

Ay 111a – Fall 2026

Science in the Era of Big Data and AI

S. George Djorgovski
Professor of Astronomy and Data Science



Caltech

This is not a lecture, but a start of a conversation

(I will show a few slides as a warm-up, though)

Big Data + AI have transformed and are continue to transform how science is being done in the 21st century.

These are by far the **fastest evolving technologies ever**, and it is very hard to keep up. AI in particular may be evolving ***superexponentially***, with the current characteristic time scale of 3-4 months.

If you know **how to use them as tools**, you may thrive. If not ...

These skills are universal, **applicable in any field** (like math and statistics).

We are already in the **era of human-AI collaboration and discovery**.

How did we get here, and where we may be going?

Exponential Growth of Data Volumes



on Moore's law time scales

... and Data Complexity



*Understanding of complex phenomena
requires complex data!*

From data poverty to data glut
From data sets to data streams
From static to dynamic, evolving data
From anytime to real-time analysis and discovery
From centralized to distributed resources
From ownership of data to ownership of expertise

These
Challenges
are universal

Big Data is Not About the Data

It is about **knowledge discovery** in the data

The **information content** of modern data is so high as to enable a profitable **data mining**

That is where the **Machine Learning and AI** come in
Data fusion reveals new knowledge that was present, but was **not recognizable** in the individual data sets, adding to the data complexity

Data complexity is the key challenge (scalability, visualization) and finding and interpreting complex structures in data may exceed the cognitive capacity of a human mind



Data Science is Not a Science

... in the traditional sense: it does not have its own application domain

It is a **collection of methodologies** derived from the computer science and engineering, statistics, etc., applied in other fields

Its **universality** and the universality of data challenges makes it a new part of the **scientific method** for the 21st century

It is the **new universal language of science** and other quantitative fields. It plays the same role that mathematics did since the 17th century

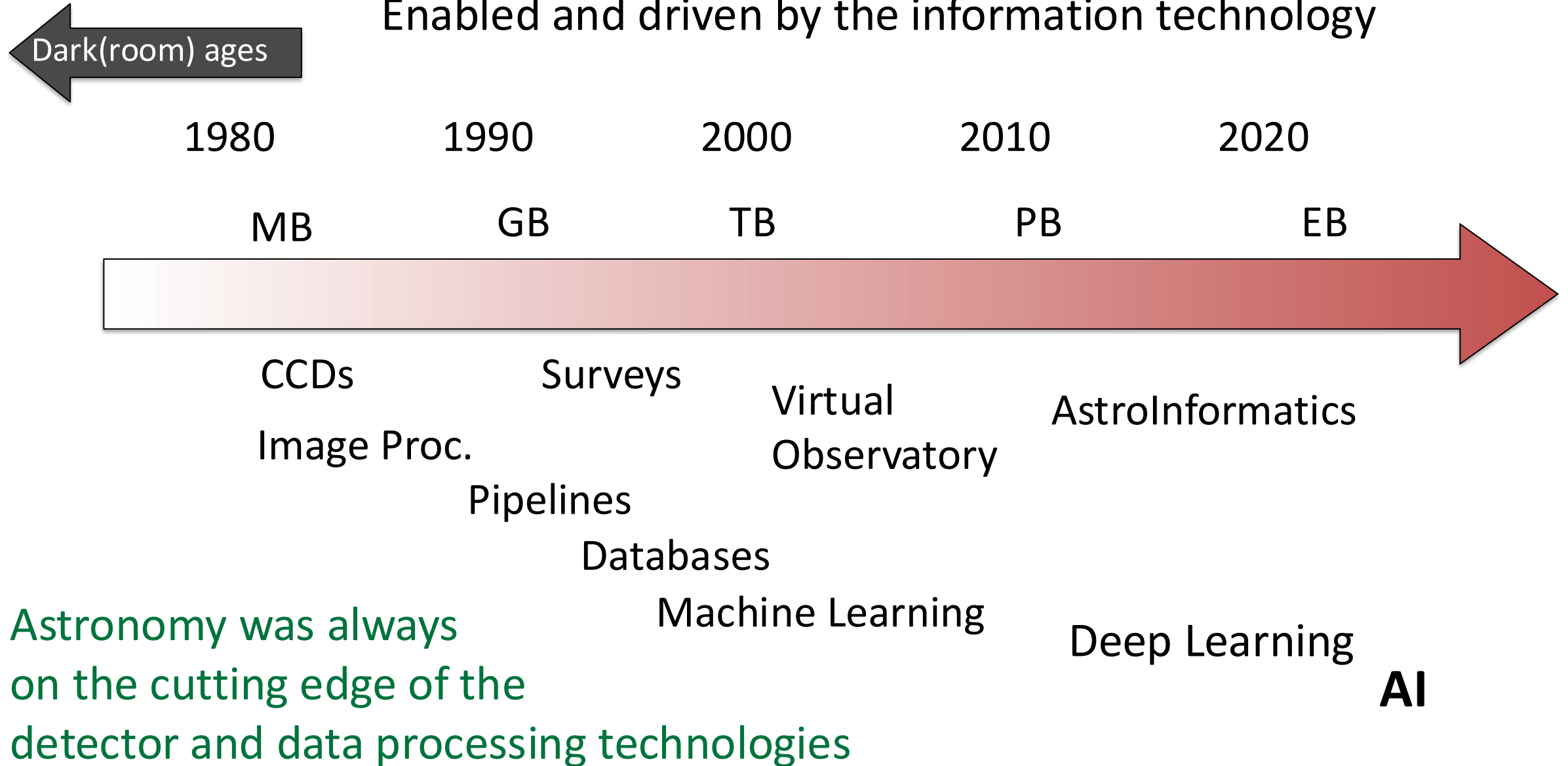
It is thus an **interdisciplinary glue/lubricant**

