that achieved similar results.

July 21, 2025

[AI] solutions were astonishing in many respects. IMO graders found them to be clear, precise and most of them easy to follow. -IMO President Prof. Dr. Gregor Dolinar

AI progress in the Anthropocene

From: [REDACTED]
<[REDACTED]>

Date: Mon, 4 Aug 2025

Subject: Fwd: Aug 12 Special AM Seminar:
"Advice for a young investigator in the first
and last days of the Anthropocene"

To: jascha.sohldickstein@gmail.com

What a bat-shit crazy abstract...

The Overton window is moving

AI progress in the Anthropocene

The competition for AGI—AI that surpasses humans at all cognitive tasks—is of fundamental geopolitical importance. -Rishi Sunak, UK Prime Minister, 2025

As profound as [cell phone] technology has been, AI will be more impactful. And it is going to come faster ... We're now starting to see these models, the what we consider to be the most advanced, reaching a really high level —
work. -Barack Obama, US president

The Overton window is moving

We believe that programmers will be replaced by AI. Within three to five years, we'll see AGI systems as smart as the best humans. And in six years, artificial superintelligence smarter than all of us combined. This is happening fast, and society isn't ready. -Eric Schmidt, former Google CEO, 2025

AI progress in the Anthropocene

The competition for AGI—AI that surpasses humans at all cognitive tasks—is of fundamental geopolitical importance. -Rishi Sunak, UK Prime Minister, 2025

As profound as [cell phone] technology has been, AI will be more impactful. And it is going to come faster ... We're now starting to see these models, these platforms be able to perform really high level — what we consider to be really high level intellectual work. -Barack Obama, US president, 2025

We believe that in the next year, the vast majority of programmers will be replaced by AI. Within three to five years, we'll see AGI systems as smart as the best humans. And in six years, artificial superintelligence smarter than all of us combined. This is happening fast, and society isn't ready. -Eric Schmidt, former Google CEO, 2025

AI progress in the Anthropocene

The
Economist

The economics of superintelligence

If Silicon Valley's predictions
are even close to being accurate,
expect unprecedented upheaval

The New York Times

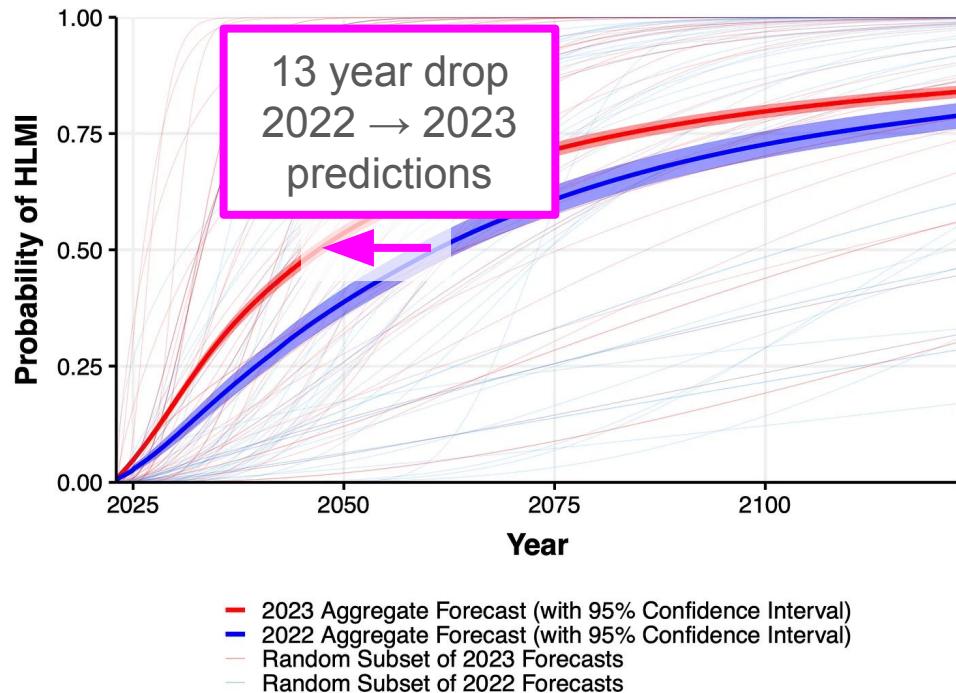
OPINION
THE EZRA KLEIN SHOW

The Government Knows A.G.I. Is Coming

March 4, 2025

AGI when?

