From Automation to Discovery:
The Next Evolution of AI Research

MATHEMATICS, HIGH PERFORMANACE SIMULATION AND AI

Prof (Dr) Simon See

Prof (Adjunct) Professor Surrey, Coventry Univ, NTU

Distinguish Fellow, Fudan Uni

And Head of AI Technology Center, NVIDIA

ANTIKYTHERA MECHANISM

ANTIKYTHERA MECHANISM


Key Inventions That Laid the Foundation of Modern Automation
1733
Flying Shuttle by John Kay
1764
Spinning Jenny by James Hargreaves
1768
Water Frame by Richard Arkwright
1769
Steam Engine by James Watt
1784
Power Loom by Edmund Cartwright
1793
Cotton Gin by Eli Whitney
1831
Mechanical Reaper by Cyrus McCormick
1807
Steamboat by Robert Fulton
1837
Telegraph by Samuel Morse
1856
Bessemer Process by Henry Bessemer
skill-mine.com

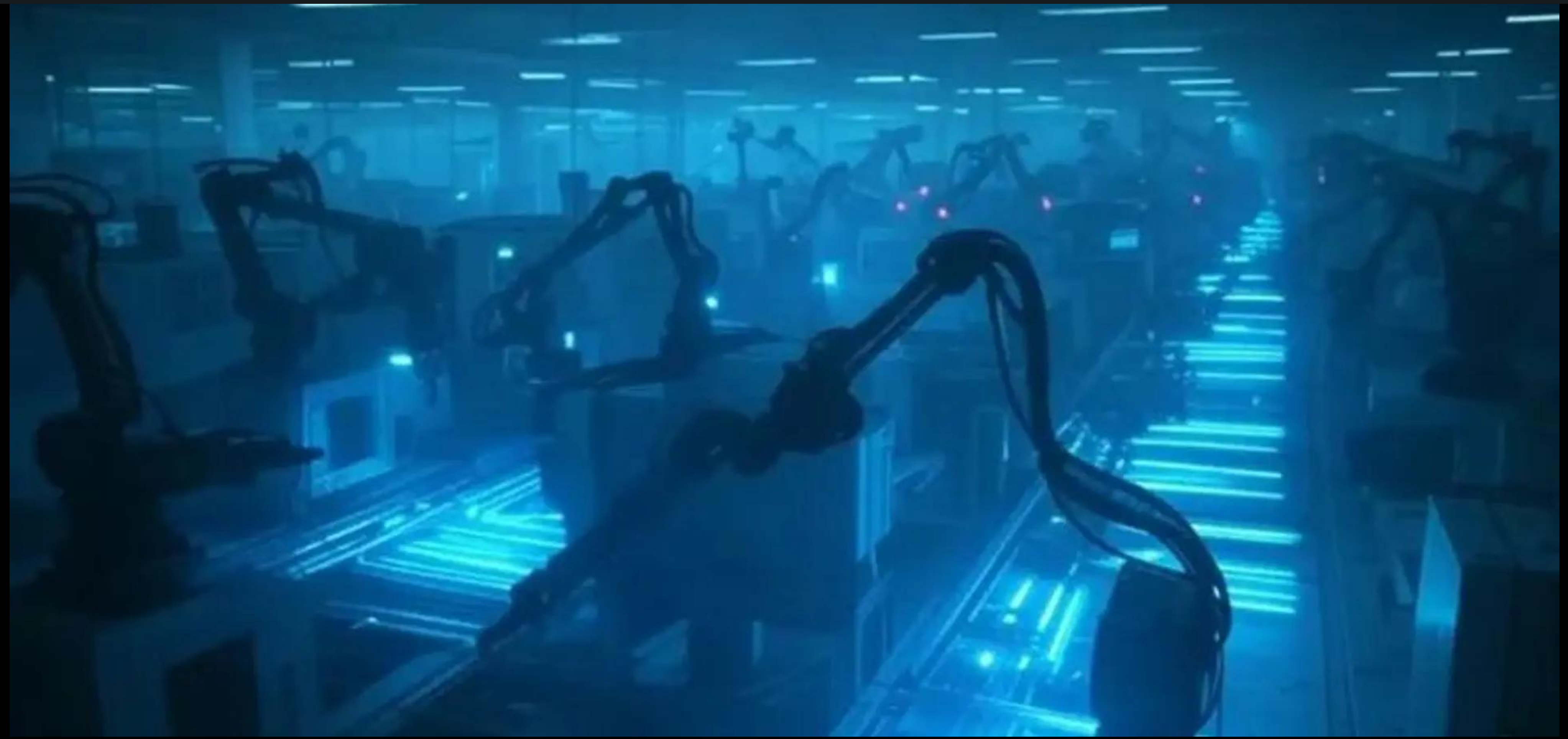

定位高端 品质优先 创新驱动

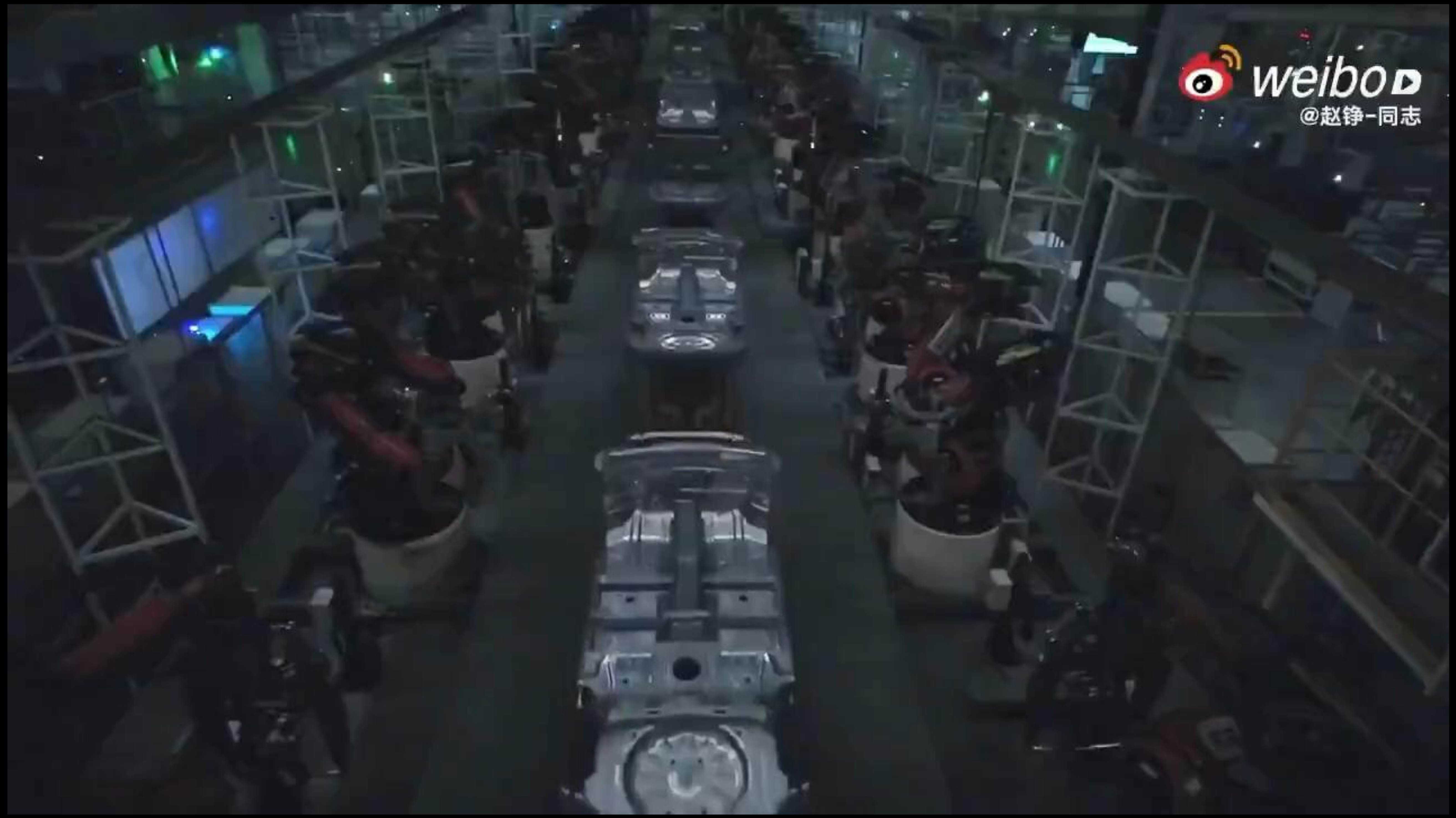

KION Group


INTERESTING ENGINEERING
FANUC America Corporation

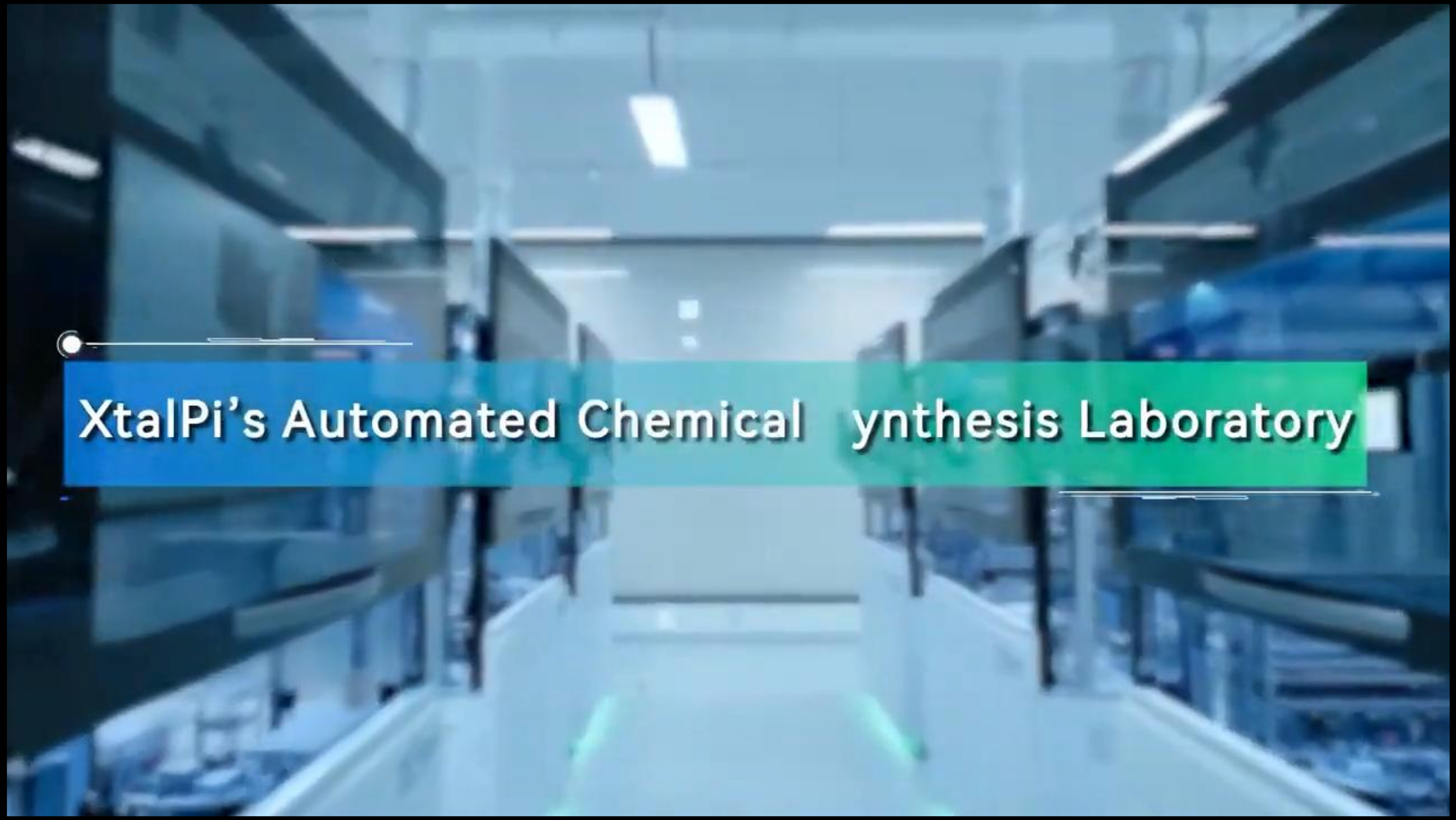

Direct INDUSTRY / Emag

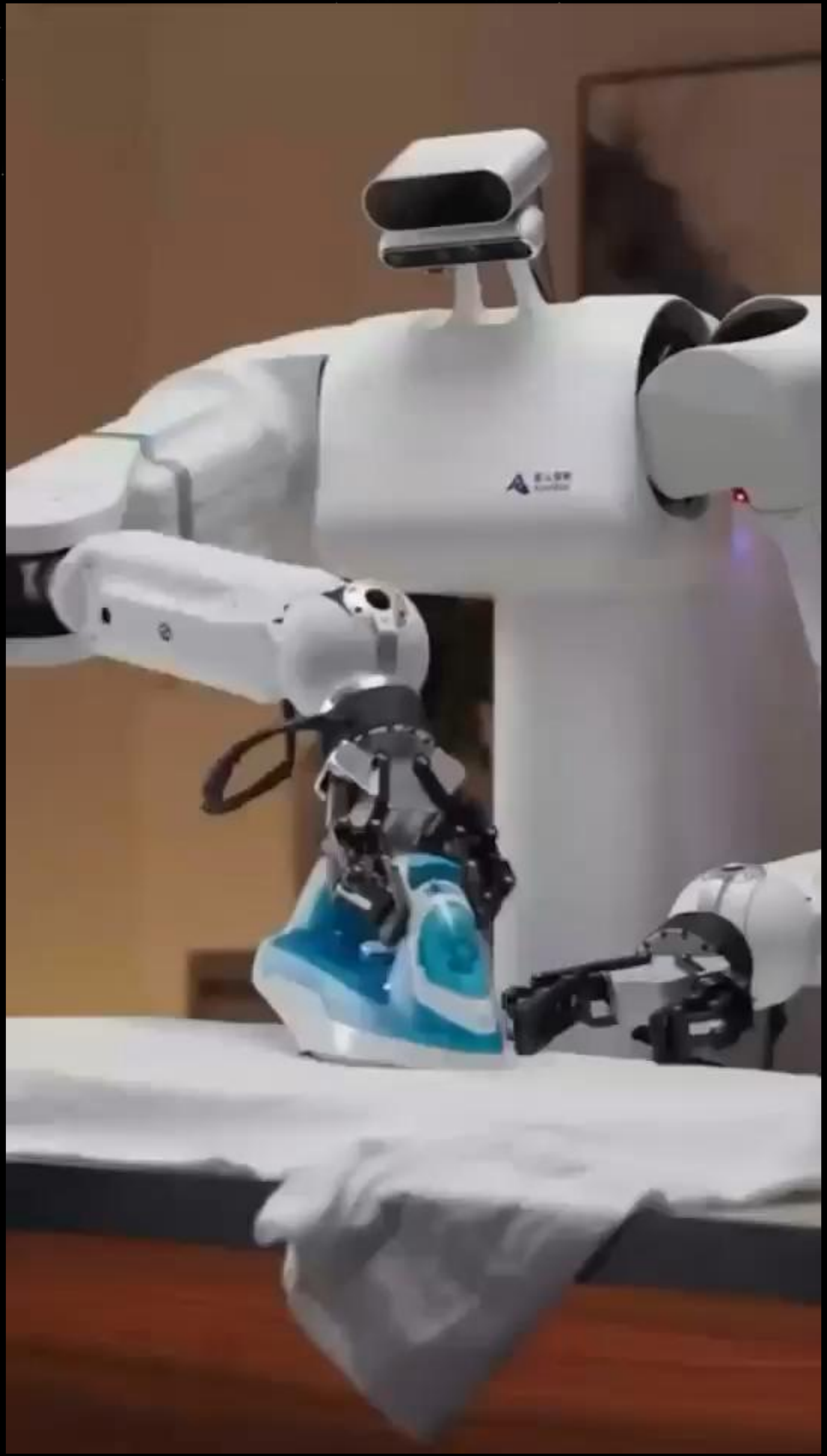
Dark Factories and the Future of Work: How AI-Driven Automation is Reshaping Manufacturing





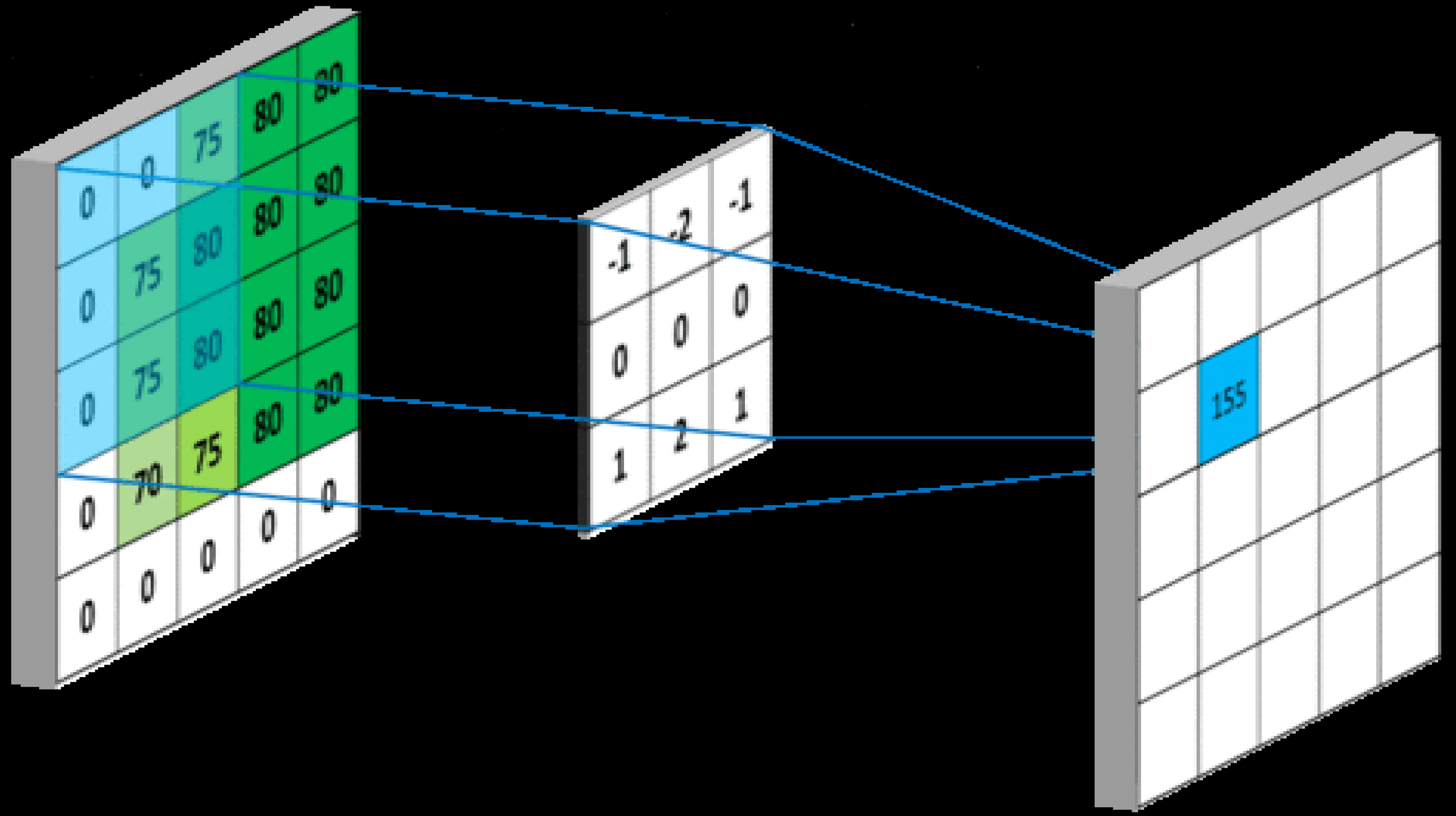


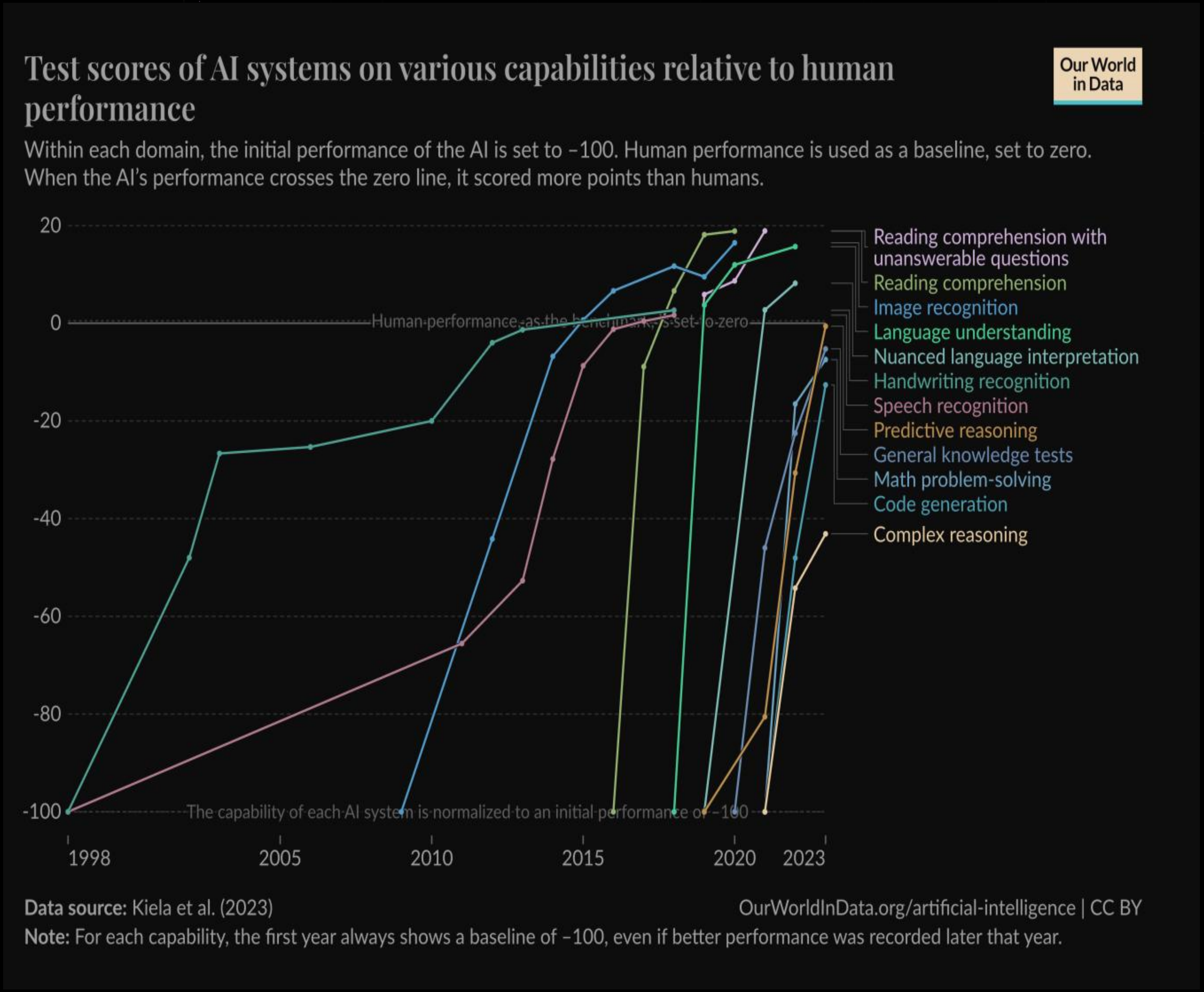

HAMILTON

XtalPi's Automated Chemical ynthesis Laboratory




*为方便观看视频已做3倍速处理

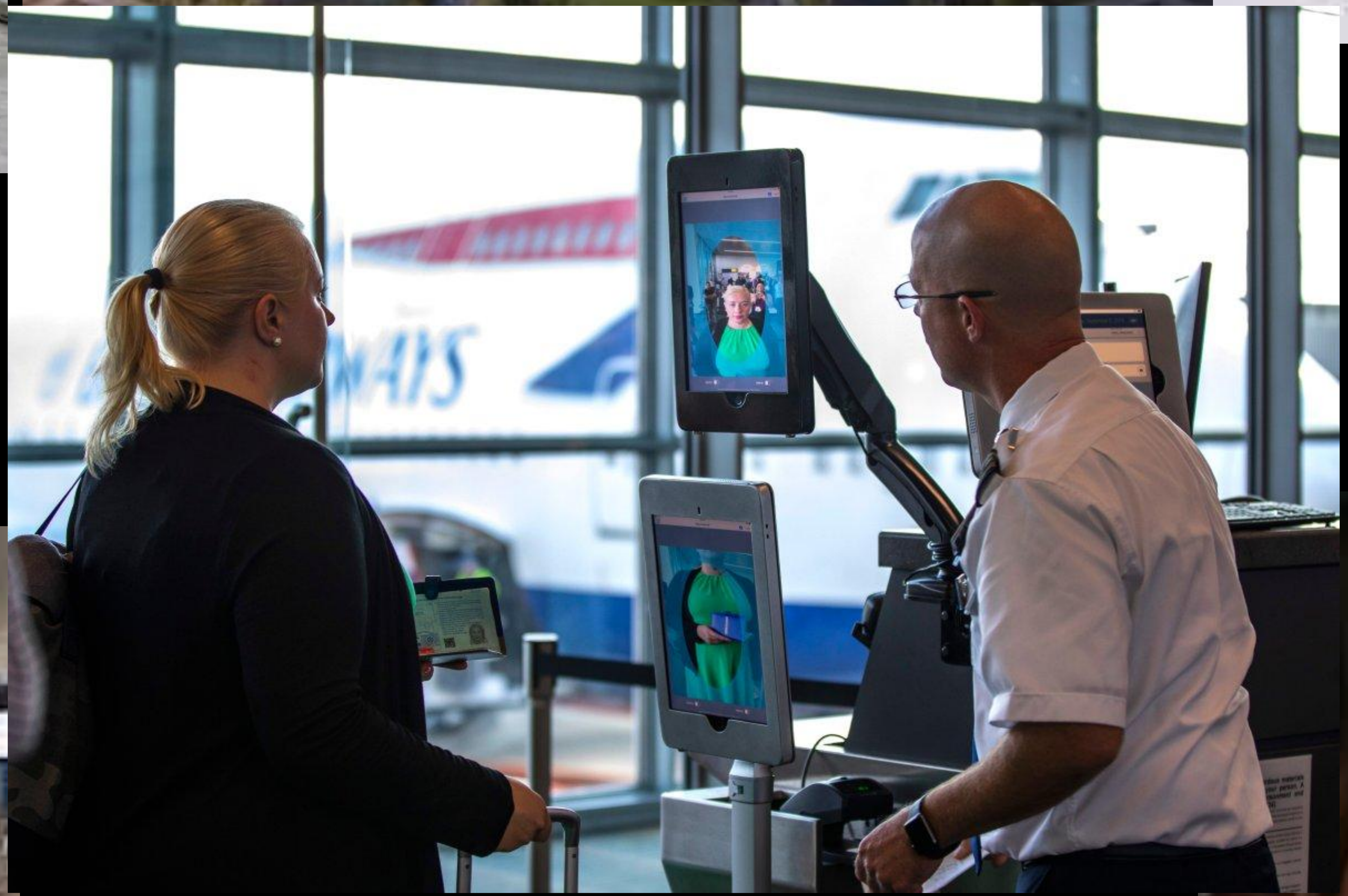
IS IT ABLE TO DO DISCOVERY ???



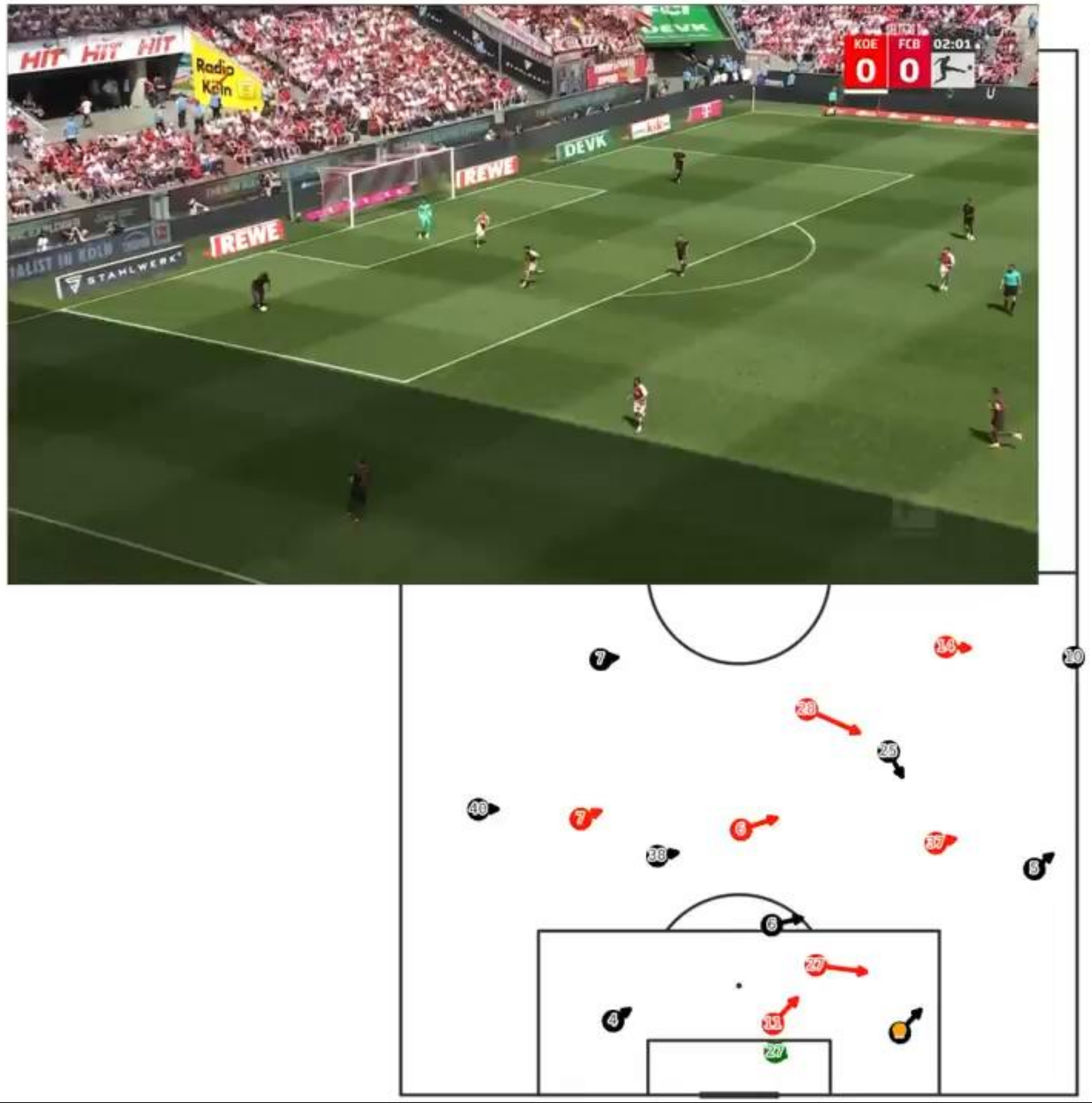

Test scores of AI systems on various capabilities relative to human performance
Our World in Data
Within each domain, the initial performance of the AI is set to −100. Human performance is used as a baseline, set to zero. When the AI's performance crosses the zero line, it scored more points than humans.
20
0
-20
-40
-60
-80
-100
Human performance, as the benchmark, is set to zero
The capability of each AI system is normalized to an initial performance of -100
1998
2005
2010
2015
2020
2023
Reading comprehension with unanswerable questions
Reading comprehension
Image recognition
Language understanding
Nuanced language interpretation
Handwriting recognition
Speech recognition
Predictive reasoning
General knowledge tests
Math problem-solving
Code generation
Complex reasoning
Data source: Kiela et al. (2023)
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.

PREDICTIVE AI


10
9
8
7

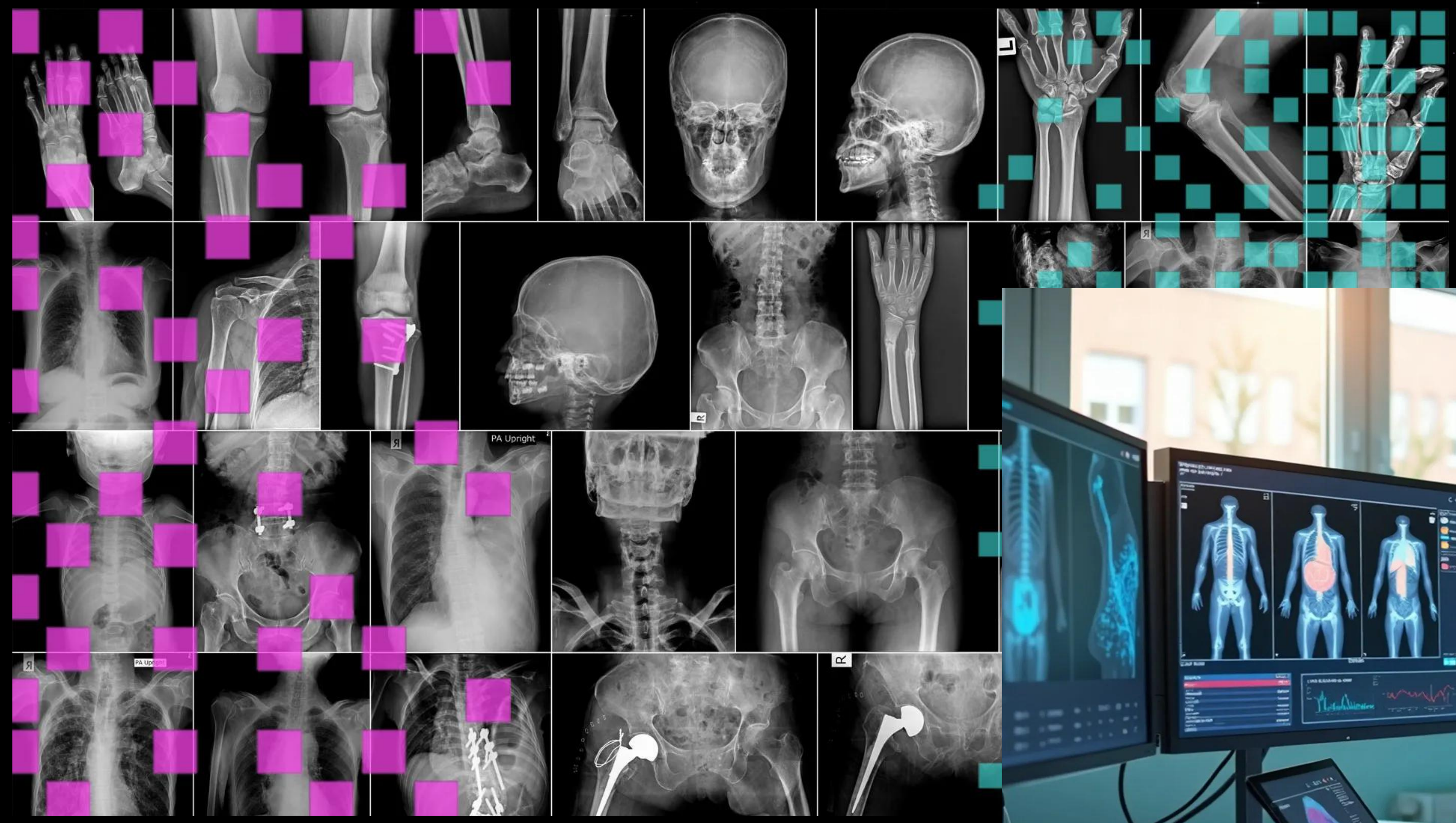


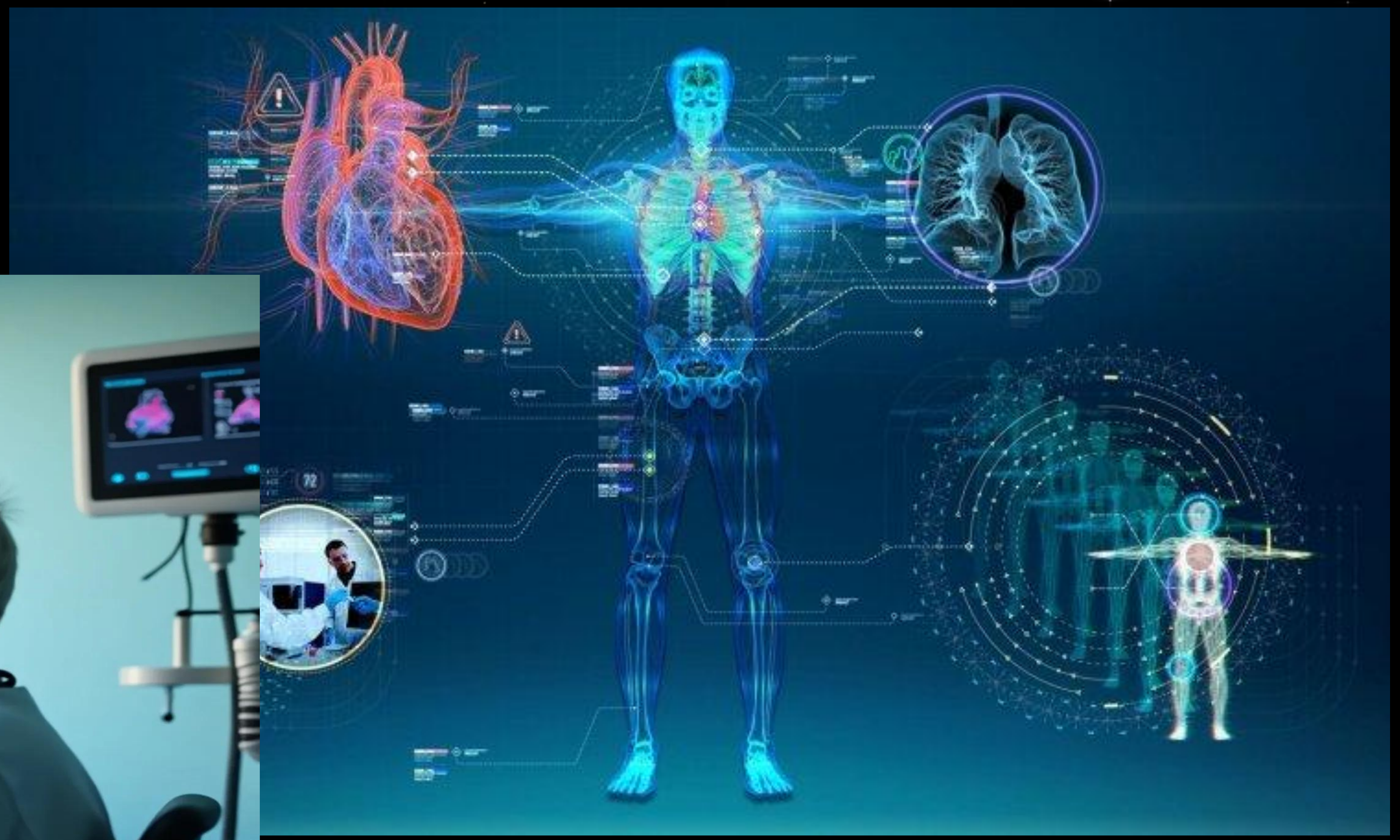

Landed 1 Thrown 3
Time Type Status Quality
02:49 L Hook Head Landed
02:50 R Hook Body Missed
02:56 L Straight Head Missed
Landed 1 Thrown 4
Time Type Status Quality
02:52 R Straight Head Missed
02:52 L Straight Head Landed
02:53 L Uppercut Head Missed
02:55 R Hook Head Missed
Balance
Front Foot Neutral Back Foot
Stance
Orthodox Squared Southpaw
Distance
Outside Mid-Range Inside
Balance
Front Foot Neutral Back Foot
Stance
Orthodox Squared Southpaw
Distance
Outside Mid-Range Inside
RND 9 of 12 2:49
jabbr.ai
DeepStrike 0.1.0-beta
TECATE
Hennessy

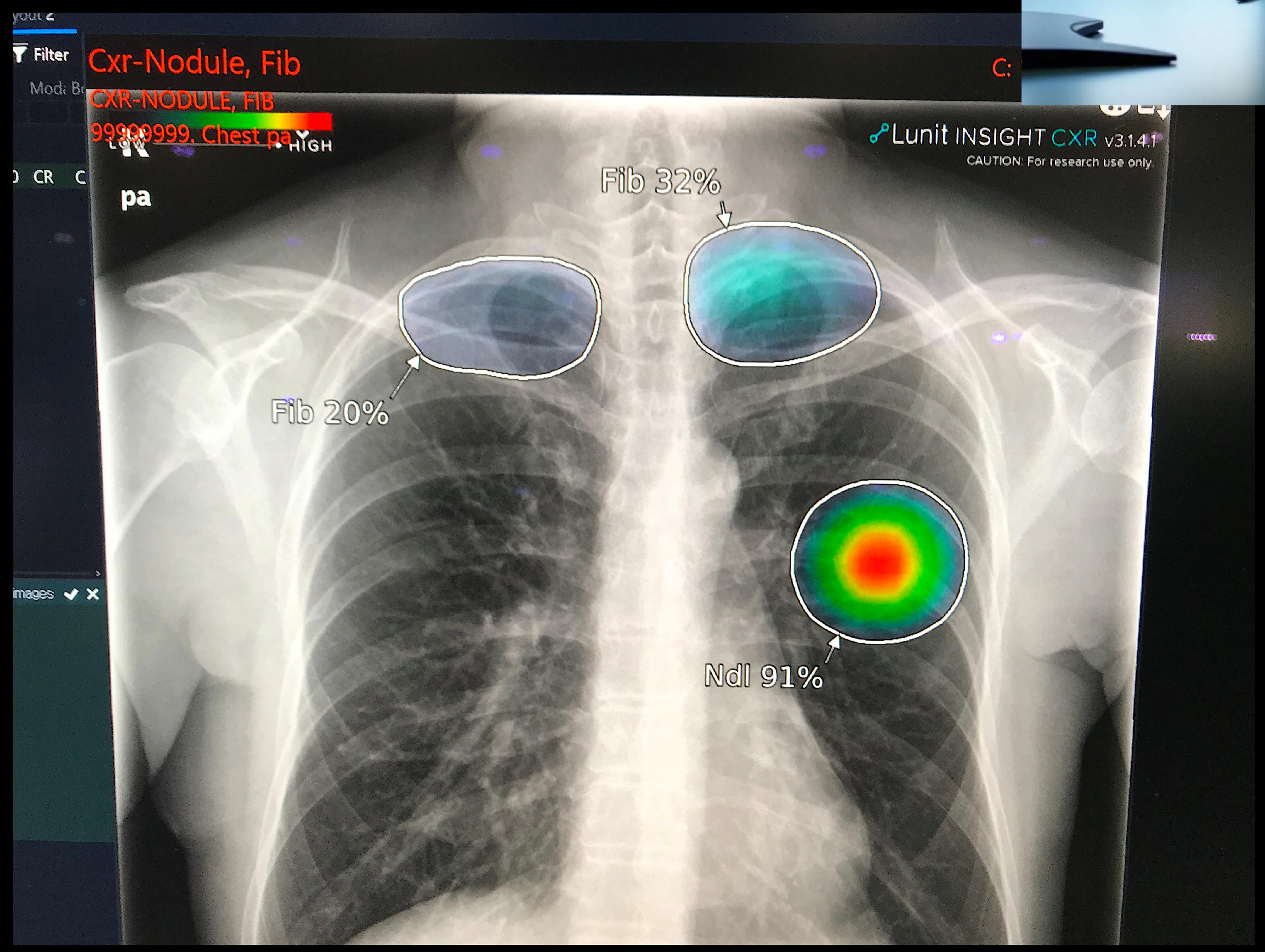


Active Pressing [v > 2.0m/s]
FC Bayern München

1. FC Köln	25	10	27	6	11	5	7	2	4	40	38
14	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
27	0 %	0 %	2 %	3 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %
2	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
20	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
11	0 %	0 %	4 %	0 %	0 %	0 %	0 %	0 %	1 %	0 %	0 %
4	0 %	0 %	0 %	0 %	0 %	0 %	4 %	0 %	0 %	0 %	0 %
37	30 %	0 %	0 %	2 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %
28	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
6	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	23 %
7	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	36 %	0 %
24	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %





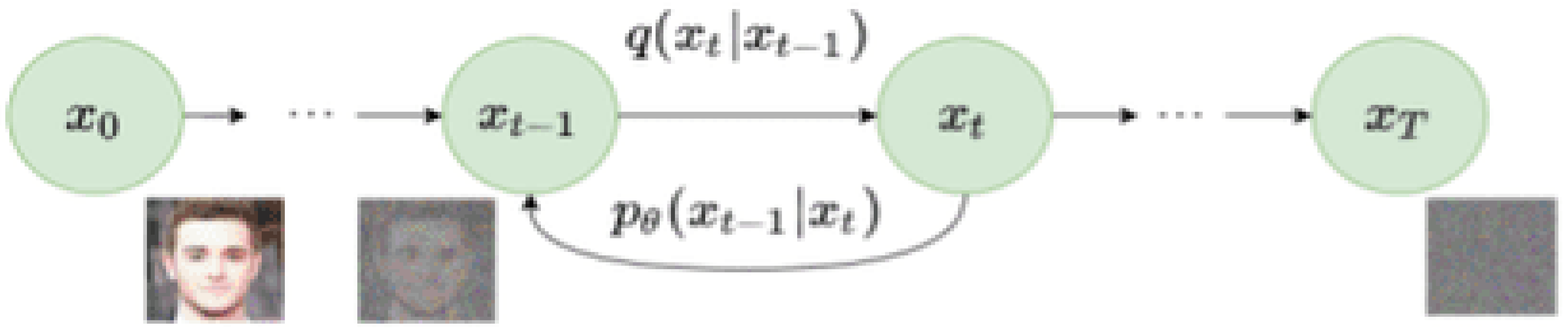

Filter
Cxr-Nodule, Fib
CXR-NODULE, FIB
99999999. Chest pa
HIGH
Lunit INSIGHT CXR v3.1.4.1
CAUTION: For research use only.
pa
Fib 32%
Fib 20%
Ndl 91%
images

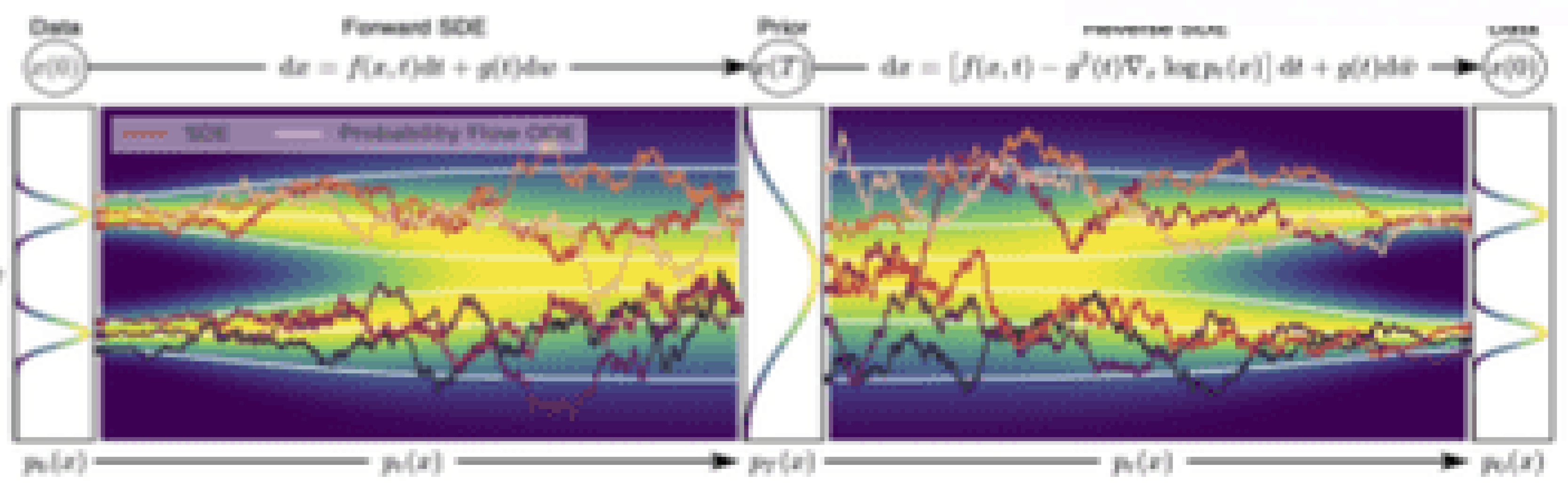
WHAT IS GENERATIVE AI?