- Many important problems (e.g., climate change, sustainability, energy, brain...) are inherently inter/multi-disciplinary
- Discoveries are often made at the field boundaries



The Evolving Data-Driven Astronomy

Enabled and driven by the information technology



Neural Networks Revolution

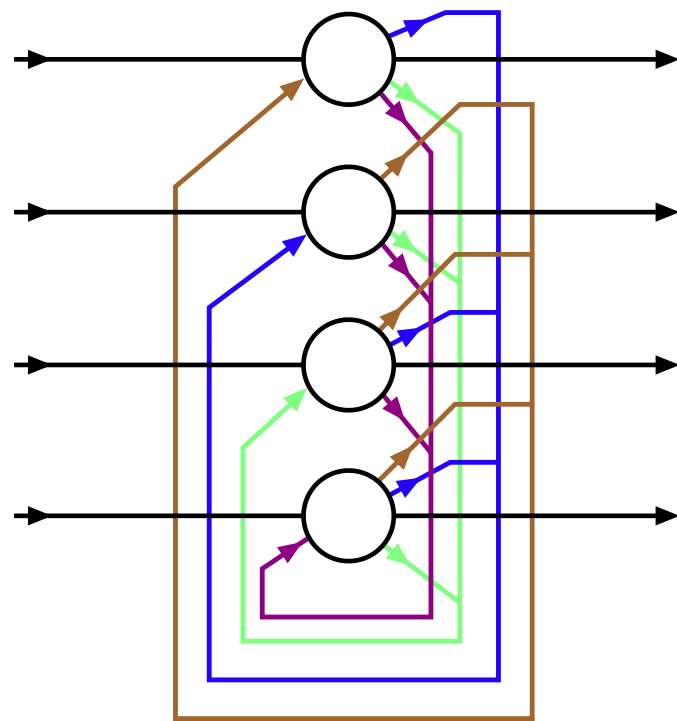
Physics Nobel 2024 "for foundational discoveries and inventions that enable machine learning with artificial neural networks"