- Consensus timeline in San Francisco AI labs
 - 2-5 years
 - "long" timelines, 10-30 years
- Most recent surveys of AI experts
- Personal
 - Current Claude feels like inept but knowledgeable + eager grad student
 - That wasn't true a year ago
 - It likely won't be true in another year



What does it mean for a machine to fly?

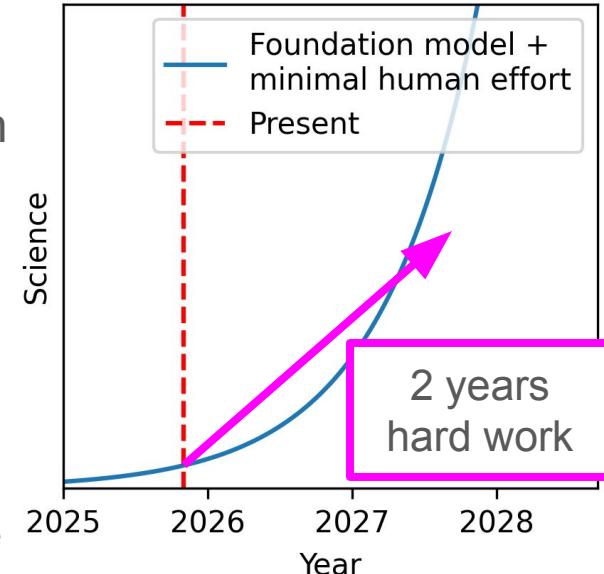
- Which of these count?
 - 1901: Machine glides 100 meters.
 - 1902: Machine glides 180 meters.
 - 1903: 59 seconds powered flight in straight line before crash.
 - 1904: 5 minutes powered flight, steering in circles.
 - 1905: 38 minutes powered+controlled flight.
- Now the precise definition seems irrelevant.

What does it mean for a machine to be generally intelligent?

- There is nuance at the cusp. (c.f. [Levels of AGI](#))
- It is AGI when the precise definition seems irrelevant.

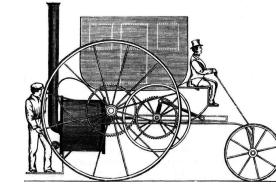
I have 2-5 years to contribute intellectually to the world. What should I do?

- Prefer ambitious, targeted, *fast* collaborations
- The timescale of the project needs to be shorter than the characteristic timescale of the exponential
 - Otherwise, it's better to wait, and use a newer foundation model to do it better+faster
 - This does not favor slow, independent, open-ended exploration :(
- The bitter lesson [Sutton, 2019]
 - General methods that leverage computation are ultimately the most effective.
 - Work on projects that will benefit from scale.
 - Avoid projects that will be solved by scale alone, before you even finish the project.



I have 2-5 years to contribute intellectually to the world. What should I do?

- Learn the new tools.
- They are often non-ergonomic. Use them anyway.
 - Horseless carriages are awkward, but still outperform carriages with horses
- You should be using foundation models to
 - Brainstorm + iterate on research ideas
 - Get writing feedback
 - Write your code
 - Iterate on analysis (Claude Code + a VSC notebook works great)
 - ...
- Being a PI is great practice for prompting LLMs, and vice versa.
 - Requires structured and clear scoping and communication of tasks.



I have 2-5 years to contribute intellectually to the world. What should I do?

- Apply your beliefs to your career (and personal) choices
 - The job of an academic will be *very different* in 10-20 years.
 - Your current career trajectory may not even exist.
 - There is a reason many academics (👉) are working in industry
- "I think we're going to have AGI in 3 years. [...] I'm planning to finish my PhD in 2 years, get a postdoc, and apply for faculty jobs after that."
- Risk-reward career tradeoffs just sacrifice reward
 - There is an unavoidable baseline of AGI risk
 - Many jobs and institutions are going to be disrupted
 - Pursue the option with the biggest upside

I have 2-5 years to contribute intellectually to the world. What should I do?

- Choose projects that shift the trajectory, rather than introducing a transient change
- Do something you will be proud of
 - After retiring to your villa in the Dyson swarm, you want to be able to tell your grandkids how you helped get the good outcome.