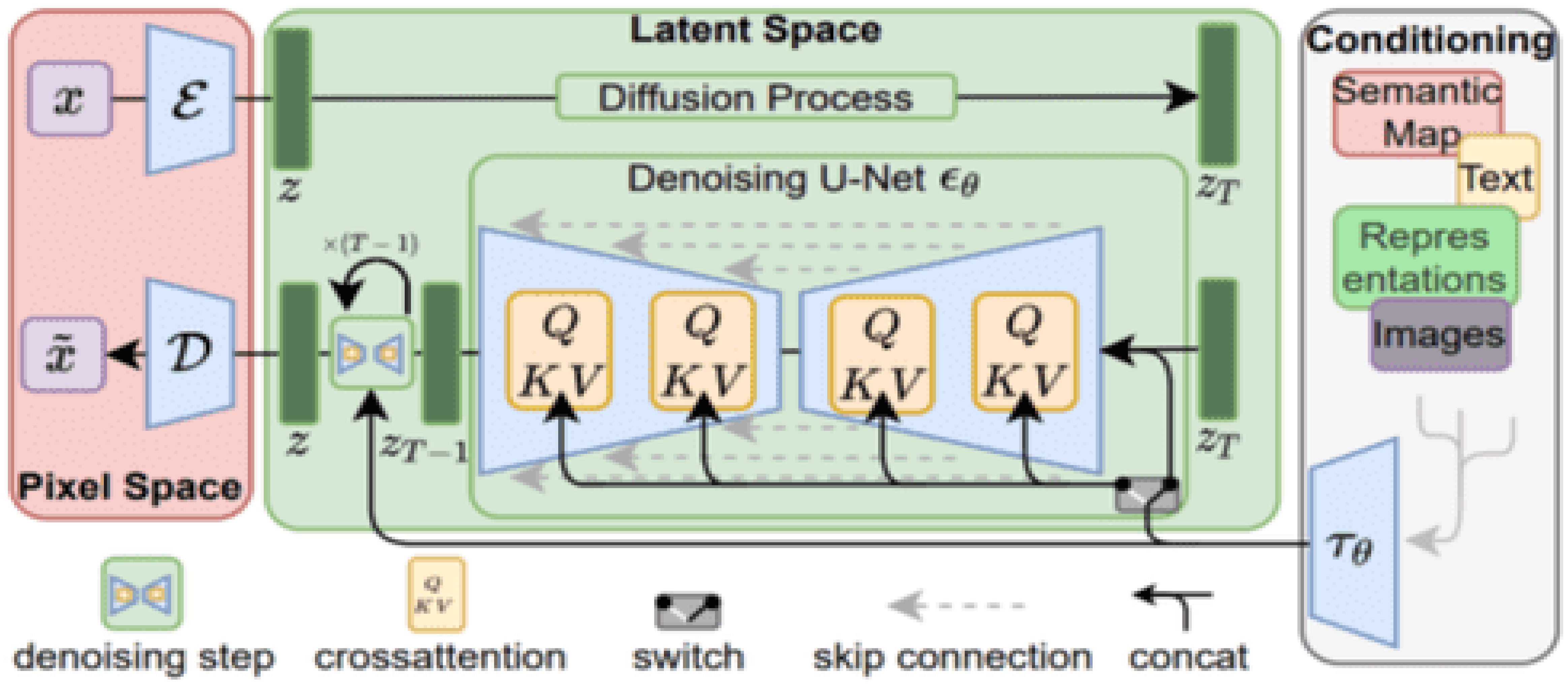
https://www.youtube.com/watch?v=wMmqCMwuM2Q&ab_channel=MITCBMM



Sora Video-to-Video







Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar\alpha_t}\mathbf{x}_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar\alpha_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

LEAP

High-resolution image reconstruction with latent diffusion models from human brain activity

Yu Takagi[1,2], Shinji Nishimoto[1,2]
[1] Osaka University, [2] NICT

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

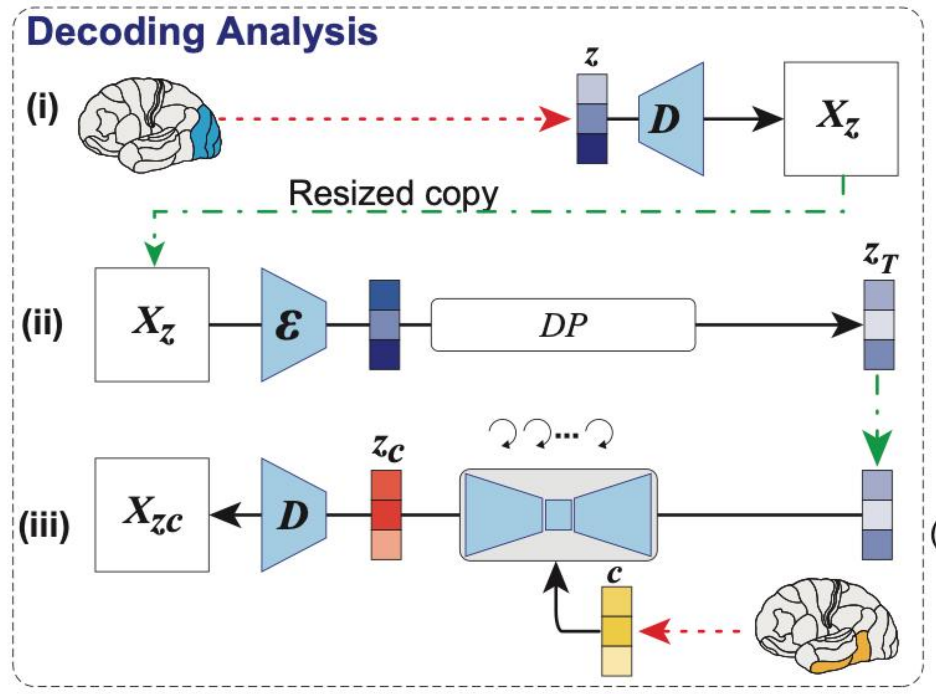
↑ Project website

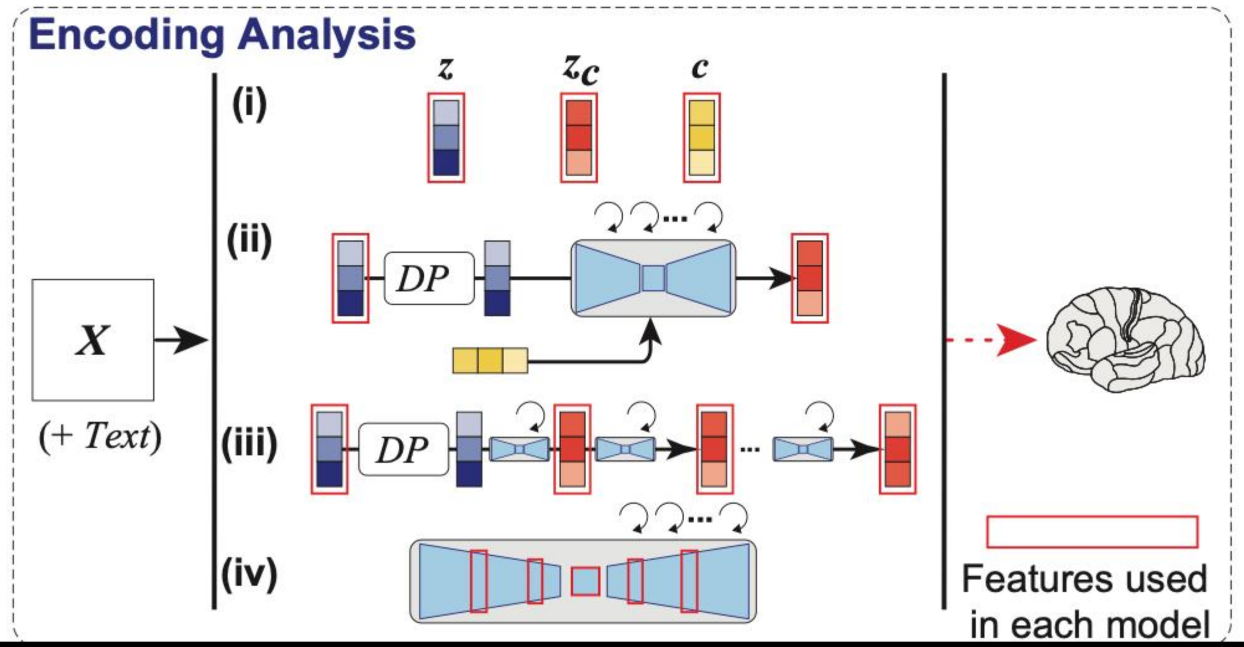
Introduction

- Latent Diffusion Models (LDMs) are deep generative models that have high text-to-image generative performance [1].
- We conducted decoding analysis [2,3]: reconstruct visual images from fMRI signals using Stable Diffusion (SD).
- We also conducted encoding analysis [4,5]: provide biological interpretations of each component of SD.

Methods

- We used Natural Scenes Dataset (NSD) [6].
- Scanned by 7-Tesla fMRI scanner over 40-hours
- Each subject viewed 10,000 MS-COCO images 3 times.



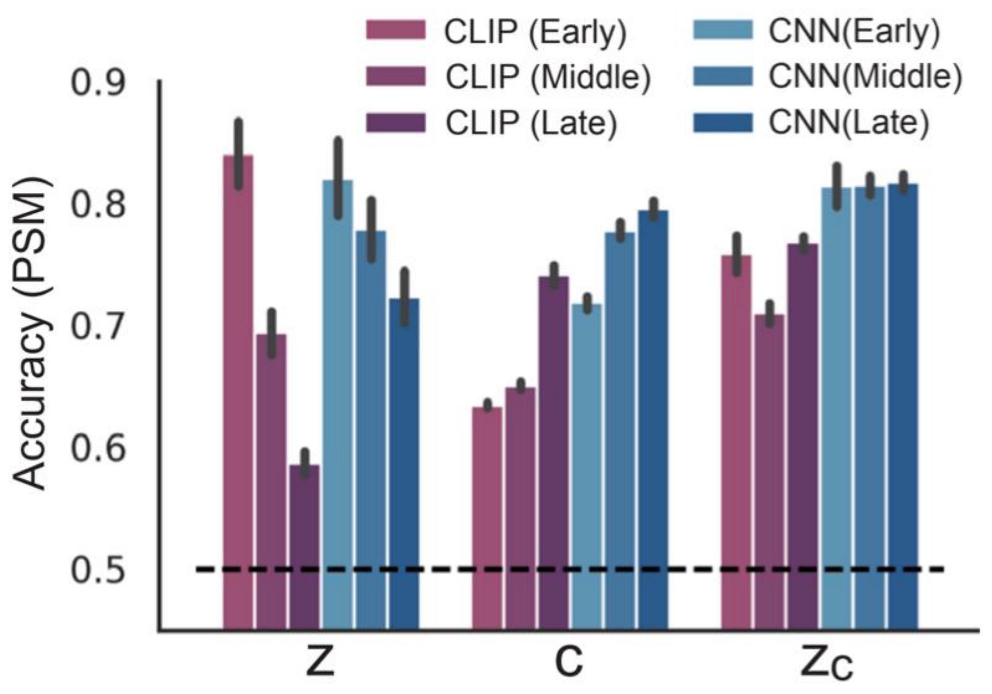


Results: Decoding analyses

Single subject result

Across subjects result

Quantitative Evaluation (2-way identification accuracy)

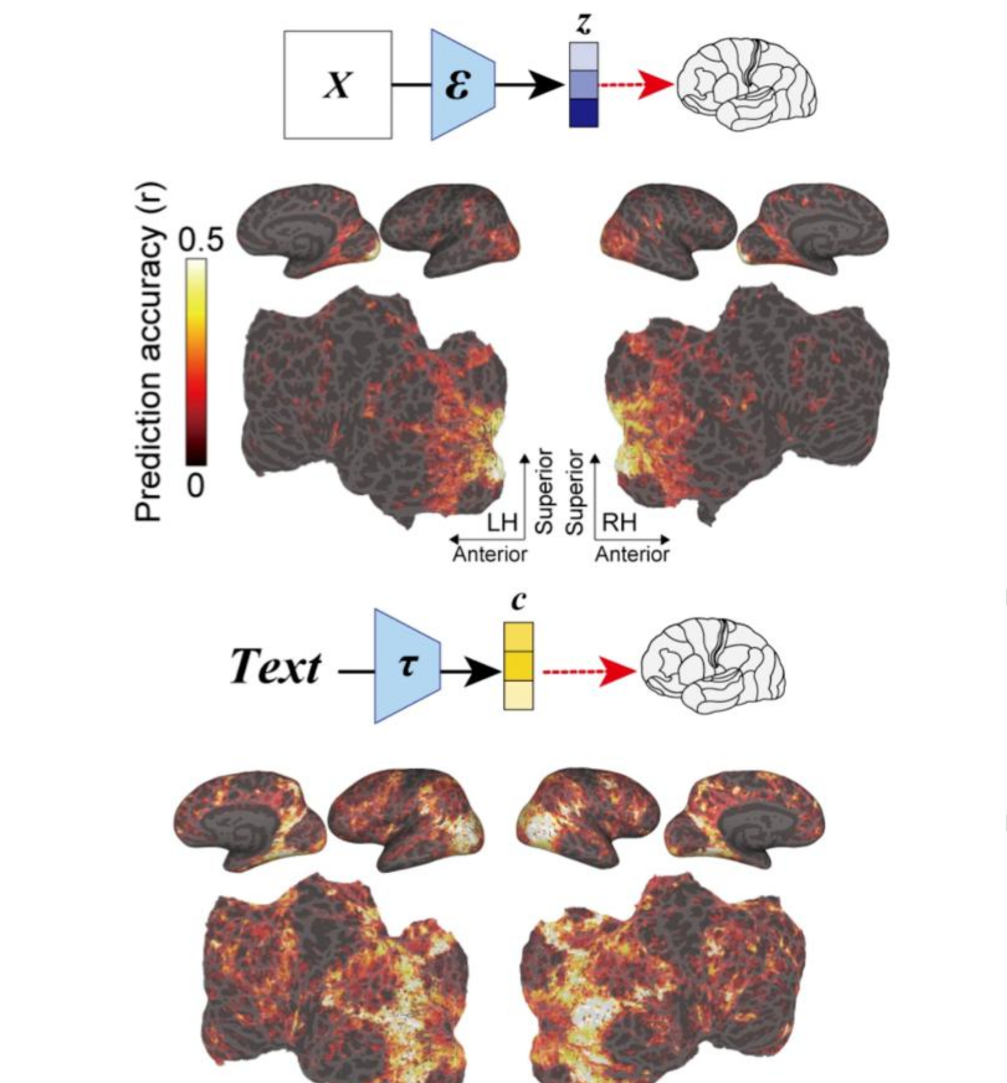


Control Analyses

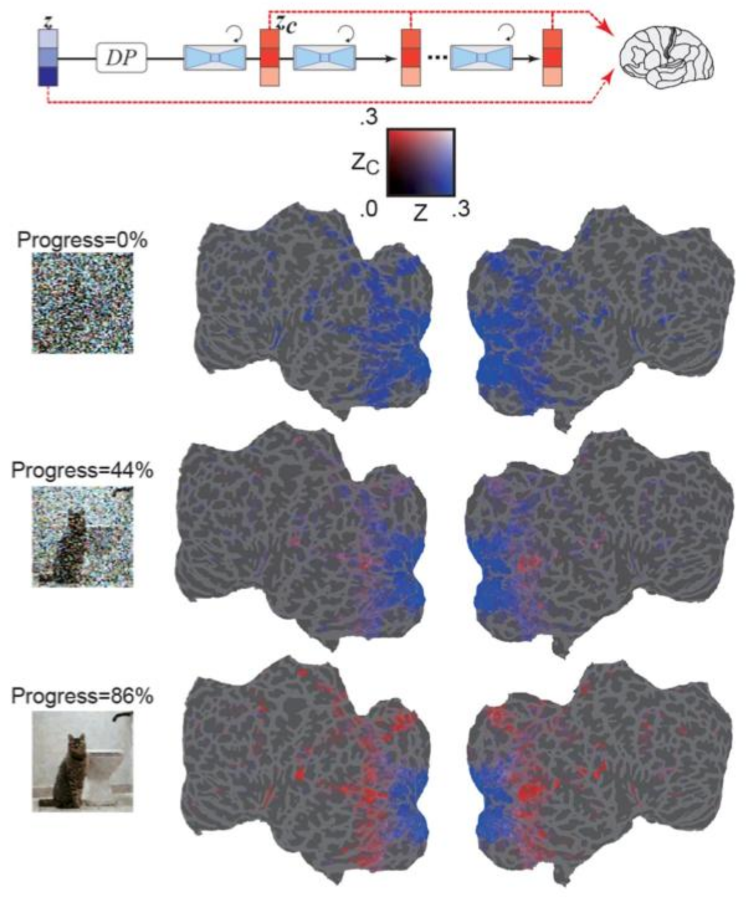
- Decoding accuracy decreases when building a semantic decoder using only image category labels, rather than caption.
- Even if we used Stable Diffusion v2.0 trained with OpenCLIP (not trained on MS-COCO) and tested on non-overlpped images, there was no change in the quantitative evaluation.

Results: Encoding analyses

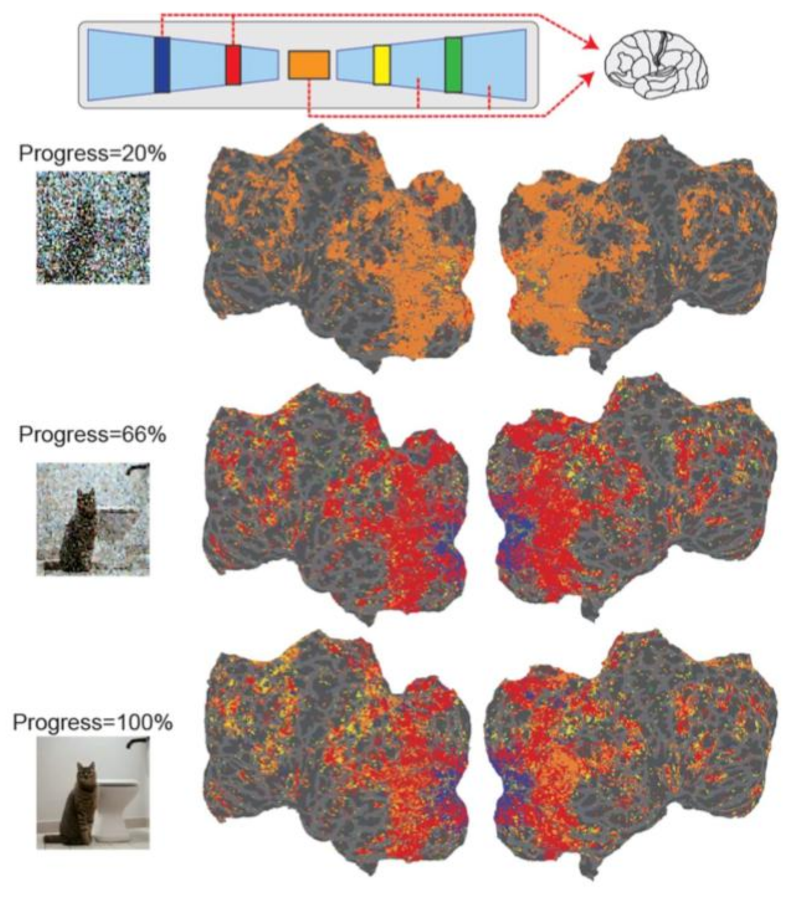
Understanding latent representations by mapping them onto the brain



Representational change through the denoising process



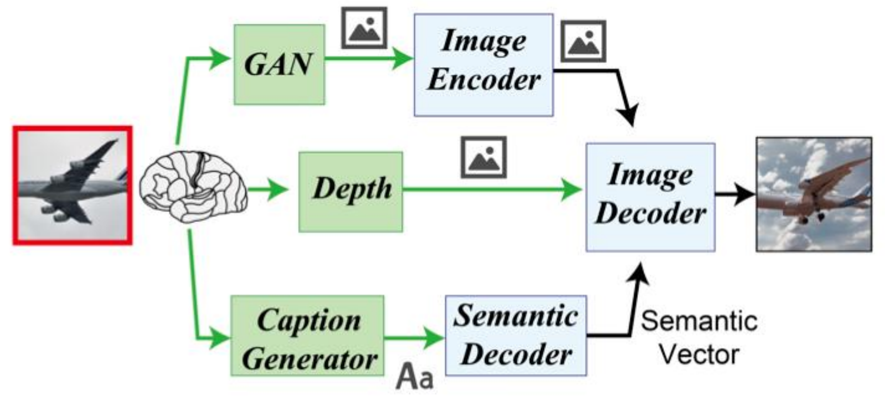
Selective engagement of different U-Net layers through the denoising process

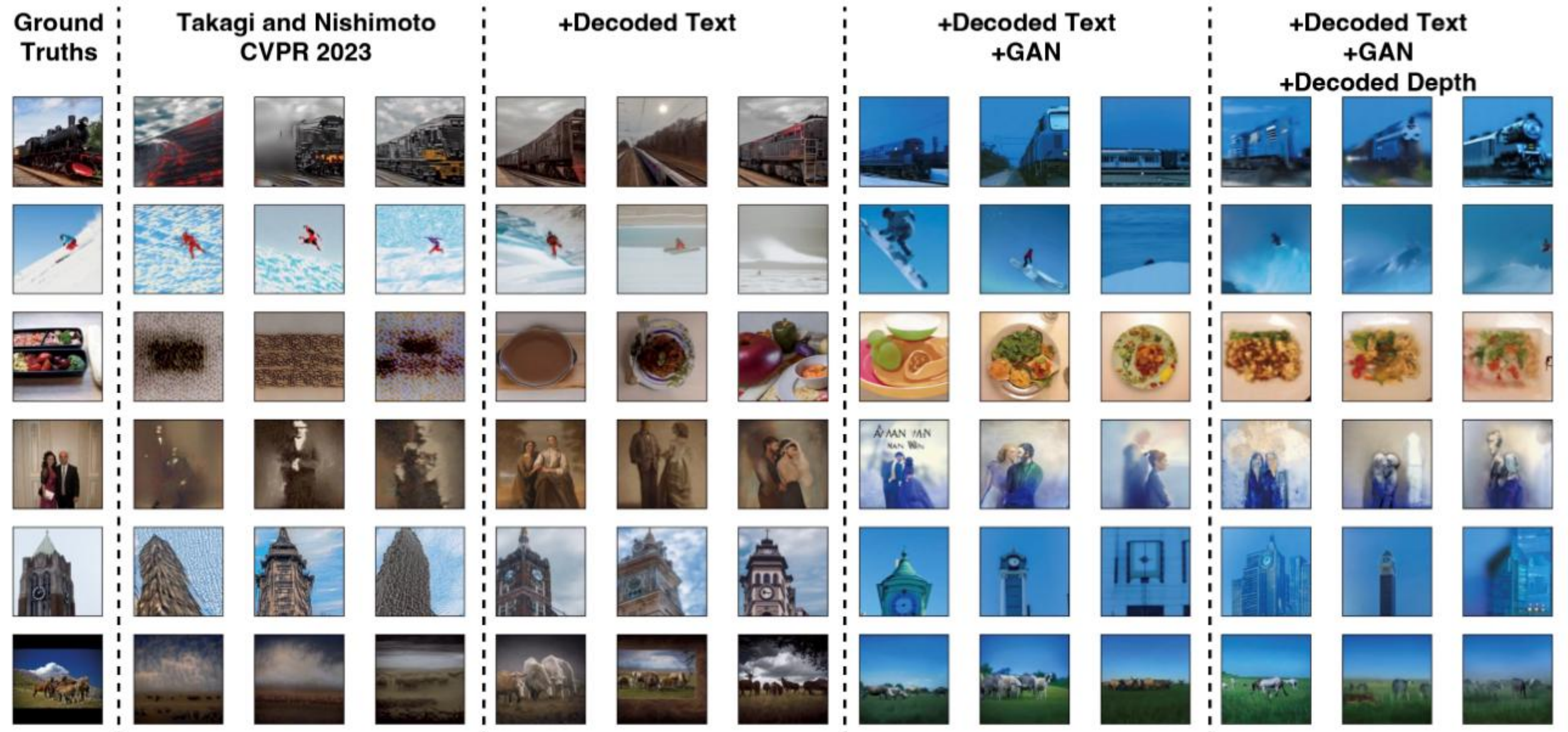


Further improvement (Takagi and Nishimoto, arXiv 2023)

We can incorporate several techniques:

- Decoded caption from brain
- Non-linear optimization (e.g. GAN)
- Decoded depth from brain





Conclusion

We propose a novel visual reconstruction method using Stable Diffusion.

We also provide a quantitative, biological interpretation for Stable Diffusion

References

[1] Rombach et al., CVPR 2022; [2] Lin et al., NeurIPS 2022
[3] Shen et al., PLoS CB 2019; [4] Nishimoto et al., Current Biology 2011; [5] Yamins et al., PNAS 2014; [6] Allen et al., Nat. Neurosci.

Introducing

Mind-Video

Cinematic Mindscapes: High-quality Video Reconstruction from Brain Activity

> We propose Mind-Video, which progressively learns spatiotemporal information from continuous fMRI data through masked brain modeling + multimodal contrastive learning + spatiotemporal attention + co-training with an augmented Stable Diffusion model that incorporates network temporal inflation.
>
> This is an extension of our previous fMRI-Image reconstruction work: MinD-Vis(CVPR2023)

Ground truth Videos Reconstructed Videos

May 15, 2023

Read Paper

View Code

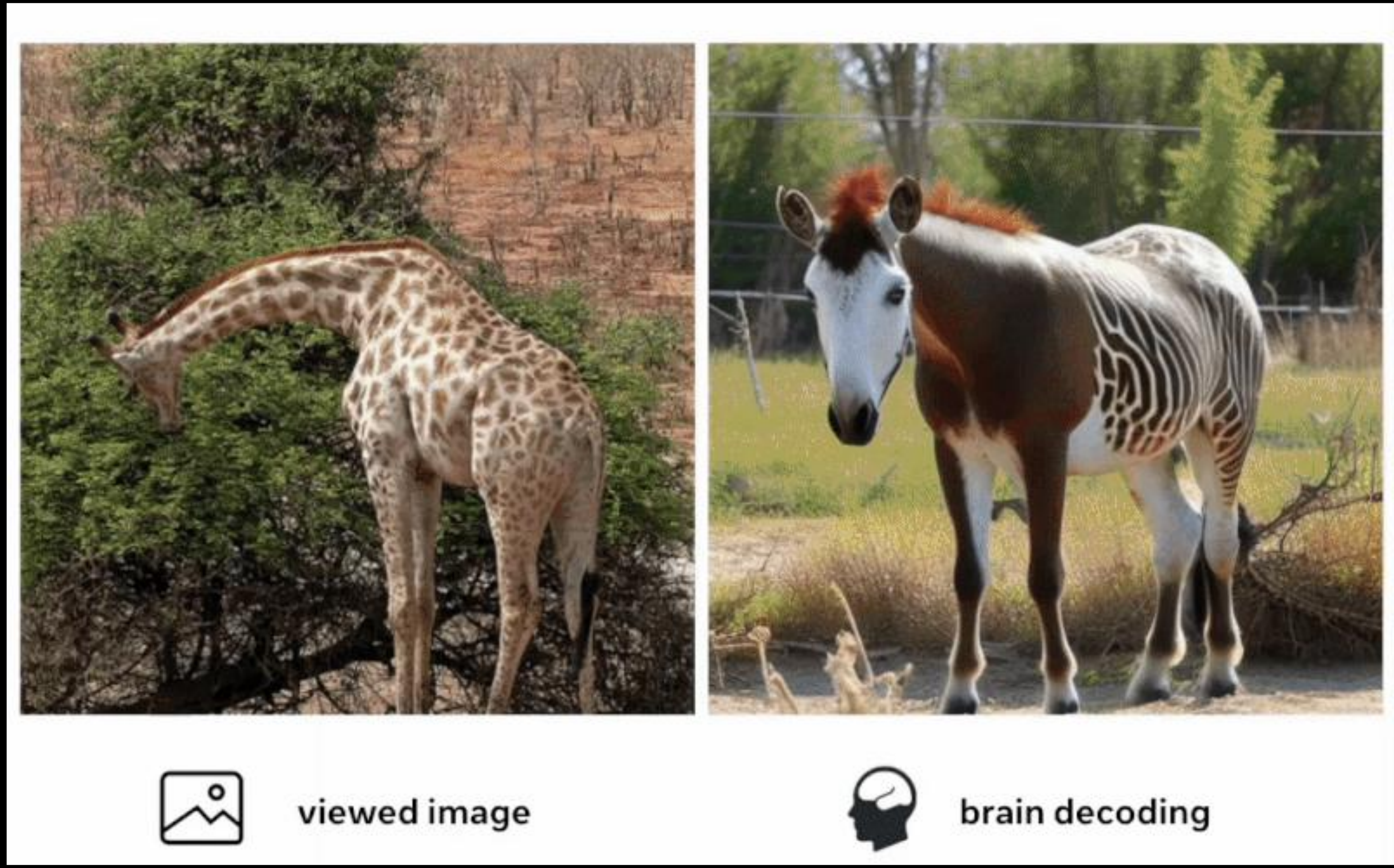
More Samples

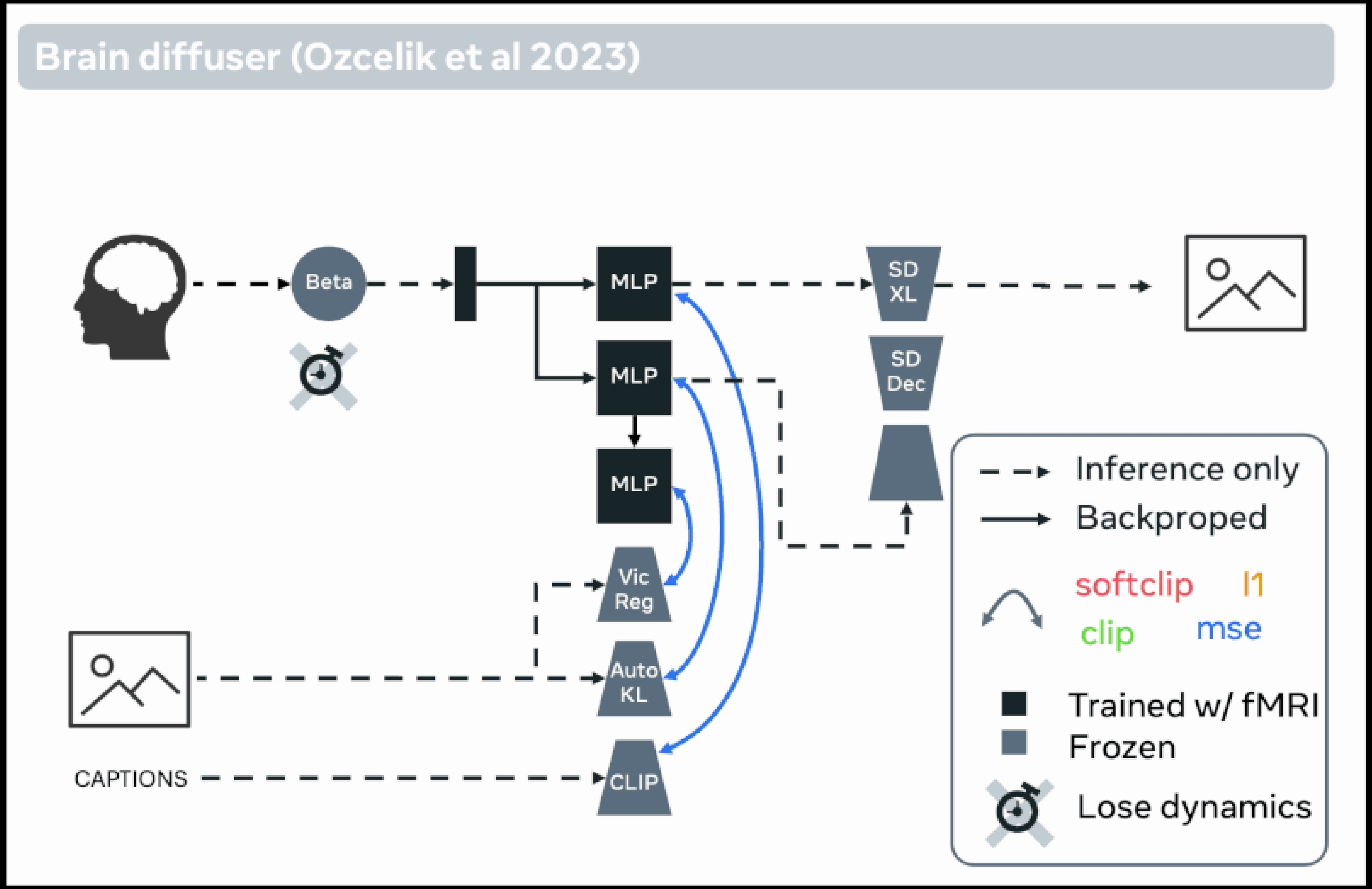
Zijiao Chen* National University of Singapore, Center for Sleep and Cognition, Centre for Translational Magnetic Resonance Research

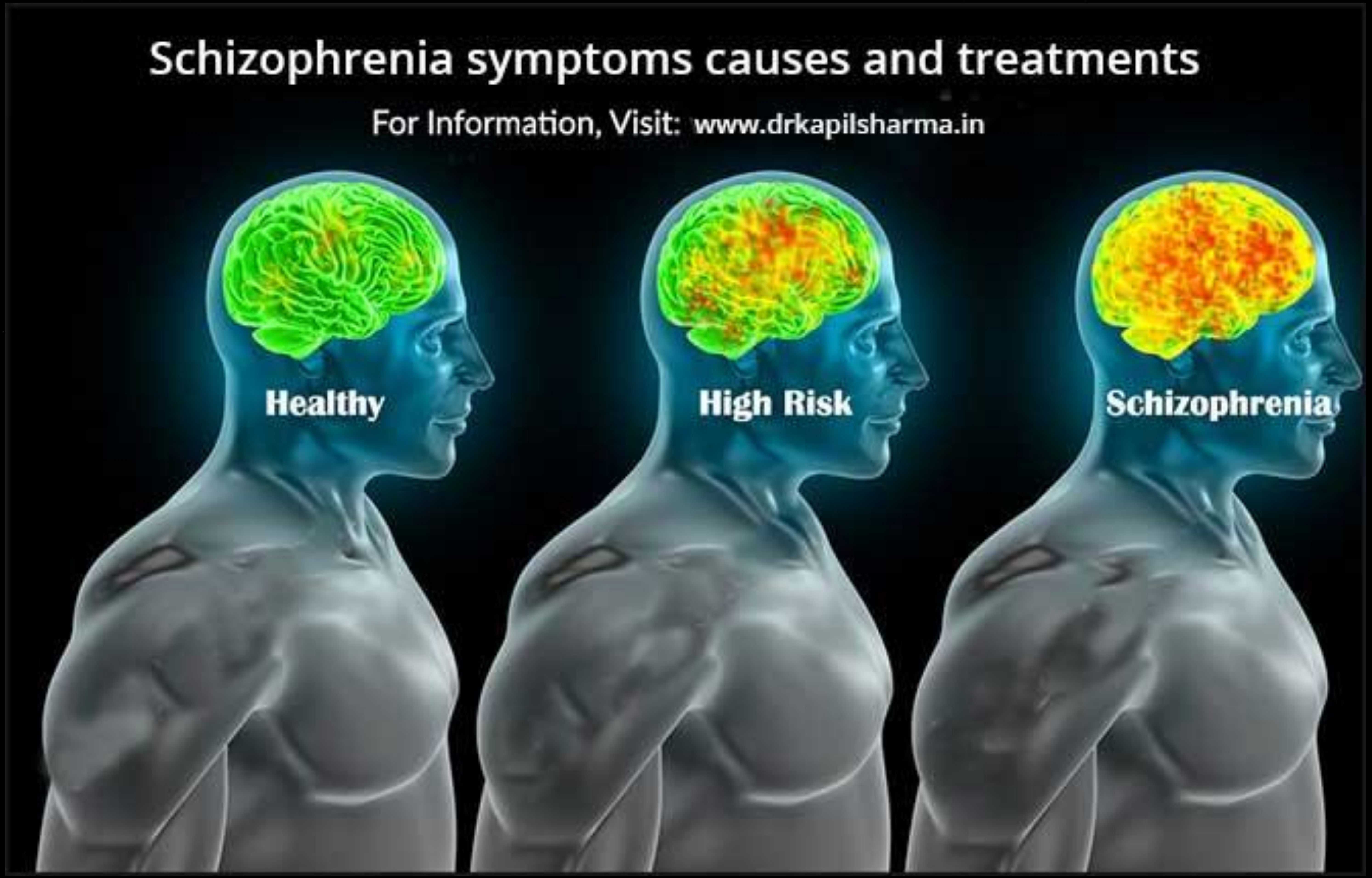
Jiaxin Qing* The Chinese University of Hong Kong, Department of Information Engineering

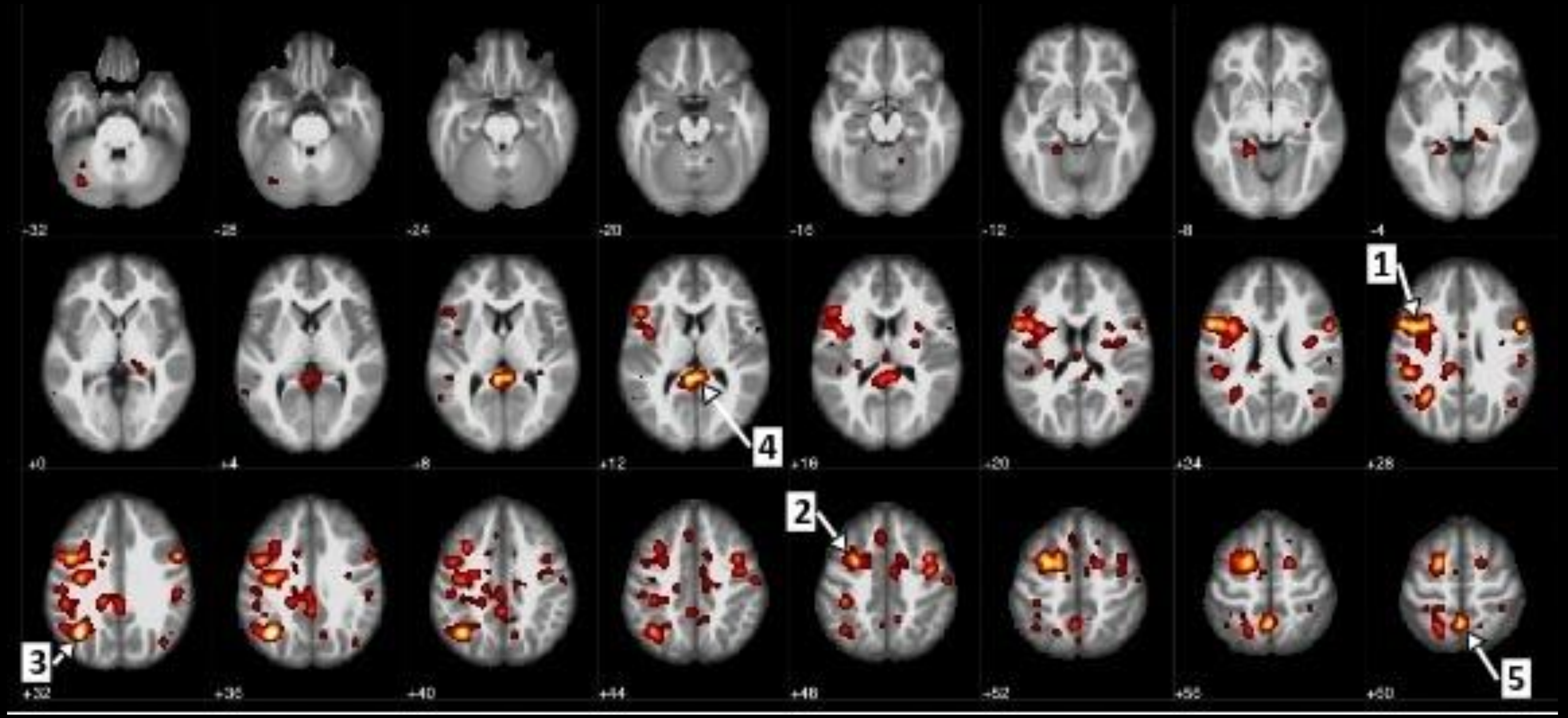
Juan Helen Zhou† National University of Singapore, Center for Sleep and Cognition, Centre for Translational Magnetic Resonance Research

"Dynadiff: Single-stage Decoding of Images from Continuously Evolving fMRI".

