# Machines of Loving Grace

Table 5 Lead Presentation

# Dario Amodei

Dario Amodei started his career studying physics and computational neuroscience - in 2016, he joined OpenAI and became VP of Research, a position that allowed him to lead the development of LLMs like GPT-2 and GPT-3. He left due to disagreements with the direction of OpenAI over AI safety; as a result, he founded Anthropic, a company dedicated to “building AI systems that are steerable, interpretable and safe.”

# Powerful AI - “A country of geniuses in a datacenter”

A powerful AI should:

- Have generalized deep intelligence on a variety of subjects (“smarter than a Nobel Prize winner”)
- Engage in any actions with all the interfaces a human can have access to
- Work autonomously on tasks, rather than passively answer questions
- Have the ability to control physical tools, although it has no physical body
- Be able to run millions of instances of itself with existing resources, acting independently or collaboratively as necessary

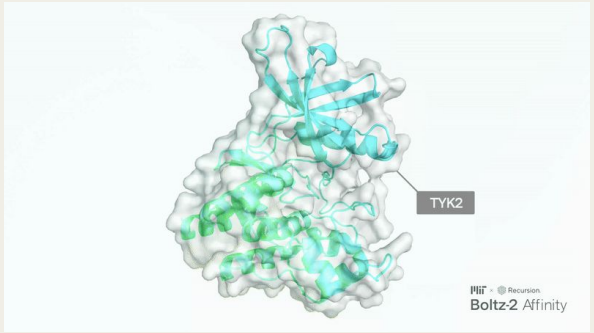
# Biology and Health

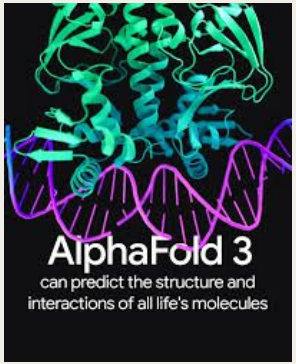
- The main solvable problem in biology and health is intrinsically complex data with no obvious catalyst to speed up real-world experiences.
- Increasing the amount of skilled researchers working on these problems could theoretically increase the rate of discovery of innovative solutions 10x or more.
- The Compressed 21st Century: after powerful AI is developed, all progress made in the whole 21st century will be made in a few years.
- Amodei predicts: prevention and treatment of all infectious diseases, elimination of cancer, prevention of genetic disease and Alzheimer's, biological freedom, doubling of the human lifespan.

# Biology and Health (So Where Are We?)

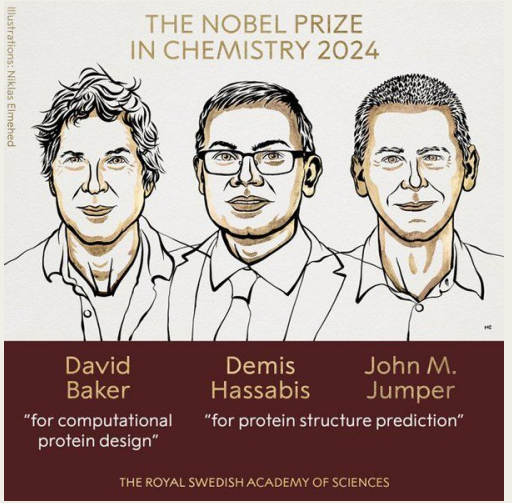
It's been a year and change since Amodei wrote this. Has any of it come true?

Kind of...


But mostly no.



# Neuroscience and Mind

- Rate of advancements will be similar to the biology methods were 100 years of progress can be made in 5-10 years for the same reasons
- Optogenetics and neural probes allow for precise reading and measurement of individual neurons or neuronal circuits
- Interpretability: how artificial computational neurons can contribute to the understanding of biological ones
- Accelerate research on the genetic basis of mental illness and drug development
- Optimistic about curing most mental illnesses, making brain alterations, accomplishing embryo screening, developing drugs

# Economic Development & Poverty

**AI can accelerate development, but it cannot replace governance, trust, or institutions.**

**What AI is well-suited for:**

- Improving global health distribution by optimizing disease modeling, targeting, and large-scale logistics
- Support economic policy & decision making with better information analysis
- Accelerate diffusion in new technologies in energy, transportation, and agriculture

**Primary constraints:**

- Corruption & weak state capacity that limit implementation
- Political legitimacy & social trust shape adoption
- Economic complexity

**Areas of optimism:**

- The possibility of eradicating/dramatically reducing infectious disease at a global scale
- An AI-enabled “second Green Revolution” to improve food security and agricultural efficiency
- Lower-cost clean energy and climate mitigation technologies that reduce development barriers

**“If AI further increases economic growth and quality of life in the developed world, while doing little to help the developing world, we should view that as a terrible moral failure,” (Amodei)**

# Peace and Governance

- AI does not structurally favor democracy. Instead, it can strengthen autocracy through better surveillance and propaganda.
- AI should be directed to a path that favors democracy
- He envisions a coalition of democracy using AI to lead the world
- AI system as an aid to help decision-making, reduce bias, and improve transparency in legal systems
- AI to be used to help provision government services—such as health benefits or social services

# Work and Meaning

- The big question of this section is: “With AIs doing everything, how will humans have meaning? For that matter, how will they survive economically?”
- Amodei argues that humans do plenty of activities for pleasure while knowing there is someone out there who can probably do it better than them
    - For example, Amodei loves video games, even though there is someone out there that is better at video games than he is.
- He believes that a sense of accomplishment is still achievable in an AI society. He offers that as long as AI is 90% better at a given task than humans, that 10% will create jobs and productivity
- At this rate, AI is on its path to become super cheap and available
- This will force society to completely reevaluate our economic setup, but we’ve done it before!
- We may see a universal basic income, large resources given out, or many other solutions that are currently hard to imagine

# Discussion Questions

1. How does Amodei's background and experiences in AI inform his perspective in this essay?
2. Is there anything you would add to Amodei's definition of "powerful AI"? Do you think this is the same as AGI, or is there some crucial difference?
3. Amodei's promising outlook seems to miss the ethical aspect. How important do you think ethics is. And how much do you think moral ethics should be involved in building an AI system.(Yuezhang)
4. Amodei claims that automating the entire research process will truly accelerate biology. Is it worth it to completely eradicate human direction in the academic research process without the guarantee of 100 years of advancement in 3-5 years? Additionally, does everyone on Earth equally benefit from biological advancement? Why or why not? (Salonee)
5. If AI is able to significantly improve health and productivity, but requires strong institutions to support those benefits, how should societies balance the priority of strengthening these institution while also aiming to deploy AI with the purpose of catching up? (Iman)
6. Amodei is optimistic about AI curing most mental illnesses through altering brains, neural network, and developing new drugs. Do you think these processes are effective and ethical, or is there more than just science behind the problems?(Carolyn)
7. Amodei acknowledges that humans enjoy a sense of accomplishment and even competition with tasks or economic labor, and this may be taken away a bit with the new AI organization. He counters that human connection is the most meaningful human need. Do you believe that human connection is enough for society? How would you envision your life if you were deprived of economic success and motivations? (Stella)