Photo: The geologic epoch after the Anthropocene

You have immense *power and leverage* over the future of AI

- AI can seem like a process beyond our individual control
- This is false
- We are still early in the exponential



Small choices early in exponential growth have huge consequences

- Proposal for standardized email (SMTP) but not chat protocol in 1980
- Free sharing of HeLa immortal human cell line
- Hobbyist creation of Linux operating system
- Introduction of invasive rabbits to Australia as food animal
- Work on Haber-Bosch nitrogen fixation process
- Choice of lead as anti-knocking agent in gasoline
- Proposal of, and choice to commercialize, PageRank algorithm
- Founding of professional organization (American Medical Association, American Bar Association, ...)
- Establishment of academic peer review system

You have immense *power and leverage* over the future of AI

- Individual, small, choices you make will determine the future commercial, social, and political landscape of AI
 - What research problem? Where do you work? Name model behavior a capability or a risk? Interpretable or inscrutable model architecture? Publicly share an ethical stance? What trajectory/tech tree does your project encourage?
- This is a responsibility as well as a power.
- Be intentional and thoughtful about projects and other choices. The world will change in remarkable ways, and *you personally have the leverage to make that change much better or much worse.*

Some project areas I like

-  AI for science
 - Often unambiguously positive in expectation
-  Science on AI models
 - Interpretability
 - Statistical physics and neuroscience and psychology and economics of AI models
-  Safety research
 - Huge importance to field size ratio
 - Foundations of field -- e.g. the right questions to ask -- are still often missing
-  Characterizing (future) capabilities and behaviors
 - Under-researched: System behavior of many AI agents (+for safety and science on AI)
-  Access, equity, fairness
 - Challenge level: pragmatism on active culture war issues
 - Fixing UX counts
-  Policy and governance
 - Technical people in policy positions are one of the scarcest resources

Take the future seriously!

- AGI is coming. Consider its implications when deciding what you work on, where you work, when you make career transitions, how you think about the important and interesting problems, how you think about the potential consequences and leverage of your work
- Prefer ambitious, targeted, *fast* collaborations
 - The timescale of the project needs to be shorter than the characteristic timescale of the exponential
- Use the new tools.
- Apply your beliefs to your career choices
 - "I think we're going to have AGI in 3 years. [...] I'm planning to finish my PhD in 2 years, get a postdoc, and apply for faculty jobs after that."
 - There is no safe option. Try for the best option.
- Do something you will be proud of
 - You have immense *power and leverage* over the future of AI
- This is a good time (the last time?) to go all in

SCRAP

A rubric to decide what to research

- *Positive impact:* if this project works flawlessly, how large is the potential benefit?
 - Small checkpoints on the way towards large impact are OK.
 - Derisk early and often!
 - Timescale of success -- will the impact still matter?
- Bitter lesson: will the project be complementary to scale?
- Opportunity cost: time and effort and resources
- Comparative advantage: why is this the project for you in particular?
- Redundancy: how many other people on the planet are doing the same thing?

A rubric to decide what to research

- *Positive impact*: if this project works flawlessly, how large is the potential benefit?
- Bitter lesson: will the project be complementary to scale?
 - Your work should still matter after intelligence is too cheap to meter
 - The impact of the right project can grow exponentially in time
 - data, best practices and frameworks, new questions, new framings, algorithms that work better with more scale
- Opportunity cost: time and effort and resources
- Comparative advantage: why is this the project for you in particular?
- Redundancy: how many other people on the planet are doing the same thing?

A rubric to decide what to research

- *Positive impact*: if this project works flawlessly, how large is the potential benefit?
- *Bitter lesson*: will the project be complementary to scale?
- *Opportunity cost*: time and effort and resources
 - Is the effort *reusable*?
 - (again) Derisk early and often!
- *Comparative advantage*: why is this the project for you in particular?
- *Redundancy*: how many other people on the planet are doing the same thing?

A rubric to decide what to research

- *Positive impact*: if this project works flawlessly, how large is the potential benefit?
- *Bitter lesson*: will the project be complementary to scale?
- *Opportunity cost*: time and effort and resources
- *Comparative advantage*: why is this the project for you in particular?
 - specific expertise
 - access to resources (e.g. compute, data)
 - collaborators with the right expertise
 - ~~I have a clever idea no one else has thought of~~
- *Redundancy*: how many other people on the planet are doing the same thing?