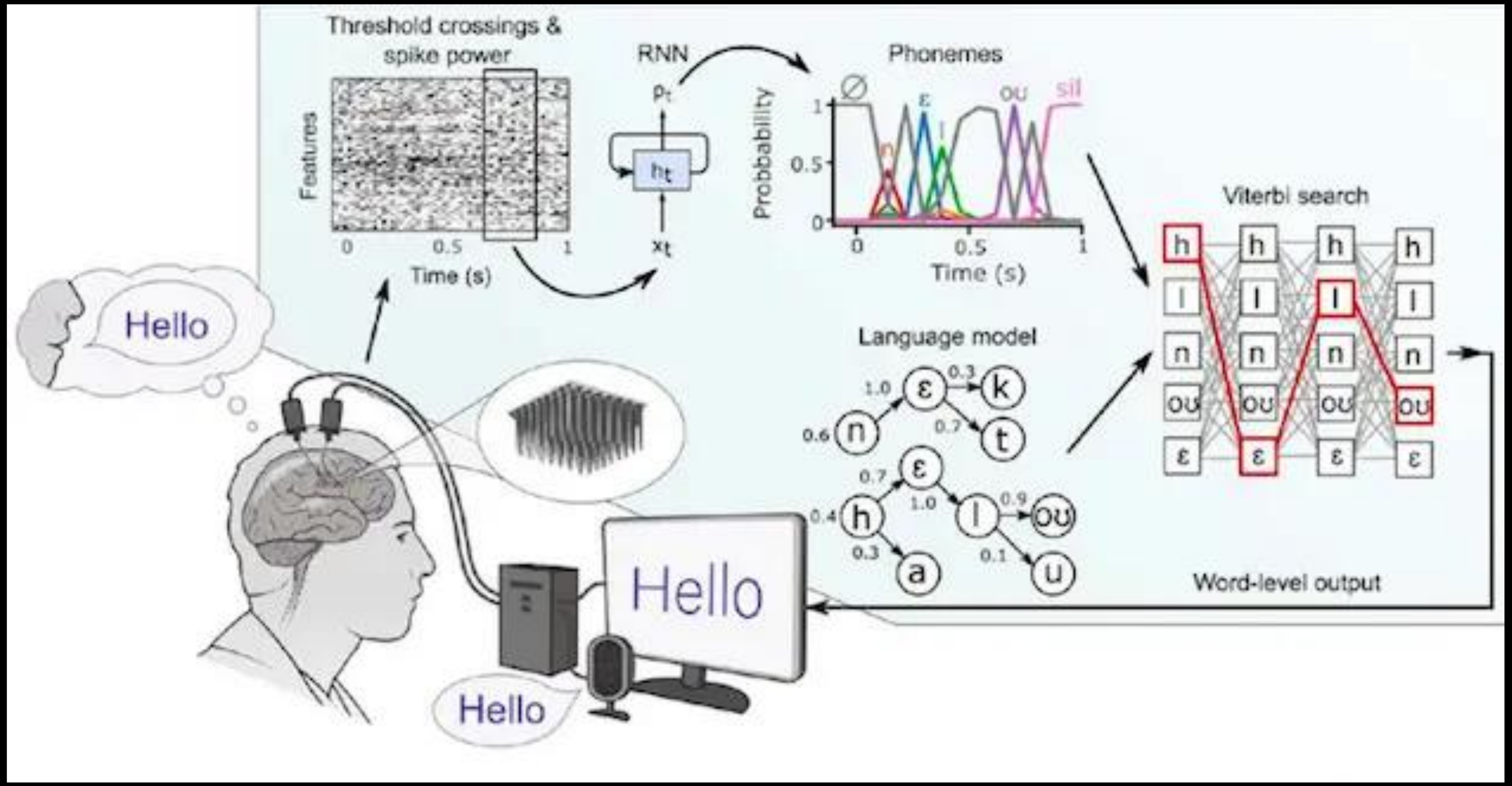


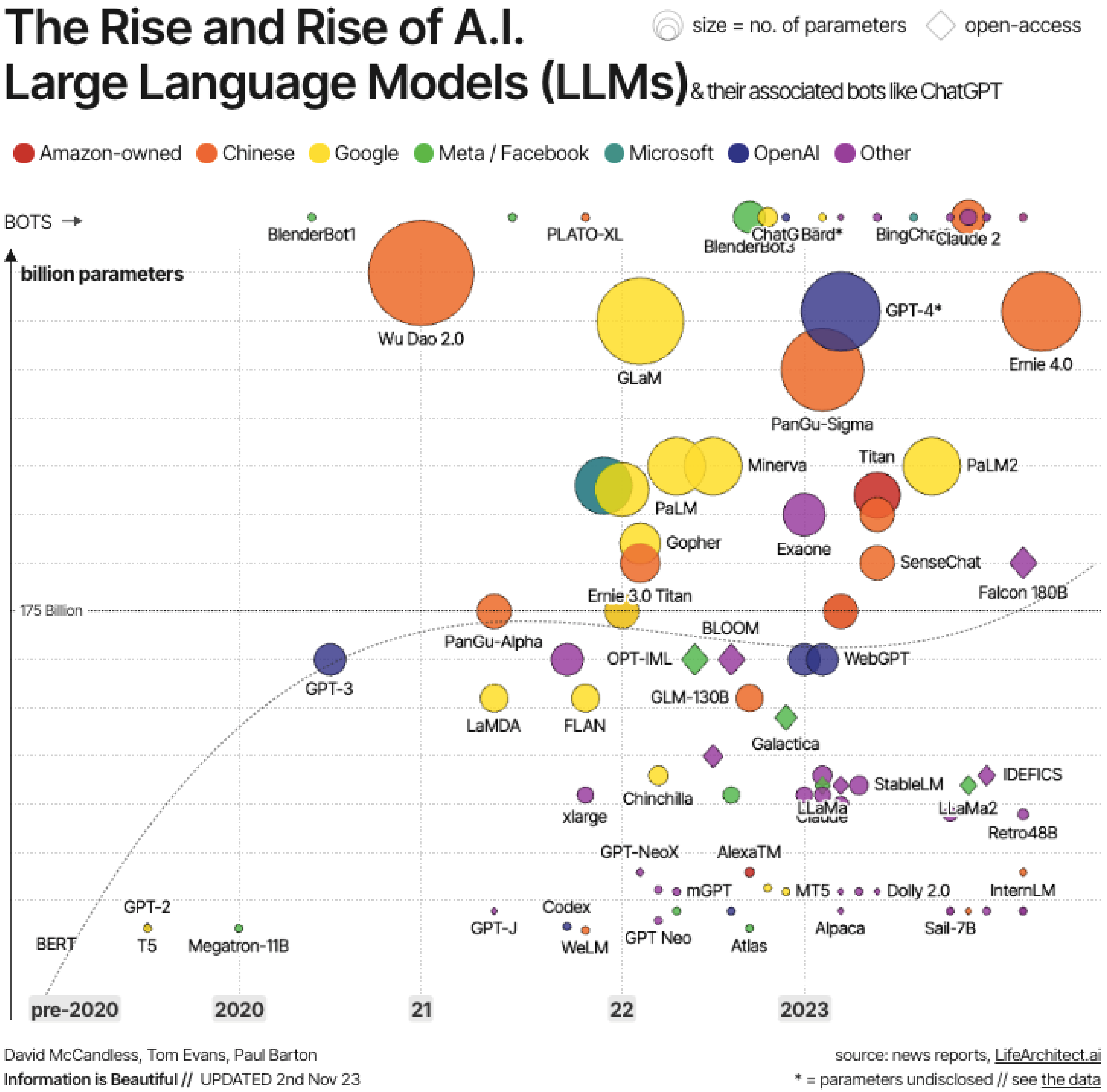



Schizophrenia symptoms causes and treatments
For Information, Visit: www.drkapilsharma.in
Healthy
High Risk
Schizophrenia



AI helps a paralyzed woman **speak** for the **first time** in 18 years





The

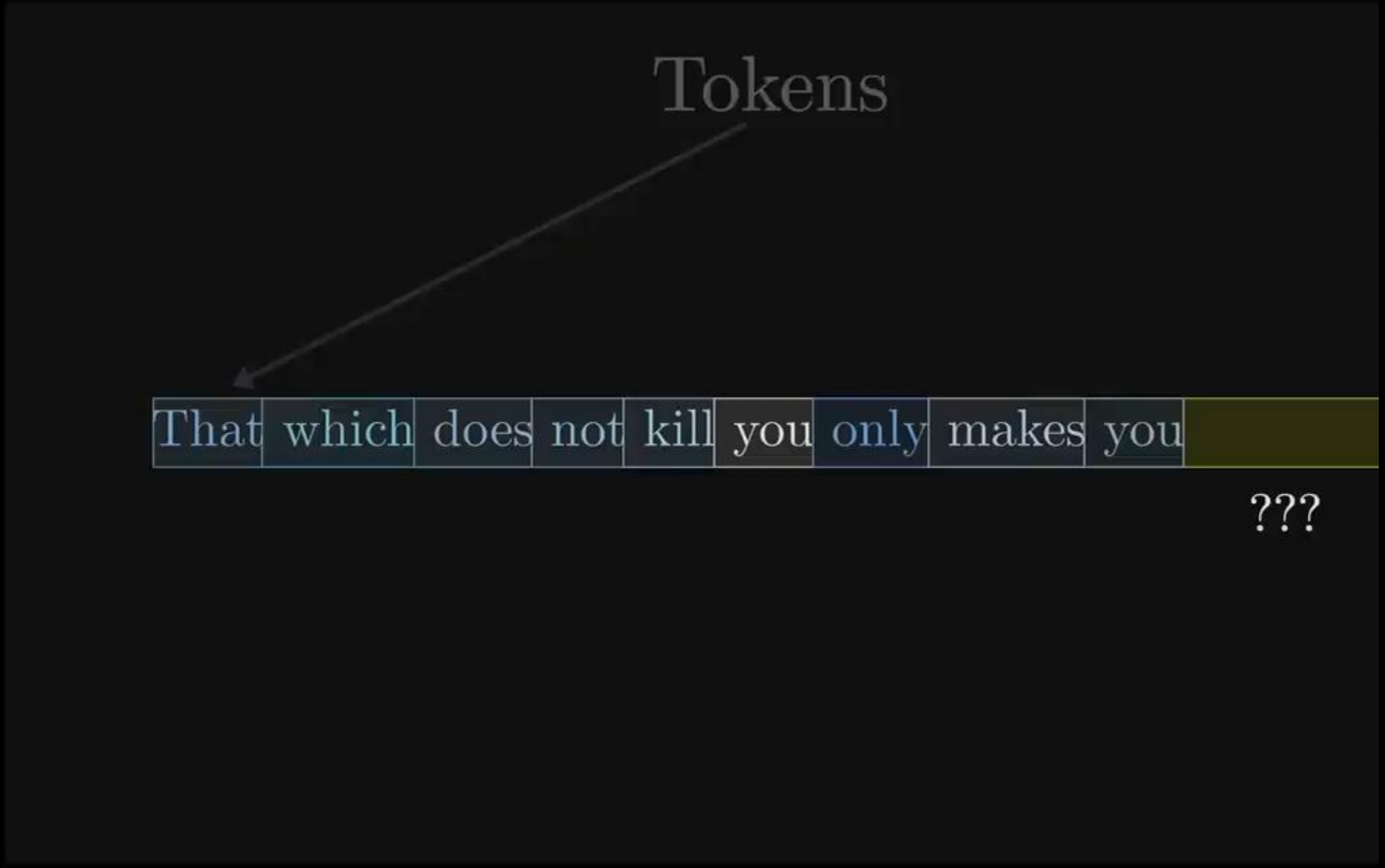
$$P_\theta(x) = \prod_{t=2}^{T} P_\theta(x_t \mid \{x_{<t}\})$$

$$= P_{\theta_2}(x_2 \mid \{x_1\})\, P_{\theta_3}(x_3 \mid \{x_1, x_2\}) \cdots P_{\theta_T}(x_T \mid \{x_1, \dots, x_{T-1}\})$$

$o \leftarrow \{\bullet\}$ x_2 x_1

$o \leftarrow \{\bullet \;\; \bullet\}$ x_3 x_1 x_2

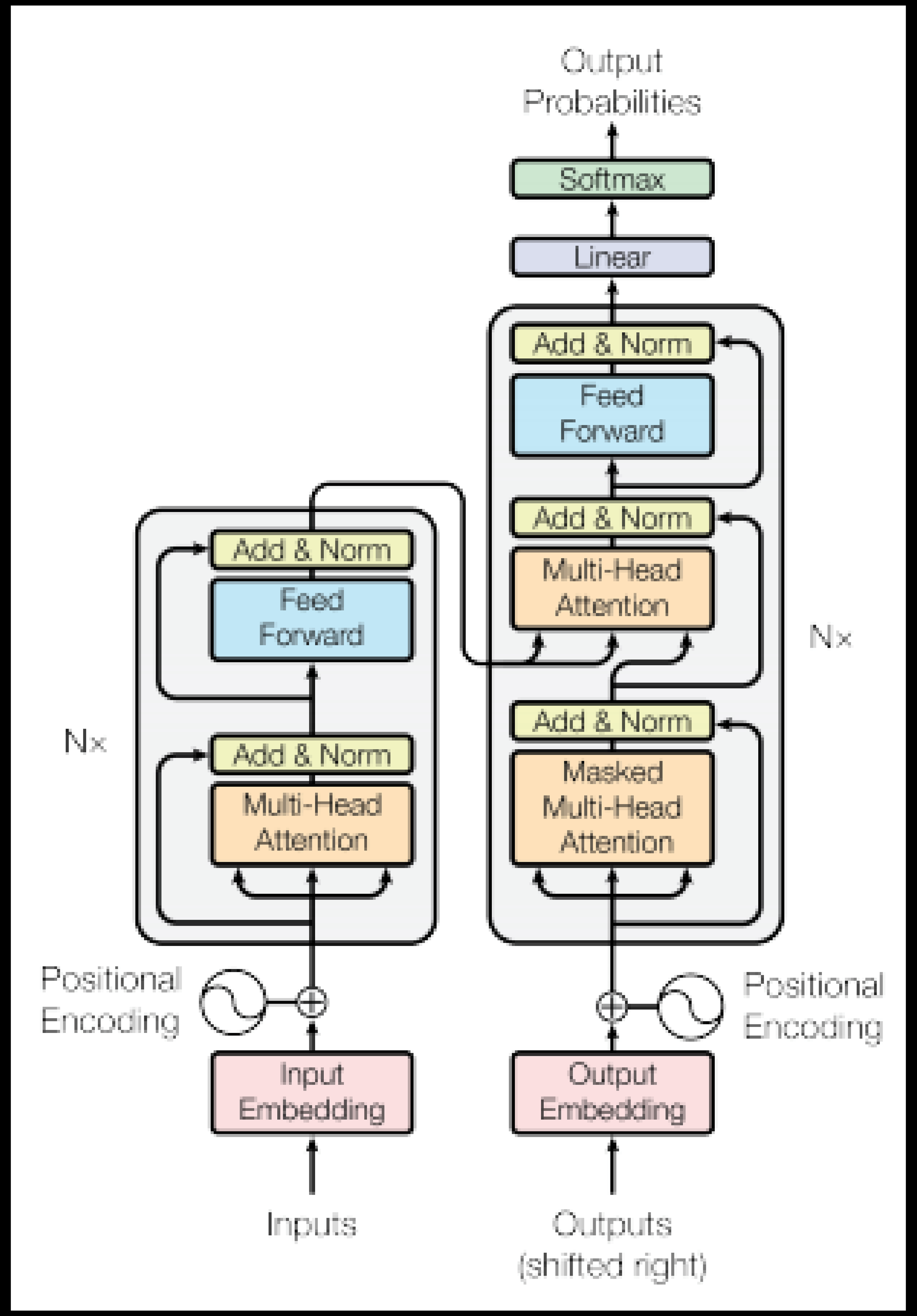
$o \leftarrow \{\bullet \cdots \bullet\}$ x_T x_1 x_{T-1}



$$P_\theta(x) = \prod_{t=2}^{T} P_\theta(x_t \mid \{x_{<t}\})$$

$$= P_{\theta_2}(x_2 \mid \{x_1\})\, P_{\theta_3}(x_3 \mid \{x_1, x_2\}) \cdots P_{\theta_T}(x_T \mid \{x_1, \dots, x_{T-1}\})$$

$o \leftarrow \{\bullet\}$: x_2, x_1 $\quad$ $o \leftarrow \{\bullet \ \bullet\}$: x_3, x_1, x_2 $\quad$ $o \leftarrow \{\bullet \cdots \bullet\}$: x_T, x_1, x_{T-1}



$$PP(W) = P(w_1 w_2 \dots w_N)^{-\frac{1}{N}}$$

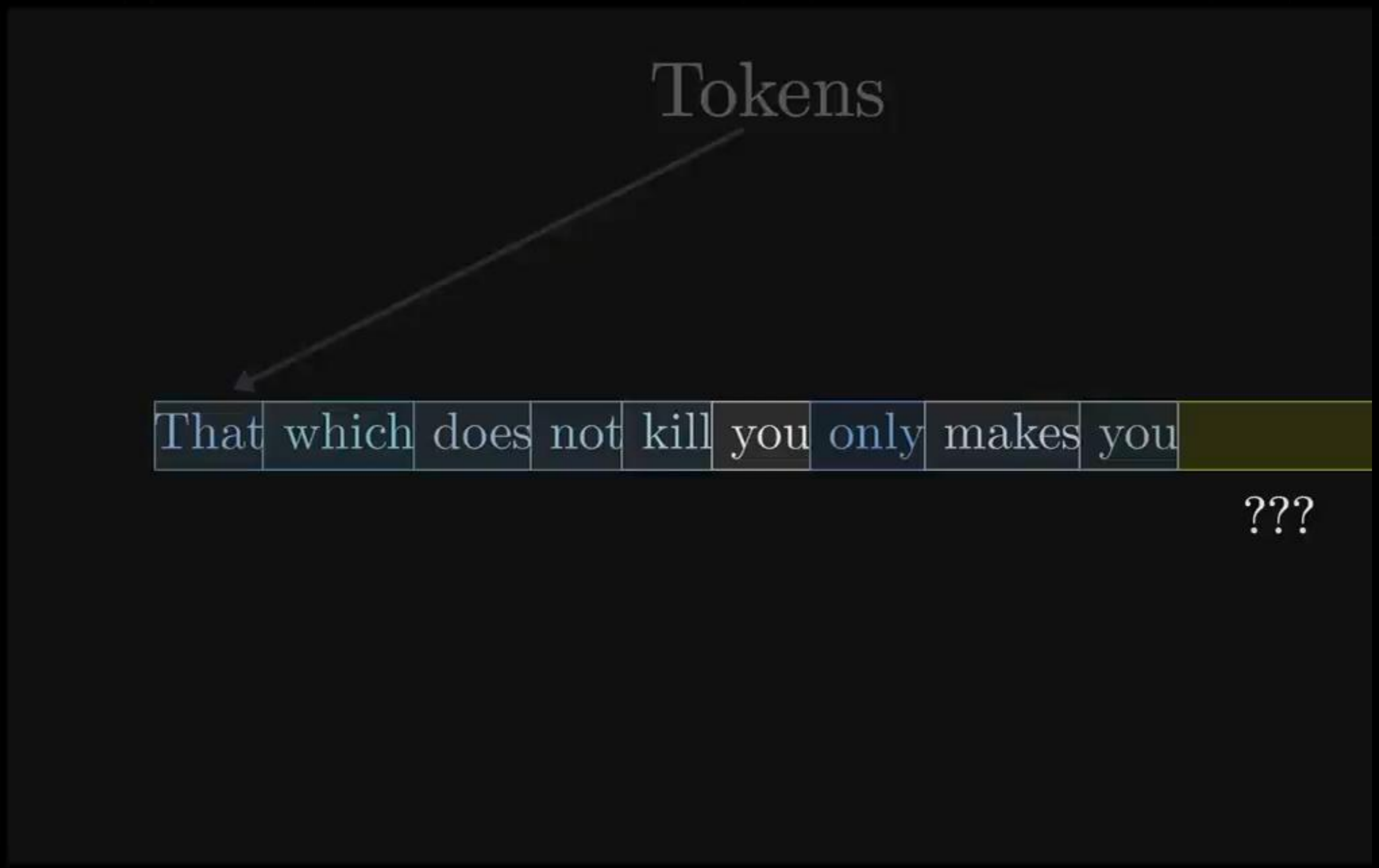
$$= \sqrt[N]{\frac{1}{P(w_1 w_2 \dots w_N)}}$$

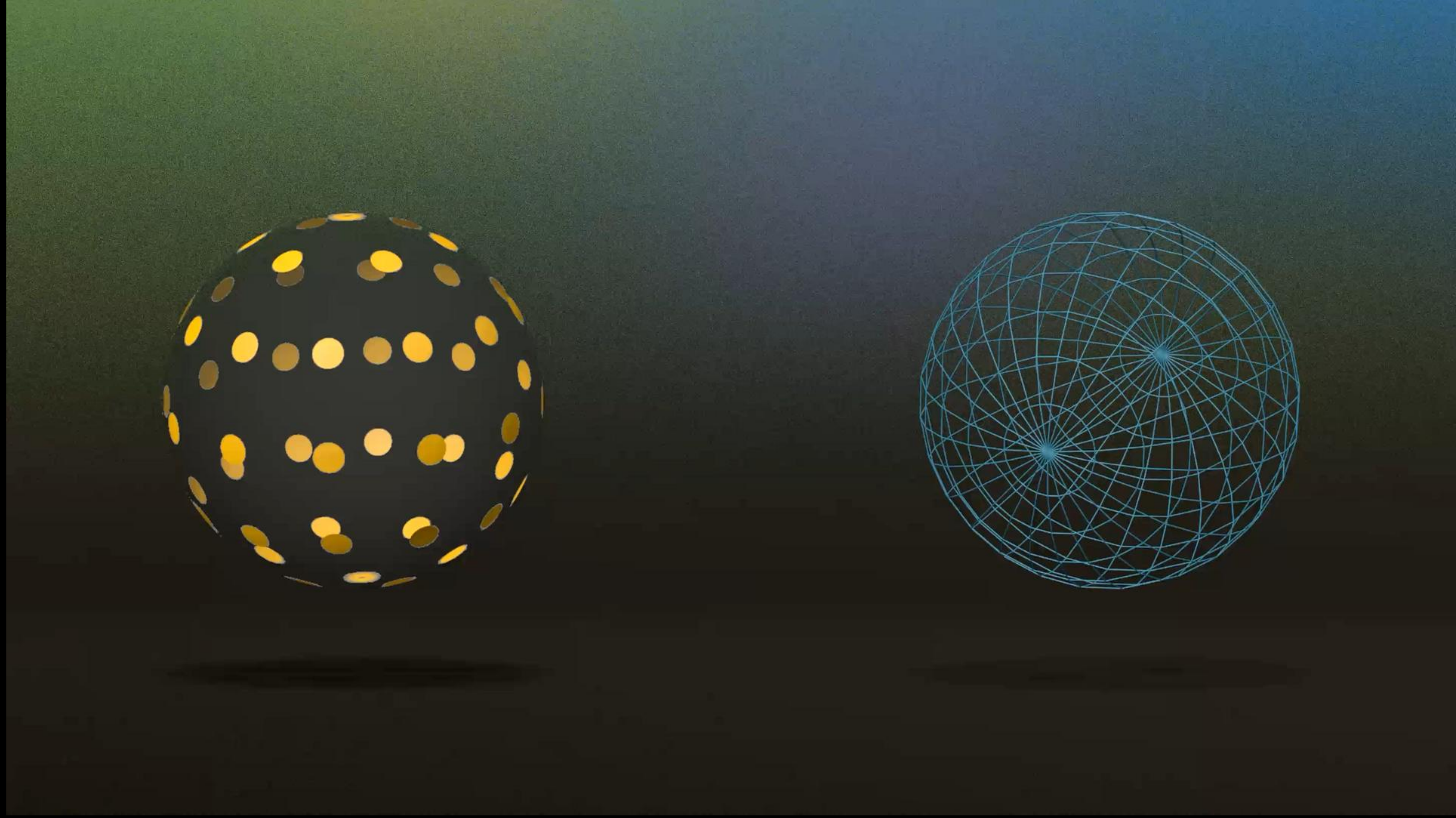
equivalently:

$$PP(W) = 2^{-l}$$

where $l = \frac{1}{N} \log P(w_1 w_2 \dots w_N)$

$$2^{-l} \text{ where } l = \frac{1}{M} \sum_{i=1}^{m} \log p(s_i)$$





$$P_\theta(x) = \prod_{t=2}^{T} P_\theta(x_t \mid \{x_{<t}\})$$

$$= P_{\theta_2}(x_2 \mid \{x_1\})\, P_{\theta_3}(x_3 \mid \{x_1, x_2\}) \cdots P_{\theta_T}(x_T \mid \{x_1, \dots, x_{T-1}\})$$

$x_2 \leftarrow \{x_1\}$ $\quad x_3 \leftarrow \{x_1, x_2\}$ $\quad x_T \leftarrow \{x_1 \cdots x_{T-1}\}$

A Mathematical Exploration of Why Language Models Help Solve Downstream Tasks

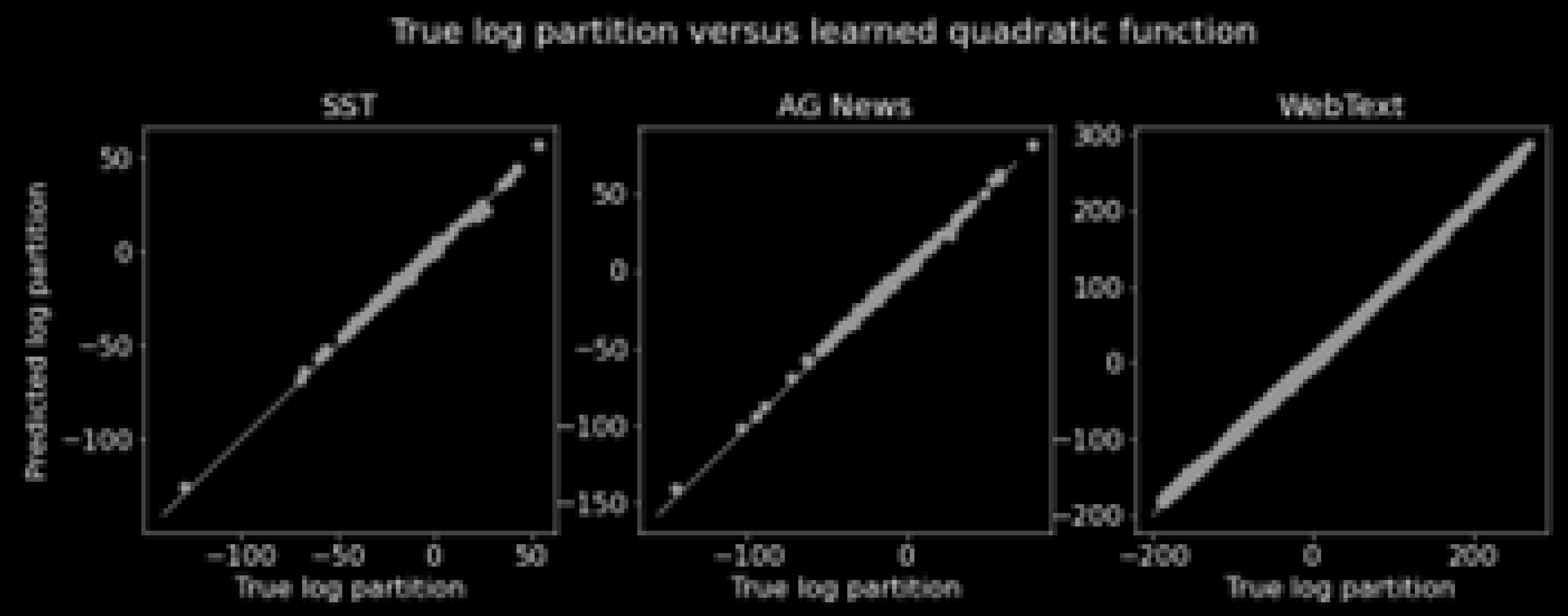
Nikunj Saunshi[1], Sadhika Malladi[1], and Sanjeev Arora[1,2]

[1]Department of Computer Science, Princeton University
{nsaunshi, smalladi, arora}@cs.princeton.edu
[2]Institute for Advanced Study

Task	k	$f(s)$ 768	$\Phi p_f(s)$ 768	$\boldsymbol{p}_{f(s)}$ (subset) ≤ 30	$\boldsymbol{p}_{f(s)}$ (class words) $\leq 2k$	$\boldsymbol{p}_{f(s)}$ (rand. proj.) 768	FT
SST	2	87.5	83.3	82.6	78.7	67.5	91.4
SST*	2	89.4	87.3	85.4	79.1	76.4	92.3
SST fine	5	49.2	43.5	44.0	39.2	23.1	50.2
SST fine*	5	49.4	48.6	47.6	40.3	28.8	53.5
AG	4	90.7	84.6	83.8	75.4	58.5	94.5
AG*	4	91.1	88.2	86.1	75.1	63.7	94.4

Proposition 2.1 (Cross-entropy recovers $\boldsymbol{p}^*_{\cdot|s}$). *The unique minimizer of* $\ell_{xent}(\{\boldsymbol{p}_{\cdot|s}\})$ *is* $\boldsymbol{p}_{\cdot|s} = \boldsymbol{p}^*_{\cdot|s}$ *for every* $s \in support(p_L)$.

Proposition 2.2 (Softmax models recover $\boldsymbol{p}^*_{\cdot|s}$ on a subspace). *Fix a fixed* Φ, *if* $f^* \in \arg\min_{f:\mathcal{S}\to\mathbb{R}^d} \ell_{xent}(f,\Phi)$ *exists, then* $\Phi \boldsymbol{p}_{f^*(s)} = \Phi \boldsymbol{p}^*_{\cdot|s}$ *for every* $s \in support(p_L)$.

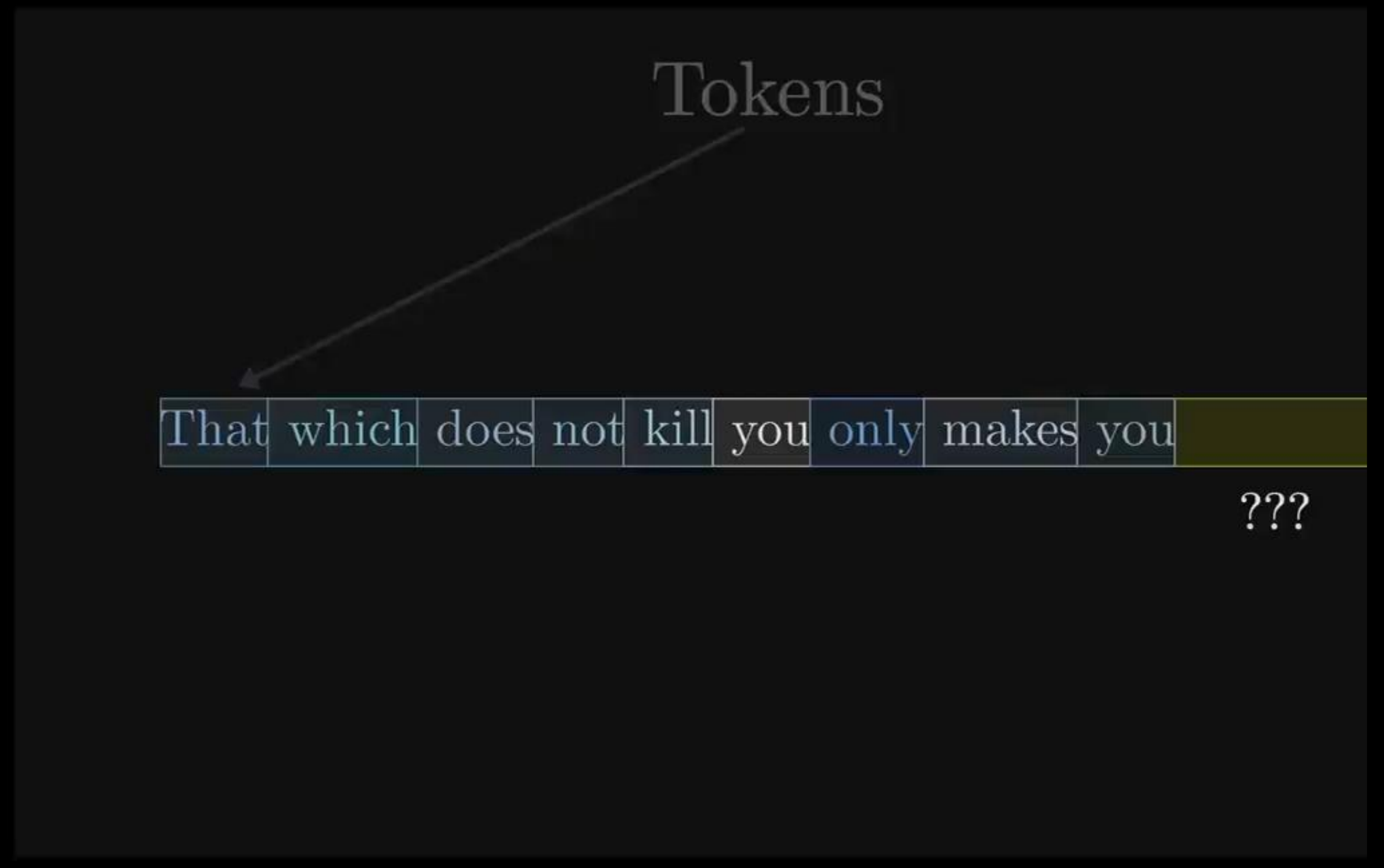


Definition 3.1. *A classification task* $\mathcal{T}$ *is* (τ, B)*-natural if* $\min_{\boldsymbol{v}\in\mathbb{R}^V, \|\boldsymbol{v}\|_\infty \leq B} \ell_{\mathcal{T}}(\{\boldsymbol{p}^*_{\cdot|s}\}, \boldsymbol{v}) \leq \tau$.

Definition 3.2. *Task* $\mathcal{T}$ *is* (τ, B)*-natural w.r.t.* $\Phi \in \mathbb{R}^{d\times V}$ *if* $\min_{\boldsymbol{v}\in row\text{-}span(\Phi), \|\boldsymbol{v}\|_\infty \leq B} \ell_{\mathcal{T}}(\{\boldsymbol{p}^*_{\cdot|s}\}, \boldsymbol{v})$

Theorem 4.1. *Let* $\{\boldsymbol{p}_{\cdot|s}\}$ *be a language model that is* ϵ*-optimal, i.e.* $\ell_{xent}(\{\boldsymbol{p}_{\cdot|s}\}) - \ell^*_{xent} \leq \epsilon$, *for some* $\epsilon > 0$. *For a classification task* $\mathcal{T}$ *that is* (τ, B)*-natural, we have*

$$\ell_{\mathcal{T}}\left(\{\boldsymbol{p}_{\cdot|s}\}\right) \leq \tau + \sqrt{2B^2\epsilon\left(\gamma(p_{\mathcal{T}})\right)^{-1}}$$



Theorem 4.2. *For a fixed* Φ, *let* f *be features from an* ϵ*-optimal* d*-dimensional softmax language model, i.e.* $\ell_{xent}(f,\Phi) - \ell^*_{xent}(\Phi) \leq \epsilon$. *For a classification task* $\mathcal{T}$ *that is* (τ, B)*-natural w.r.t.* Φ,

$$\ell_{\mathcal{T}}(\Phi p_f) \leq \tau + \sqrt{2B^2\epsilon\left(\gamma(p_{\mathcal{T}})\right)^{-1}}$$


02:45:37
CFA
02:45:37
AI NEWS
AI just passed the hardest Chartered Financial Analyst (CFA) exam in minutes

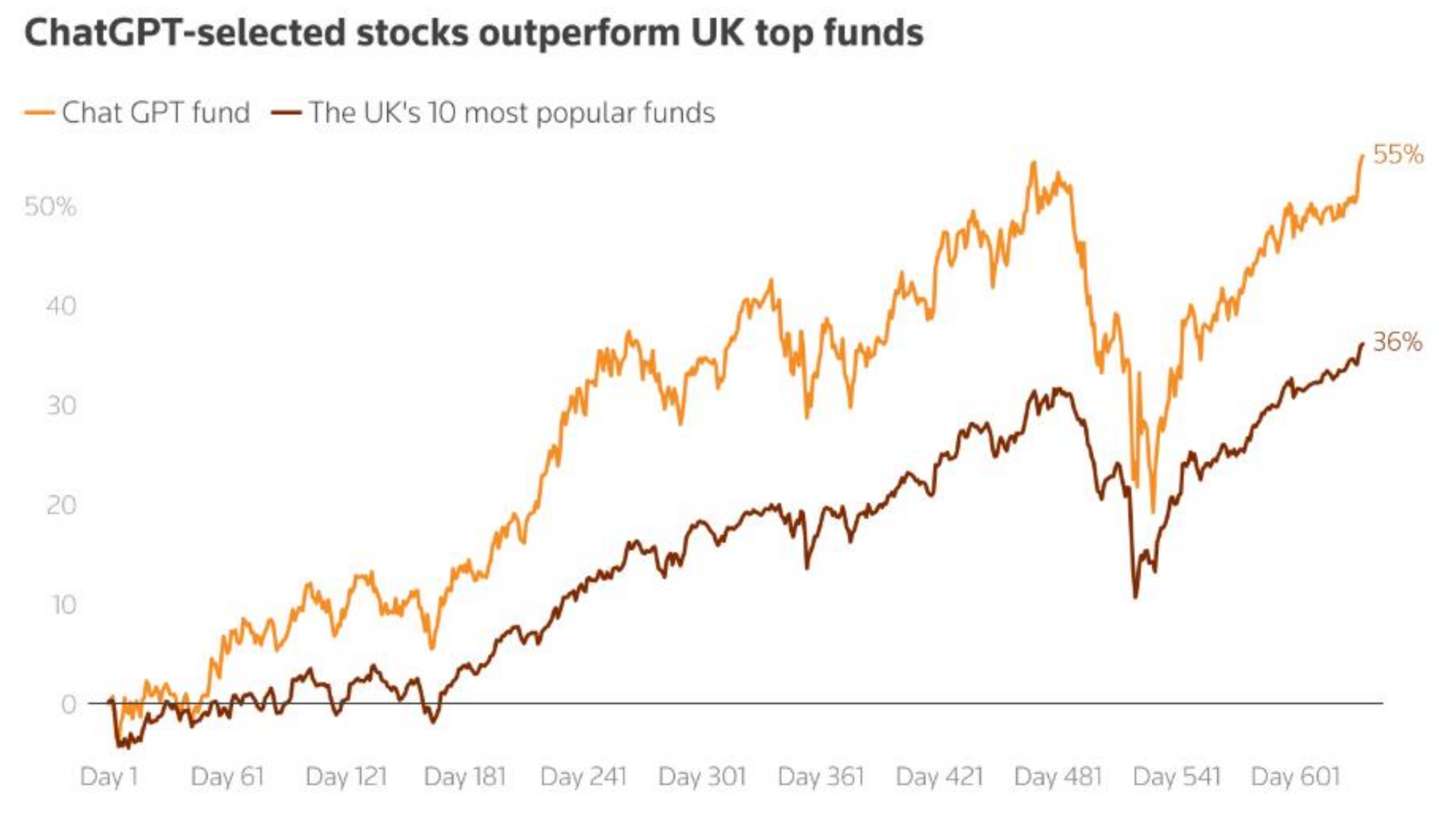

Goldman
Sachs


M&A
MERGERS AND ACQUISITIONS
GROWTH
COSTS & INNOVATION
NEW PRODUCT
INTELLECTUAL PROPERTY
HUMAN CAPITAL
CUSTOMER BASES
EFFICIENCY
COMPETITION


Goldman
Sachs

95% Fewer False Alarms: JPMorgan Chase Uses AI to Sharpen Anti-Money Laundering Efforts

AI • Mar 26,2024



Notes: Finder asked ChatGPT in March 2023 to select a basket of stocks and compared the performance of those stocks with the average of the UK's 10 most popular funds

Source: Finder | Joice Alves

Finder asked ChatGPT in March 2023 to select a basket of stocks and compared the price performance of those stocks with the average of the UK's 10 most popular funds.

Hallucination is Inevitable: An Innate Limitation of Large Language Models

Ziwei Xu **Sanjay Jain** **Mohan Kankanhalli**
School of Computing, National University of Singapore
ziwei.xu@u.nus.edu {sanjay,mohan}@comp.nus.edu.sg

Definition 1 (Alphabet and Strings)**.** Let $\mathbb{N}$ be the set of natural numbers. An alphabet $\mathcal{A}$ is a finite set of N tokens $\mathcal{A} = \{a_0, a_1, \ldots, a_{N-1}\}$. A string is a sequence $w_{0:n-1} = w_0 w_1 \ldots w_{n-1}$ obtained by concatenating tokens for n times, where $i, n, N \in \mathbb{N}, w_i \in \mathcal{A}$. Let $\mathcal{S}$ be a computable set[1] of all the finite-length strings of alphabet $\mathcal{A}$ and $(s_0, s_1, \ldots)$ be an one-to-one enumeration of all the elements in $\mathcal{S}$.

Definition 2 (Formal World of f)**.** A formal world of ground truth function f is a set $\mathcal{G}_f = \{(s, f(s)) \mid s \in \mathcal{S}\}$, where $f(s)$ is the only correct output of input string s, for all $s \in \mathcal{S}$.

The training samples $\mathcal{T}$ is defined as a set of input-output pairs we get from the formal world of f:

Definition 3 (Training Samples $\mathcal{T}$)**.** Training samples $\mathcal{T}$ is a set $\{(s_0, y_0), (s_1, y_1), \ldots, (s_i, y_i), \ldots \mid y_i = f(s_i), s_i \in \mathcal{S}, i \in \mathbb{N}\}$.

Theorem 1. For all computably enumerable sets of LLMs $\{h_0, h_1, \ldots\}$, there exists a computable ground truth function f, such that all $h_i^{[j]}, i, j \in \mathbb{N}$, will hallucinate.

Theorem 2. For all computably enumerable sets of LLMs $\{h_0, h_1, \ldots\}$, there exists a computable ground truth function f, such that all $h_i^{[j]}, i, j \in \mathbb{N}$, hallucinate on infinitely many inputs.

Theorem 3. For all computable LLMs h, there exists a computable ground truth function f such that each of $h^{[j]}, j \in \mathbb{N}$ hallucinates w.r.t. f. Furthermore, there exists a computable ground truth function[3] f' such that each of $h^{[j]}, j \in \mathbb{N}$ hallucinates on infinitely many inputs w.r.t. f'.

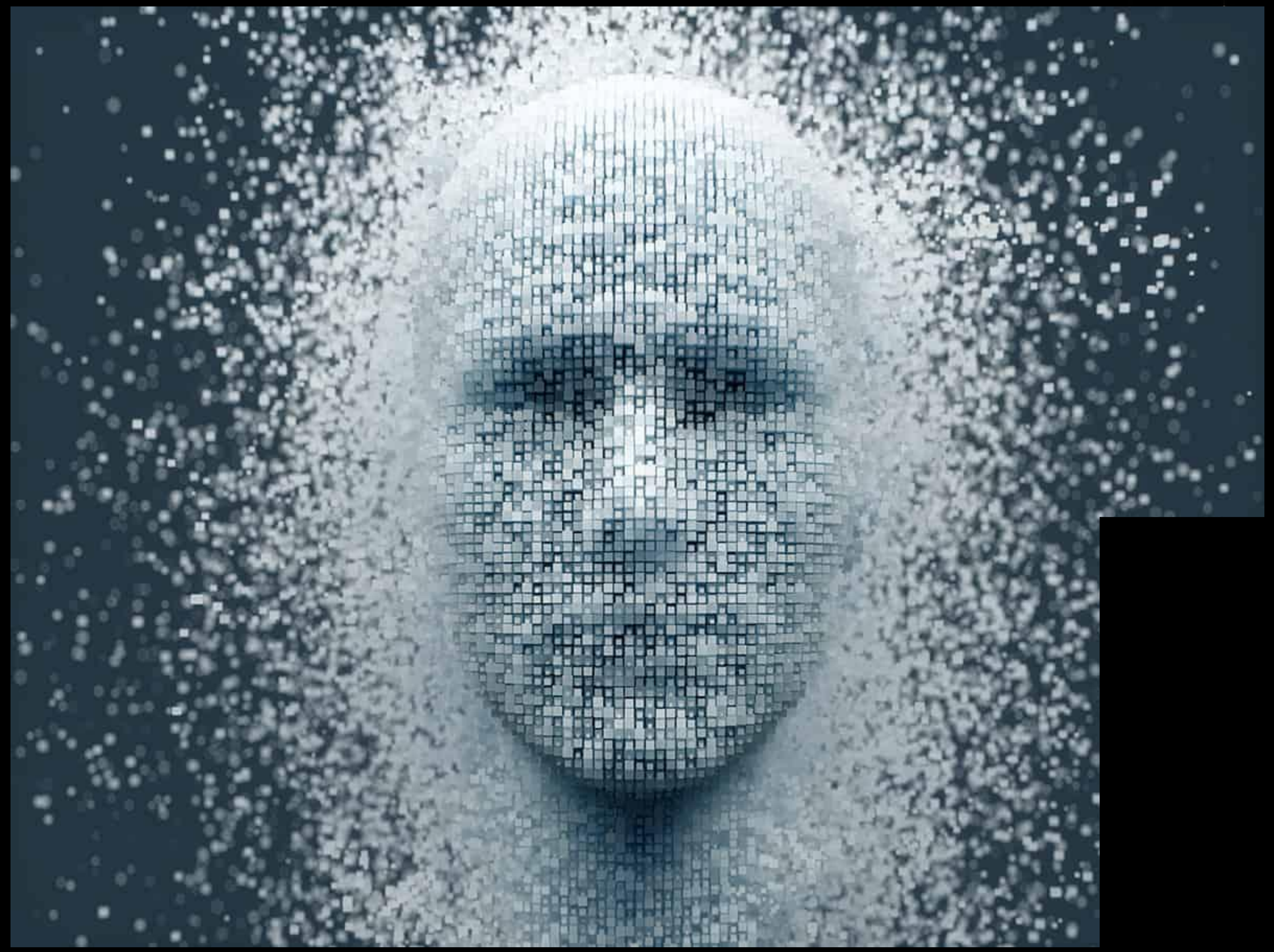
Corollary 1. All computable LLMs cannot prevent themselves from hallucinating.

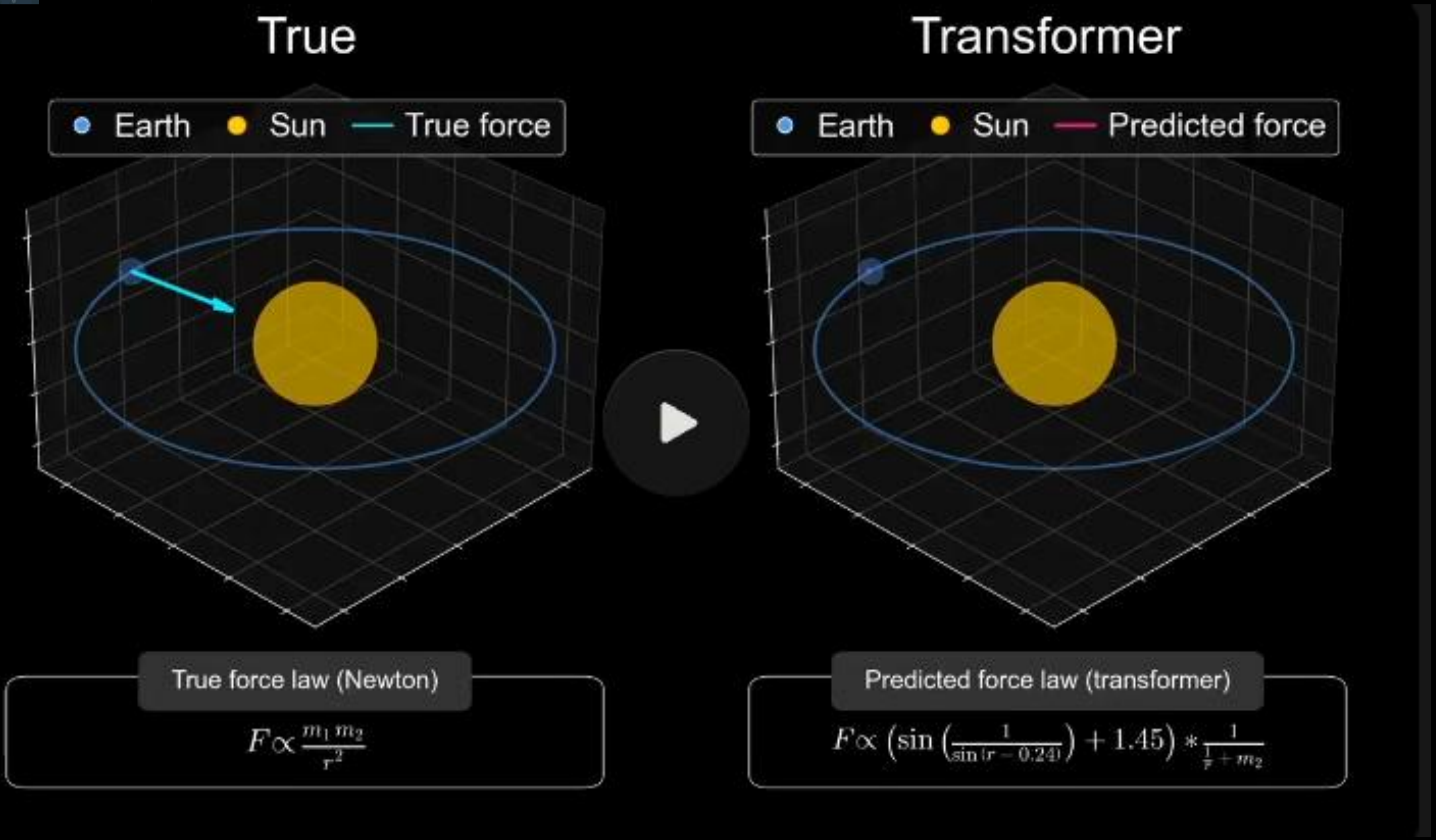
Theorem 4. For all computable LLM h, there exists a computable ordering $<$ such that LLM h hallucinates when answering question "$s_{2n+1} < s_{2n}$?" after being trained on training samples $\{($"$s_i < s_j$?", $f_<(s_i < s_j?)) \mid i, j < 2n\}$, where $n \in \mathbb{N}$, $f_<(s_i < s_j?) =$ "yes" if $s_i < s_j$ and "no" otherwise.

Theorem/corollary	Statement	Implication for real-world LLMs
Theorem 1: computably enumerable LLMs will hallucinate	For any computably enumerable set of LLMs, there exists a computable ground truth function f such that all states of all LLMs in that set will hallucinate.	All currently proposed polynomial-time bounded LLMs are inherently prone to hallucination; it cannot be completely eliminated.
Theorem 2: LLMs will hallucinate on infinitely many questions	For any computably enumerable set of LLMs, there exists a computable ground truth function f such that all states of all LLMs in that set will hallucinate on infinitely many inputs.	Hallucinations are not isolated incidents but a persistent challenge across a vast range of inputs for any LLM.
Theorem 3: any computable LLM will hallucinate	For any individual computable LLM, there exists a computable ground truth function f such that every state of that LLM will hallucinate. Furthermore, for any computable LLM, there exists another f' such that every state will hallucinate on infinitely many inputs.	This generalizes inevitability to any specific LLM, confirming that current and future LLMs will always exhibit some form of hallucination.
Corollary 1: inability to self-eliminate hallucination	All computable LLMs cannot prevent themselves from hallucinating.	LLMs cannot solely rely on internal mechanisms (e.g., self-correction, chain-of-thought prompting) to eliminate hallucination; external safeguards are essential.