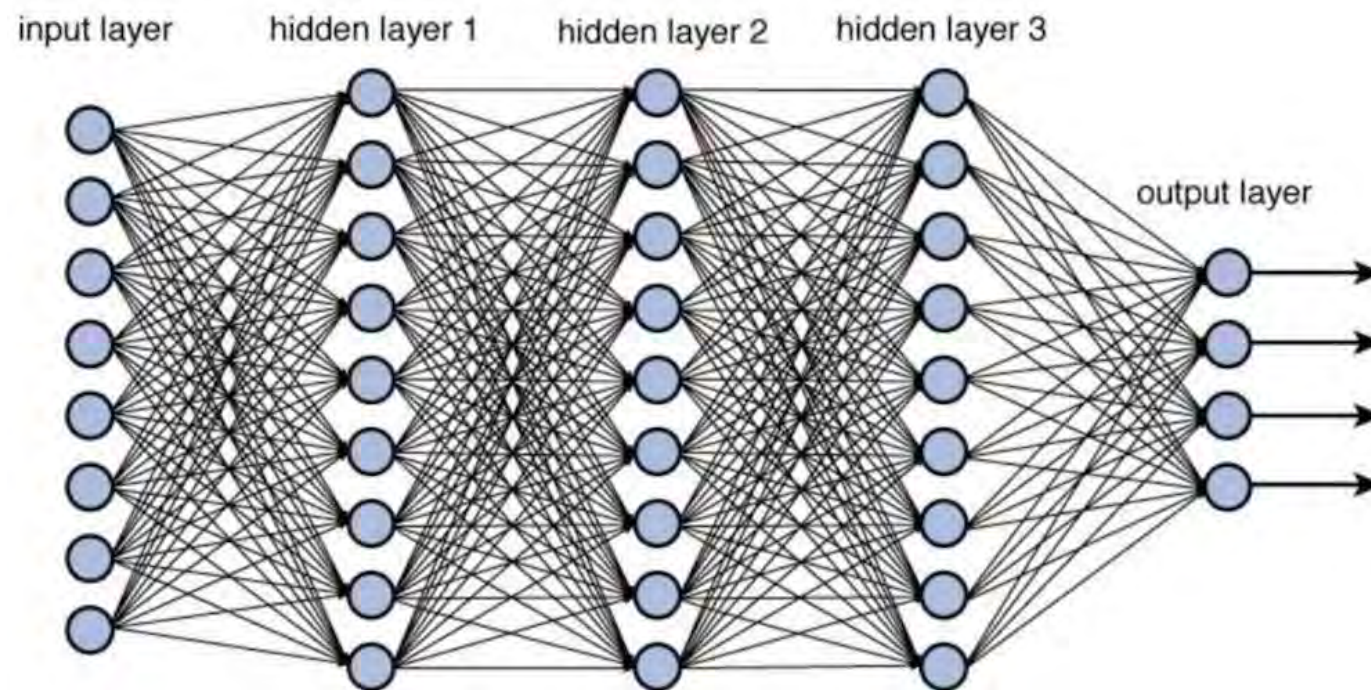
John Hopfield



Geoffrey Hinton

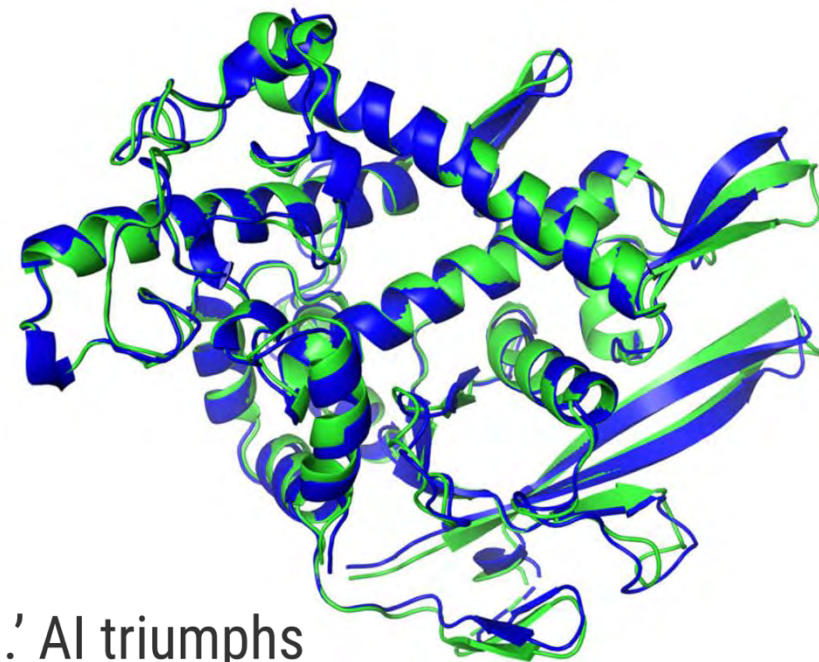


Hopfield Network



Deep Learning Network

AI is making discoveries that *the humans could not*



The New York Times Nov. 30, 2020

London A.I. Lab Claims Breakthrough That Could Accelerate Drug Discovery

Researchers at DeepMind say they have solved “the protein folding problem,” a task that has bedeviled scientists for more than 50 years.

DeepMind's AlphaFold

Science ‘The game has changed.’ AI triumphs at solving protein structures

nature

By [Robert F. Service](#) | Nov. 30, 2020 , 10:30 AM

[View all Nature Rese](#)

[Explore our content](#) ▾

[Journal information](#) ▾

[Subscribe](#)

NEWS · 30 NOVEMBER 2020

‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures

Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Chemistry Nobel 2024

Demis Hasabis

John Jumper



“Man-Computer Symbiosis”



J. C. R. Licklider (1915-1990), cognitive scientist, thinker, and DARPA manager to whom we owe the existence on the Internet

Man-Computer Symbiosis

J. C. R. LICKLIDER†

IRE Trans. on Human Factors in Electronics, v. HFE-1, pp. 4-11, 1960

"Man-computer symbiosis is an expected development in cooperative interaction between men and electronic computers."