A rubric to decide what to research

- *Positive impact:* if this project works flawlessly, how large is the potential benefit?
- *Bitter lesson:* will the project be complementary to scale?
- *Opportunity cost:* time and effort and resources
- *Comparative advantage:* why is this the project for you in particular?
- *Redundancy:* how many other people on the planet are doing the same thing?
 - Even if you publish first, your time was wasted! The world should be in a different state because of your efforts.
 - Find ideas that no one else recognizes as important

A rubric to decide what to research

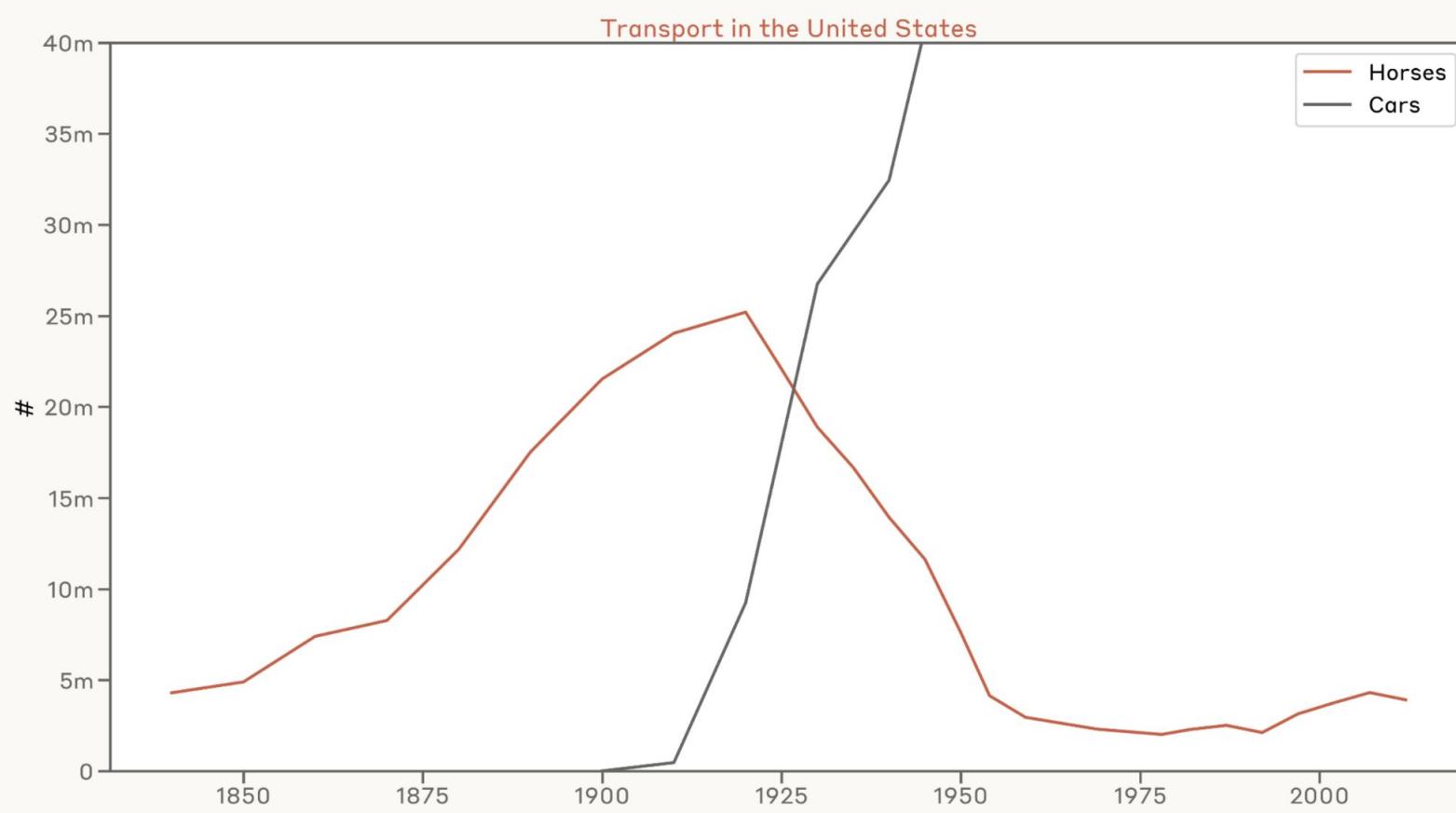
- *Positive impact*: if this project works flawlessly, how large is the potential benefit?
- *Bitter lesson*: will the project be complementary to scale?
- *Opportunity cost*: time and effort and resources
- *Comparative advantage*: why is this the project for you in particular?
- *Redundancy*: how many other people on the planet are doing the same thing?