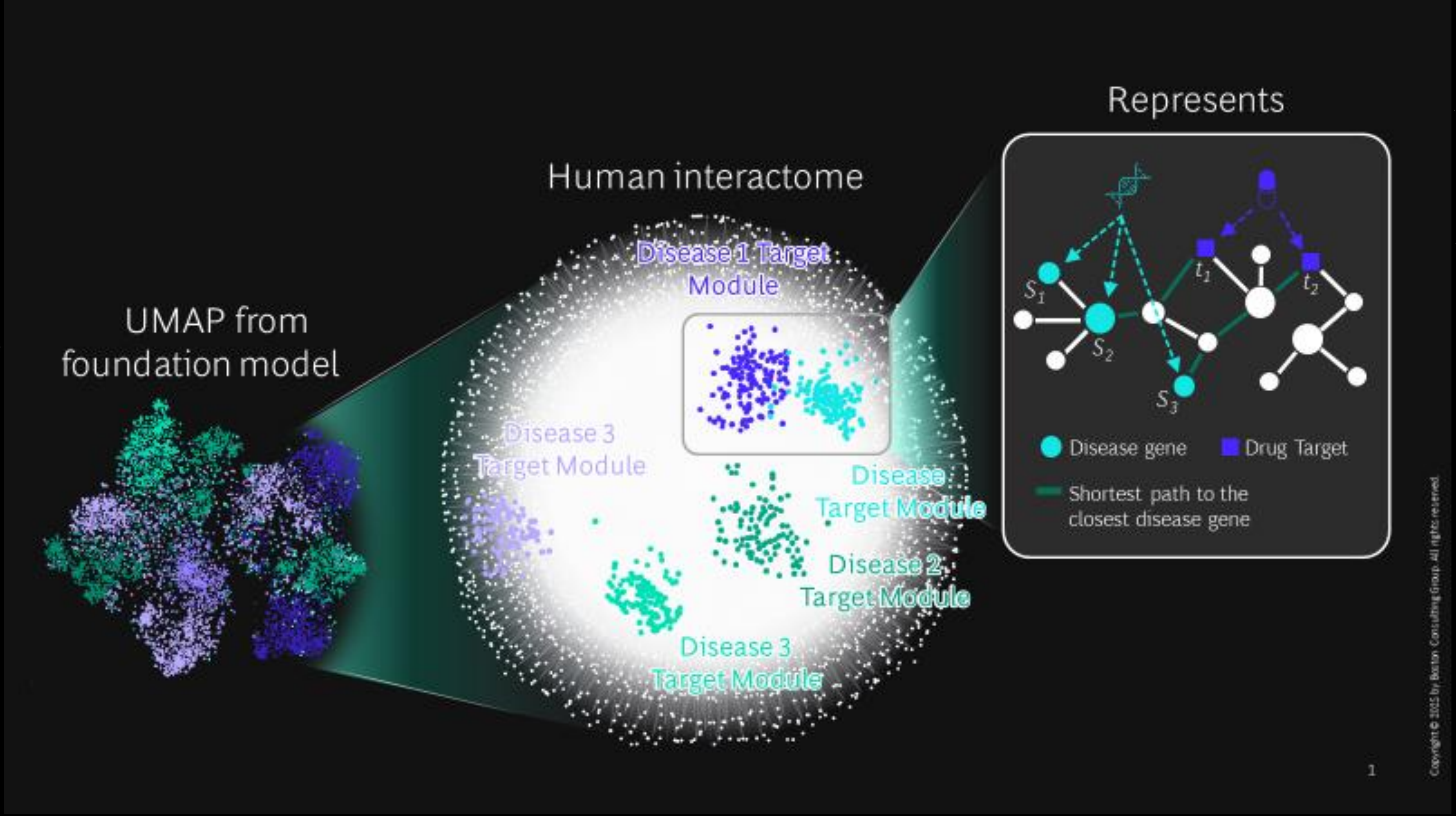
Table 3: Identified hallucination-prone problems for different sets of LLMs. $O(n^k)$ and $O(2^n)$ indicate polynomial-time and exponential-time complexity, respectively.

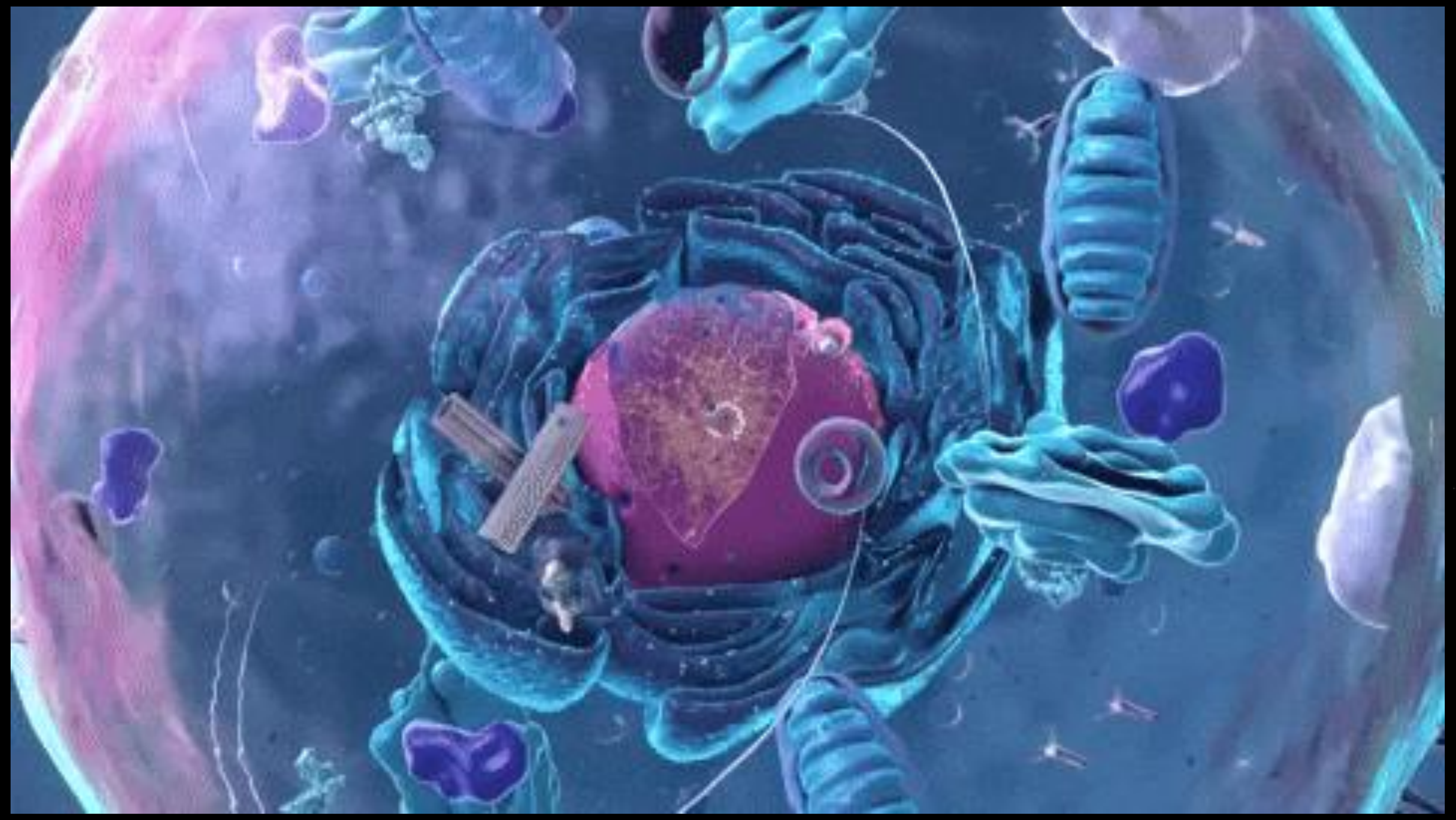
LLMs	Problem	Hallucination Guaranteed by	Extra Assumption
$O(n^k)$ time bounded LLMs (e.g., all existing LLMs)	Combinatorial List	Requiring $\Omega(2^n)$ solution	None
	Subset Sum	Being NP-complete	P $\neq$ NP
	Boolean Satisfiability (SAT)	Being NP-complete [13]	P $\neq$ NP
	Entailment of Propositional Logic	Being co-NP-complete [2]	P $\neq$ NP
$O(2^n)$ time bounded LLMs, $O(n^k)$ time bounded LLMs	Presburger Arithmetic [56, 65]	Requiring $\Omega(2^{2^{cn}})$ solution [18]	None
All Computable LLMs (incl. all LLMs above)	Learning all Computable Linear Orders	Theorem 4	None
	Solving all Computable Problems	Theorem 3	None
	Entailment of First-Order Logic	Being Undecidable [68]	None

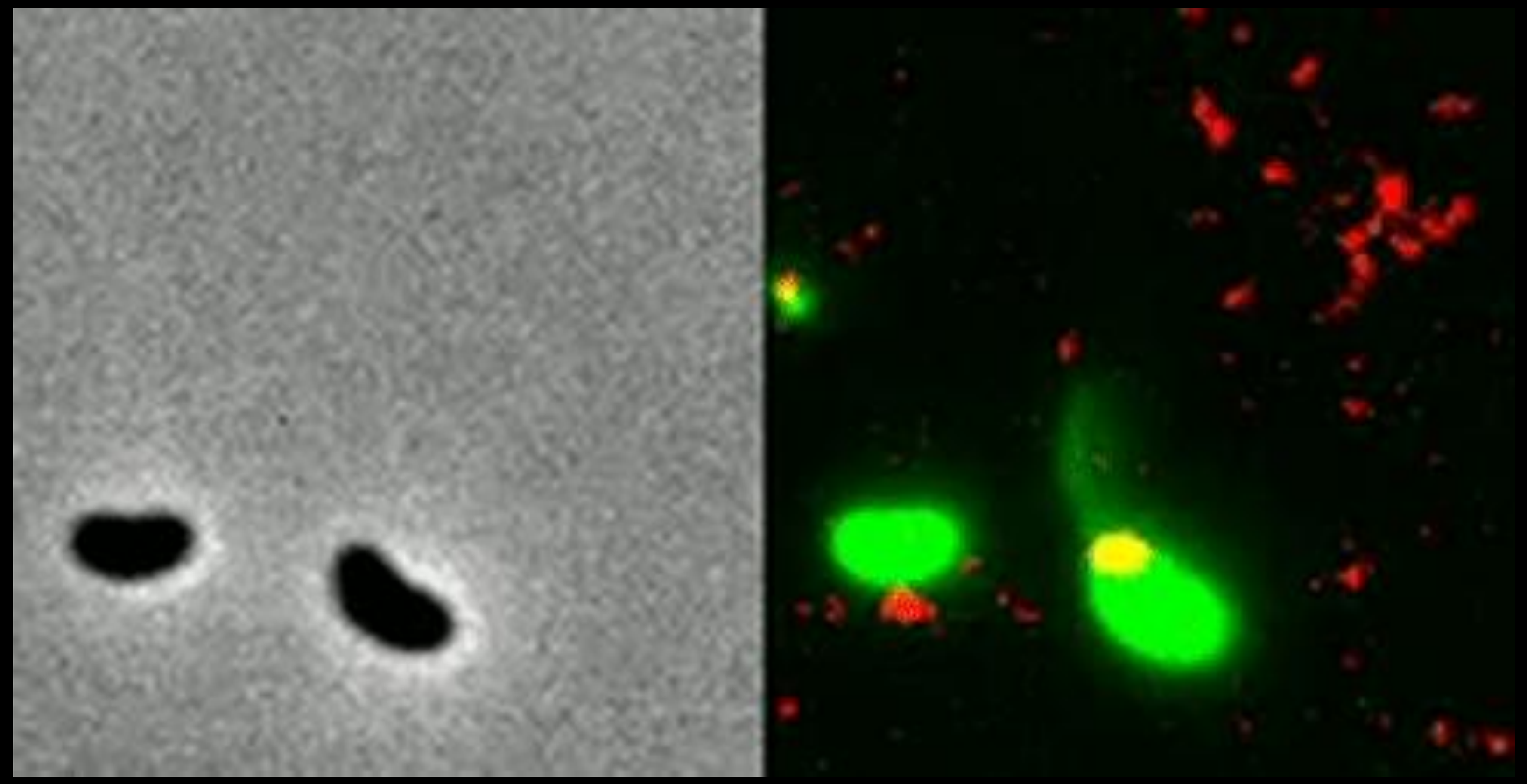



True
Transformer
Earth
Sun
True force
Earth
Sun
Predicted force
True force law (Newton)
$F \propto \frac{m_1 m_2}{r^2}$
Predicted force law (transformer)
$F \propto \left(\sin\left(\frac{1}{\sin(r - 0.24)}\right) + 1.45\right) * \frac{1}{\frac{1}{r} + m_2}$


Text-to-Text
Image-to-Text
Foundation Models
Text-to-Video
Text-to-Image


UMAP from foundation model
Human interactome
Disease 1 Target Module
Disease 3 Target Module
Disease Target Module
Disease 2 Target Module
Disease 3 Target Module
Represents
S_1
S_2
S_3
t_1
t_2
Disease gene
Drug Target
Shortest path to the closest disease gene
1

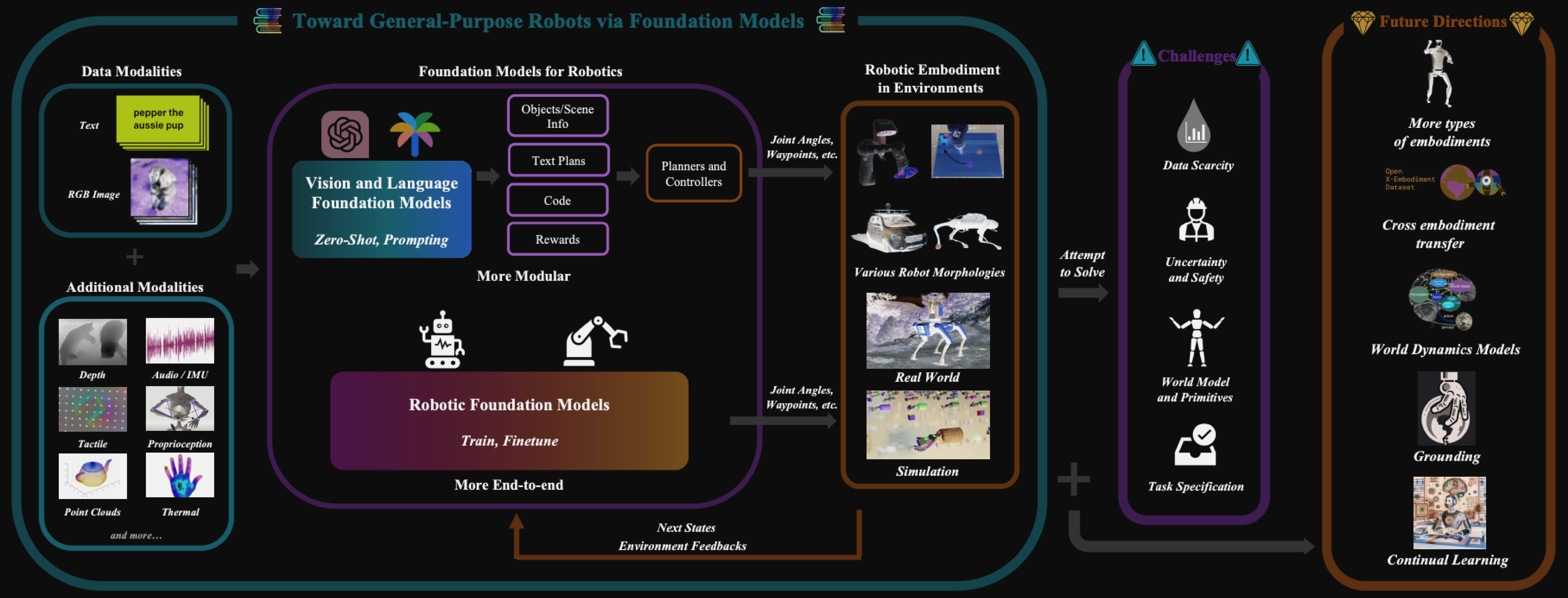


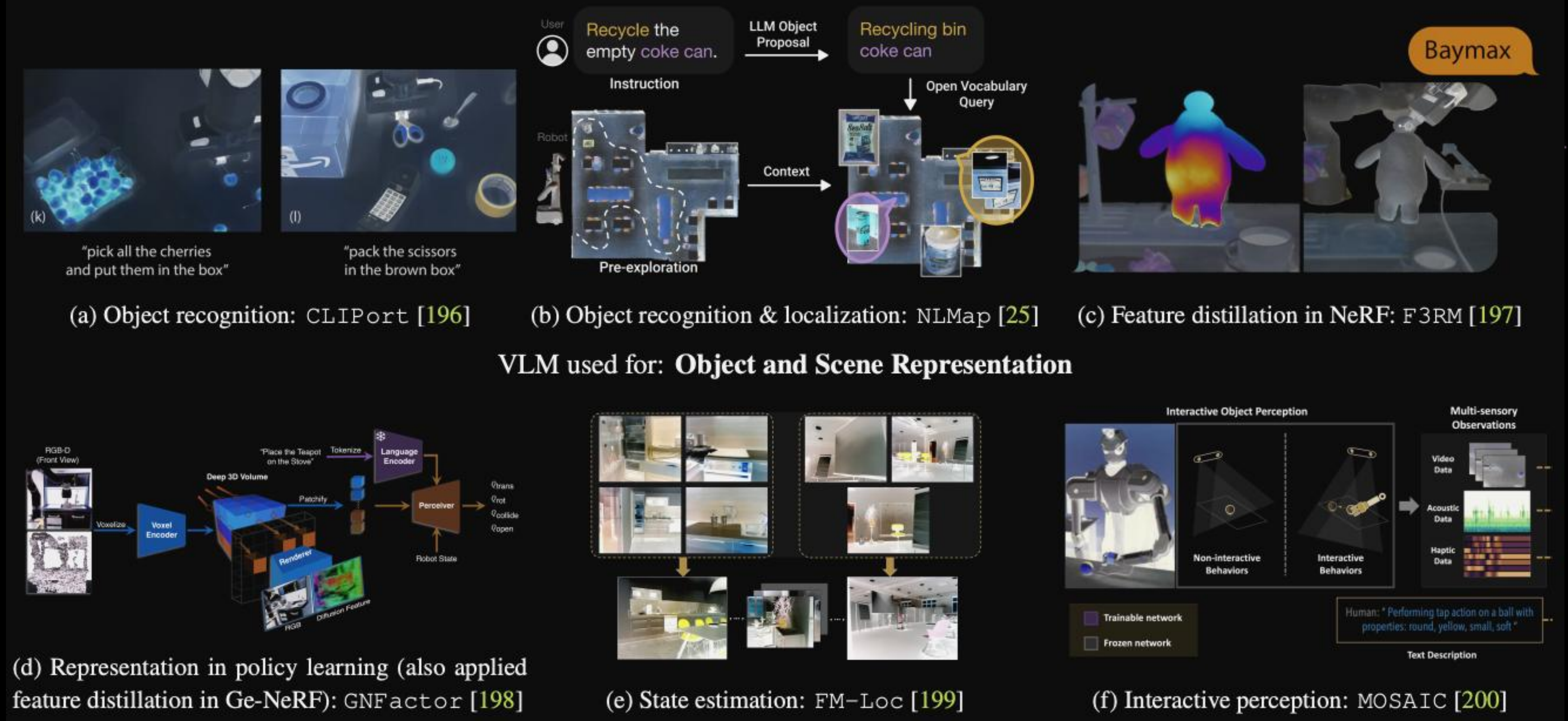


Toward General-Purpose Robots via Foundation Models: A Survey and Meta-Analysis

Yafei Hu[1*] Quanting Xie[1*] Vidhi Jain[1*]
Jonathan Francis[1,2] Jay Patrikar[1] Nikhil Keetha[1] Seungchan Kim[1] Yaqi Xie[1] Tianyi Zhang[1]
Shibo Zhao[1] Yu Quan Chong[1] Chen Wang[3] Katia Sycara[1] Matthew Johnson-Roberson[1]
Dhruv Batra[4,5] Xiaolong Wang[6] Sebastian Scherer[1] Zsolt Kira[4] Fei Xia[7†] Yonatan Bisk[1,5†]

[1]CMU [2]Bosch Center for AI [3]SAIR Lab [4]Georgia Tech [5]FAIR at Meta [6]UC San Diego [7]Google DeepMind



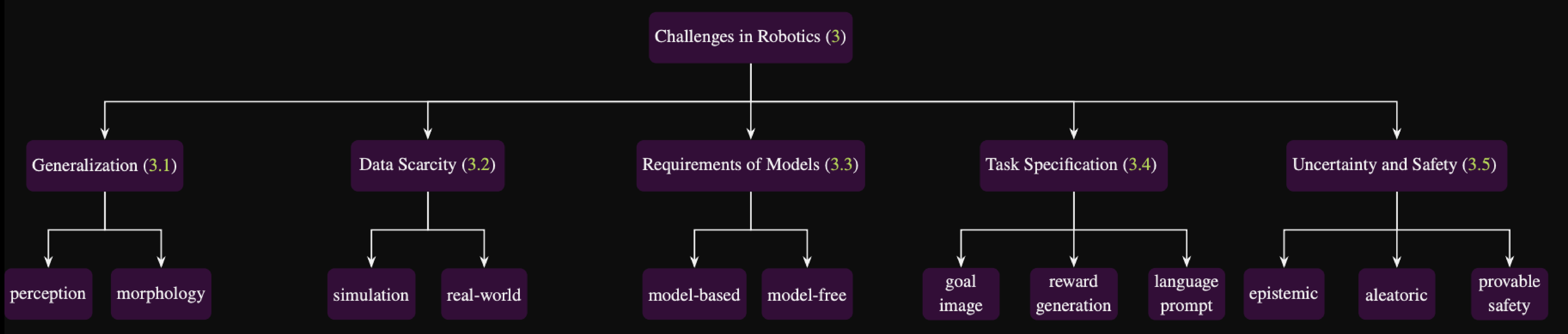


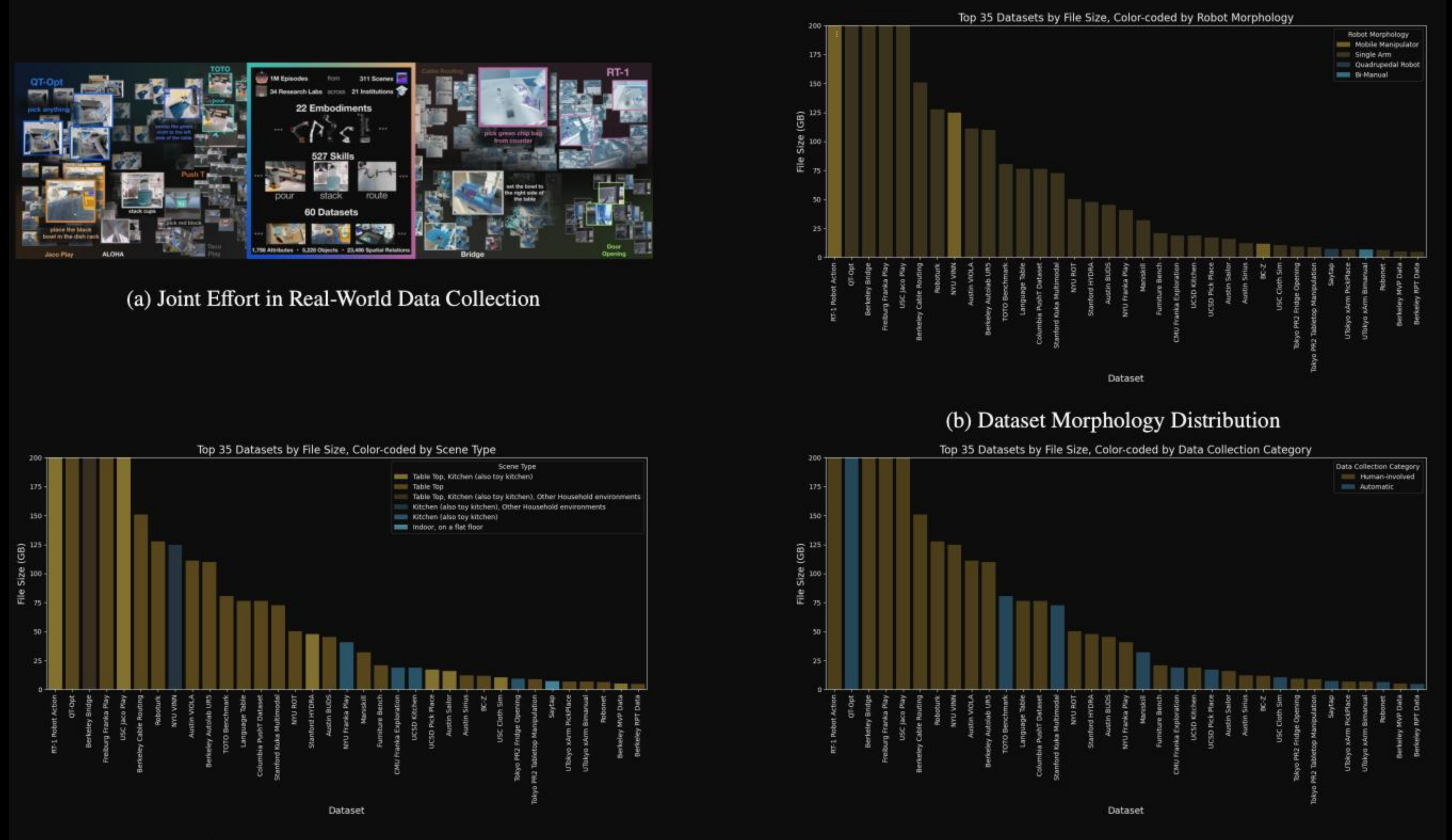
(a) Object recognition: CLIPort [196] (b) Object recognition & localization: NLMap [25] (c) Feature distillation in NeRF: F3RM [197]

VLM used for: **Object and Scene Representation**

(d) Representation in policy learning (also applied feature distillation in Ge-NeRF): GNFactor [198] (e) State estimation: FM-Loc [199] (f) Interactive perception: MOSAIC [200]

VFM and VLM used for: **Policy learning, State estimation and Interactive perception**





(a) Joint Effort in Real-World Data Collection (b) Dataset Morphology Distribution

(c) Scene Type Classification (d) Data Collection Methods

Figure 4: Comprehensive visualizations of the Open-X Embodiment Dataset encompassing data collection methods, robot morphologies, and scene types.

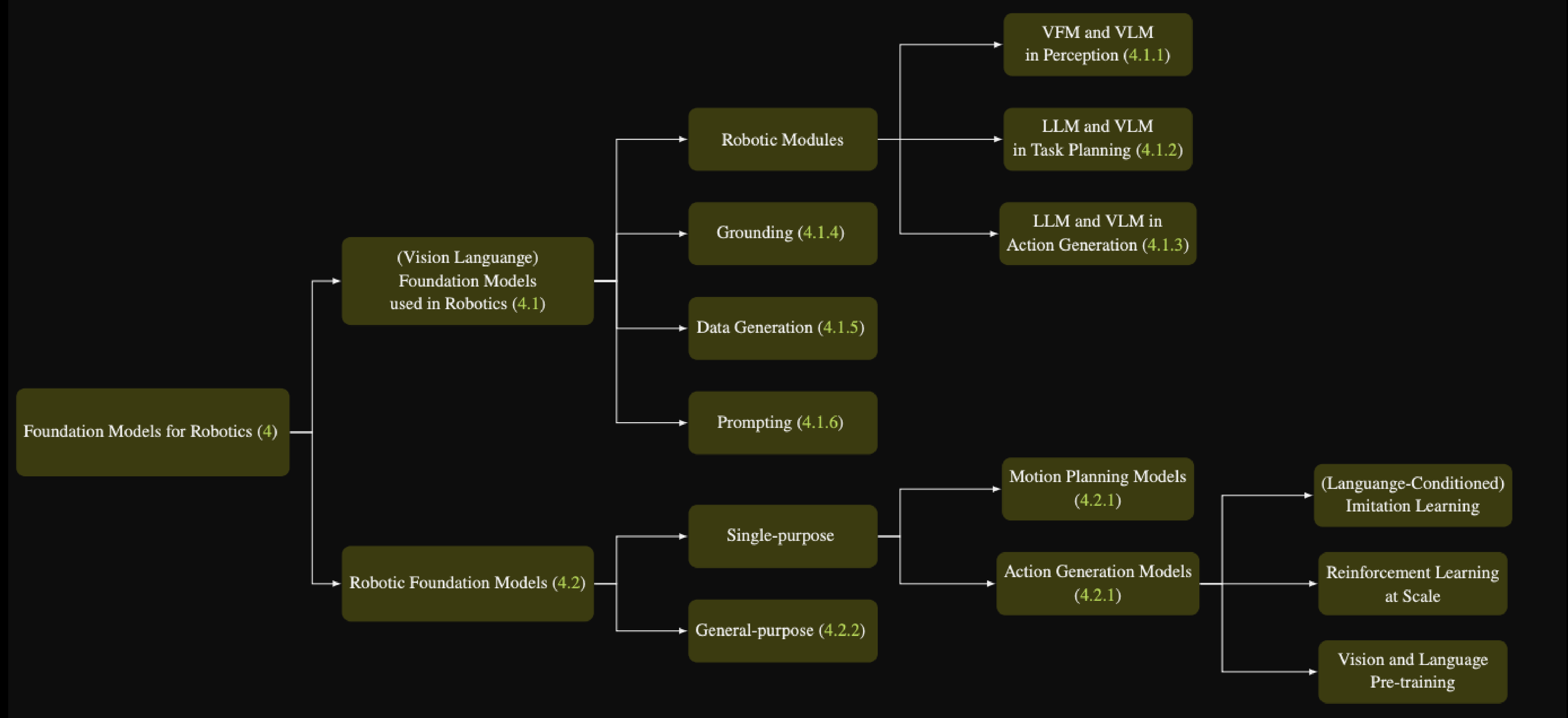
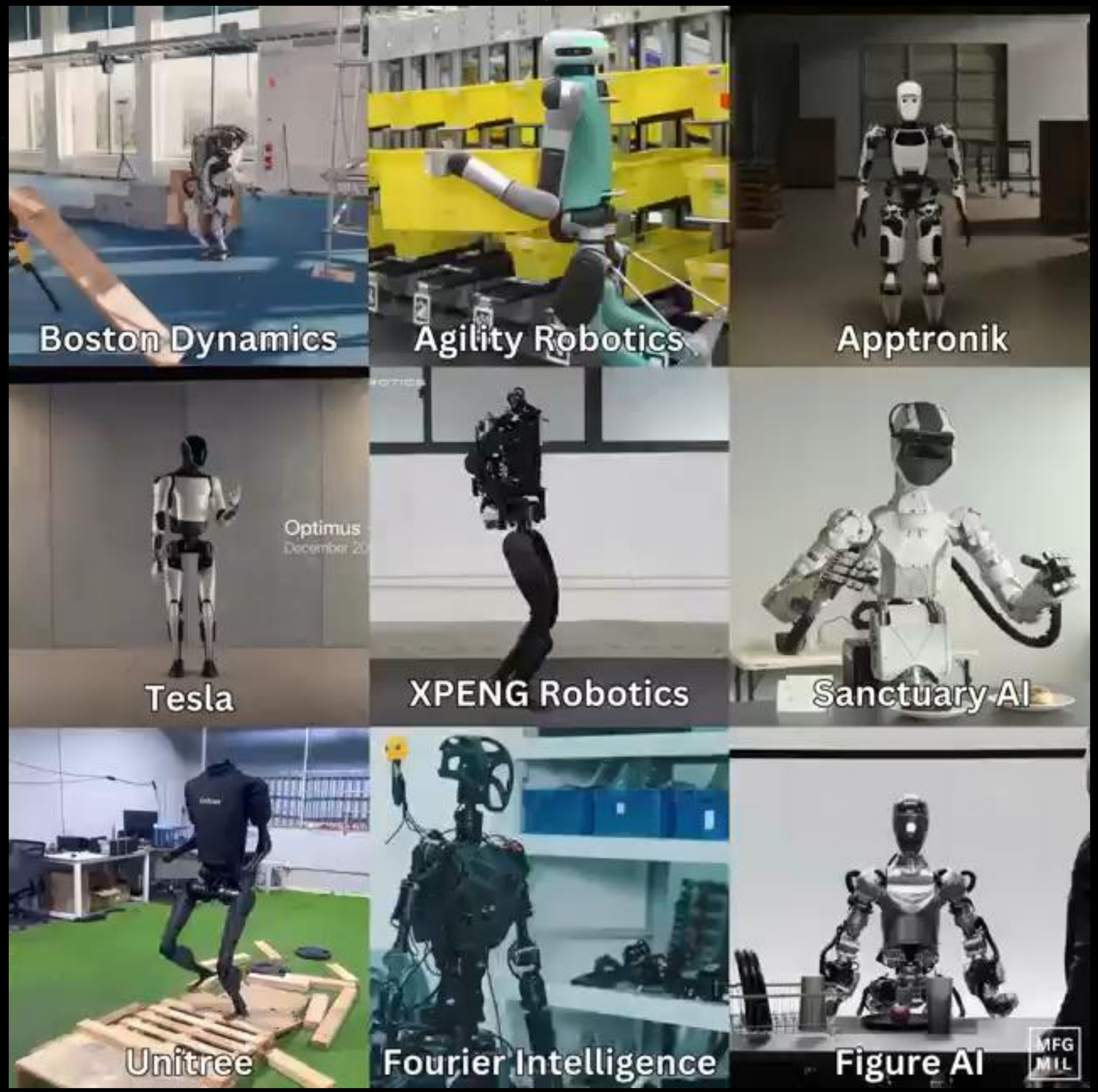
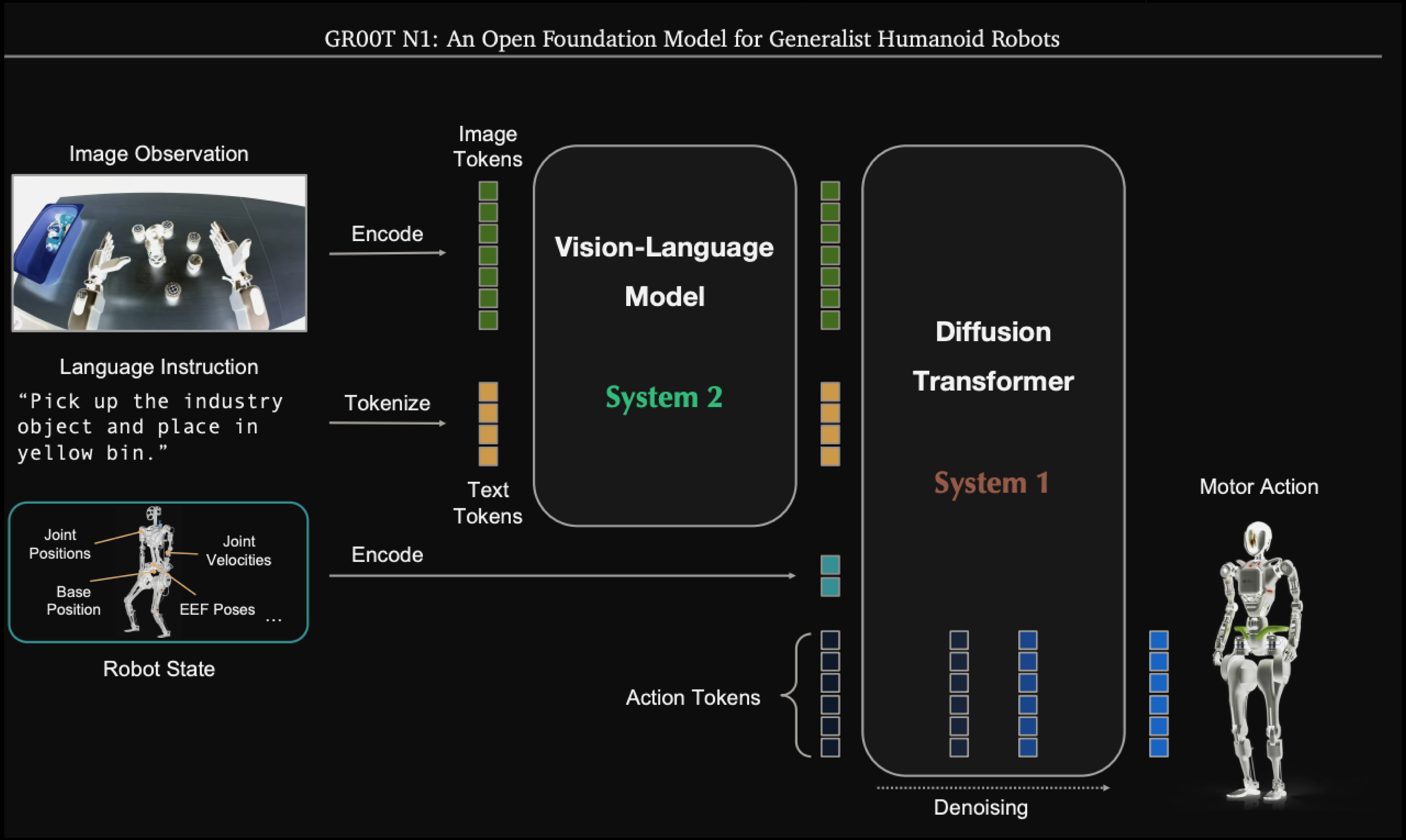


Figure 5: Conceptual Framework of Foundation Models in Robotics: The figure illustrates a structured taxonomy of foundational models, categorized into two primary segments: the application of existing foundation models (vision and language models) to robotics, and the development of robotic-specific foundation models. This includes distinctions between vision and language models used as perception tools, in planning, and in action, as well as the differentiation between single-purpose and general-purpose robot foundation models.


Toward General-Purpose Robots via Fou
A Survey and Meta-Analy
Yafei Hu[1*] Quanting Xie[1*] Vidhi
Jonathan Francis[1,2] Jay Patrikar[1] Nikhil Keetha[1] Seungchan K
Shibo Zhao[1] Yu Quan Chong[1] Chen Wang[3] Katia Sycara
Dhruv Batra[4,5] Xiaolong Wang[6] Sebastian Scherer[1] Zsolt Kira[4]
[1]CMU [2]Bosch Center for AI [3]SAIR Lab [4]Georgia Tech [5]FAIR at Meta
Toward General-Purpose Robots via Foundation Models
Data Modalities
Foundation Models for Robotics
Robotic Embodiment in Environments
Vision and Language Foundation Models
Zero-Shot, Prompting
Objects/Scene Info
Text Plans
Code
Rewards
Planners and Controllers
More Modular
Robotic Foundation Models
Train, Finetune
More End-to-end
Additional Modalities
Various Robot Morphologies
Real World
Simulation
Next States
Environment Feedbacks
(a) Joint Effort in Real-World Data Collection
(b) Dataset Morph
(c) Scene Type Classification
(d) Data Colle
Figure 4: Comprehensive visualizations of the Open-X Embodiment Dataset encompassing data collection and scene types.
Recycle the empty coke can.
LLM Object Proposal
Recycling bin coke can
Instruction
Baymax
distillation in NeRF: F3RM [197]
tive perception: MOSAIC [200]
rception
Task Specification (3.4)
Uncertainty and Safety (3.5)
goal image
reward generation
language prompt
epistemic
aleatoric
provable safety
in planning, and in action, as well as the differentiation between single-purpose and general-purpose robot foundation models.
Boston Dynamics
Agility Robotics
Apptronik
Tesla
Optimus
XPENG Robotics
Sanctuary AI
Unitree
Fourier Intelligence
Figure AI

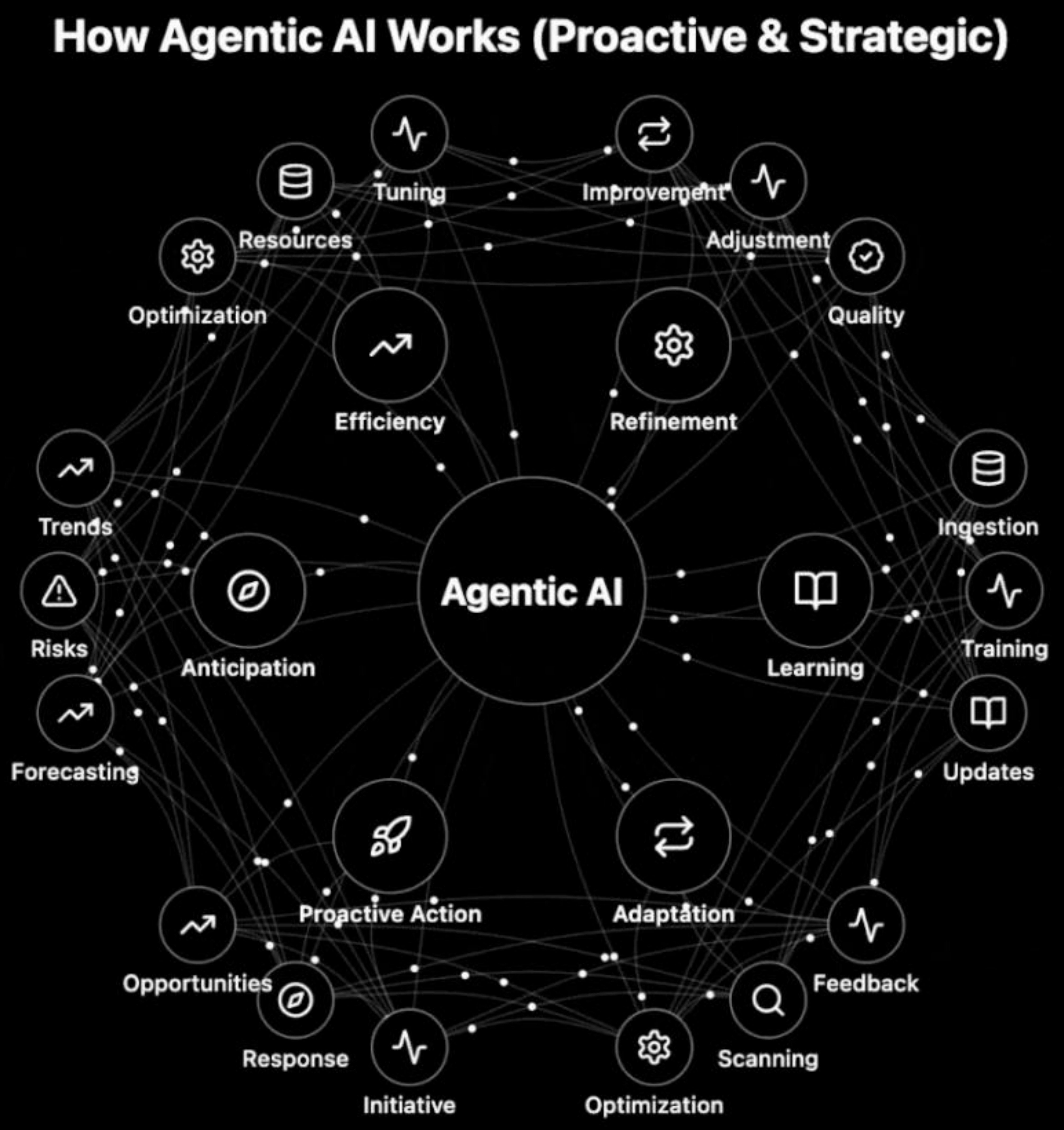

Boston Dynamics
Agility Robotics
Apptronik
Tesla
XPENG Robotics
Sanctuary AI
Unitree
Fourier Intelligence
Figure AI


GR00T N1: An Open Foundation Model for Generalist Humanoid Robots
Image Observation
Language Instruction
"Pick up the industry object and place in yellow bin."
Joint Positions
Joint Velocities
Base Position
EEF Poses ...
Robot State
Encode
Tokenize
Encode
Image Tokens
Text Tokens
Vision-Language Model
System 2
Diffusion Transformer
System 1
Action Tokens
Denoising
Motor Action

AGENTS

How do AI agents work?

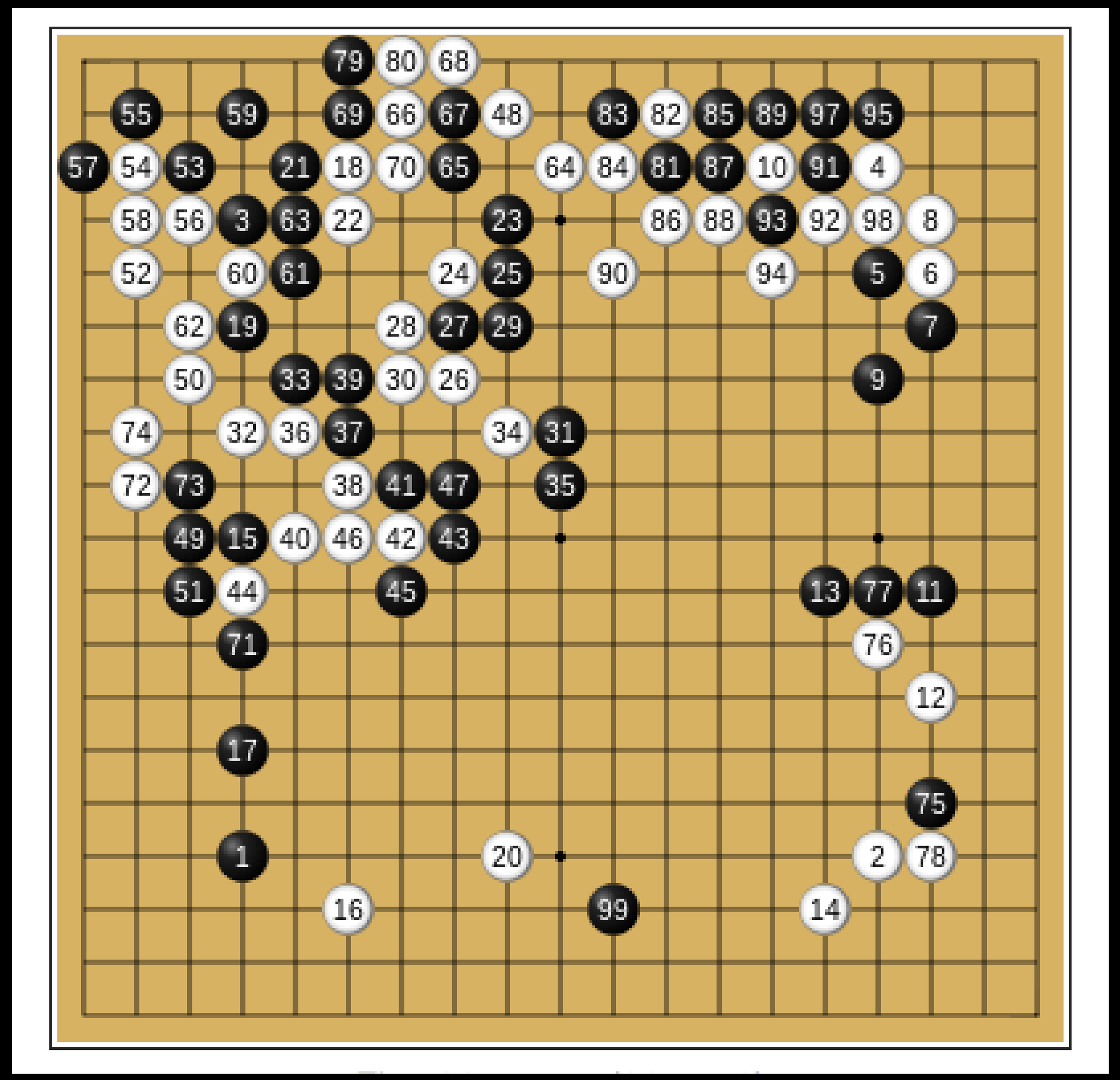
- **Goals**: Humans set goals for AI agents, but the agents decide how to achieve them.
- **Data collection**: AI agents collect data from their environment.
- **Decision making**: AI agents use the data they collect to make decisions about how to act.
- **Action**: AI agents perform actions to achieve their goals.