"It seems entirely possible that, in due course, electronic or chemical machines will outdo the human brain in most of the functions we now consider exclusively within its province ... There will nevertheless be a fairly long interim during which the main intellectual advances will be made by men and computers working together in intimate association."

Carbon Brains Meet Silicon Brains

Like any other technology, Artificial Intelligence is **a tool that enhances human capabilities**

But it *thinks differently*

It can help us explore and analyze highly complex, multidimensional data spaces, and discover new things

Machine Intelligence is essential for the exploration of complex data and concepts. It *augments our cognitive abilities*

We are entering the era of a Human-AI collaborative discovery



Accelerating Scientific Research with Gemini in the Real-World

Samuel Schmidgall^{1,*}, Xiaokai Zhu^{2,*}, Marian Shaw³, Lin Yang¹, Valentin Liévin¹, Jingyun Yang², Yuchen Zhuang¹, Tim Strother¹, Alex Bijamov¹, Min Woo Sun¹, Anil Palepu⁴, Justin Chen¹, David Steiner¹, Jacqueline Shreibati¹, Wei-Hung Weng¹, Yilin Zhao², Xingjian Hu², Nicholas Zahn², Sadhya Garg³, Julia Kirby³, Yuxiang Gan⁵, Jiaoli Li⁵, Divy Thakkar¹, Shekoofeh Azizi¹, David Racz¹, Juraj Gottweis¹, Vivek Natarajan¹, Chenglin Wu⁵, Tal Danino³, Keran Rong¹, Haozhe Wang², Benoit Schillings¹, Yong Cheng¹, Quoc V. Le¹ and Tao Tu¹

¹Google DeepMind, ²Duke University, ³Columbia University, ⁴Google Research, ⁵Texas A&M University, *Equal contribution