A rubric to decide what to research

- *Positive impact:* if this project works flawlessly, how large is the potential benefit?
- *Bitter lesson:* will the project be complementary to scale?
- *Opportunity cost:* time and effort and resources
- *Comparative advantage:* why is this the project for you in particular?
- *Redundancy:* how many other people on the planet are doing the same thing?
- What is the absolute weirdest project (research or non-research) you would enjoy doing? The sweet spot:
 - You can clearly articulate why the project is a good idea
 - Everyone looks at you funny when you first explain it

Should you do academic research?

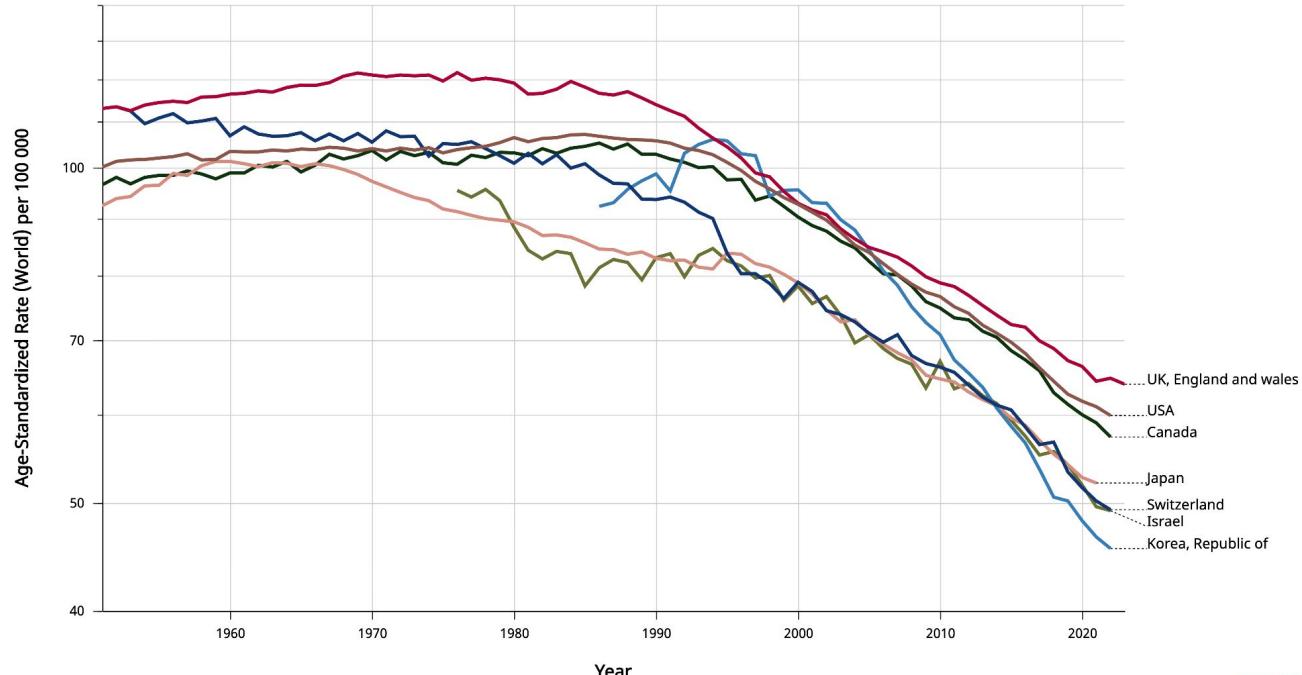
- What are you optimizing for?
 - Love of science? (Scientific) fame? Respect? Money? Interesting problems? Amazing friends and colleagues? Positive impact on world?
- If this was your last job, would you do something different?
 - Are you in a holding pattern for a future career?
- What would you do instead? What is your best alternative?
 - startup, industry research, industry engineering, advocacy, family, ...?
- How much leverage over the future will you have in each case?



[Andy Jones, personal communication, 2025]

Cancer deaths are exponentially falling, in the rich world

Age-standardized rate (World) per 100 000, mortality, both sexes, age [0-74]





Reasoning – robot navigation



I have 2-5 years to contribute intellectually to the world. What should I do?

- Apply your beliefs to your career (and personal) choices
 - The job of an academic will be *very different* in 10-20 years.
 - Your current career trajectory may not even exist.
- "I think we're going to have AGI in 3 years. [...] I'm planning to finish my PhD in 2 years, get a postdoc, and apply for faculty jobs after that."
- There is a reason many academics () are working in industry
 - More leverage on the future
 - More access to cutting edge problems
 - More resources
 - (More money)