Introducing

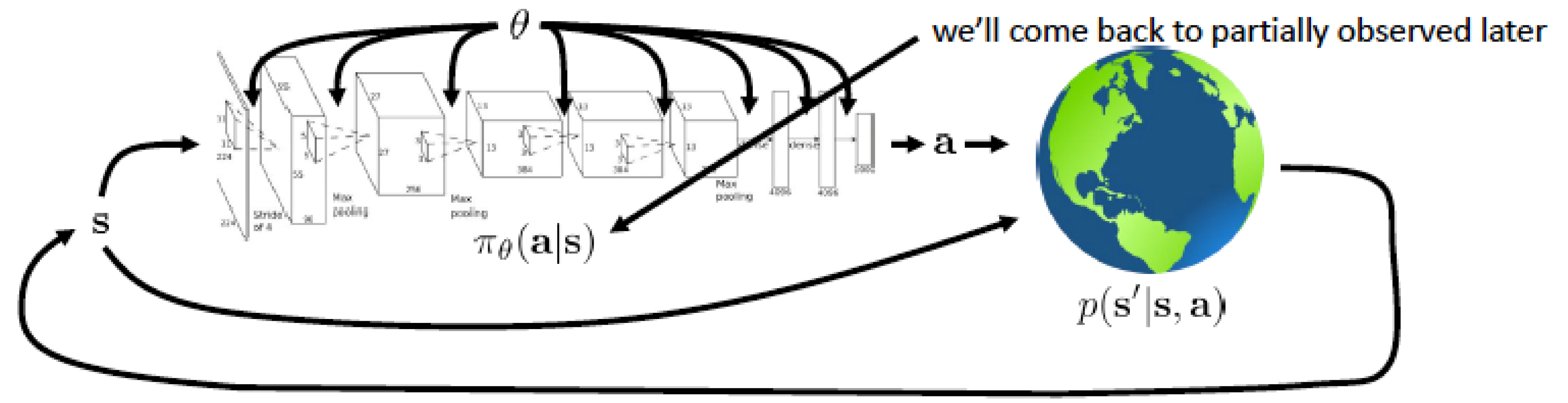
Perplexity Assistant

AlphaGo



AI: ALPHAGO

REINFORCEMENT LEARNING



$$\underbrace{p_\theta(\mathbf{s}_1,\mathbf{a}_1,\ldots,\mathbf{s}_T,\mathbf{a}_T)}_{\pi_\theta(\tau)} = p(\mathbf{s}_1)\prod_{t=1}^{T}\underbrace{\pi_\theta(\mathbf{a}_t|\mathbf{s}_t)p(\mathbf{s}_{t+1}|\mathbf{s}_t,\mathbf{a}_t)}_{\text{Markov chain}}$$

$p_\theta(\mathbf{s}_t,\mathbf{a}_t)$ state-action marginal

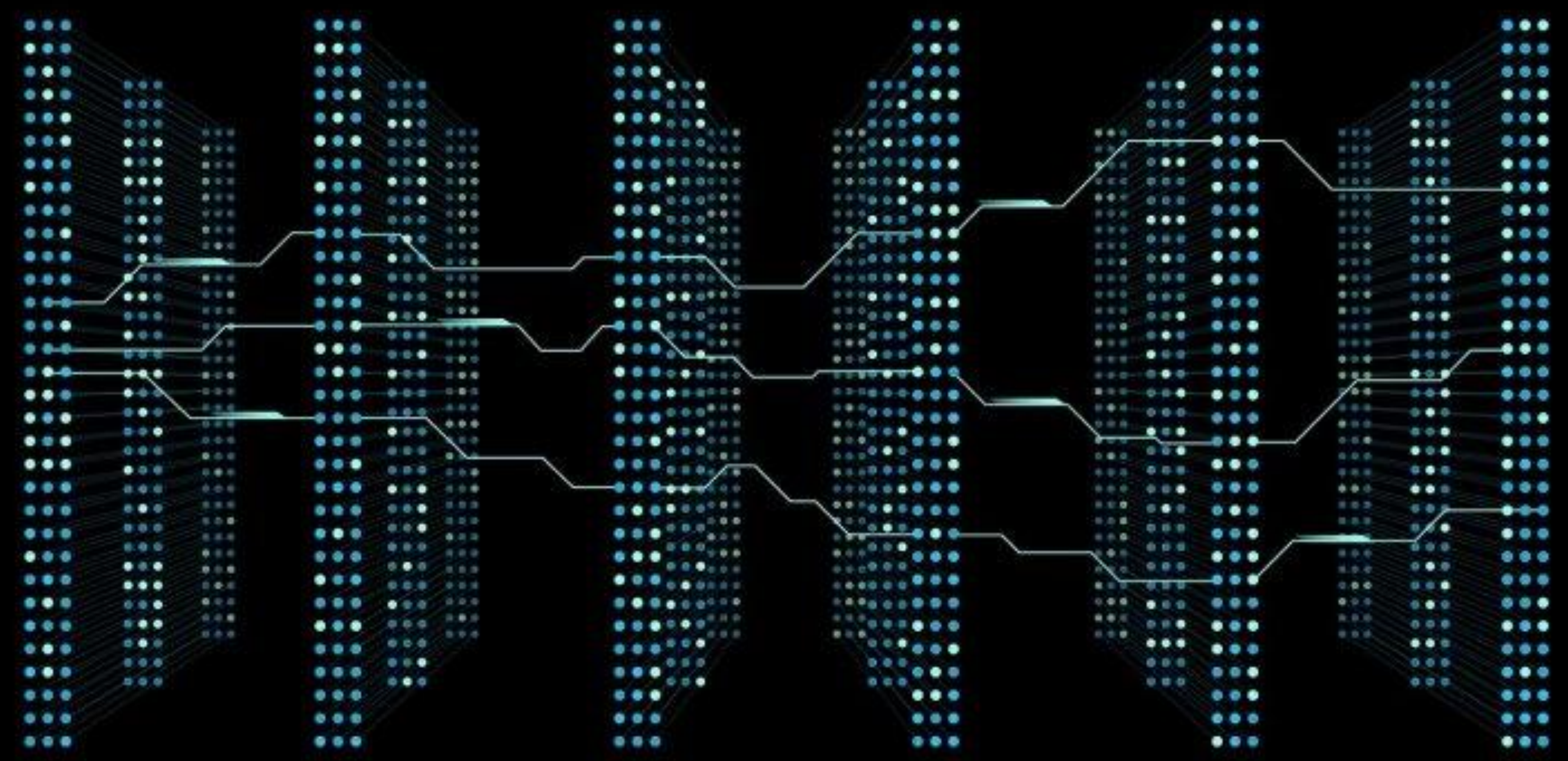
$p_\theta(\mathbf{s},\mathbf{a})$ stationary distribution

$$\theta^\star = \arg\max_\theta E_{(\mathbf{s},\mathbf{a})\sim p_\theta(\mathbf{s},\mathbf{a})}[r(\mathbf{s},\mathbf{a})]$$

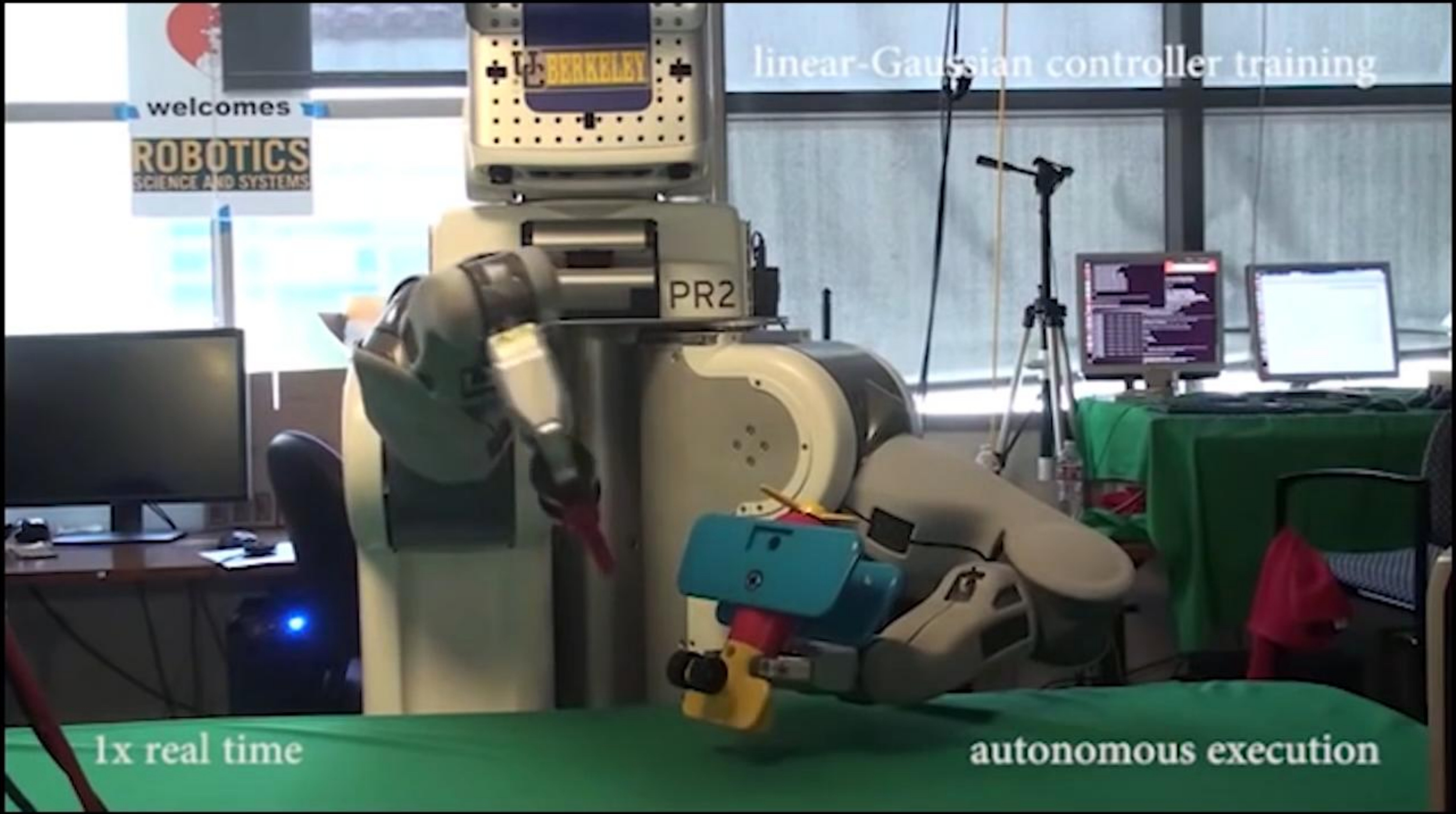
infinite horizon case

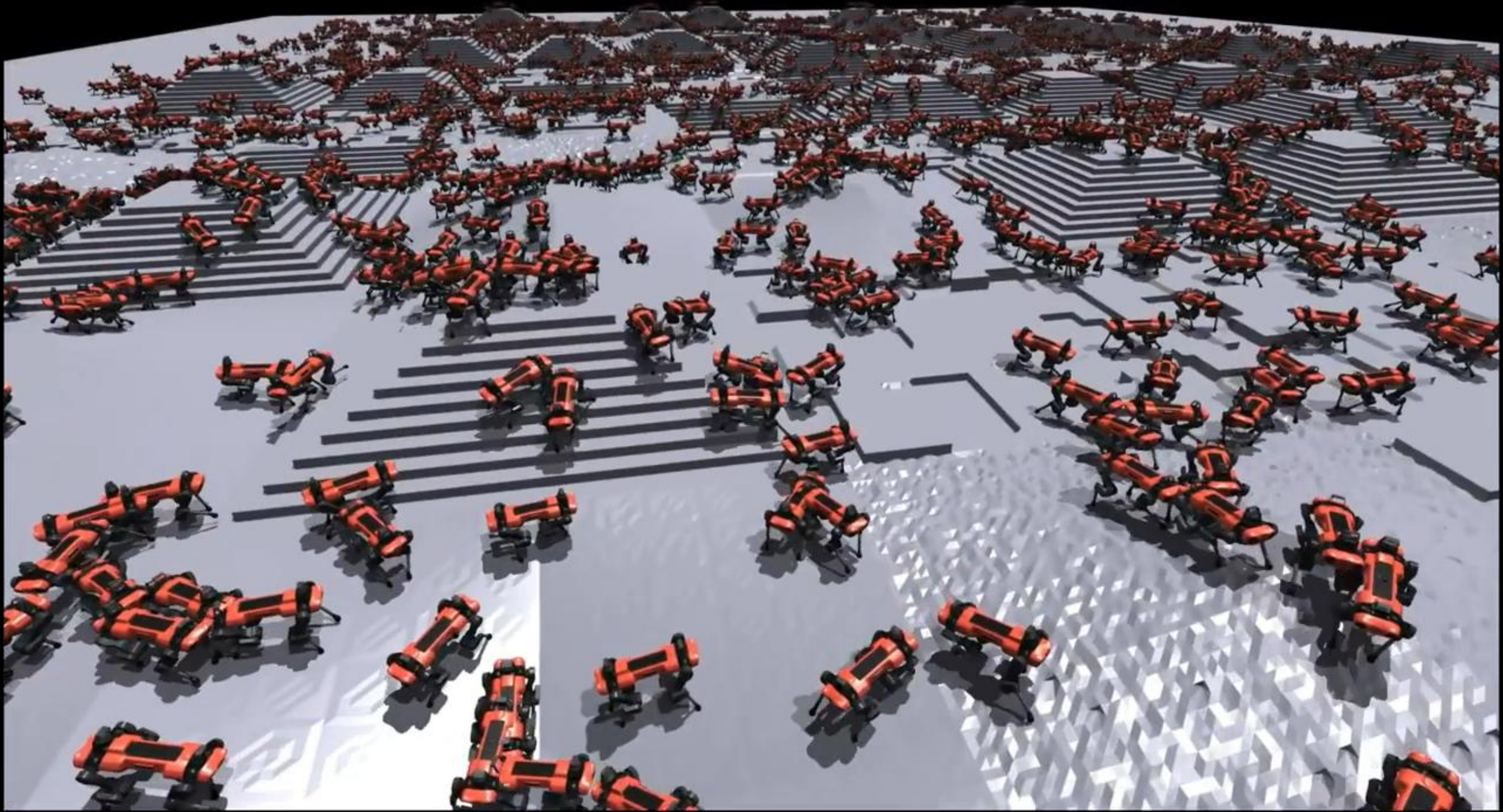
$$\theta^\star = \arg\max_\theta \sum_{t=1}^{T} E_{(\mathbf{s}_t,\mathbf{a}_t)\sim p_\theta(\mathbf{s}_t,\mathbf{a}_t)}[r(\mathbf{s}_t,\mathbf{a}_t)]$$

finite horizon case



REINFORCEMENT LEARNING





Model	AIME 2024		MATH-500	GPQA Diamond	LiveCode Bench	CodeForces
	pass@1	cons@64	pass@1	pass@1	pass@1	rating
OpenAI-o1-mini	63.6	80.0	90.0	60.0	53.8	1820
OpenAI-o1-0912	74.4	83.3	94.8	77.3	63.4	1843
DeepSeek-R1-Zero	71.0	86.7	95.9	73.3	50.0	1444

Table 2 | Comparison of DeepSeek-R1-Zero and OpenAI o1 models on reasoning-related benchmarks.

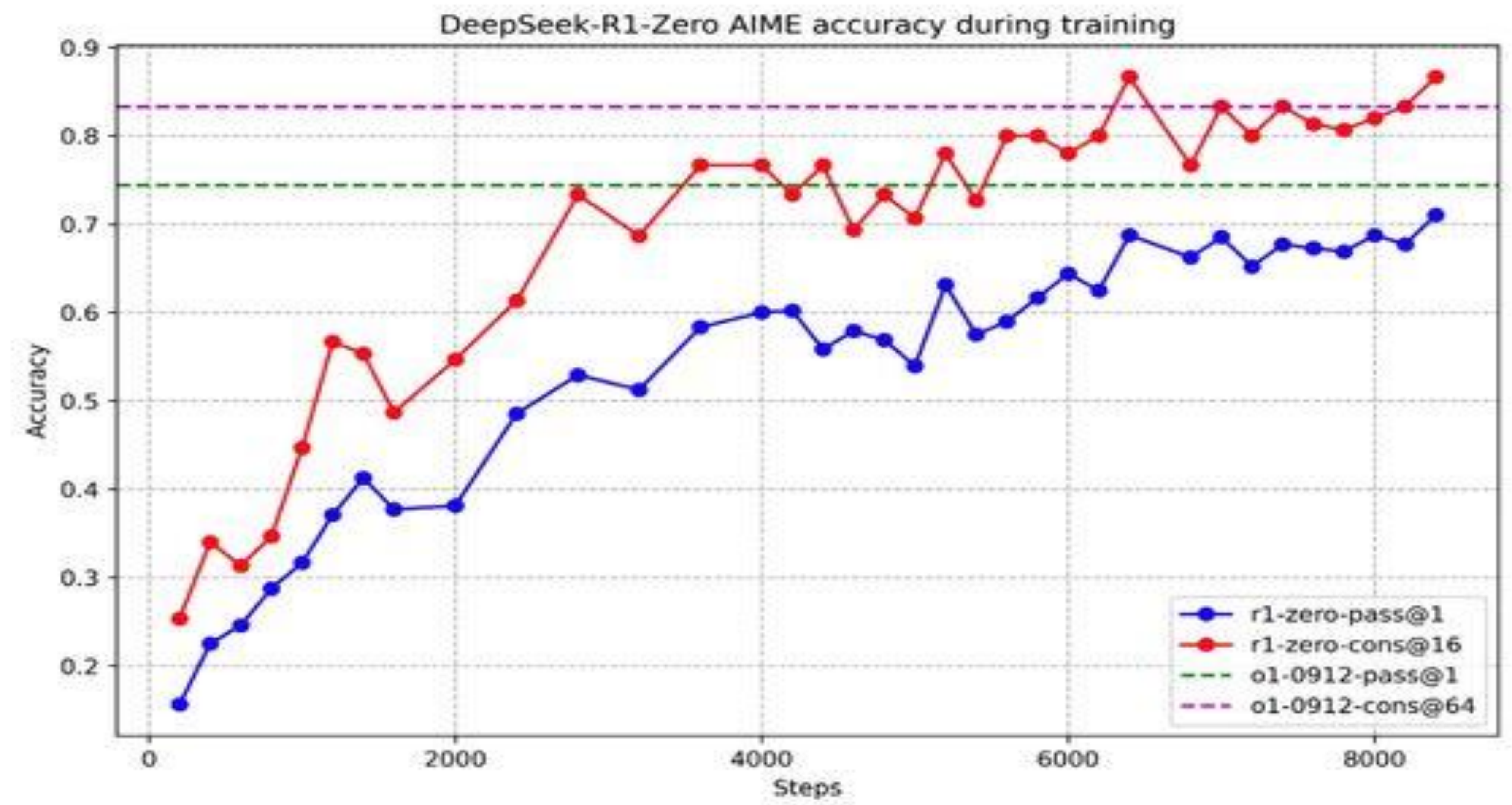
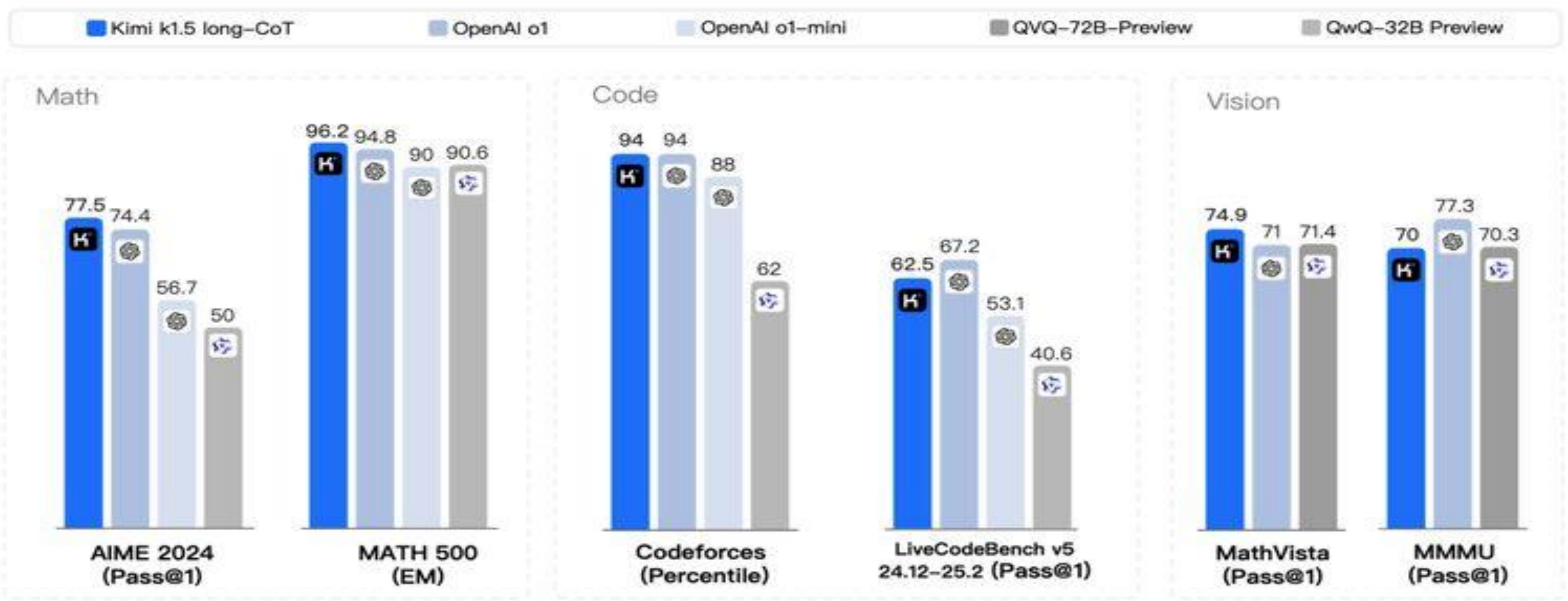
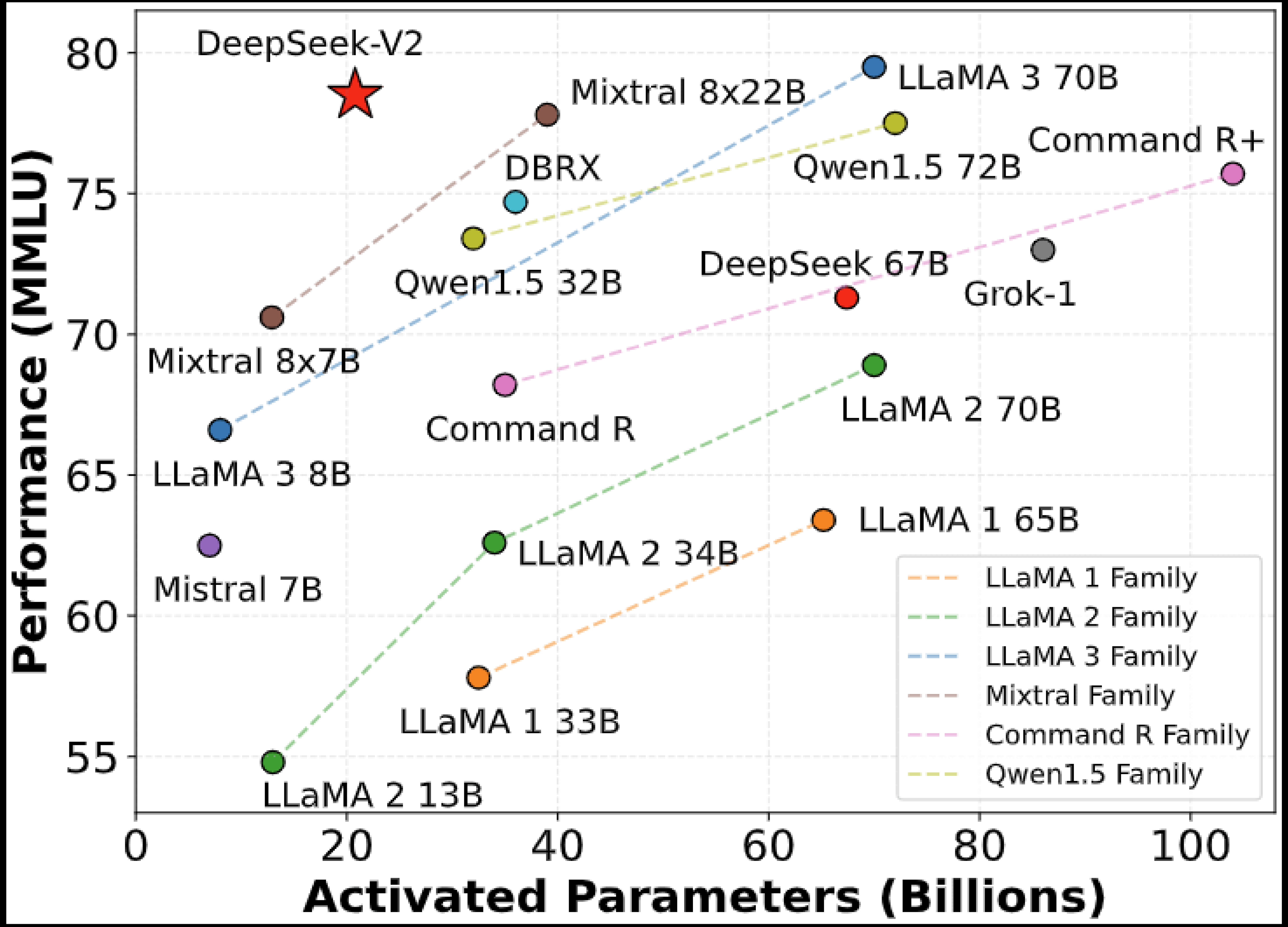
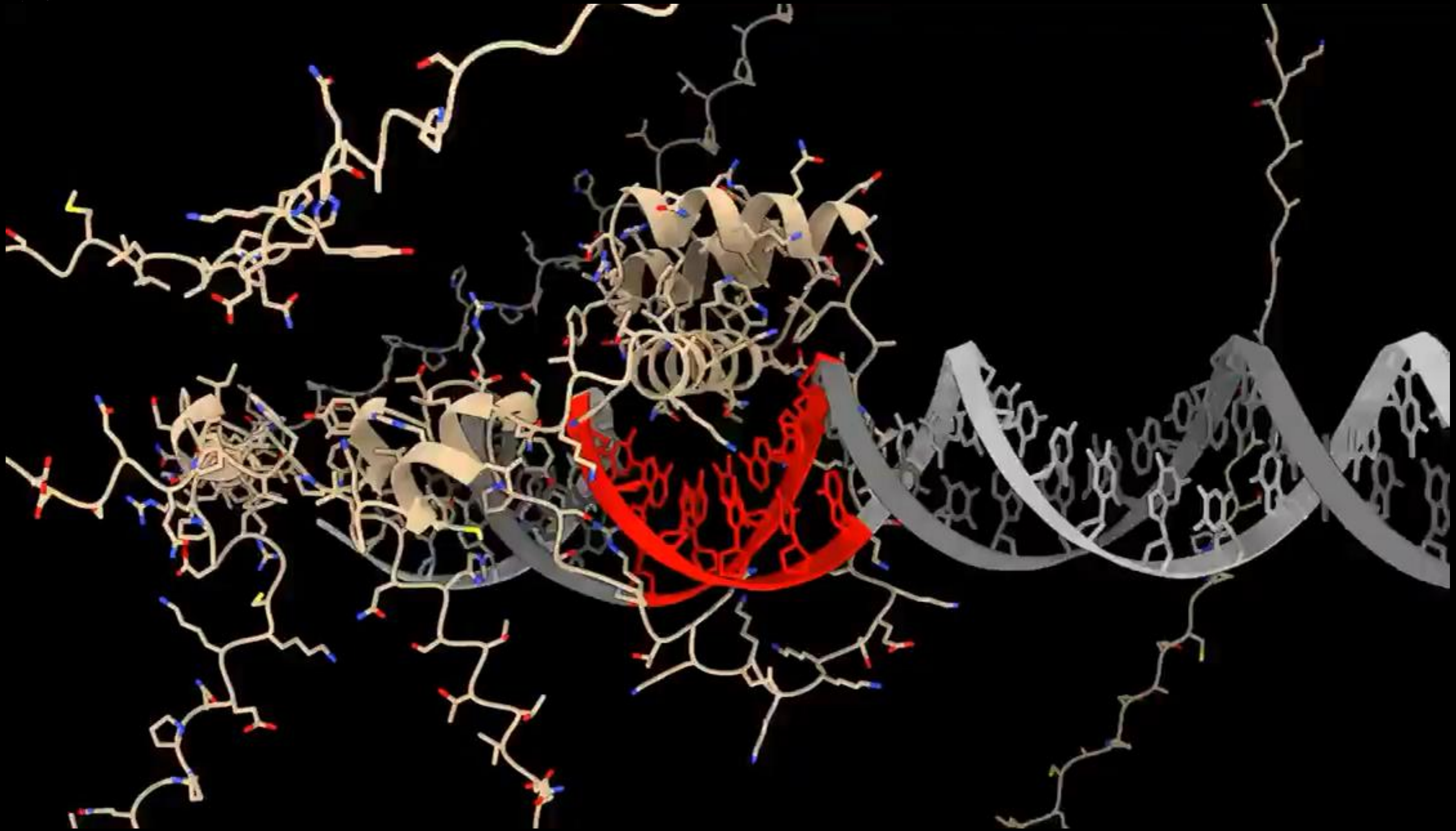


Figure 2 | AIME accuracy of DeepSeek-R1-Zero during training. For each question, we sample 16 responses and calculate the overall average accuracy to ensure a stable evaluation.

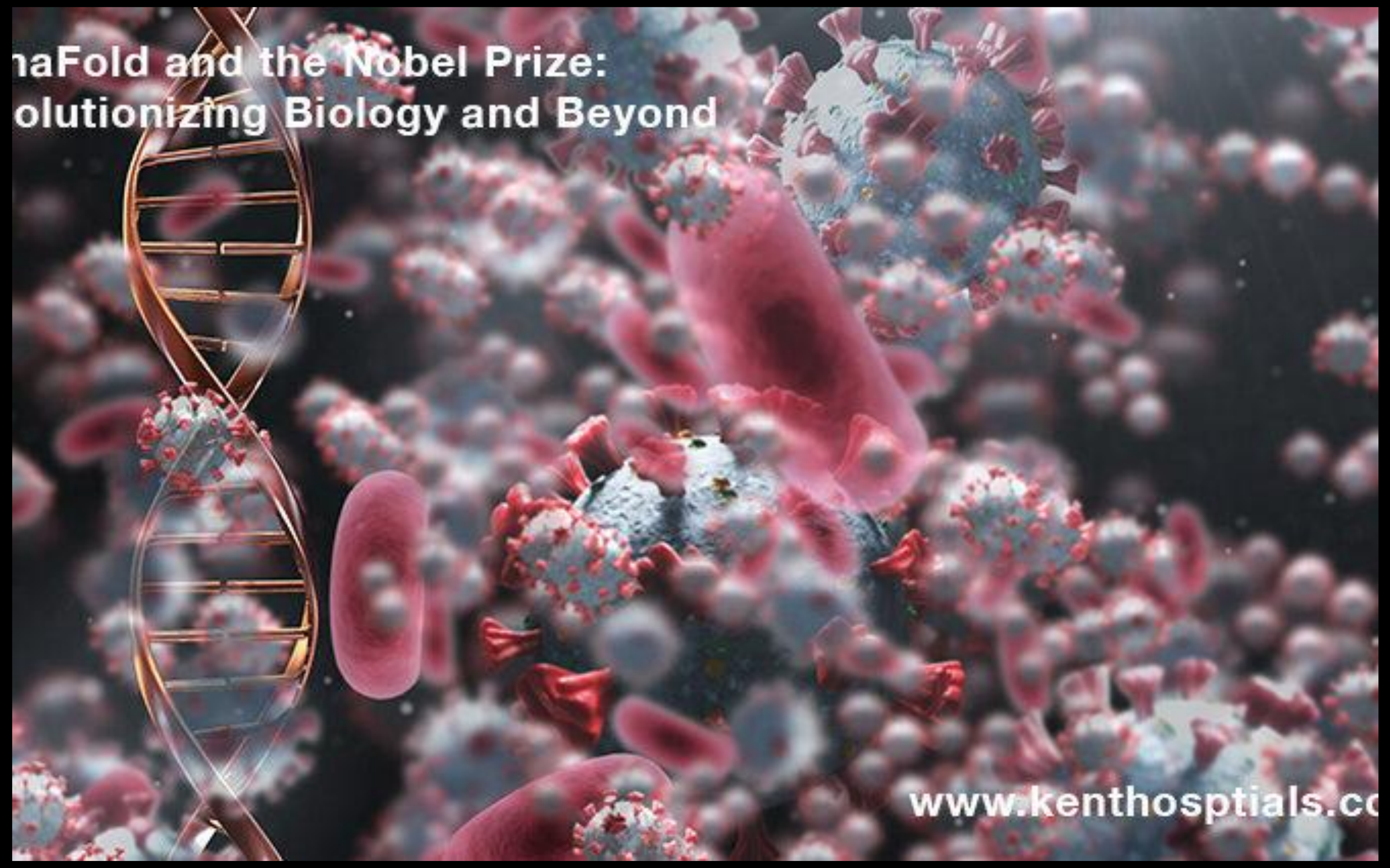
deepseek

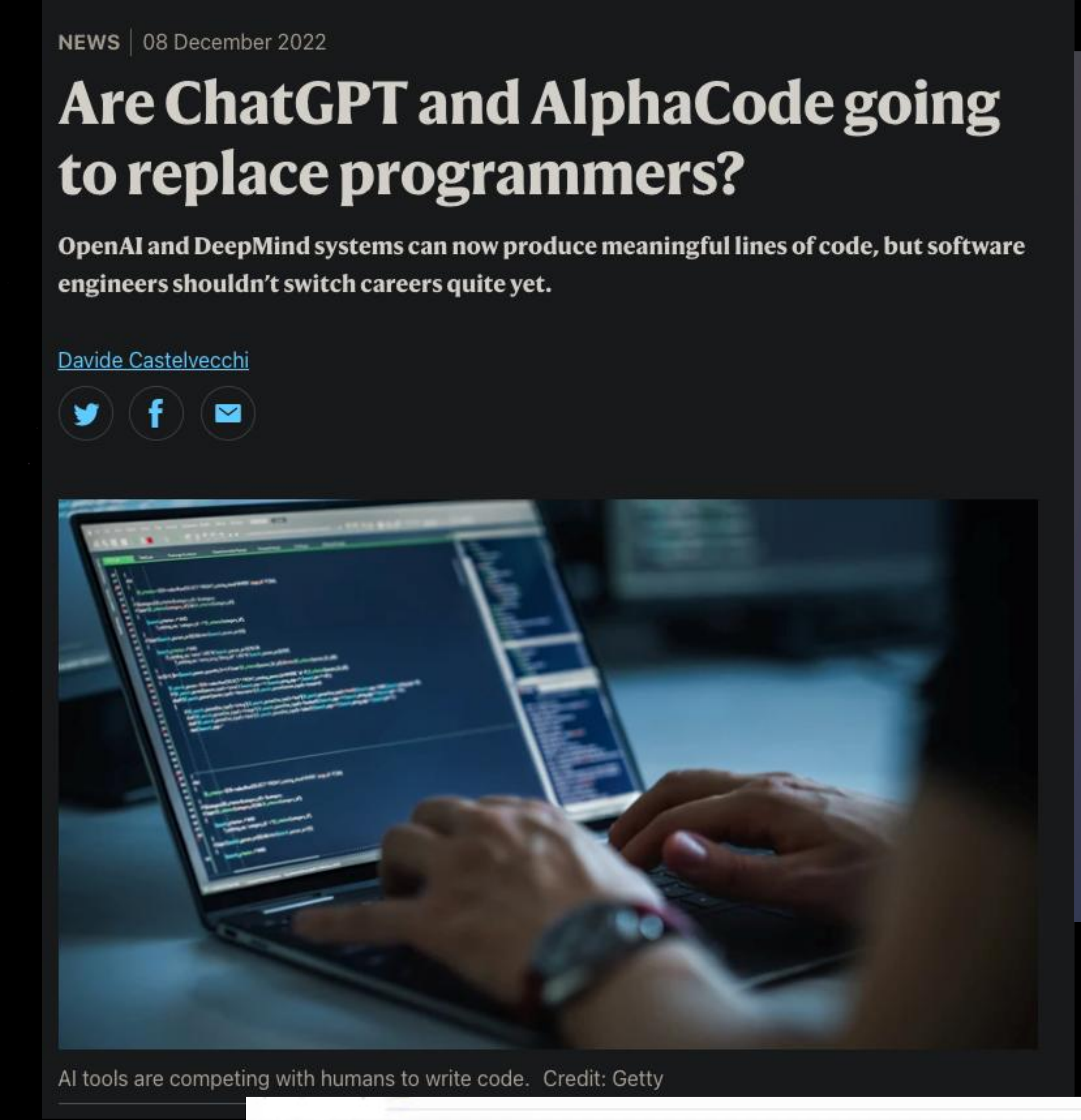







DeepMind CEO Wins Nobel Prize In
Chemistry For Groundbreaking
AI Innovation
WORLD
NEWS


naFold and the Nobel Prize:
olutionizing Biology and Beyond
www.kenthosptials.c

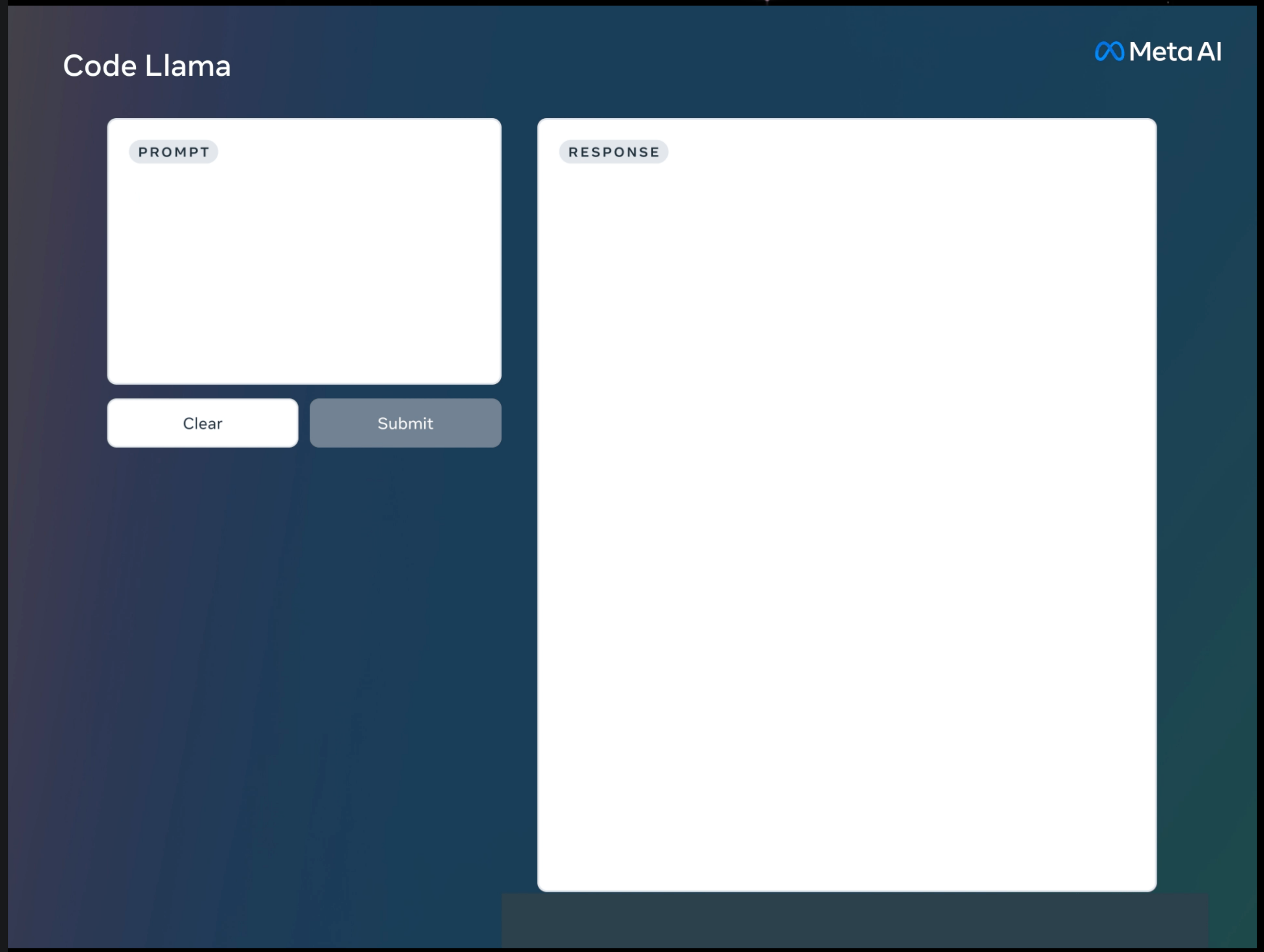

NEWS | 08 December 2022

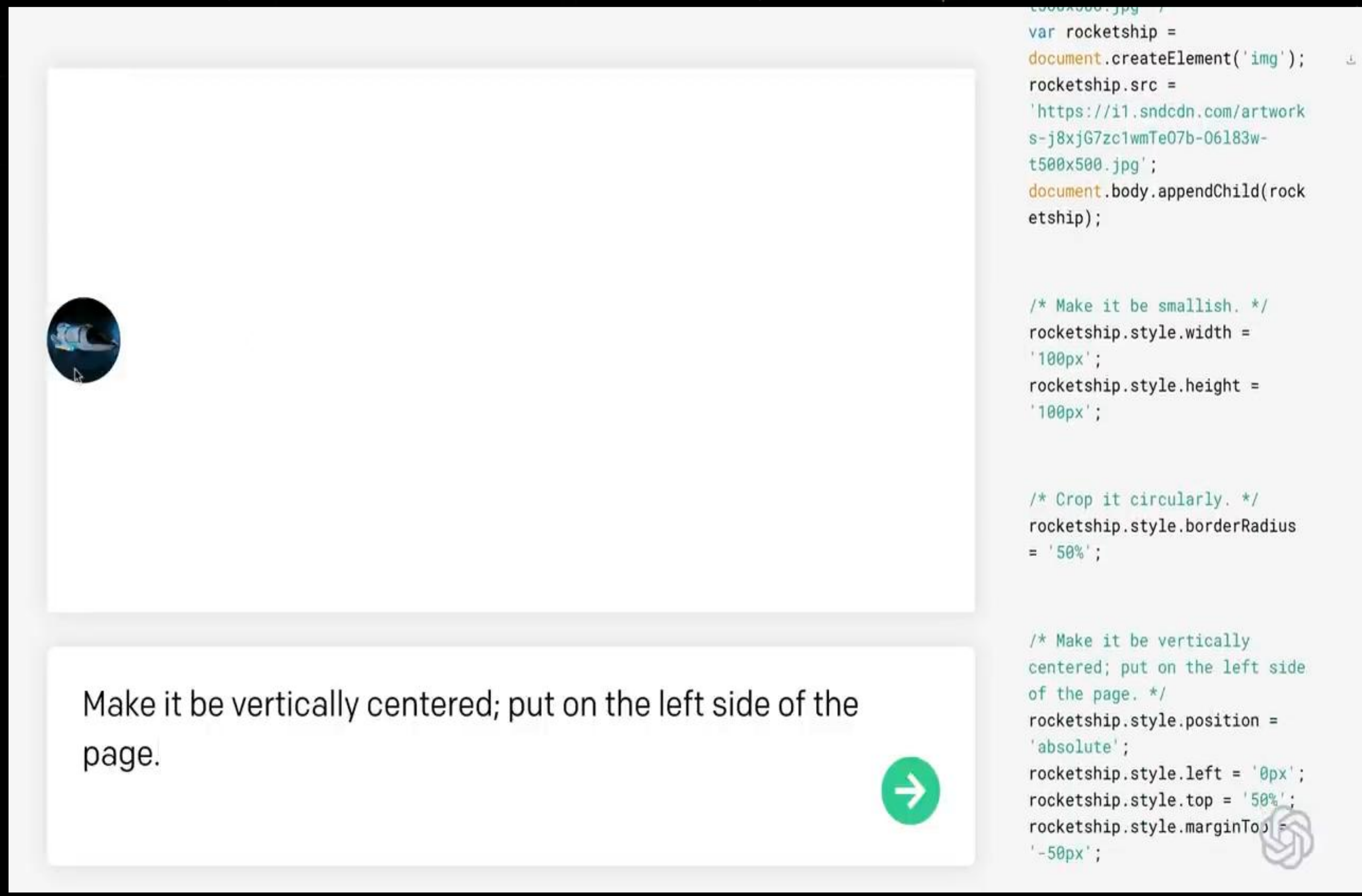
Are ChatGPT and AlphaCode going to replace programmers?