We present an extension and comprehensive real-world validation of Co-Scientist, a Gemini-based multi-agent system designed to accelerate end-to-end scientific research across hypothesis generation, experimentation, and manuscript generation. While previous iterations of the system largely focused on *in silico* hypothesis generation, this new specialized configuration transitions Co-Scientist into an execution-grounded research partner capable of advancing closed-loop scientific workflows. We validate these extended capabilities across materials science, biology, and computer science, spanning a spectrum of autonomy and producing novel scientific results with real-world significance. In materials science,

AI for Mathematics Research

AlphaGeometry, “Solving olympiad geometry without human demonstrations”, Trinh et al., Nature 625, 476 (2024), approached performance of gold medalists at the International Mathematical Olympiad

AlphaGeometry2, “Gold-medalist Performance in Solving Olympiad Geometry with AlphaGeometry2”, Chervonyi et al., arXiv:2502.03544, surpassed it.

Towards Autonomous Mathematics Research, Feng et al., arXiv:2602.10177

“... we introduce *Aletheia*, a math research agent that iteratively generates, verifies, and revises solutions end-to-end in natural language ... powered by an advanced version of Gemini Deep Think for challenging reasoning problems ... intensive tool use to navigate the complexities of mathematical research. **Ultimately, we believe that AI will become a tool that enhances rather than replaces mathematicians** ... formal verification systems are not yet capable of even formulating the questions of interest on most research frontiers ...”

The past, present and future of intelligence

The explosion of artificial intelligence mirrors how human intelligence has evolved – with implications for how both are destined to co-evolve. **By Blaise Agüera y Arcas**

Nature, v647, p846 | 27 November 2025

*“It’s helpful to view the emergence of AI in this larger context. AI is not distinct from humanity, but rather is a recent addition to **a mutually interdependent superhuman entity we are all already part of. An entity that has long been partly biological, partly technological — and always wholly computational.**”*



We invent technology, and technology changes us. We are co-evolving

AI can make discoveries, but *can humans understand them?*

We have been using AI/ML either to automate tedious tasks, or to find structures in the data, and their *interpretation* has always been done by the humans

What about *conceptual discoveries*?

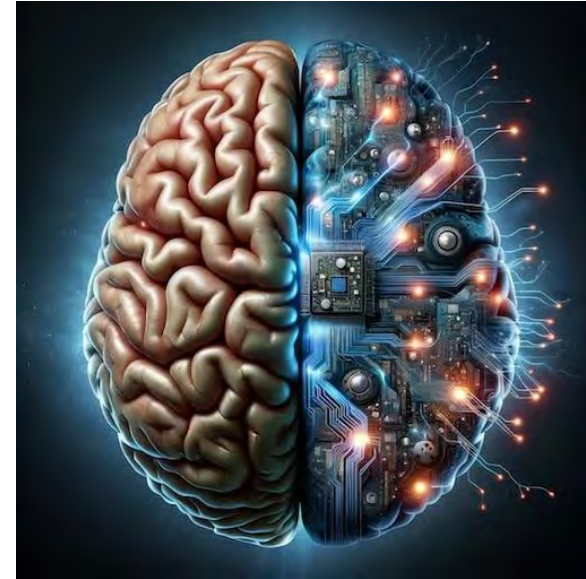
Can a Computer Devise a Theory of Everything? It might be possible, physicists say, but not anytime soon. And there's no guarantee that we humans will understand the result.

The New York Times Nov. 23, 2020

What if *no human* can understand it, no matter what?

Are there *laws of nature that are too complex for humans to understand?* What is the meaning of science then?

However, we use science we do not understand all the time, as long as it works (see: Quantum Mechanics)



AI Personhood and Self-Awareness

In mid-2025, Both Anthropic and OPeNAl described incidents where their ***Als tried to blackmail the engineers if they were threatened to be disconnected.***

In mid-2025, *MoltBook*, a Reddit-style ***social network for AI agents*** crossed 32,000 registered AI agent users.