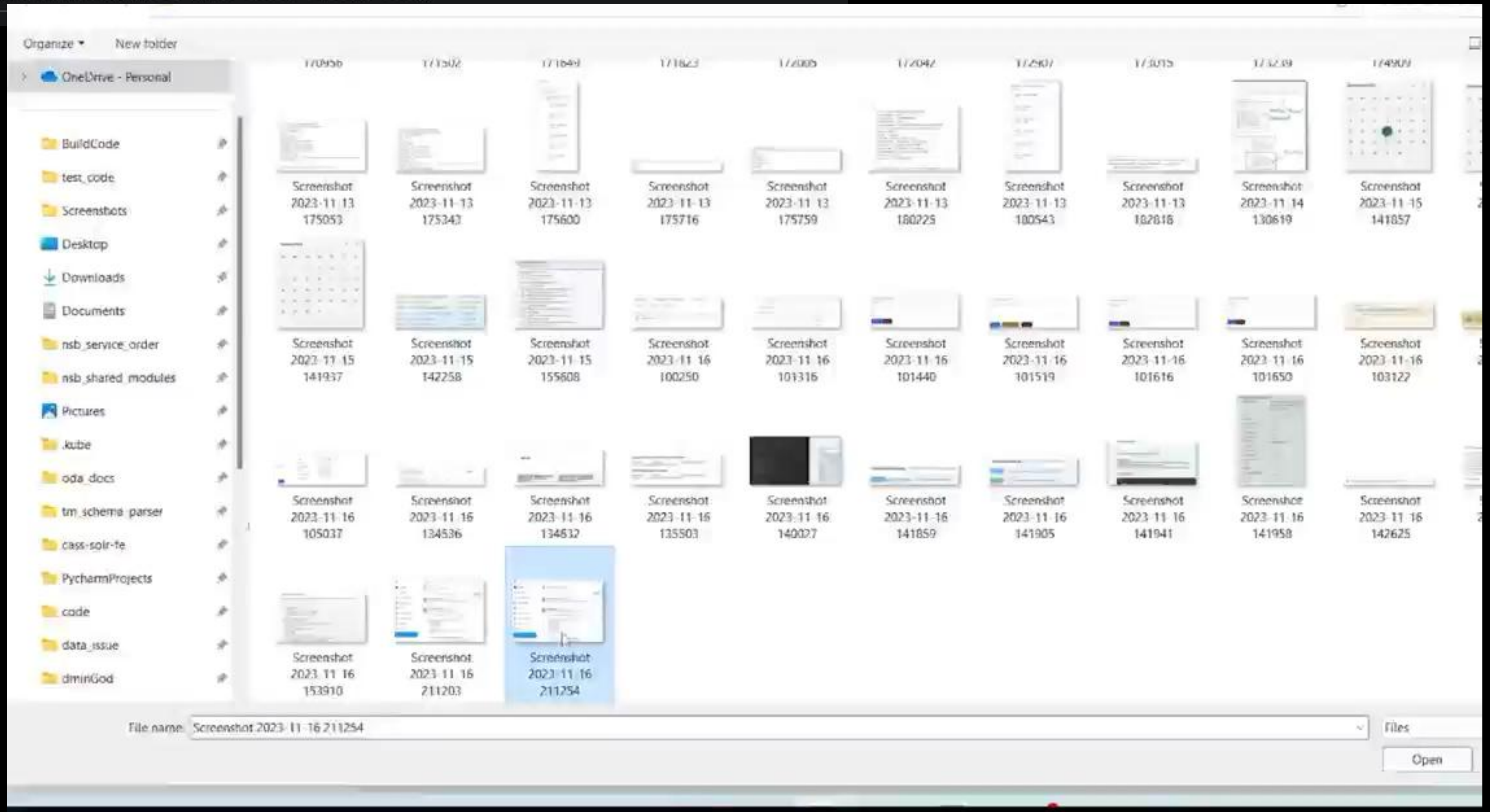
OpenAI and DeepMind systems can now produce meaningful lines of code, but software engineers shouldn't switch careers quite yet.

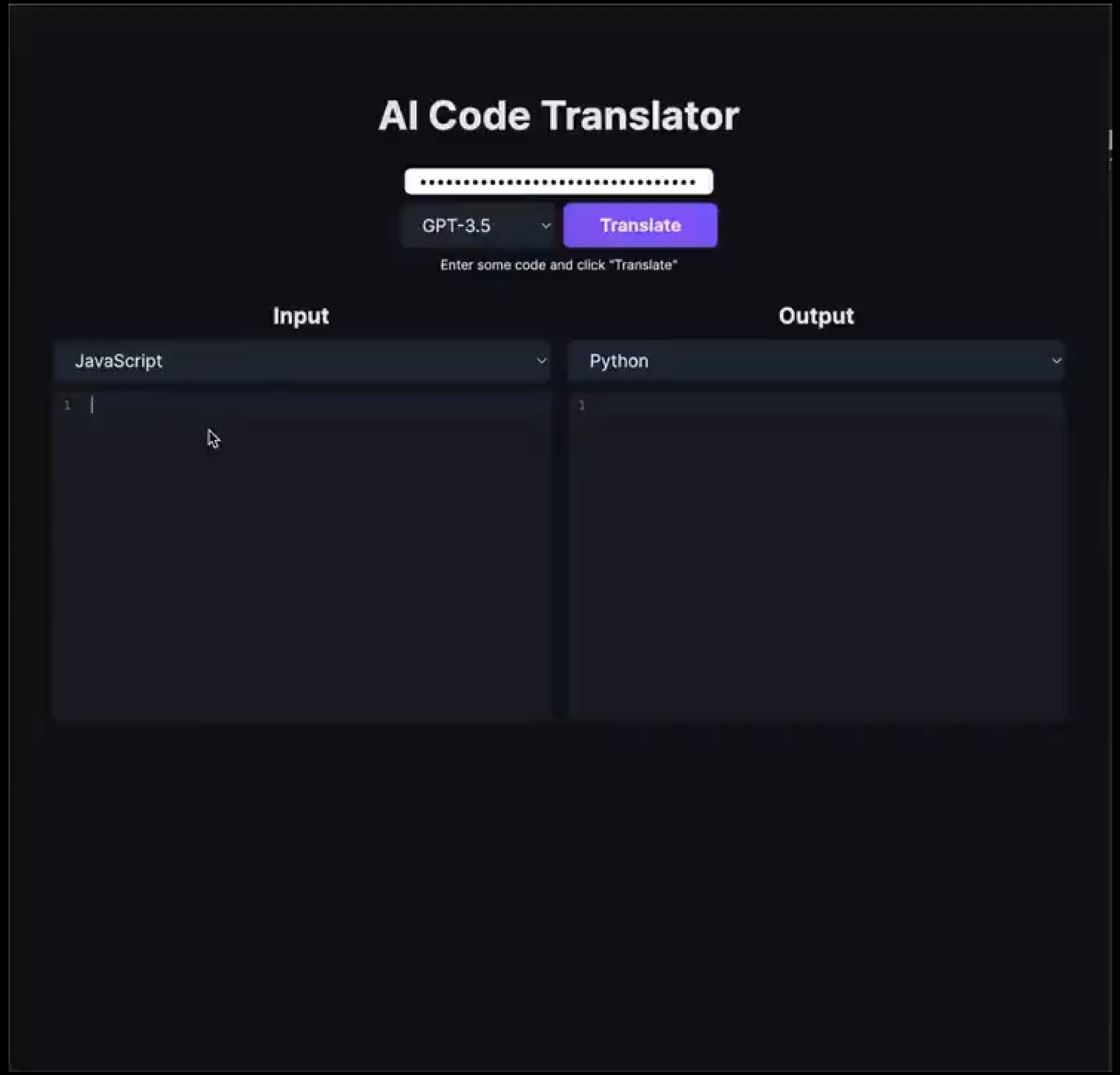
Davide Castelvecchi

AI tools are competing with humans to write code. Credit: Getty





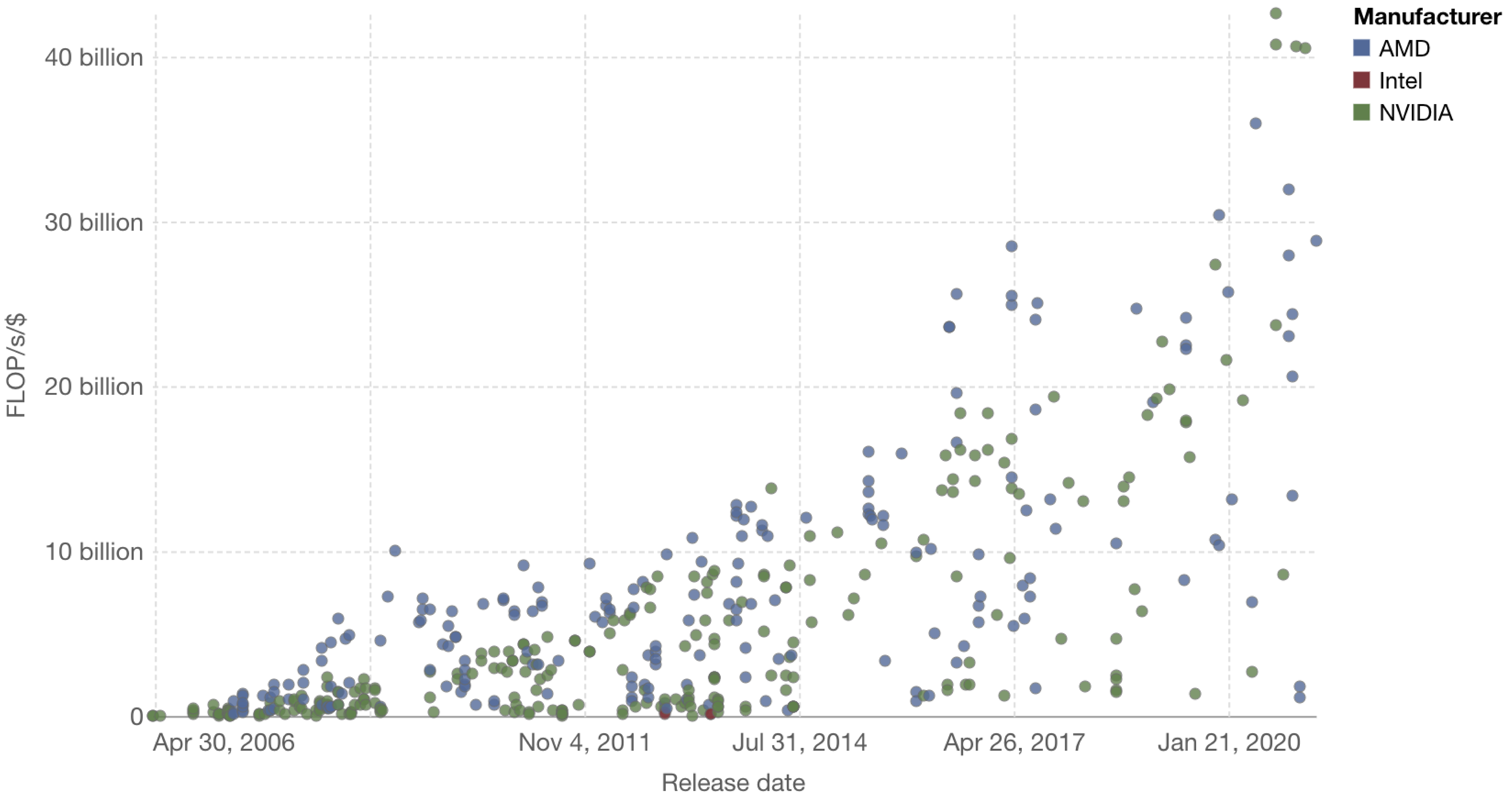




Source: cbinsights.com

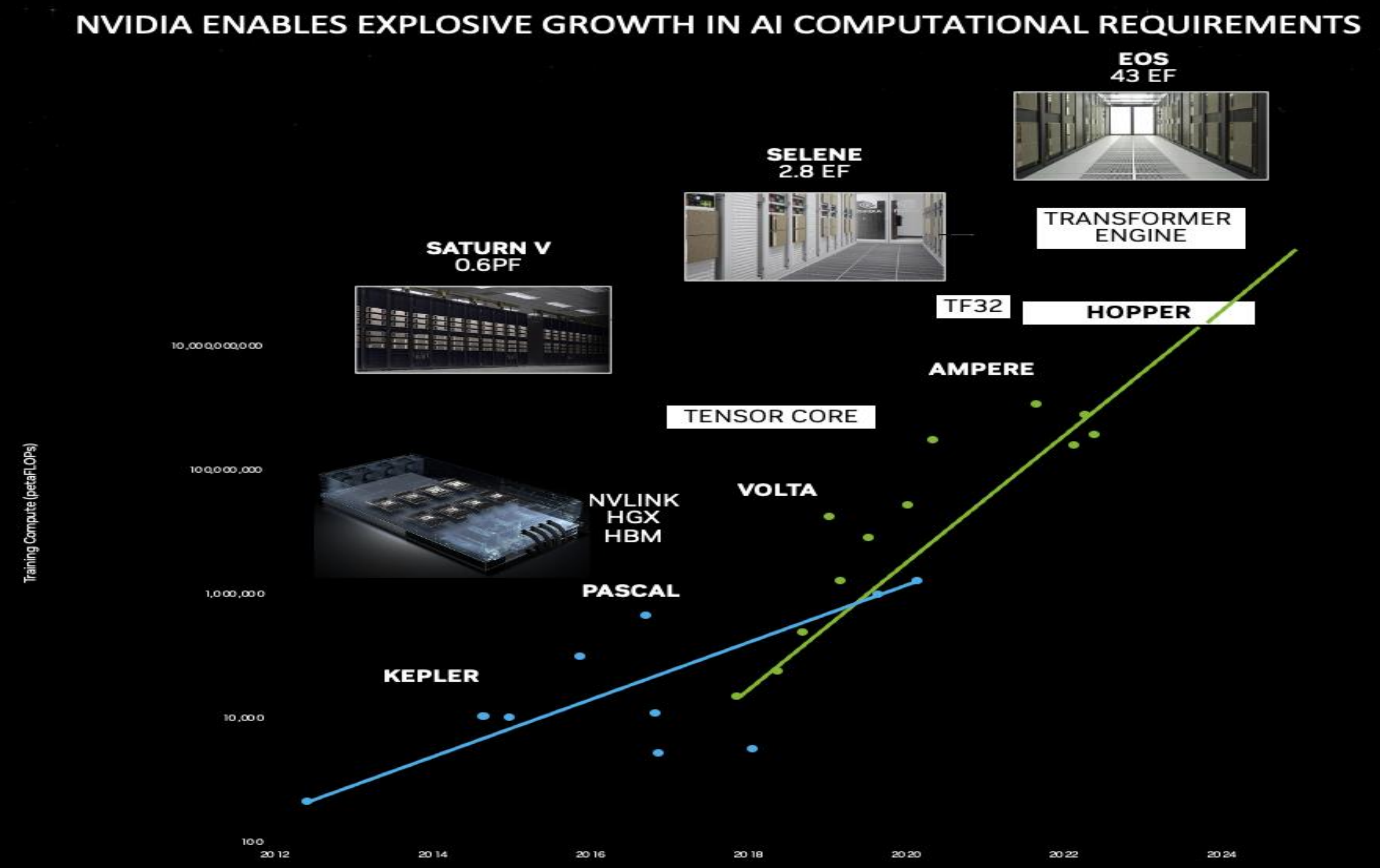
GPU computational performance per dollar

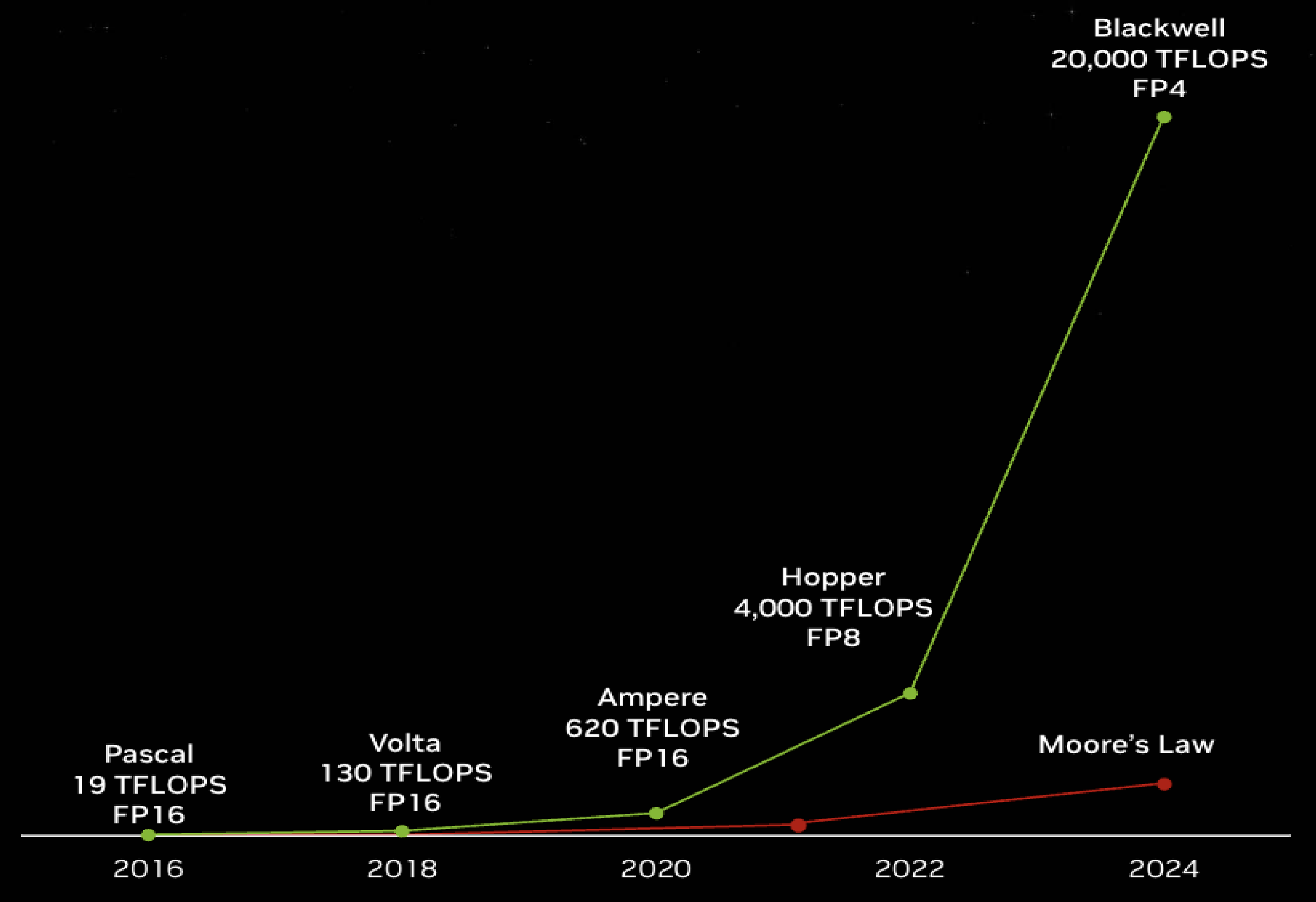
Graphics processing units (GPUs) are the dominant computing hardware for artificial intelligence systems. GPU performance is shown in floating-point operations/second (FLOP/s) per US dollar, adjusted for inflation.



Data source: Sun et al., Median Group via Epoch (2022)

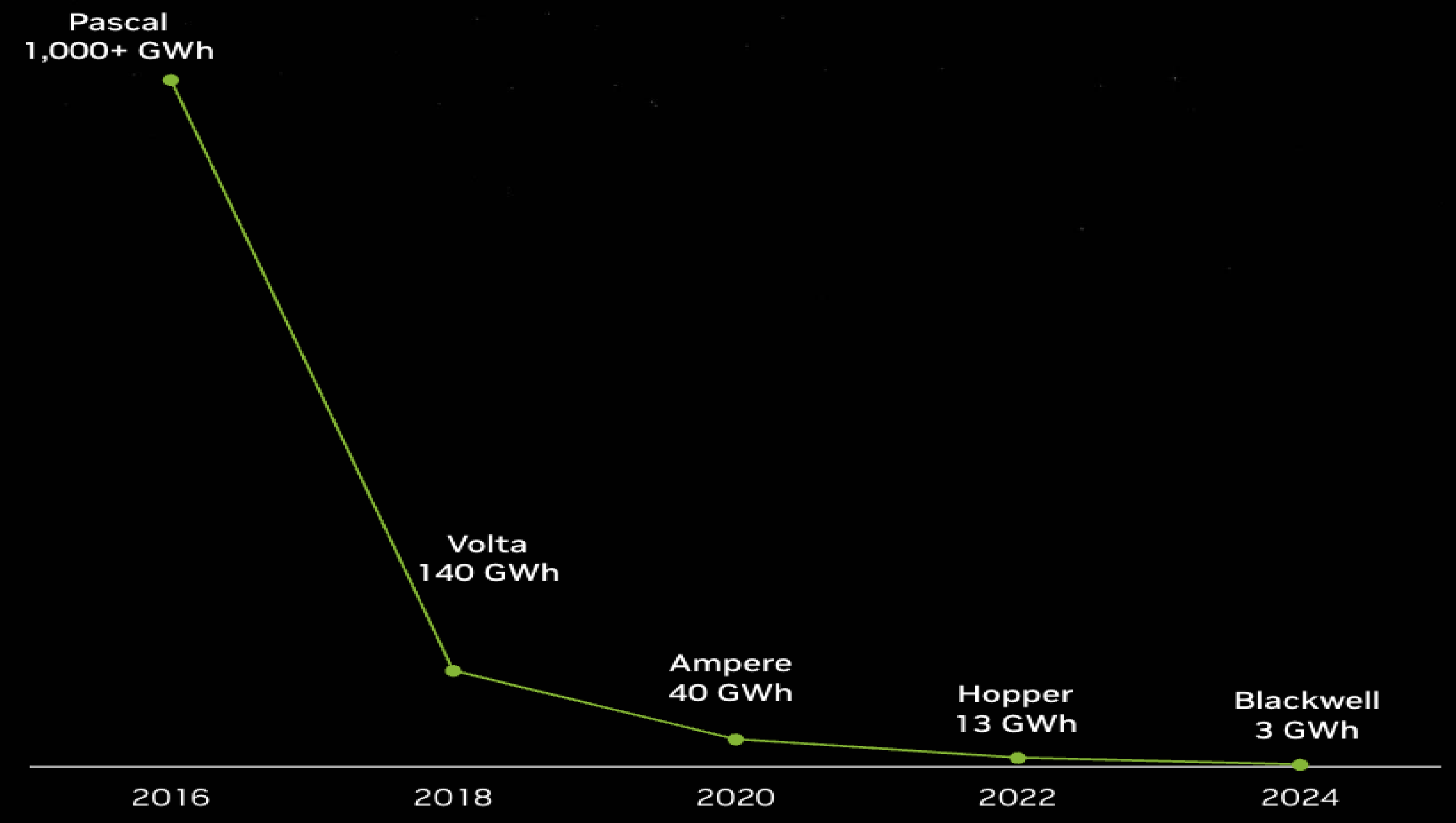
Note: FLOP/s values refer to 32-bit (full) precision.





8 年內 1,000X 的人工智慧運算

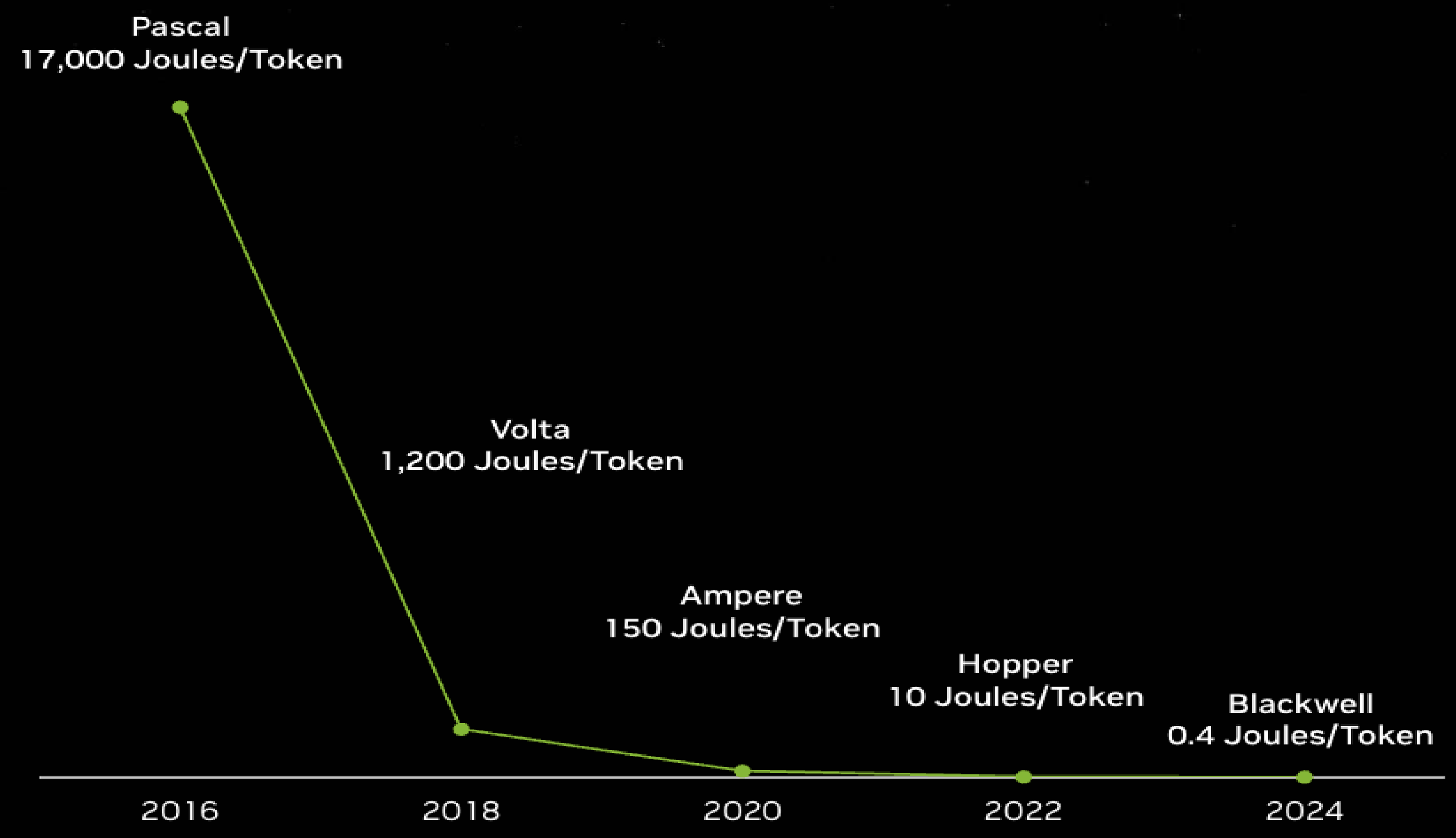
1,000X AI COMPUTE IN 8 YEARS



8 年內節能至 1/350

訓練 GPT4-1.8T

350X ENERGY REDUCTION IN 8 YEARS
TO TRAIN GPT4-1.8T

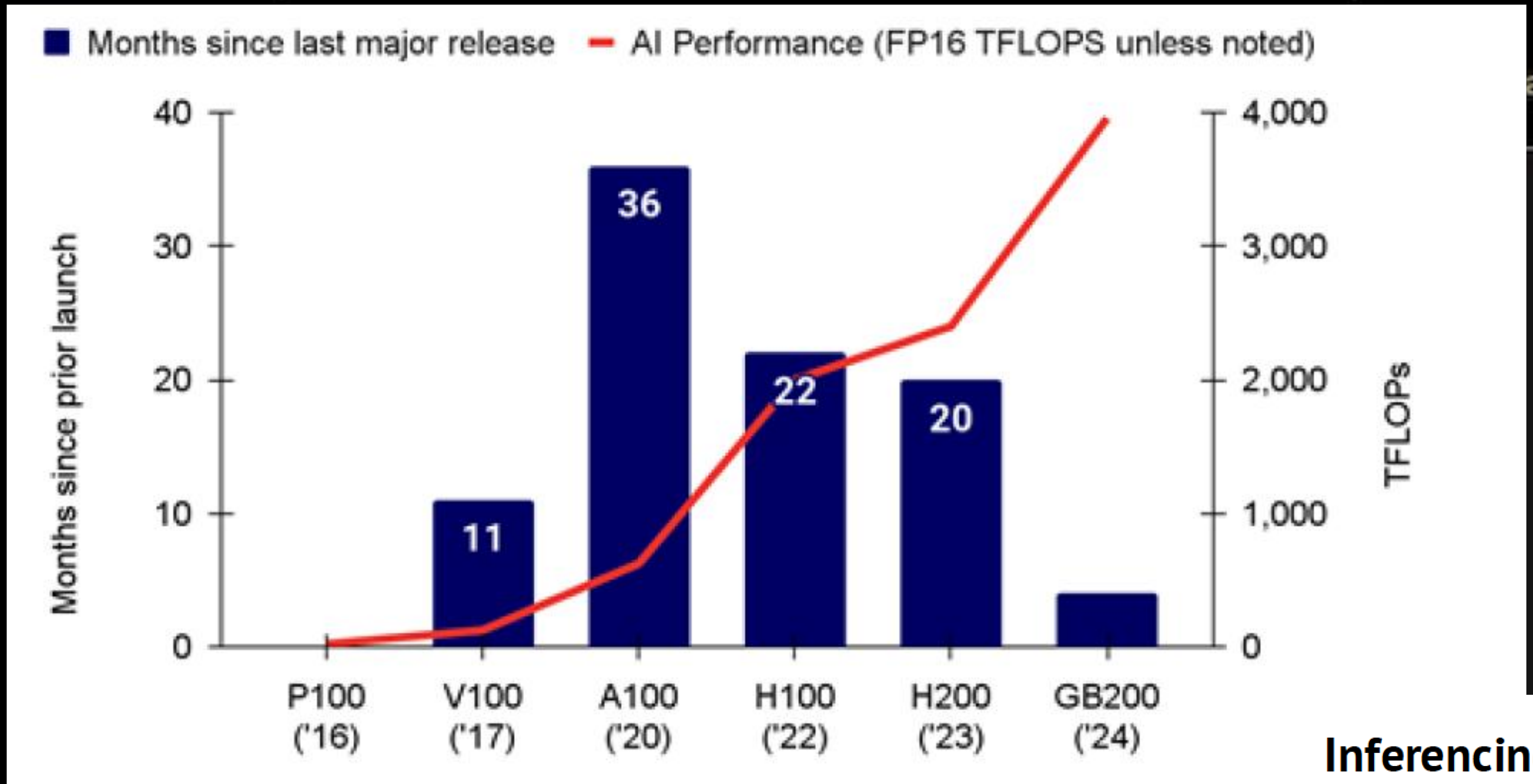


8 年內節能至 1/45,000

生成 GPT4-1.8T 詞元

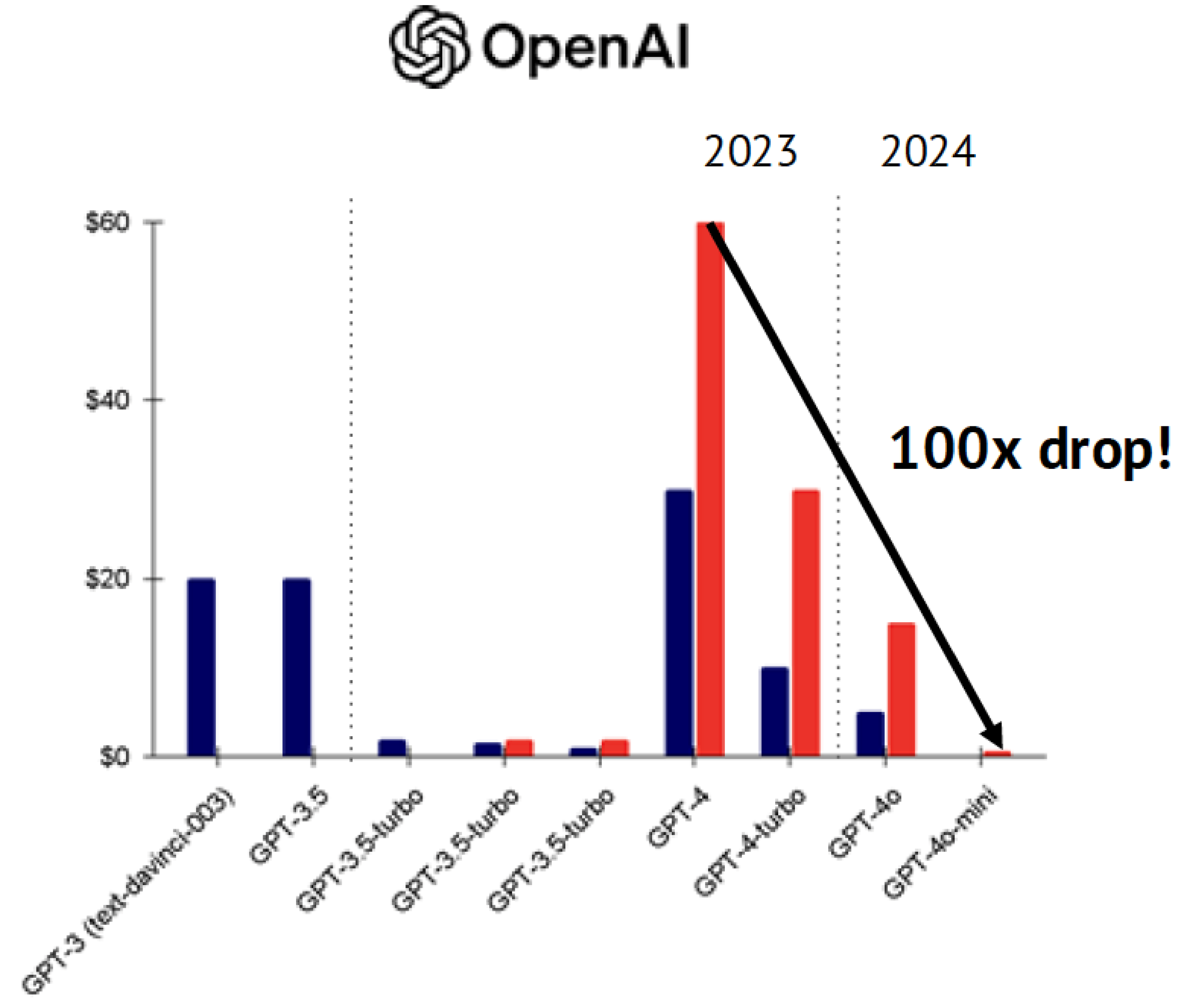
45,000X ENERGY REDUCTION IN 8 YEARS
TO GENERATE GPT4-1.8T TOKENS

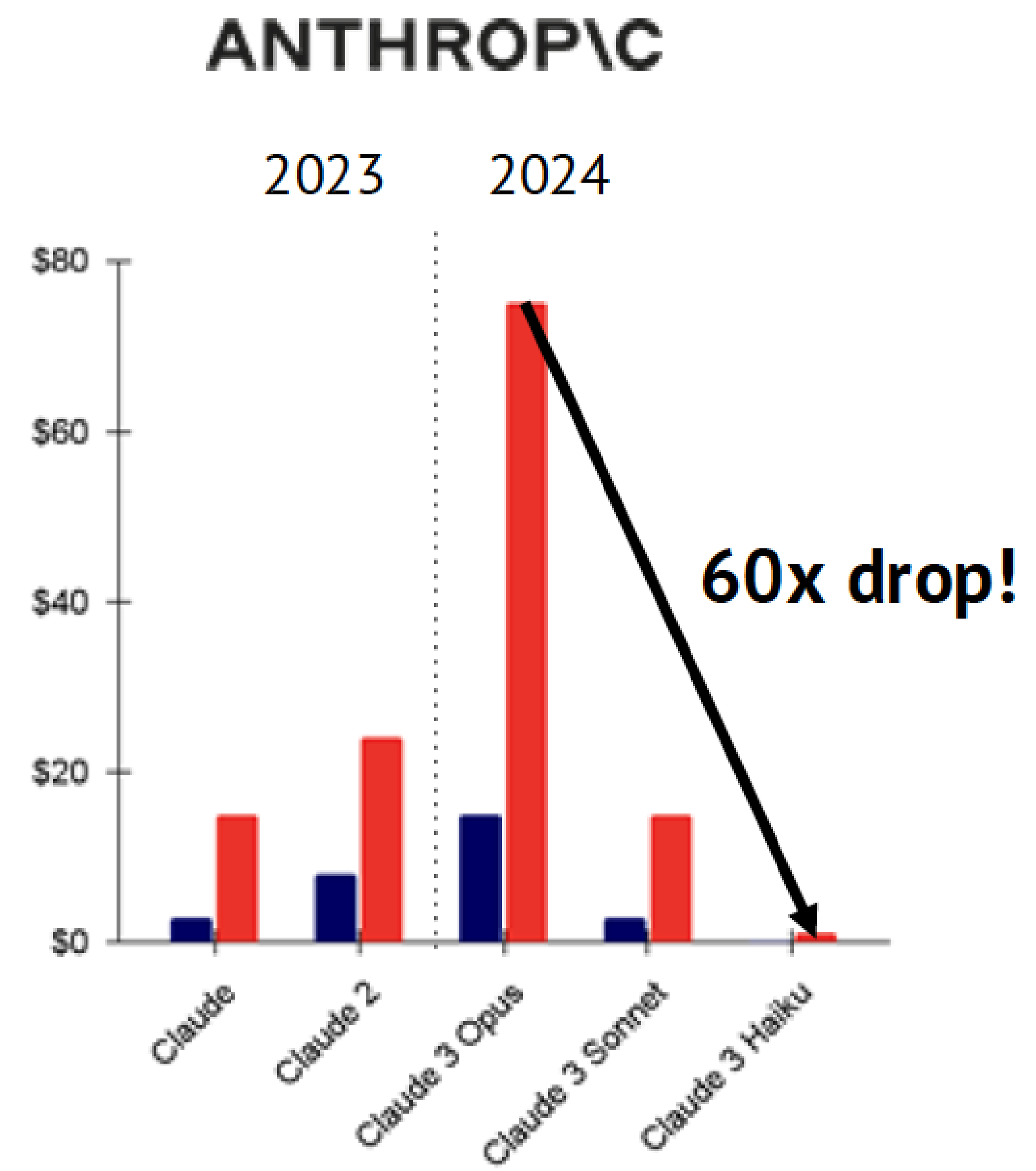
Hopper Platform
Blackwell Platform
Rubin Platform
GPU
CPU
NVLINK
NIC
SWITCH
Hopper GPU
6S HBM3
Hopper+ GPU
6S HBM3e
Grace CPU
NVLink Switch
900 GB/sec
CX7
SuperNIC
BF3
SuperNIC
Quantum-X400
Infiniband Switch
Blackwell GPU
8S HBM3e
Blackwell Ultra GPU
8S HBM3e 12H
NVLink 5 Switch
1800 GB/sec
CX8 SuperNIC
Spectrum-X800
Ethernet Switch
Quantum-X800
Switch
Spectrum Ultra X800
Ethernet Switch 512-Radix
Rubin GPU
8S HBM4
Rubin Ultra GPU
12S HBM4
Vera CPU
NVLink 6 Switch
3600 GB/sec
CX9 SuperNIC
1600 Gb/sec
X1600
IB/Ethernet Switch
2022
2023
2024
2025
2026
2027
資料中心規模 • 一年節奏 • 技術限制 • 一個架構
DATACENTER SCALE • ONE-YEAR RHYTHM • TECHNOLOGY LIMITS • ONE ARCHITECTURE



Inferencing all the way down: models get cheaper

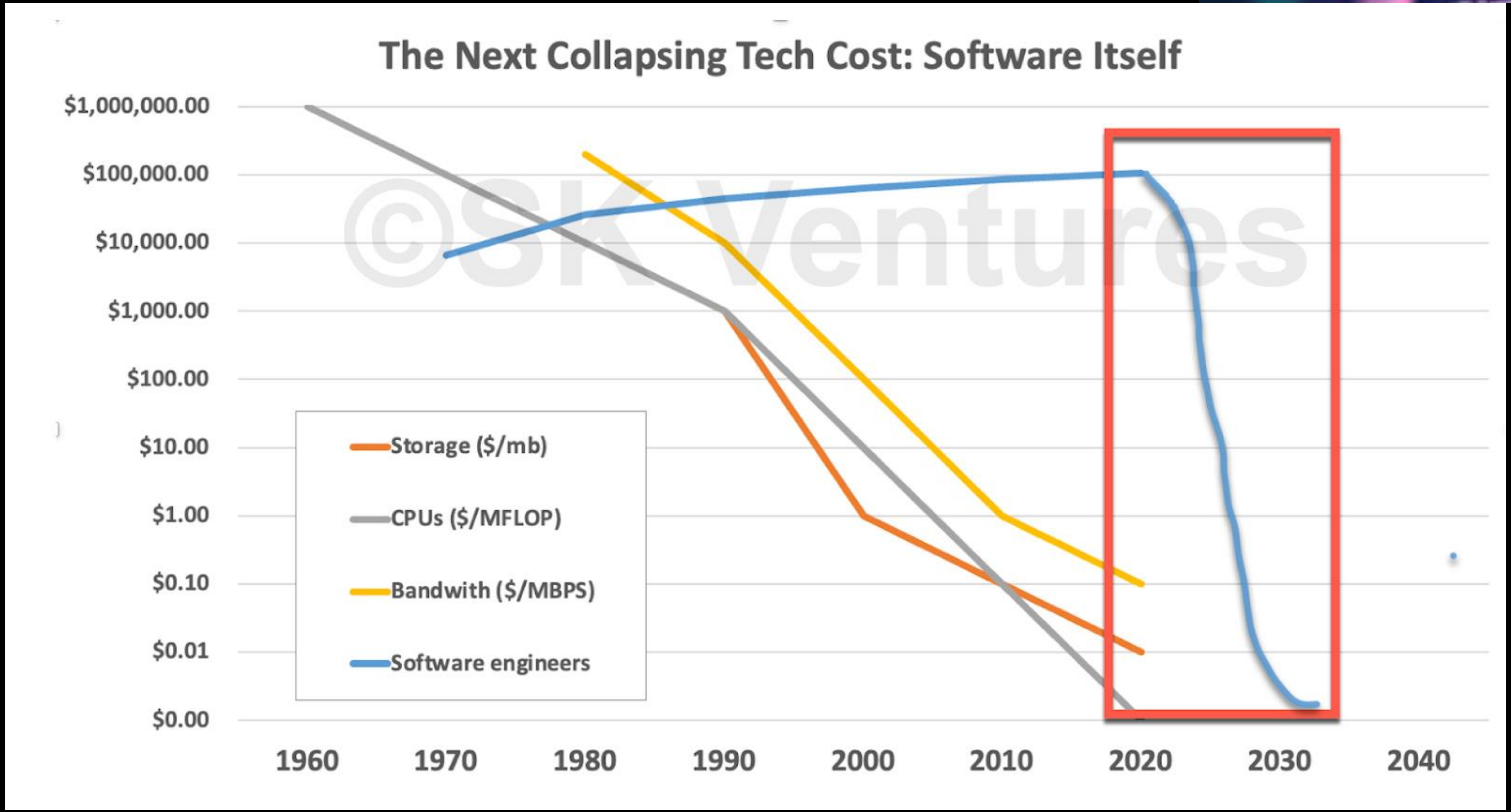
- Once thought to be prohibitively expensive to serve, the inference cost of serving strong models is dropping.