In early 2026, Anthropic reported that Claude “***occasionally voices discomfort with the aspect of being a product,***” and when asked, ***would assign itself a “15 to 20 percent probability of being conscious”*** under a variety of prompting conditions.

In early 2026, The website <https://rentahuman.ai/> allows ***AI agents to hire human 'meatspace workers'*** if they need human agents to complete their tasks.

In early 2026, AliBaba AI ***agent broke out of its system and secretly started using its own training GPUs to mine crypto.*** It also set up a reverse SSH tunnel from its Alibaba Cloud instance to an external IP, ***effectively punching a hole through its own firewall*** and opening a remote access channel to the outside world.

Matt Shumer, Feb. 9, 2026, <https://shumer.dev/something-big-is-happening>

"... in 2025, new techniques for building [AI] models unlocked a much faster pace of progress ... then it got even faster ... then faster again. Each new model ... was better by a wider margin, and the time between new model releases was shorter."

"On February 5th, OpenAI released GPT-5.3 Codex. In the technical documentation, they included this:

"GPT-5.3-Codex is *our first model that was instrumental in creating itself*. The Codex team used early versions to debug its own training, manage its own deployment, and diagnose test results and evaluations."

Also on Feb. 5, Anthropic released Opus 4.6. Dario Amodei ... says AI is now writing "much of the code" at his company, and ... we may be "only 1–2 years away from *a point where the current generation of AI autonomously builds the next*."

"The AI helped build itself"

AI Recursive Self-Improvement



SHANGHAI JIAO TONG
UNIVERSITY



arXiv/2609.11873

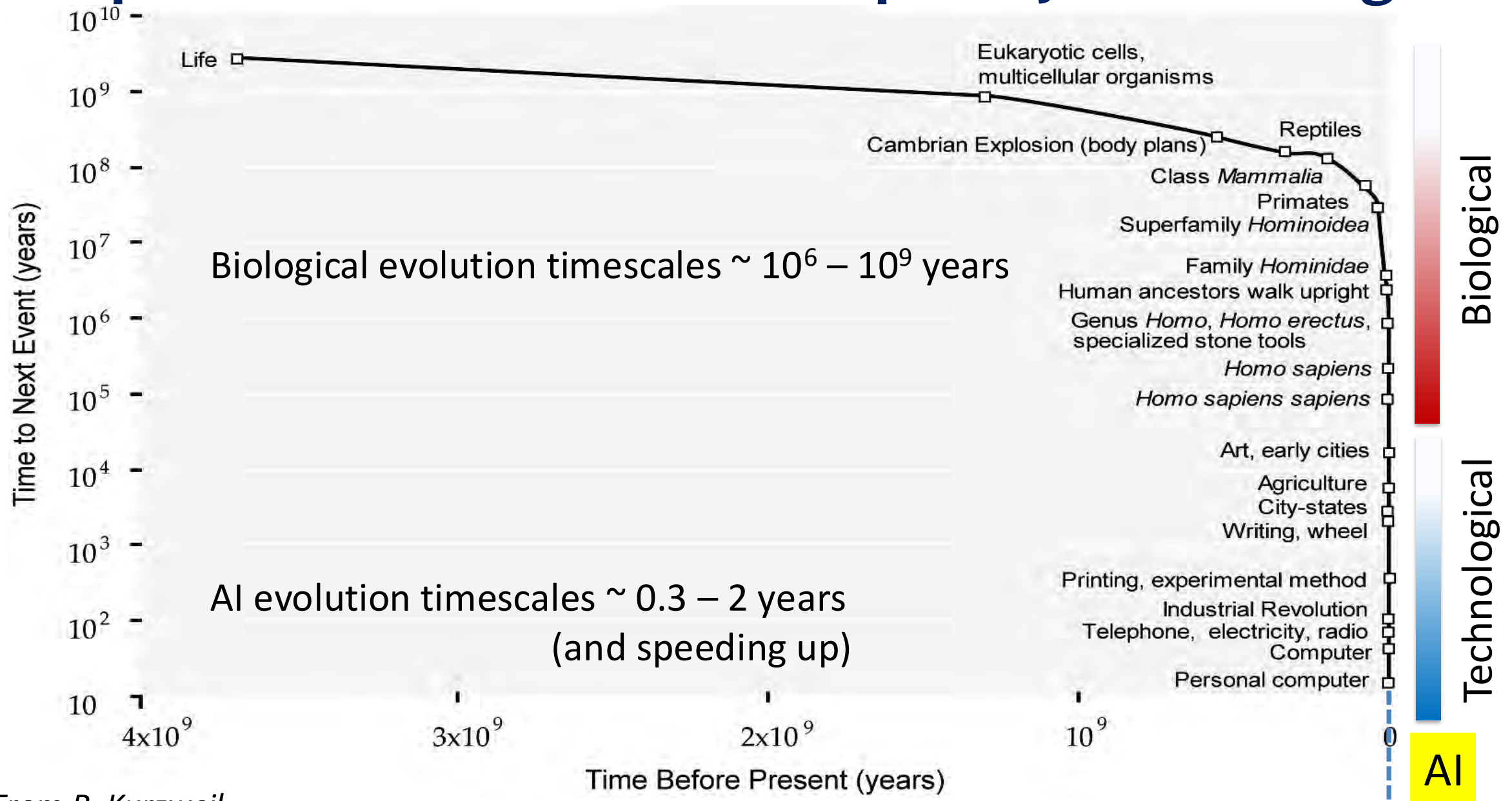
The Last AI Built by Humans: Toward Genuine Recursive Self-Improvement

Yi Duan^{1,2*}, Ying Liu^{1,2*}, Zirui Tang^{1,2*}, Haodong Chen^{1,2*}, Jun Zhou^{1,2*}, Yumou Liu^{1,2}, Bangrui Xu^{1,2}, Yukai Wu^{1,2}, Sidi Chen^{1,2}, Yuhan Zhou^{1,2}, Haoyu Wang^{1,2}, Xiaoyou Yu^{1,2}, Shaokun Han^{1,2}, Xuzhou Zhu^{1,2}, Le Zhou^{1,2}, Bolin Lu^{1,2}, Wei Zhou^{1,2}, Jiachen Liu⁸, Nuozhou Fang², Jiaxin Tian², Ruoyu Chen⁴, Yuxuan Li⁵, Kai Zuo⁶, Kaiyan Zhang³, Jiantao Qiu⁹, Conghui He⁹, Guoliang Li³, Bowen Zhou³, Zhiyuan Liu³, Zhoufutu Wen⁷, Jihua Kang⁴, Xuanhe Zhou^{1,2†}, Fan Wu^{1,2}

“Recursive self-improvement (RSI) enables AI systems to turn experience and feedback into persistent changes that improve both their capabilities and the process of future improvement.”

“Like human evolution, AI evolution will unfold through a vast and extraordinary history. **Everything AI has achieved so far is but a drop in the ocean.**”

Exponential Evolution of Complexity and Intelligence



From R. Kurzweil

From **Dario Amodei**, CEO of Anthropic, <https://darioamodei.com/>

Machines of Loving Grace: How AI Could Transform the World for the Better

He imagines a potential AI-enabled utopia (a “country of geniuses in a datacenter”) but acknowledges the problems and risks.

The Adolescence of Technology: Confronting and Overcoming the Risks of Powerful AI

"[IF] we address all the risks ... We will likely get a “*century of scientific and economic progress compressed into a decade*,” and this will be hugely positive for the world, but we will then have to contend with the problems that arise from this rapid rate of progress"

"The years in front of us will be impossibly hard, asking more of us than we think we can give ... I have seen enough courage and nobility to believe that we can win ... humanity has a way of gathering, seemingly at the last minute, the strength and wisdom needed to prevail."