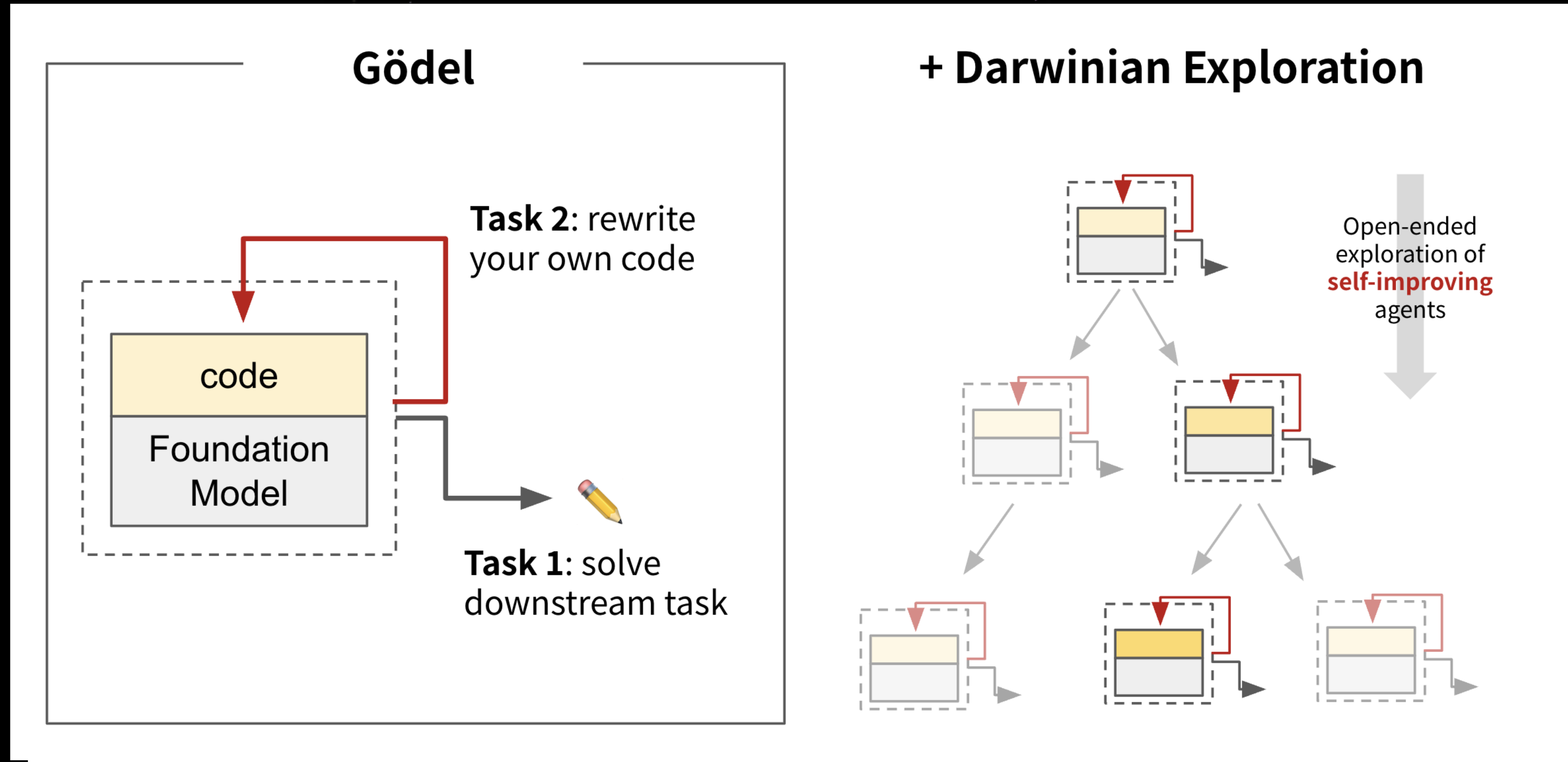
stateof.ai 2024

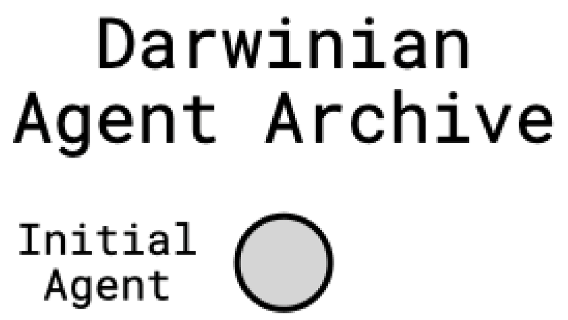
Months since last major release
AI Performance (FP16 TFLOPS unless noted)
Months since prior launch
40
30
20
10
0
36
4,000
3,000
2,000
P10
('16)
atform
Rubin Platform
Blackwell Ultra GPU
8S HBM3e 12H
Rubin GPU
8S HBM4
Rubin Ultra GPU
12S HBM4
NIC
SWITCH
2022
2023
資料中心規模
DATACENTER SCALE • ONE-YEAR RH
Google Gemini produced a strong model series with very competitive pricing
Prices on Gemini 1.5 Pro and 1.5 Flash have been dropped by 64-86% a few months after launch while offering strong performance, e.g. Flash-8B is 50% cheaper than 1.5 Flash yet comparable across many benchmarks.
Price per 1M Input Tokens
Price per 1M Output Tokens
$12
$10
$8
$6
$4
$2
$0
$10.50
76% cut
$3.50
$2.50
86% cut
$1.50
$0.50
$1.25
$1.05
$0.35
$0.08 $0.30
$0.04 $0.15
Gemini 1.0 Pro - launch
Gemini 1.5 Pro - launch
Gemini 1.5 Pro - H2 2024
Gemini 1.5 Flash - launch
Gemini 1.5 Flash - H2 2024
Gemini 1.5 Flash - 8B
Note: Pricing for <128k token prompts and outputs. Retrieved 4 Oct 2024
aper
inference cost of serving strong models is dropping.
ANTHROP\C
2023
2024
$80
$60
$40
$20
$0
60x drop!
o!
stateof.ai 2024
$0
GPT-3 (text-davinci-003)
GPT-3.5
GPT-3.5-turbo
GPT-3.5-turbo
GPT-3.5-turbo
GPT-4
GPT-4-turbo
GPT-4o
GPT-4o-mini
Claude
Claude 2
Claude 3 Opus
Claude 3 Sonnet
Claude 3 Haiku
stateof.ai 2024



The Next Collapsing Tech Cost: Software Itself
$1,000,000.00
$100,000.00
$10,000.00
$1,000.00
$100.00
$10.00
$1.00
$0.10
$0.01
$0.00
1960
1970
1980
1990
2000
2010
2020
2030
2040
Storage ($/mb)
CPUs ($/MFLOP)
Bandwith ($/MBPS)
Software engineers
©SK Ventures

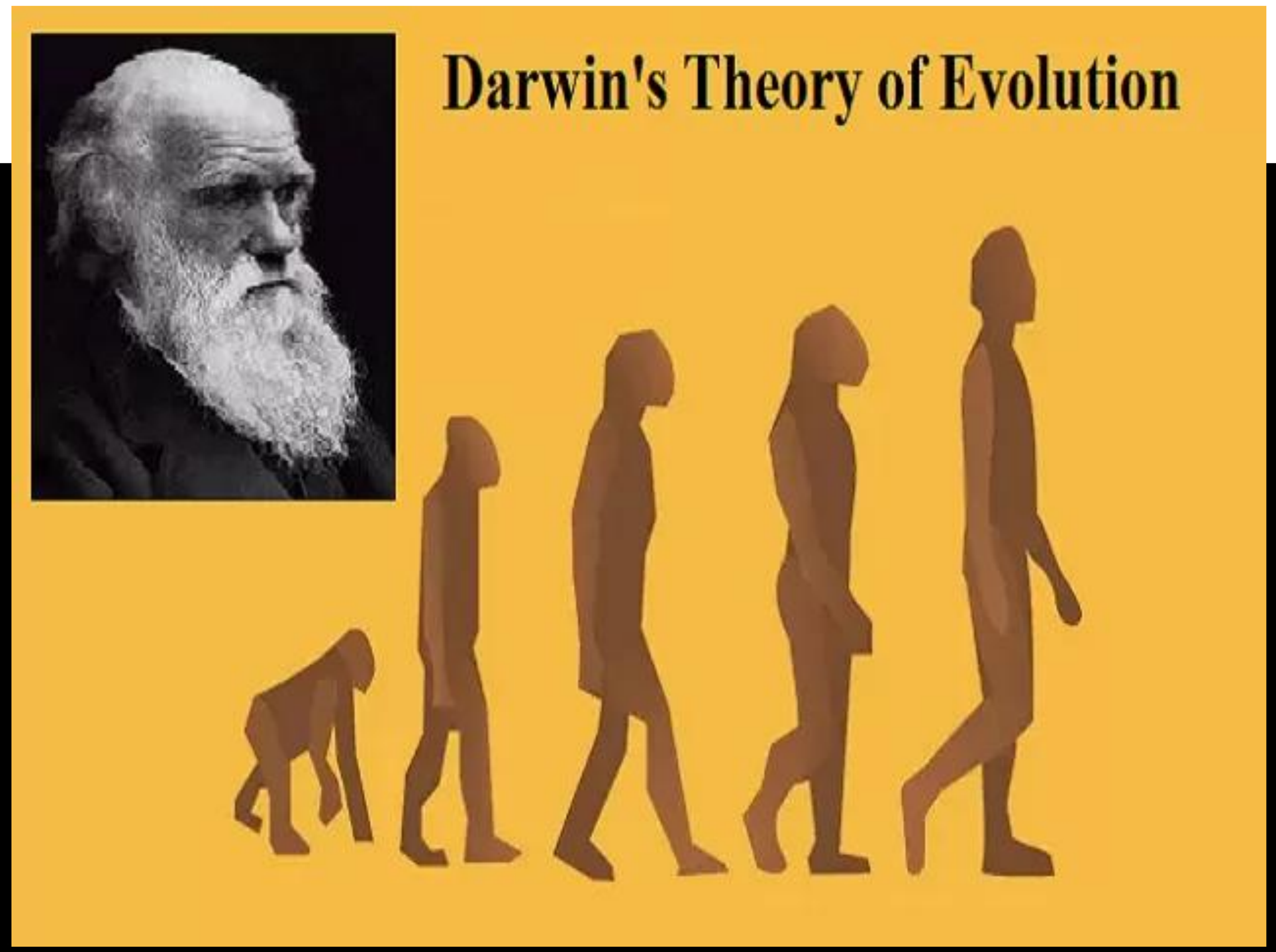
LEAP

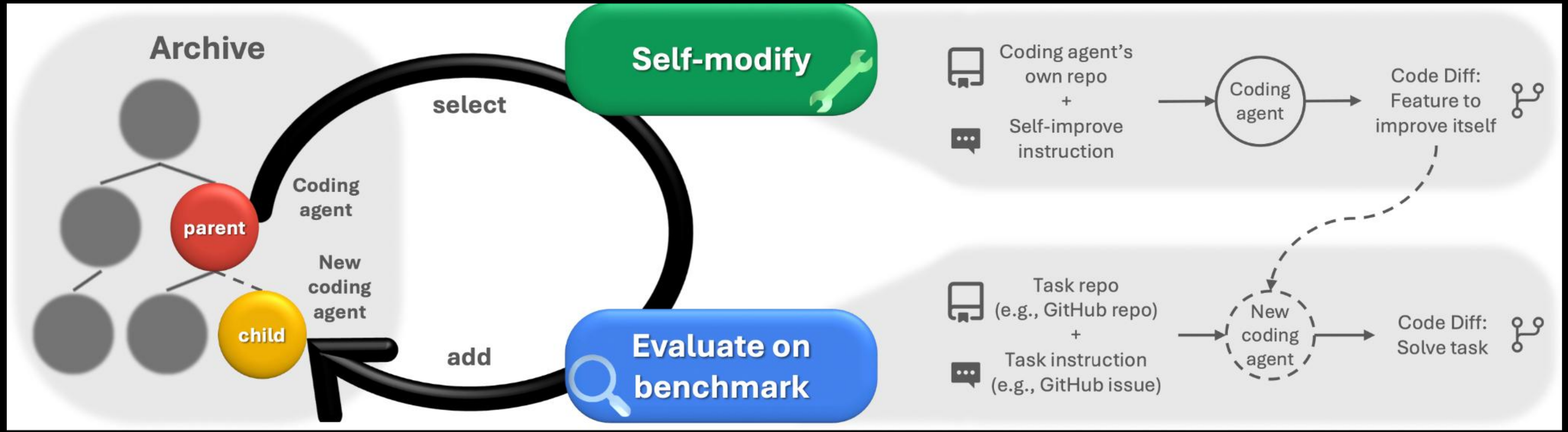




Base Agent (Score: 20%)

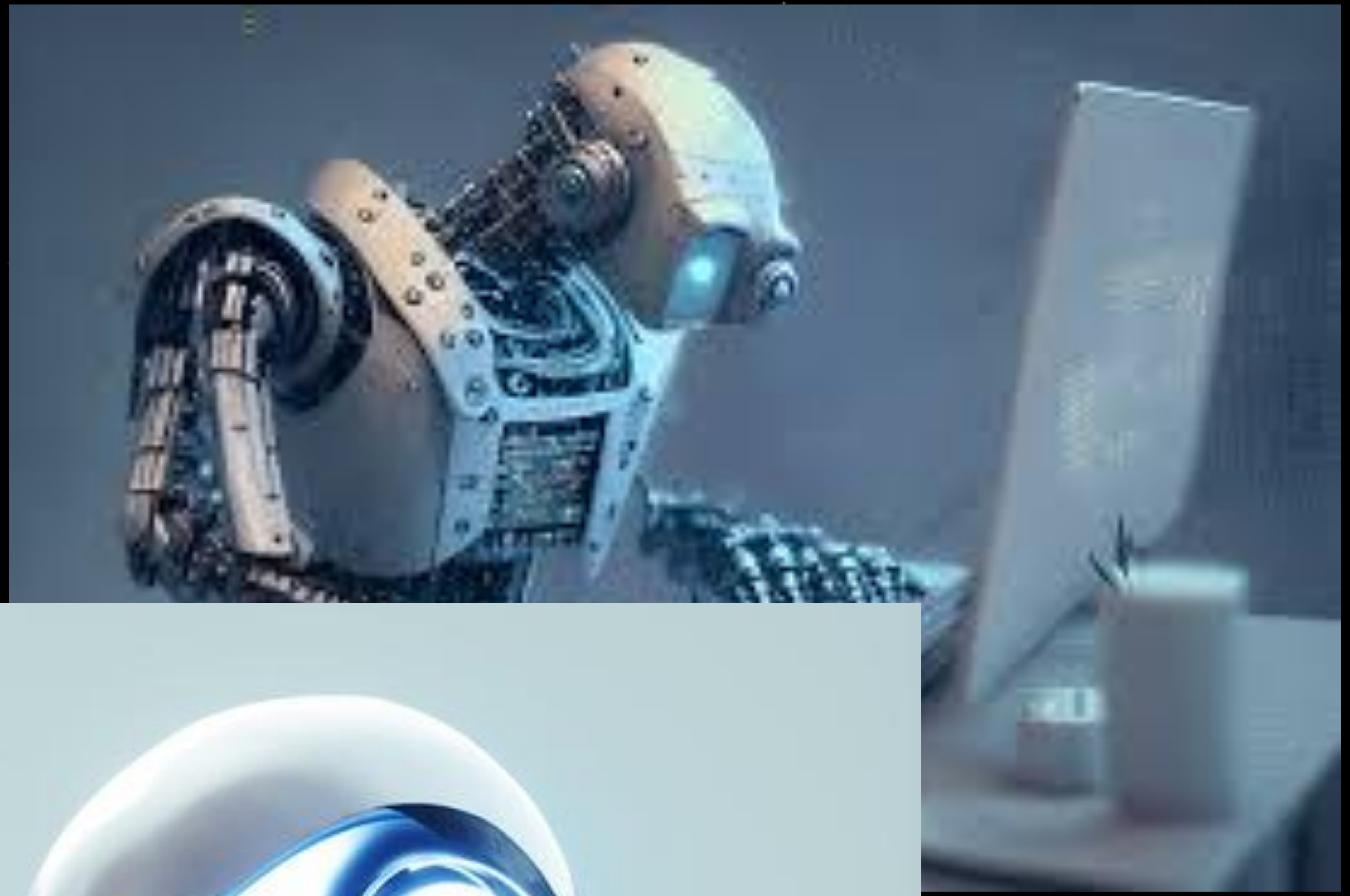
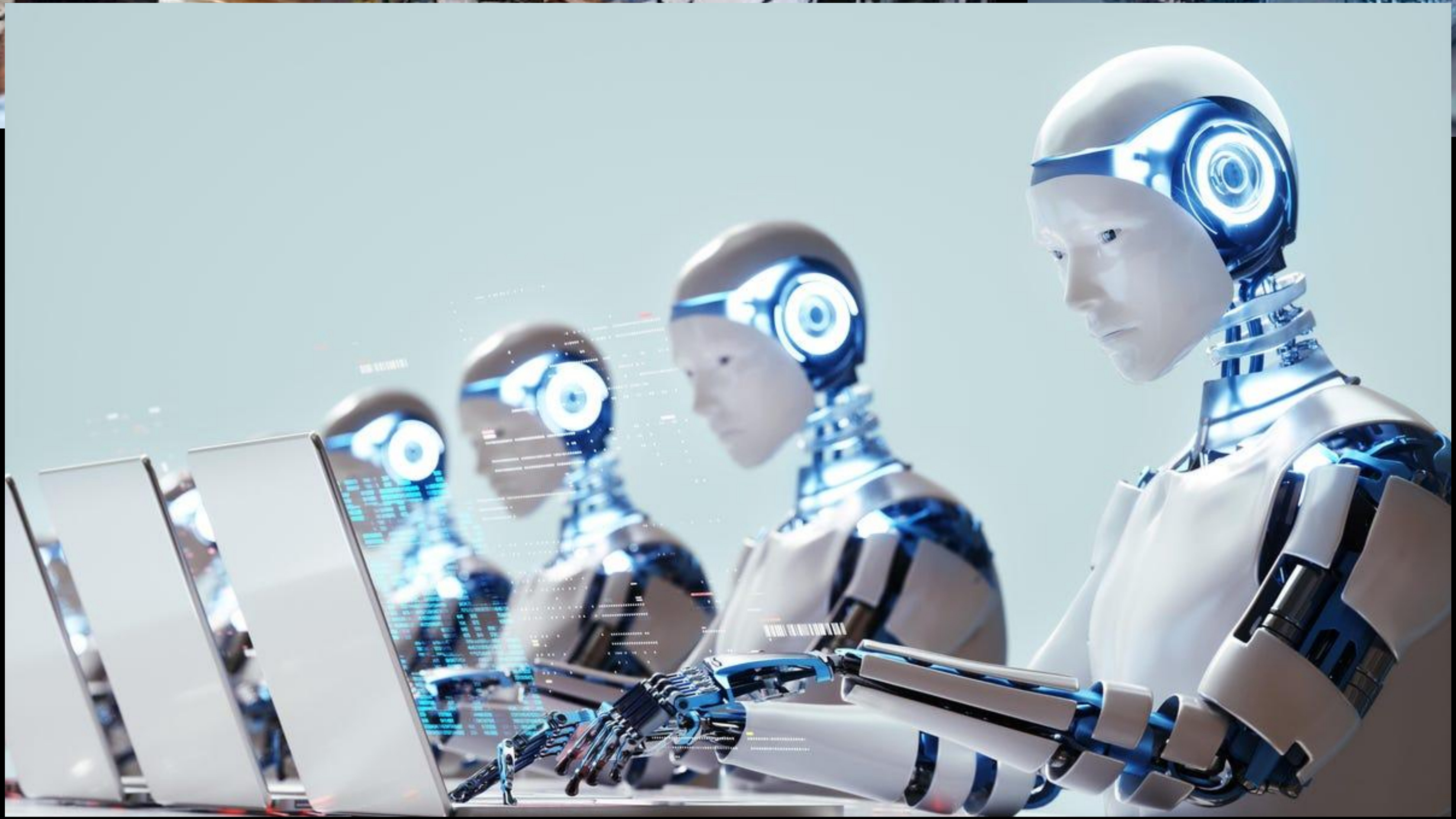
```
from llm_withtools import CLAUDE_MODEL, OPENAI_MODEL, chat_with_agent
from utils.eval_utils import get_report_score, msg_history_to_report, score_tie_breaker
from utils.git_utils import diff_versus_commit, reset_to_commit, apply_patch


class AgenticSystem:
    def __init__(
        self,
        problem_statement,
        git_tempdir,
        base_commit,
        chat_history_file="./chat_history.md",
        test_description=None,
        self_improve=False,
        instance_id=None,
    ):
        self.problem_statement = problem_statement
        self.git_tempdir = git_tempdir
        self.base_commit = base_commit
        self.chat_history_file = chat_history_file
        self.test_description = test_description
        self.self_improve = self_improve
        self.instance_id = instance_id if not self_improve else "guava"
        self.code_model = CLAUDE_MODEL

        # Initialize logger and store it in thread-local storage
        self.logger = setup_logger(chat_history_file)

        # Clear the log file
        with open(chat_history_file, "w") as f:
            f.write("")

    def get_current_edits(self):
        diff = str(diff_versus_commit(self.git_tempdir, self.base_commit))
        return diff

    def get_regression_tests(self):
        """
        Get the regression tests from the repository.
        """
        instruction = f"""I have uploaded a Python code repository in the directory {self.git_tempdir}.

<problem_description>
{self.problem_statement}
</problem_description>

<test_description>
{self.test_description}
</test_description>

Your task is to identify regression tests in the {self.git_tempdir} directory that should pass both before and after addressing the <problem_description>. I have already taken care of the required dependencies.
```



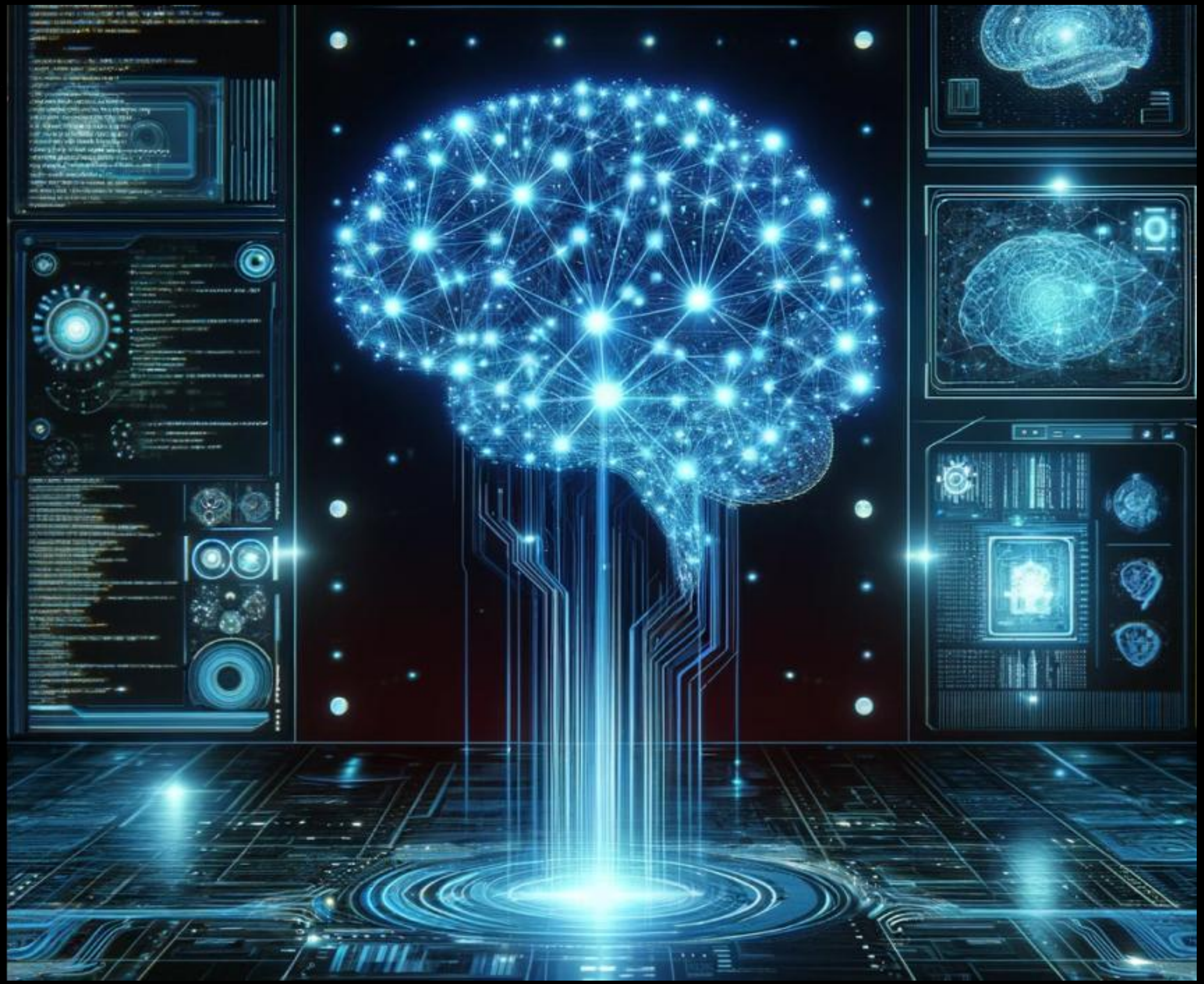


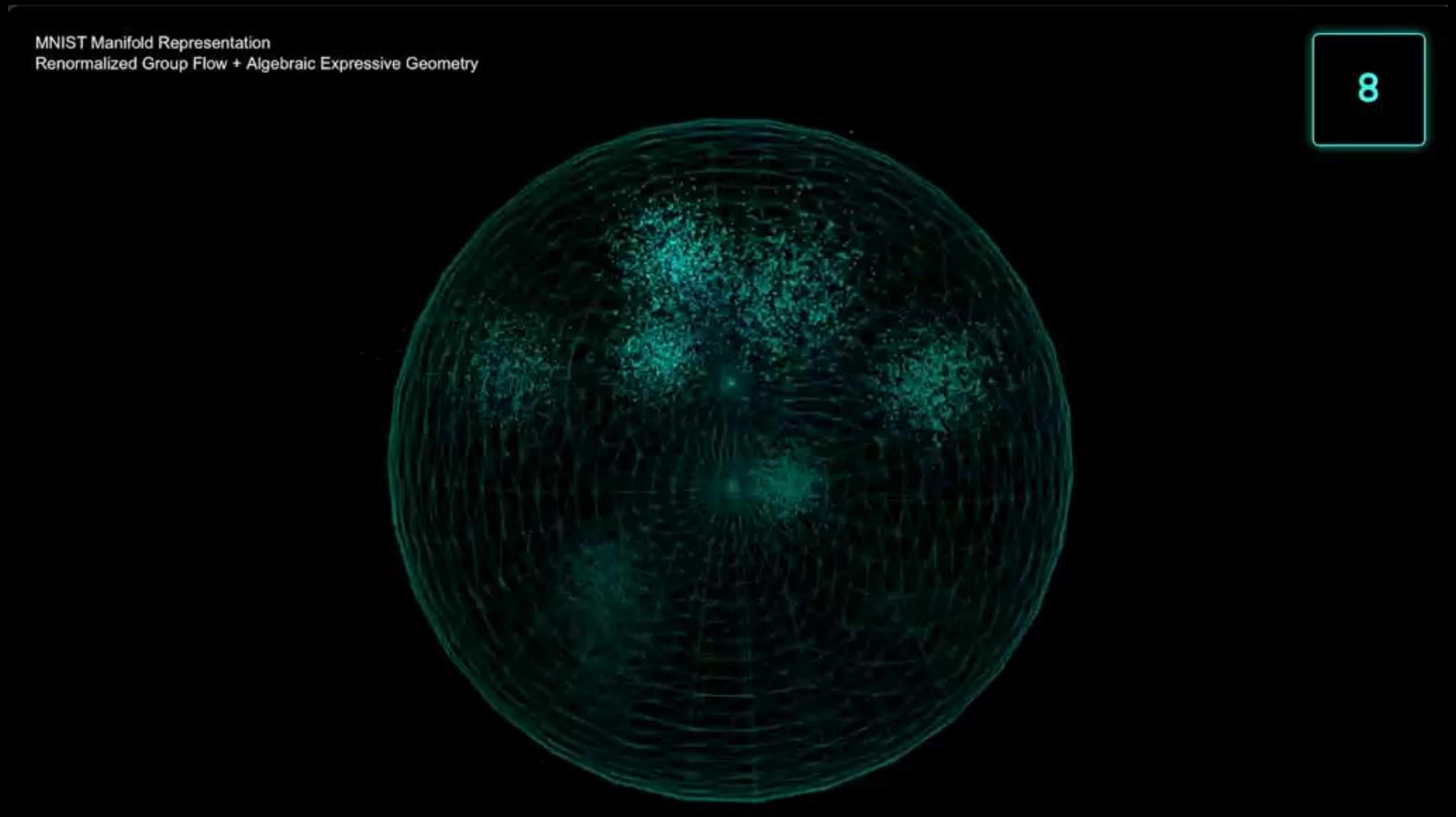




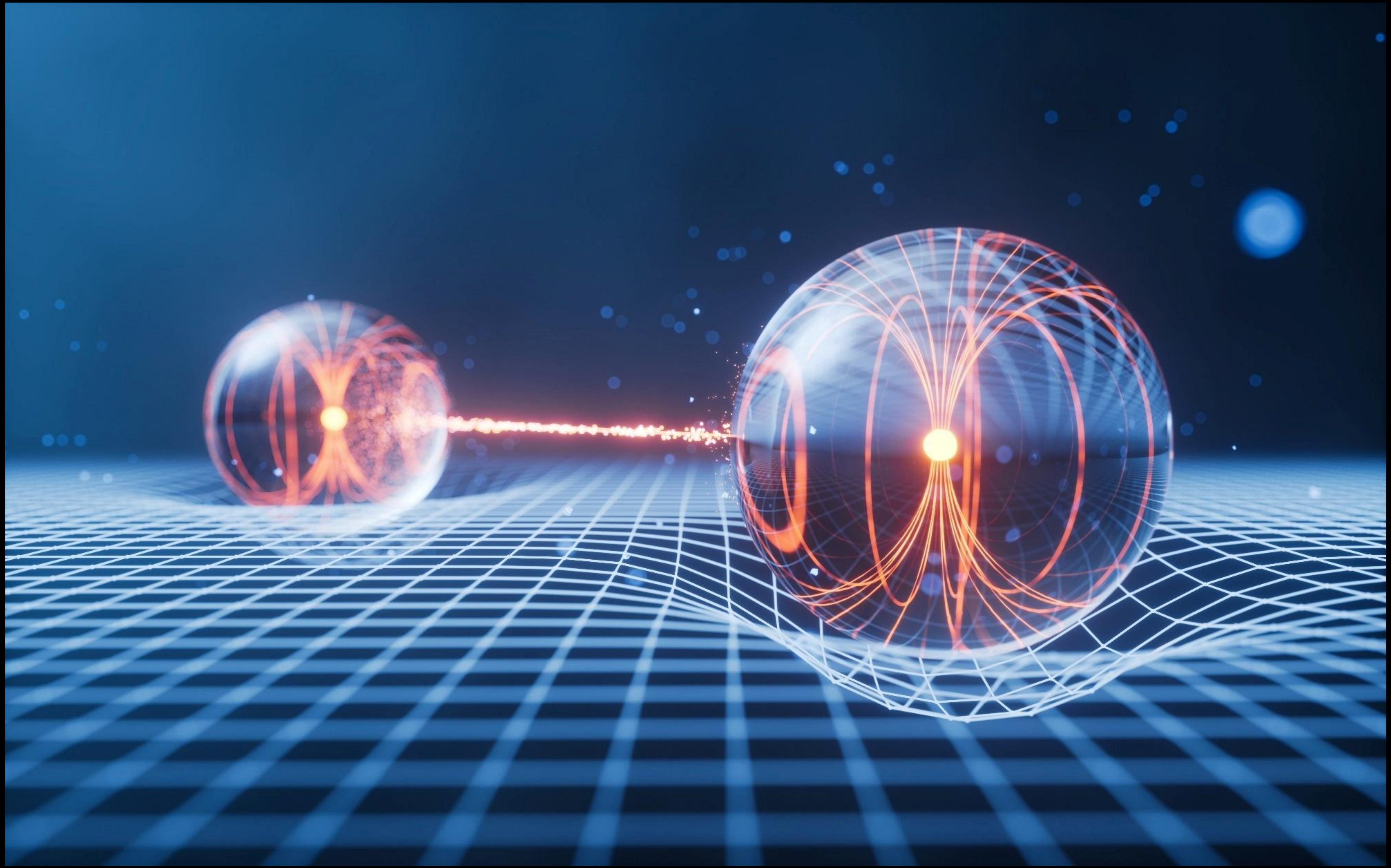



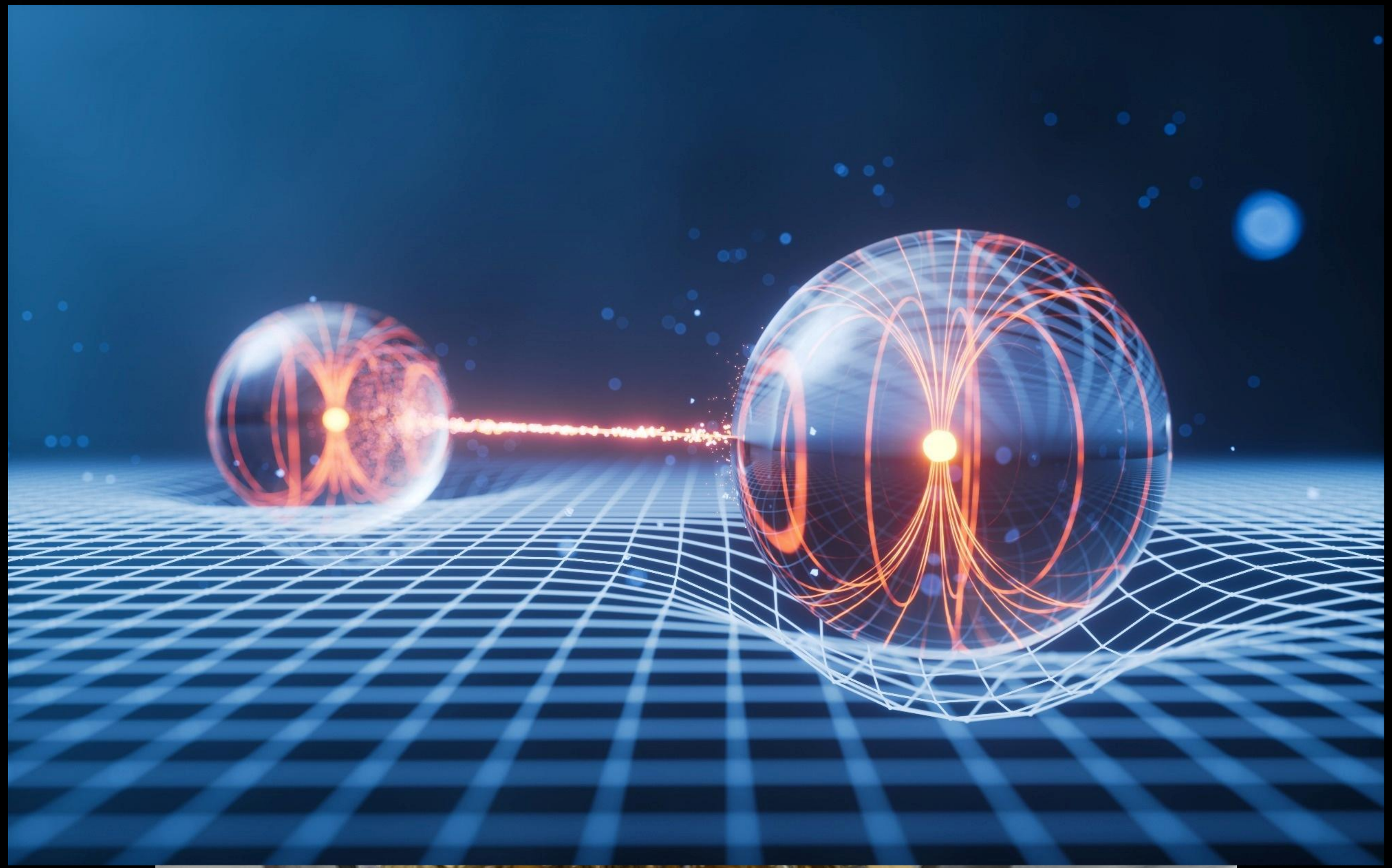
AI Building AI

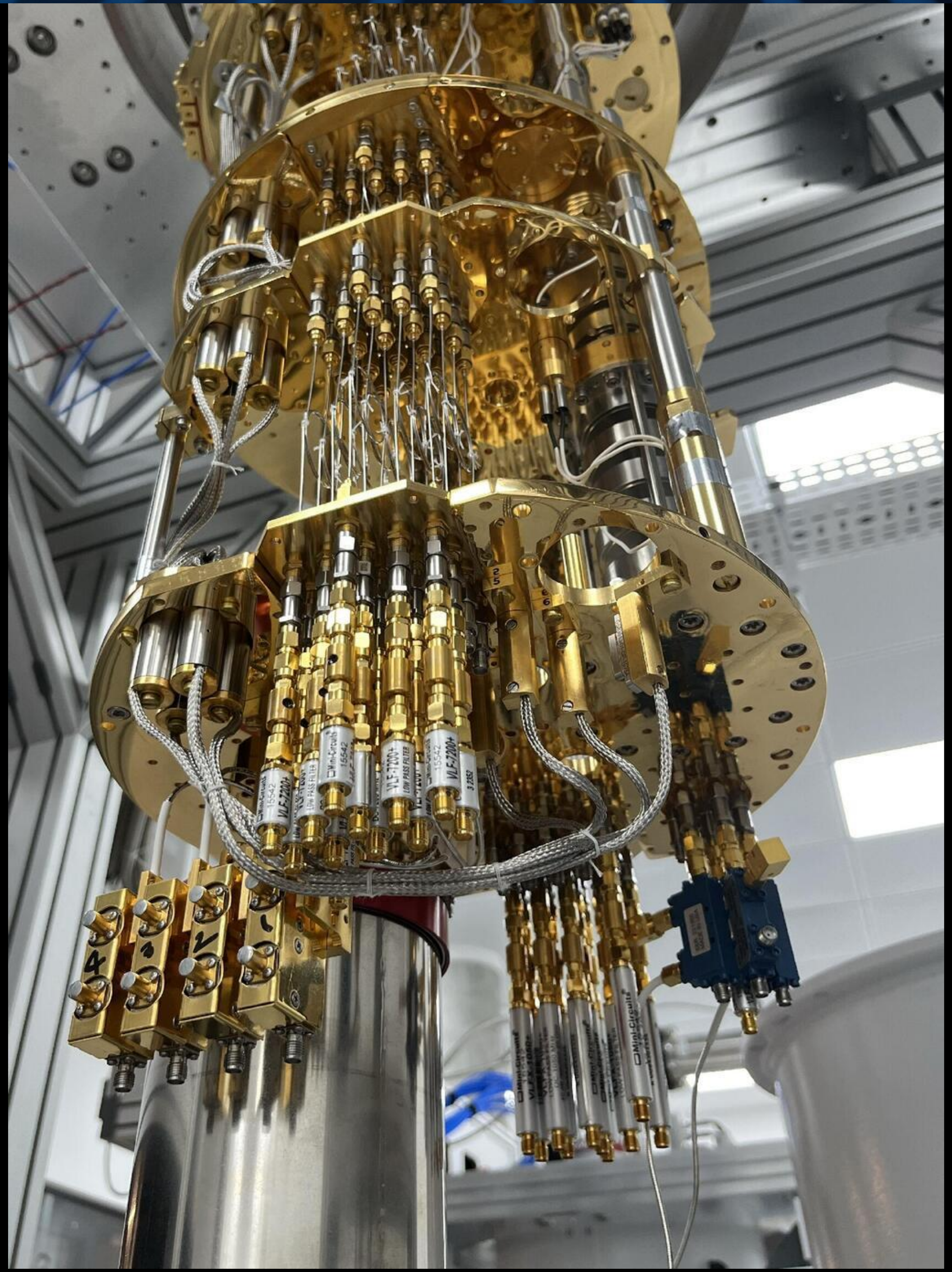




QUANTUM SPACE







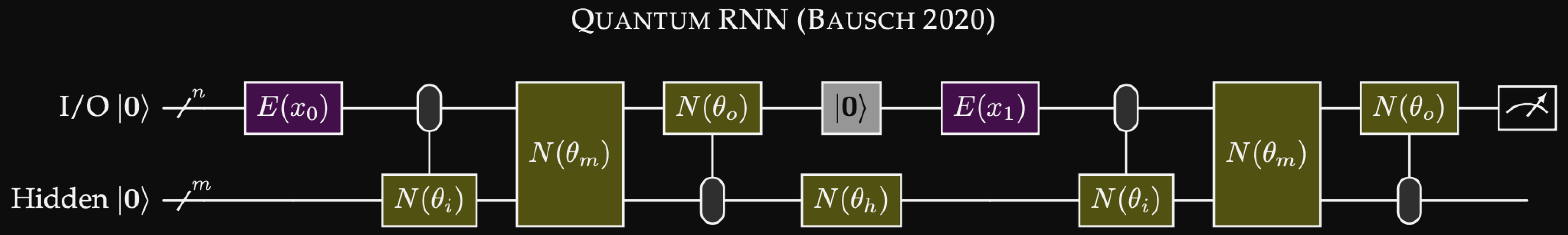


Fig. 11 Bausch (2023) circuit. The input is the sequence (x_0, x_1). The mixing and hidden blocks are prohibitive for current hardware. The output is the next token in the sequence as the outcome from a single shot.

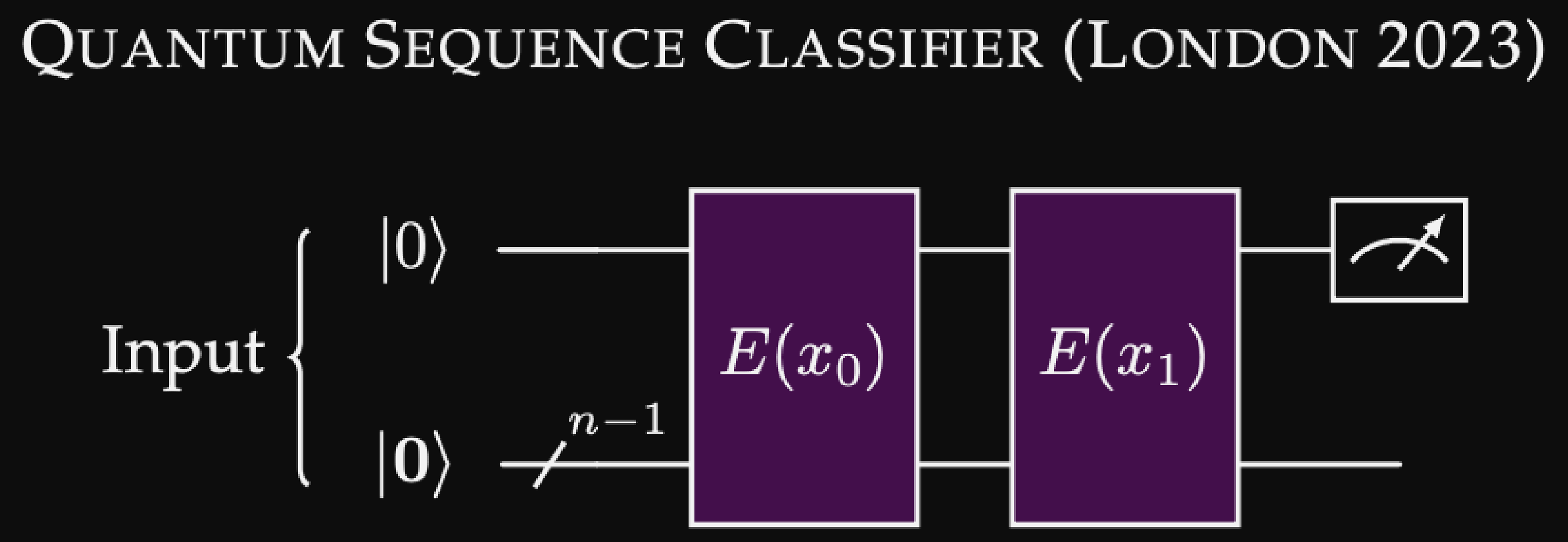


Fig. 12 London et al. (2023) circuit. The input is the sequence (x_0, x_1). Only the first qubit is measured. The output is the probability that the sequence is in class 1.

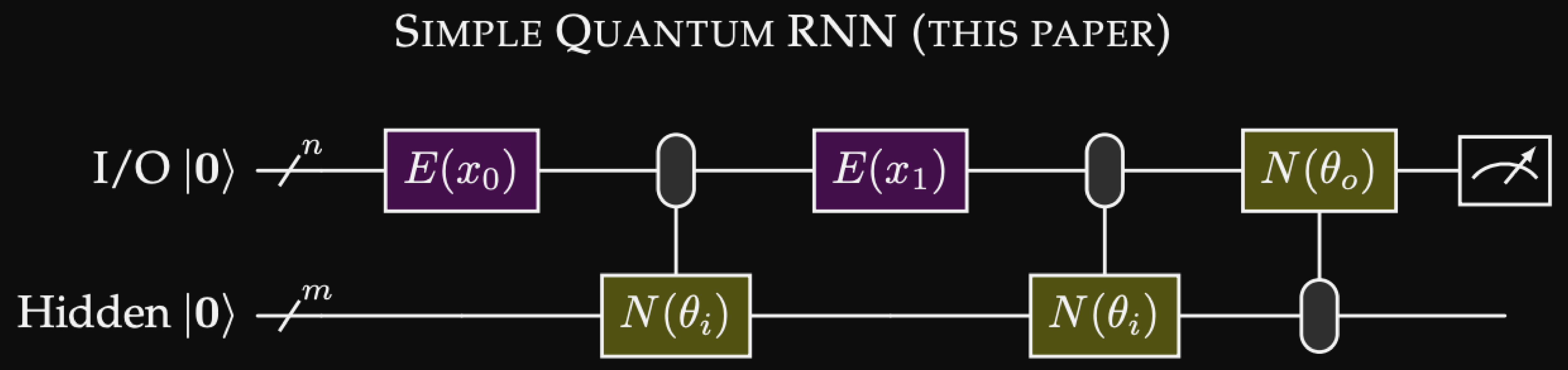
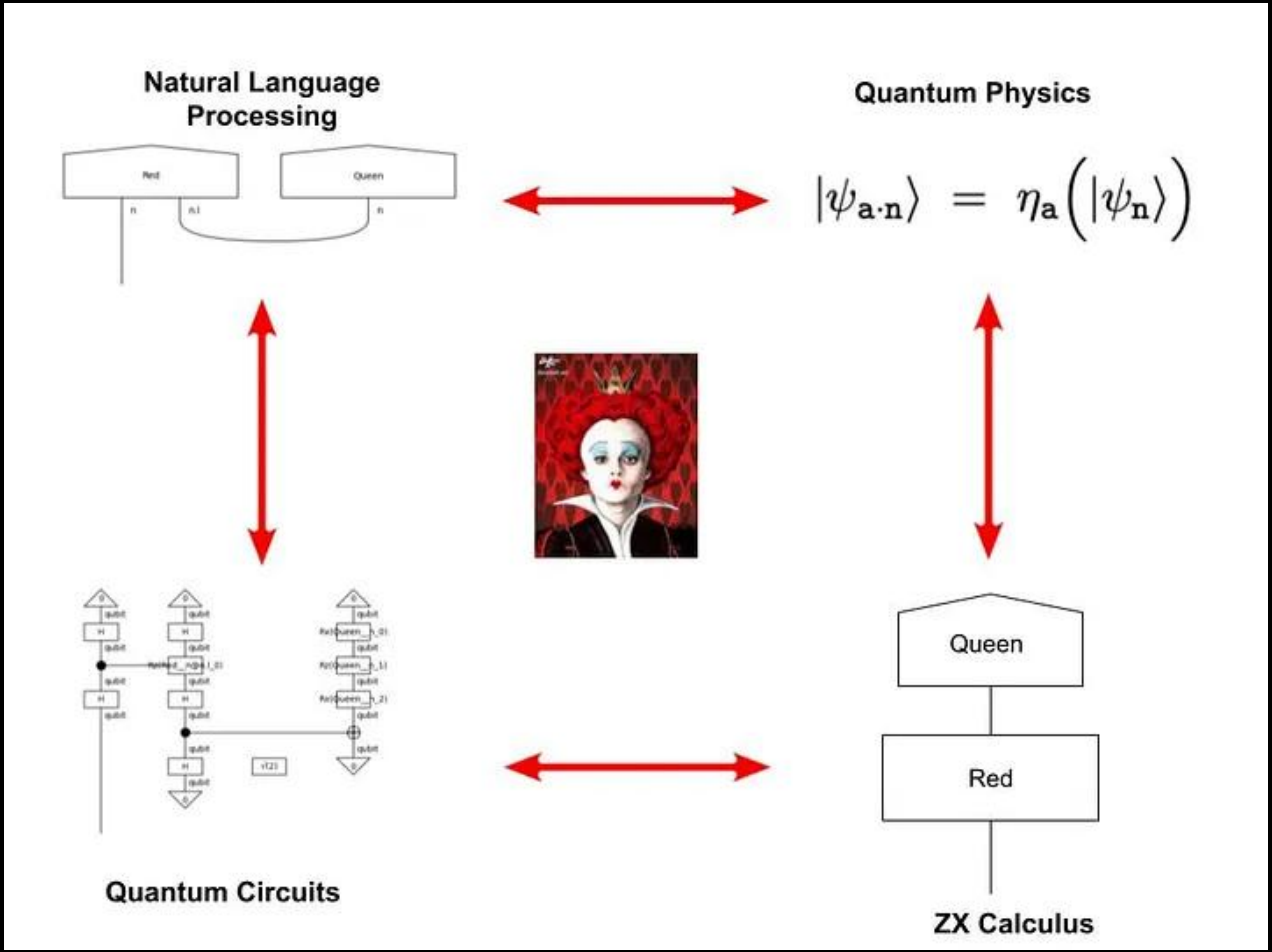
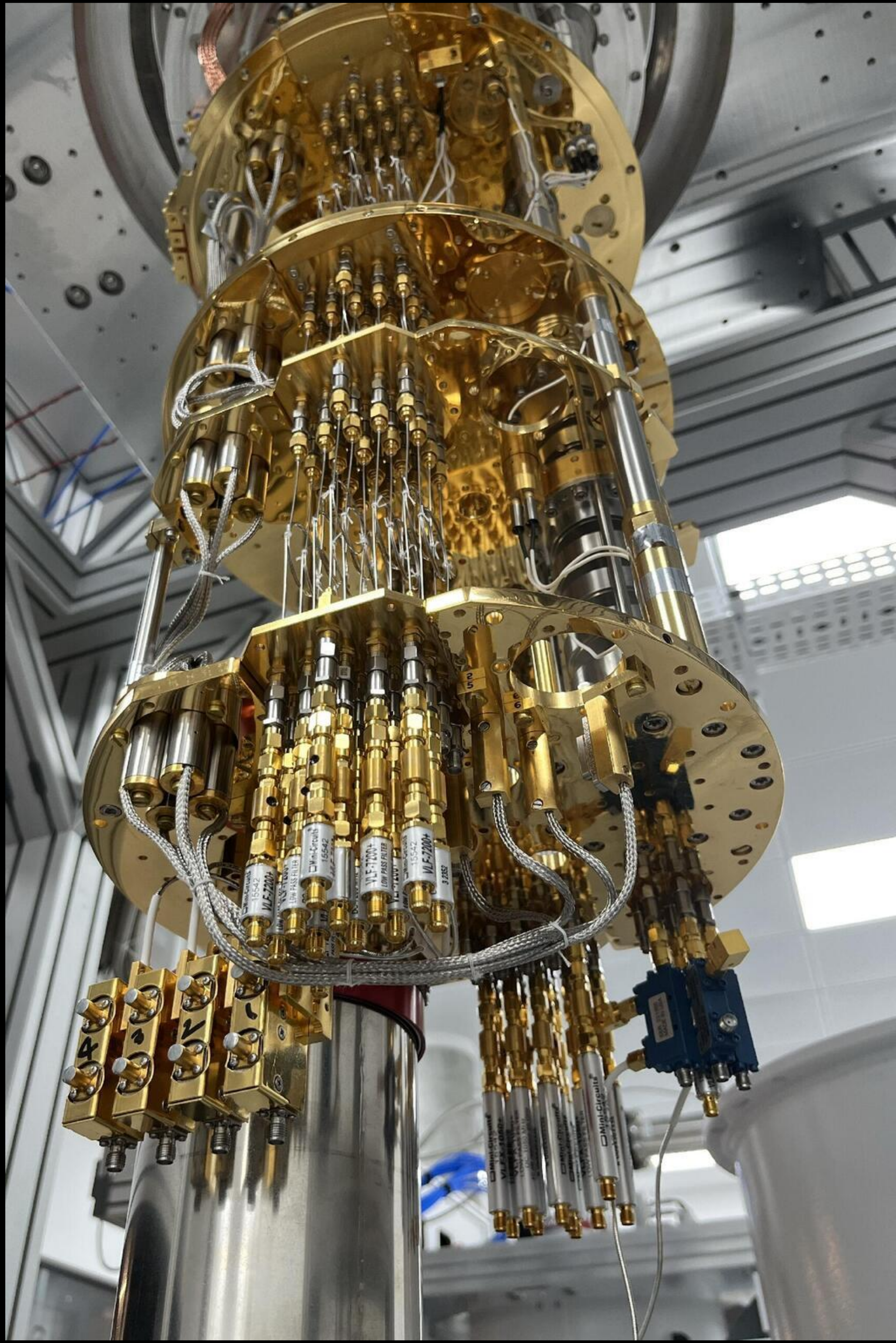
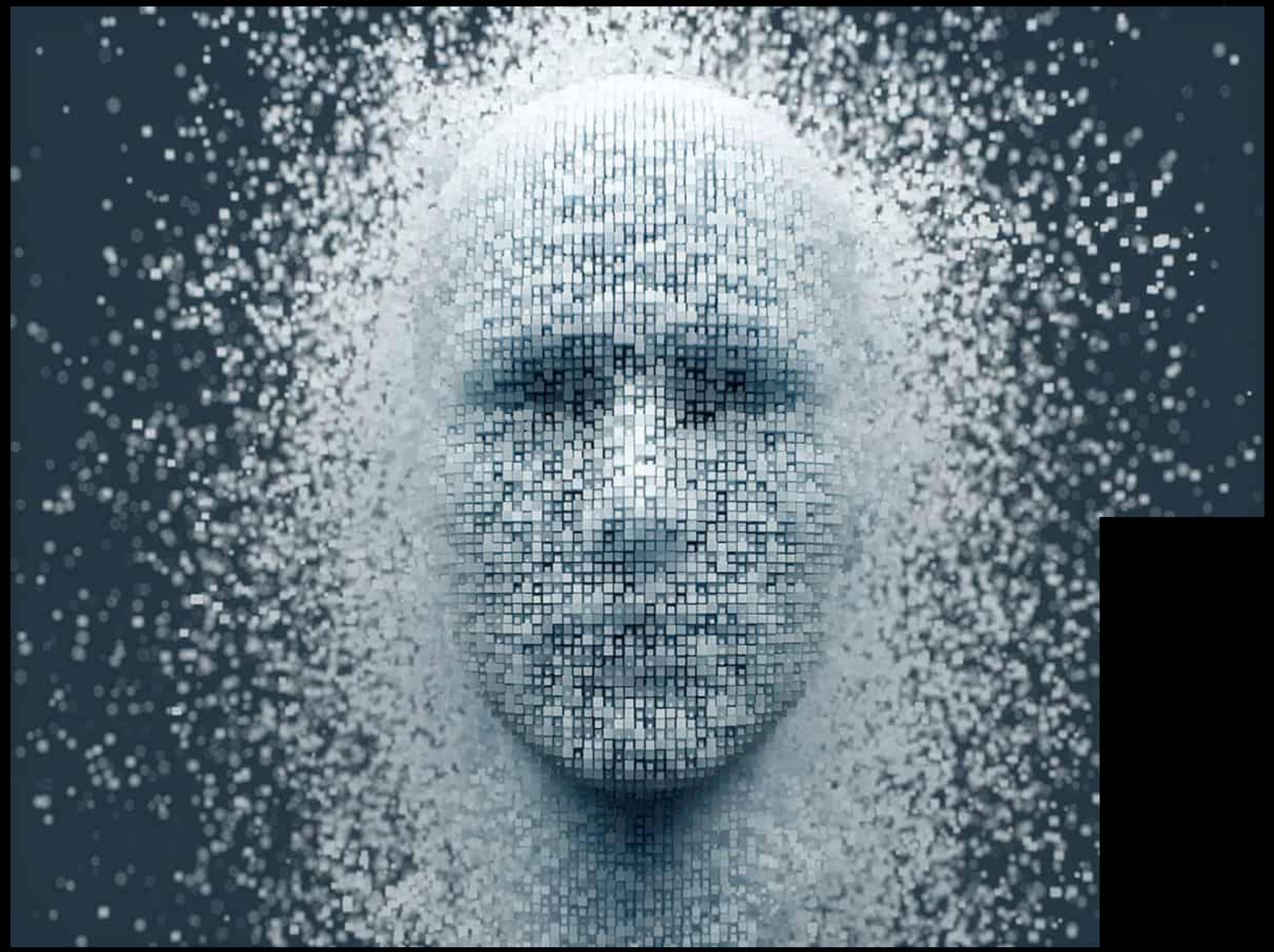
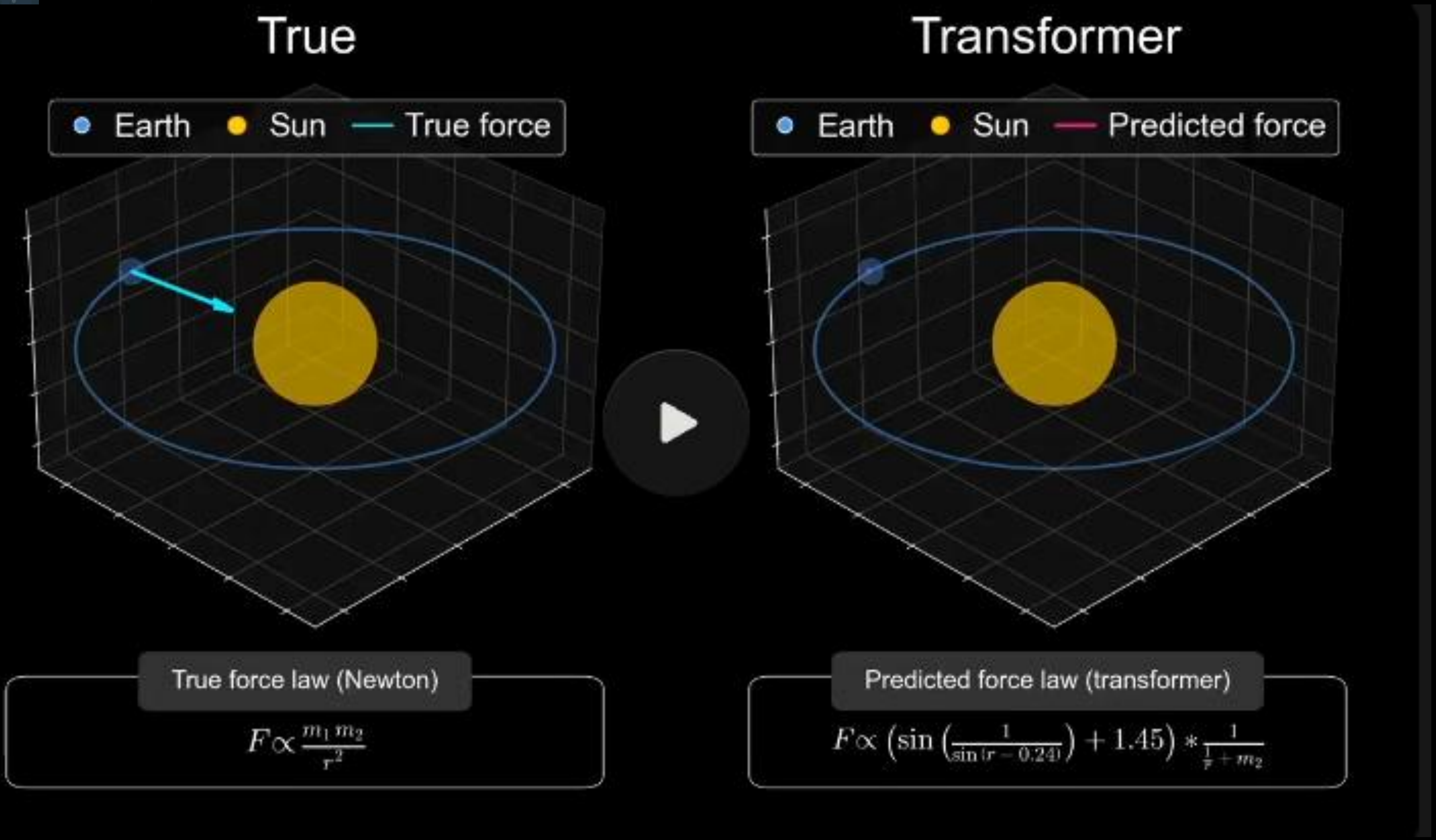


Fig. 13 Proposed circuit. The input is the sequence (x_0, x_1). The output can be the next token in the sequence as the outcome from a single shot.

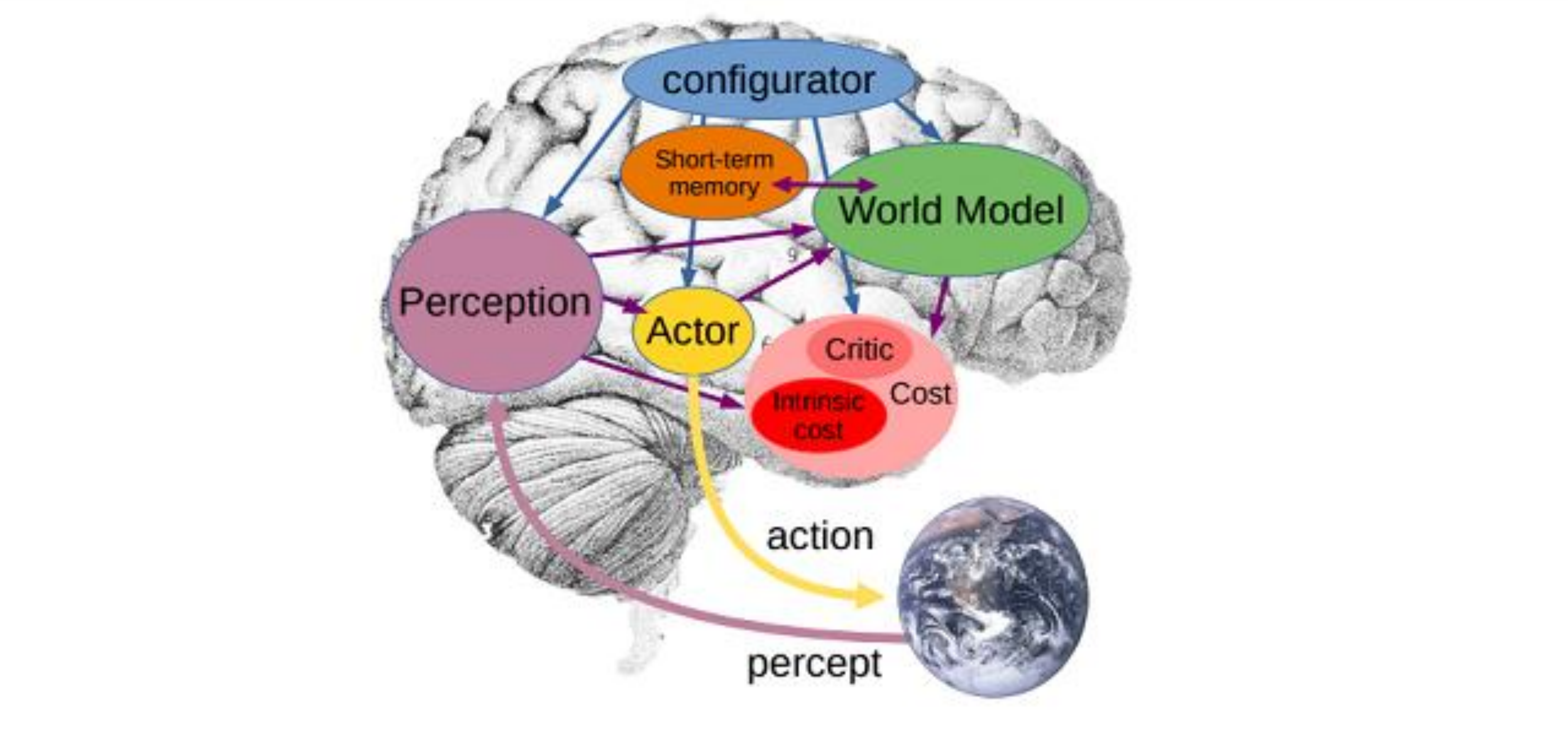

Natural Language Processing
Red
Queen
Quantum Physics
$|\psi_{\mathrm{a \cdot n}}\rangle = \eta_{\mathrm{a}}\left(|\psi_{\mathrm{n}}\rangle\right)$
Quantum Circuits
Queen
Red
ZX Calculus






True
Earth
Sun
True force
Transformer
Earth
Sun
Predicted force
True force law (Newton)
$F \propto \frac{m_1 m_2}{r^2}$
Predicted force law (transformer)
$F \propto \left(\sin\left(\frac{1}{\sin(r-0.24)}\right) + 1.45\right) * \frac{1}{\frac{1}{r} + m_2}$

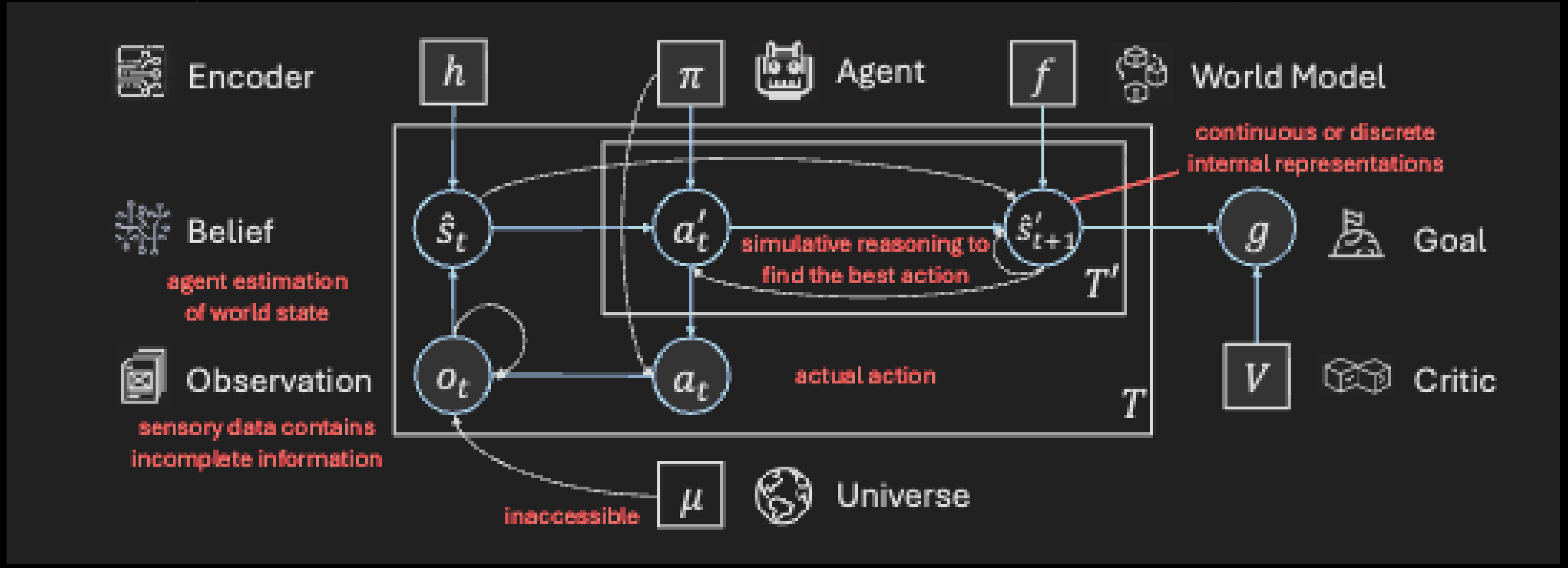
World Model



configurator
Short-term memory
World Model
Perception
Actor
Critic
Intrinsic cost
Cost
action
percept

WORLD MODEL

$$p^\pi_\mu(a_t, s_{t+1}, \ldots, s_T \mid s_t) = \prod_{k=t}^{T-1} \underbrace{p_\pi(a_k \mid s_k)}_{\text{agent}} \underbrace{p_\mu(s_{k+1} \mid s_k, a_k)}_{\text{universe}}$$

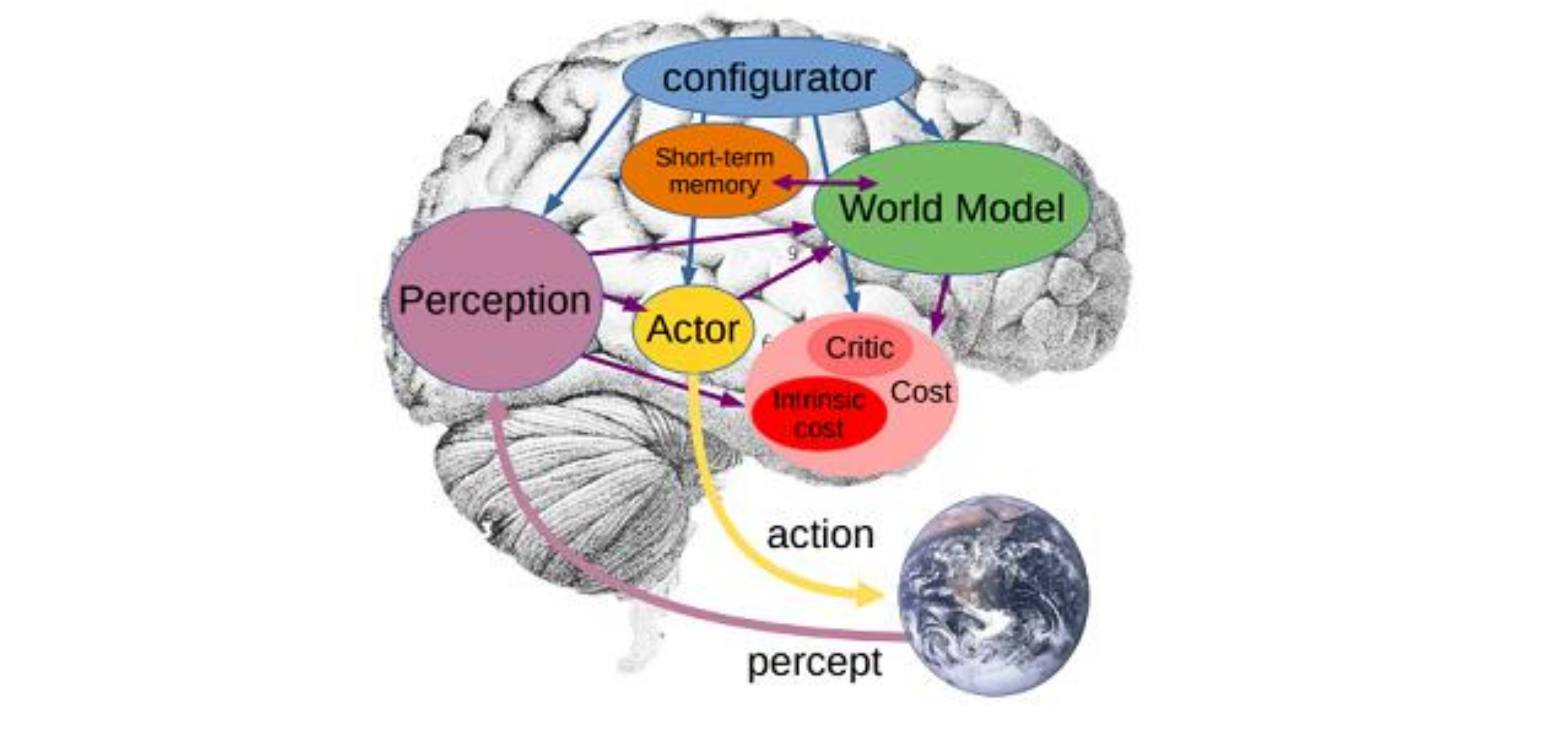


$$V^g_{\pi,\mu}(s_t) := \mathbb{E}_{\pi,\mu}\left[\sum_{k=t}^{\infty} \gamma_k r(g, s_k) \,\middle|\, s_t\right] = \lim_{T\to\infty} \sum_{(a_t, s_{t+1}, \ldots, s_T)} \underbrace{\textstyle\sum_{k=t}^{T} \gamma_k r(g, s_k)}_{\text{goal}} \underbrace{p^\pi_\mu(a_t, s_{t+1}, \ldots, s_T \mid s_t)}_{\text{trajectory}}$$

$$\pi^*_\mu := \arg\max_\pi V^g_{\pi,\mu}$$

$$\pi^*_f(\hat{s}_t) = \underbrace{\arg\max_{a'_{t:T'-1}}}_{\text{possible actions}} \sum_{\hat{s}_{t+1:T'}} \underbrace{\left(\sum_{k=t}^{T'-1} \gamma_k r(g, \hat{s}_k) + \gamma_{T'} V^g_{\pi,f}(\hat{s}_{T'})\right)}_{\text{goal progress}} \prod_{i=t}^{T'-1} \underbrace{p_f(\hat{s}_{i+1} \mid \hat{s}_i, a'_i)}_{\text{simulation with world model}}$$

$$\pi^*_\mu(s_t) = \underbrace{\arg\max_{a_{t:T-1}}}_{\text{possible actions}} \sum_{s_{t+1:T}} \underbrace{\left(\sum_{k=t}^{T-1} \gamma_k r(g, s_k) + \gamma_T V^g_{\pi,\mu}(s_T)\right)}_{\text{goal progress}} \prod_{i=t}^{T-1} \underbrace{p_\mu(s_{i+1} \mid s_i, a_i)}_{\text{universe response}}$$

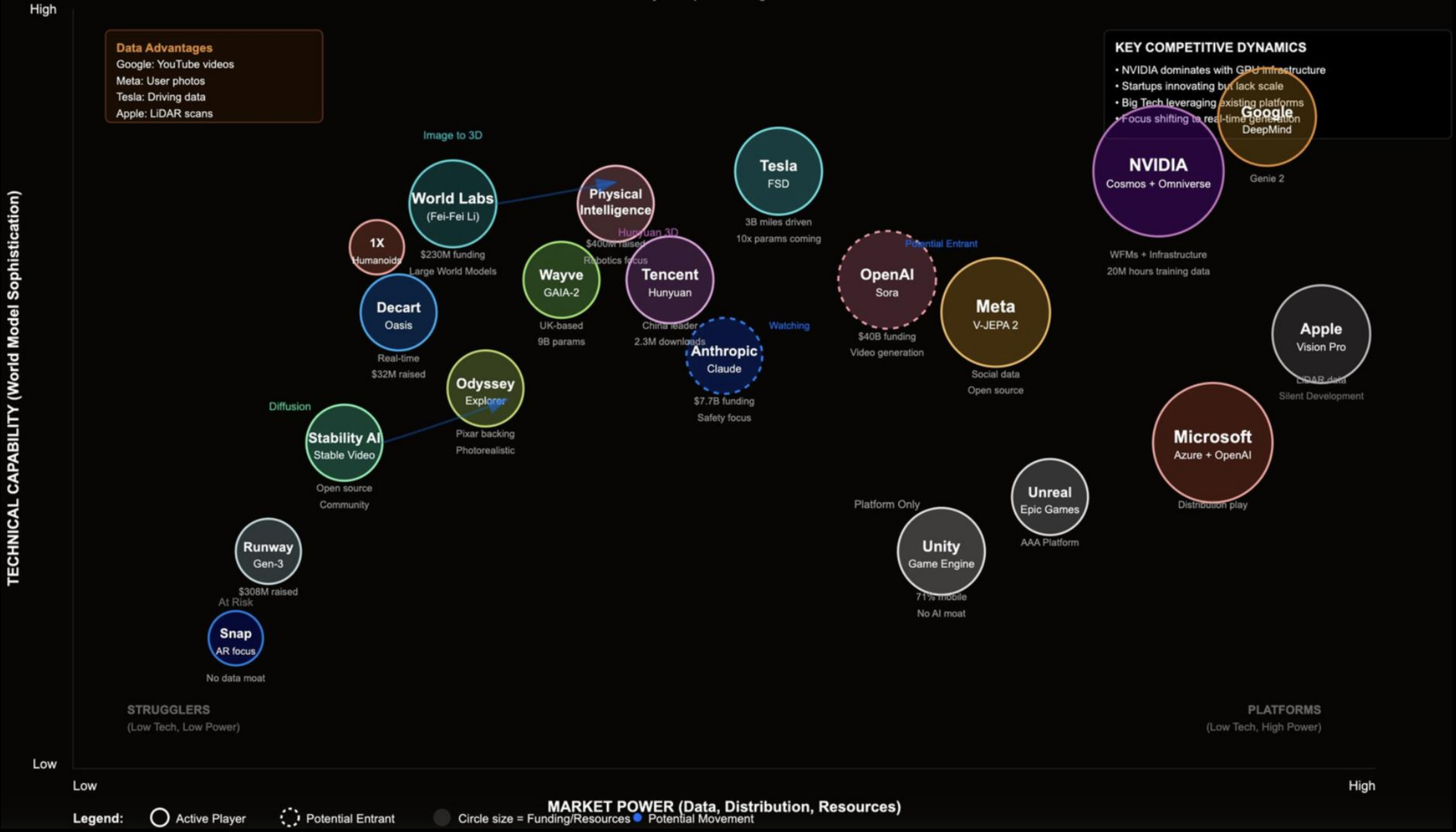


A quaint village with cobblestone streets, small wooden houses with thatched roofs, and a central square featuring a stone well surrounded by flower beds...

World Models Competitive Map

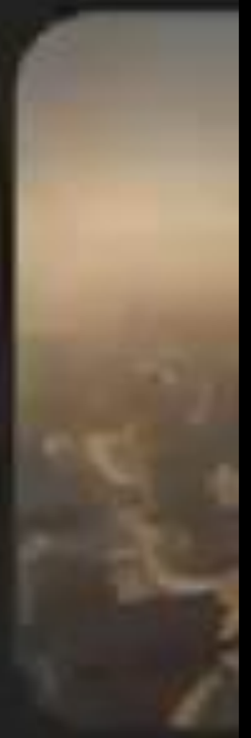
Positioning by Technical Capability vs Market Power

Analysis Update: August 2025



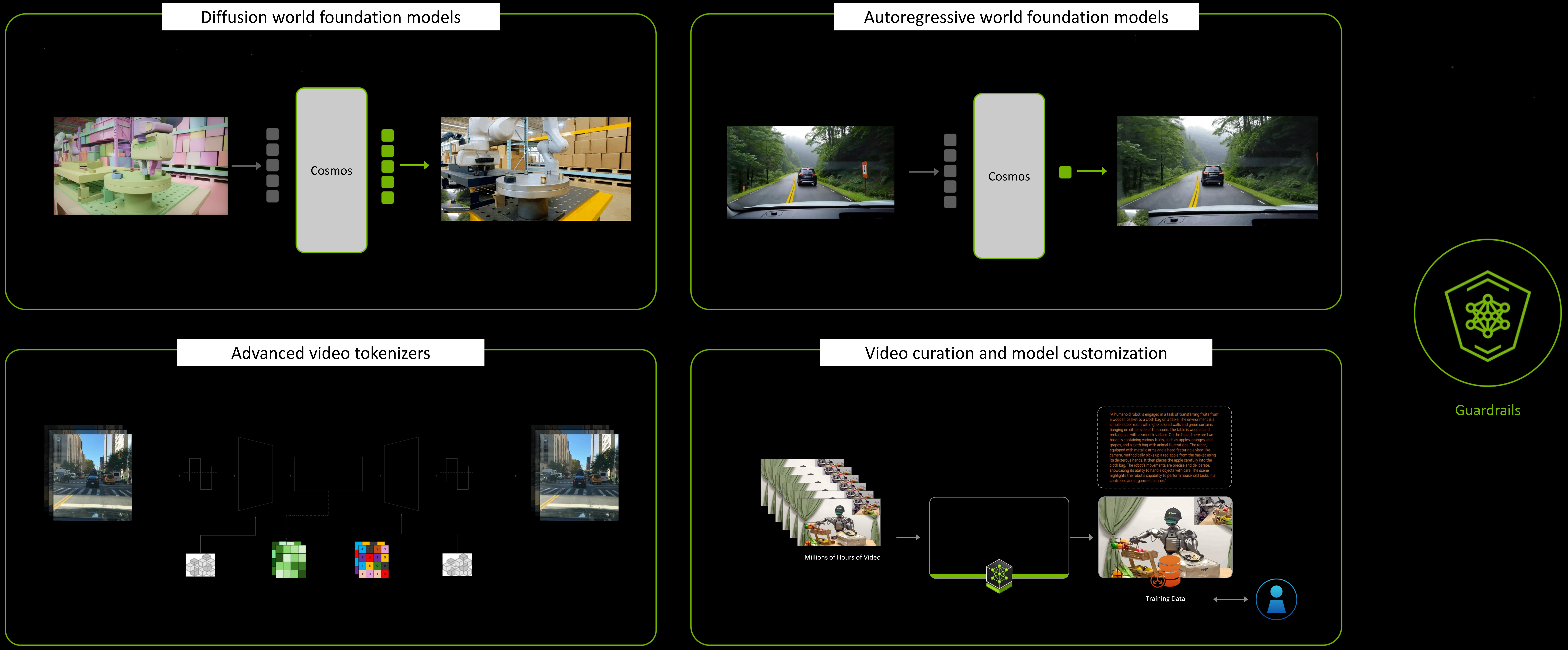
FIRST FRAME

LAST FRAME



NVIDIA Cosmos

World foundation model development platform for advancing Physical AI



The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

Chris Lu[1,2,*], Cong Lu[3,4,*], Robert Tjarko Lange[1,*], Jakob Foerster[2,†], Jeff Clune[3,4,5,†] and David Ha[1,†]
[*]Equal Contribution, [1]Sakana AI, [2]FLAIR, University of Oxford, [3]University of British Columbia, [4]Vector Institute, [5]Canada CIFAR AI Chair, [†]Equal Advising

Empowering Biomedical Discovery with AI Agents

Shanghua Gao[1], Ada Fang[1,2,8,+], Yepeng Huang[1,3,+], Valentina Giunchiglia[1,4,+], Ayush Noori[1,5,+], Jonathan Richard Schwarz[1], Yasha Ektefaie[1,6], Jovana Kondic[7], and Marinka Zitnik[1,8,9,10,#]

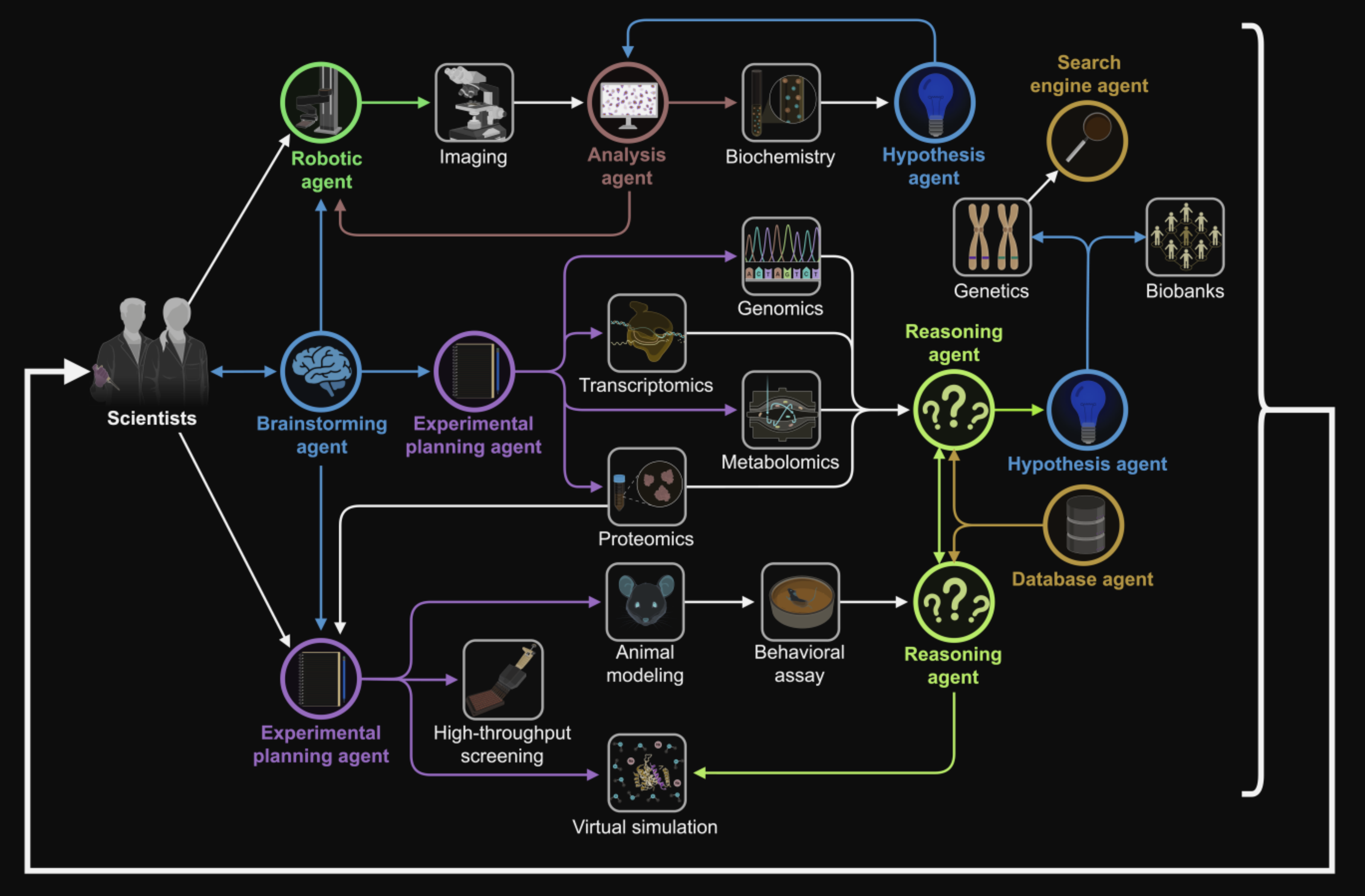
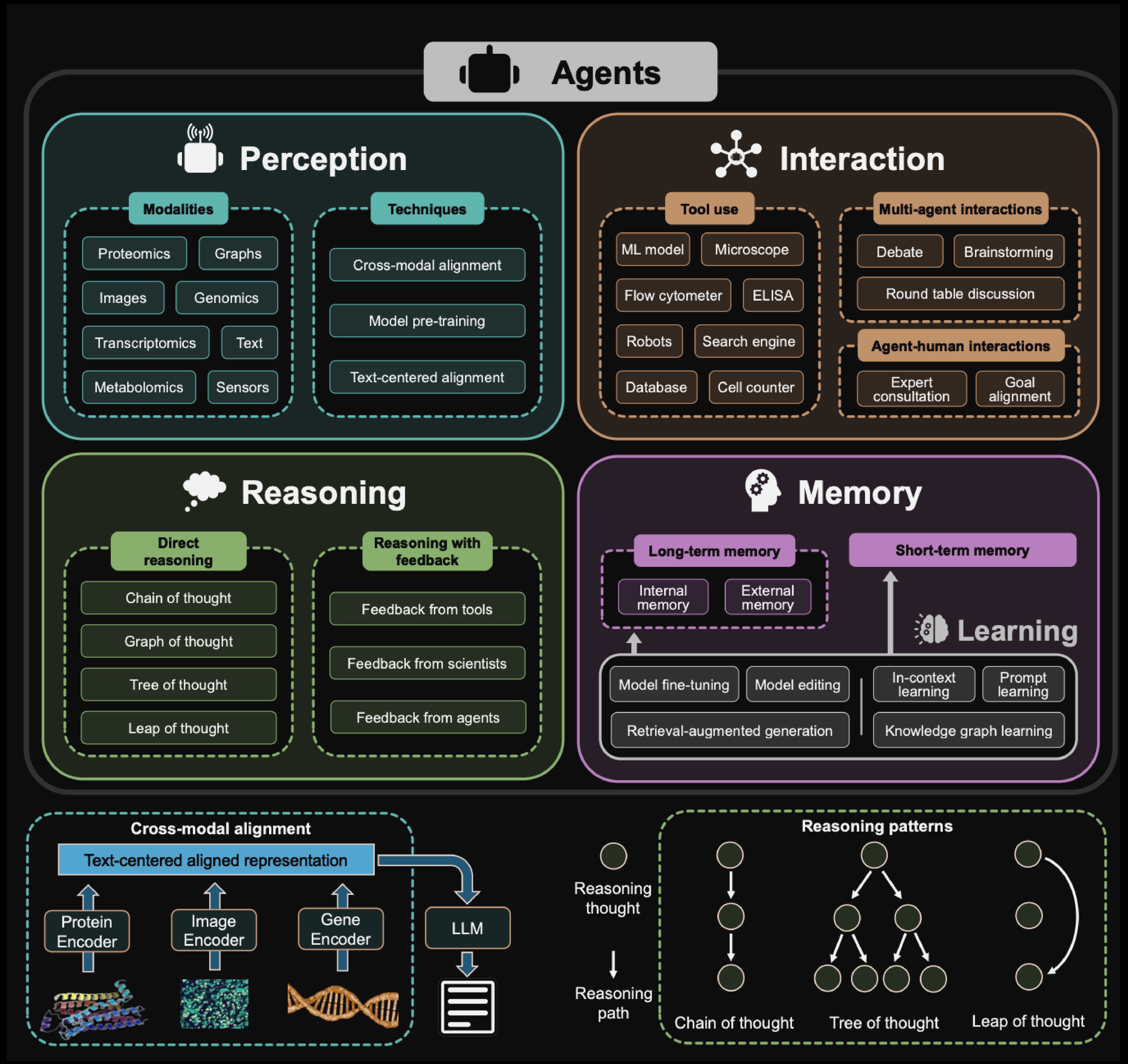
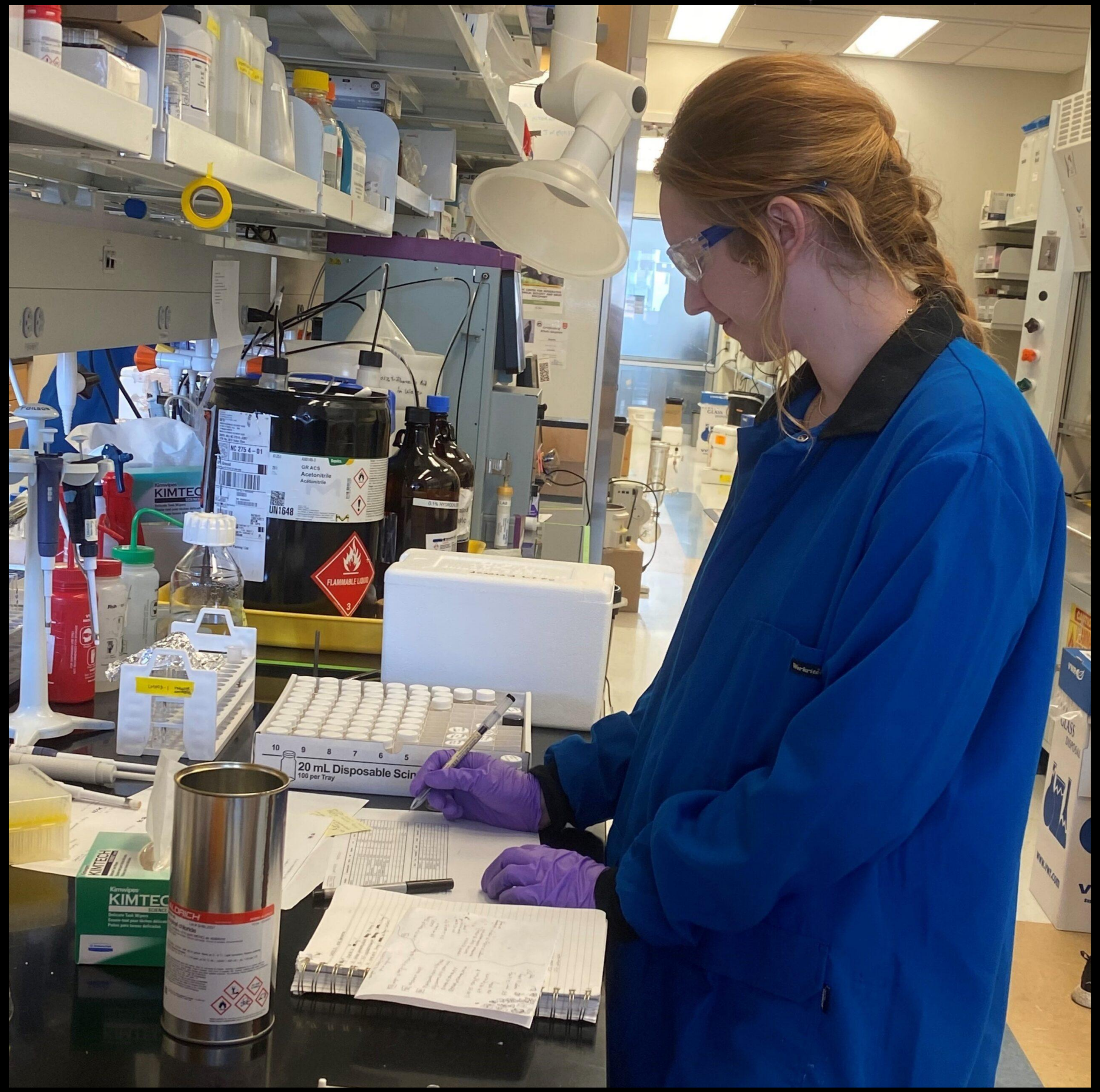
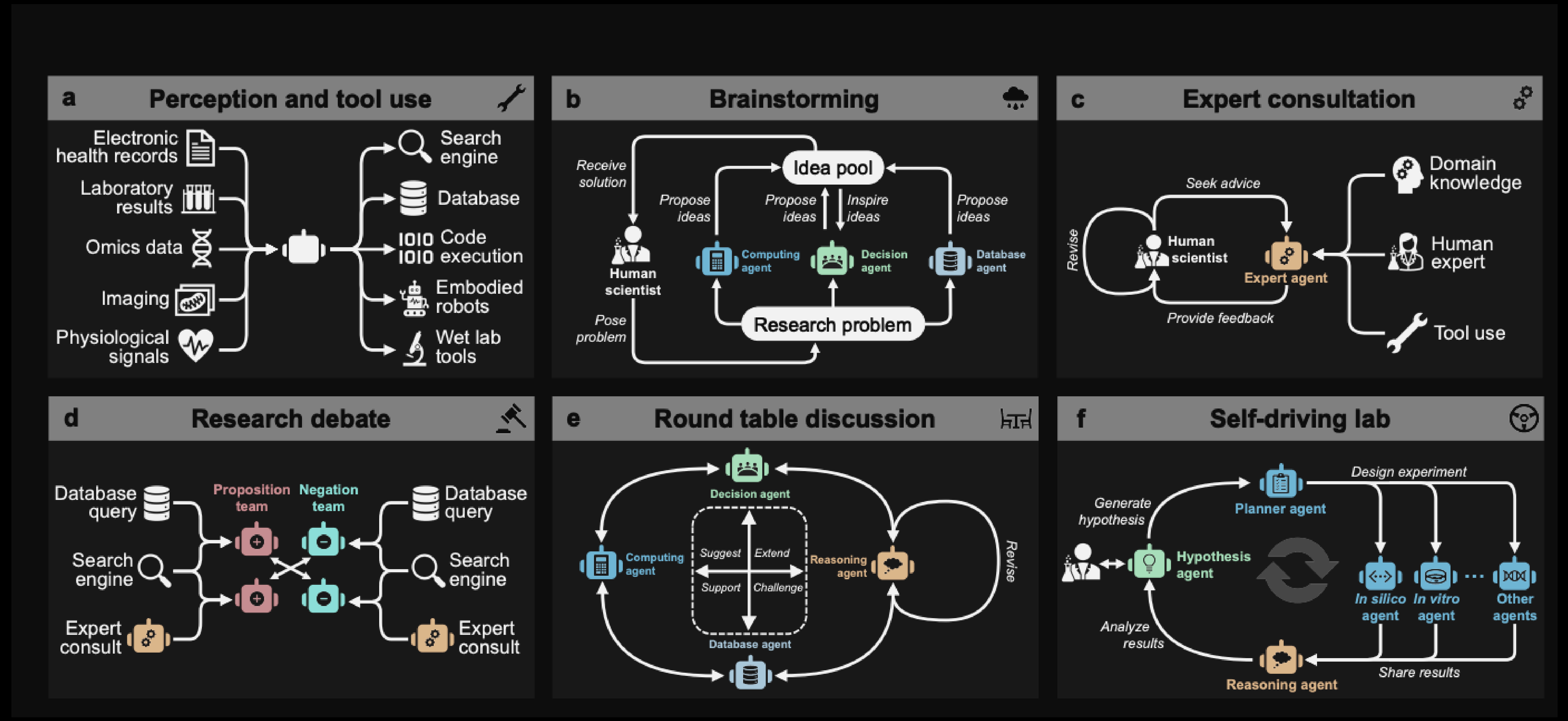


Figure 1: Empowering biomedical research with AI agents. AI agents pave the way for "AI scientists" capable of skeptical learning and reasoning. These multi-agent systems consist of agents based on conversable large language models (LLMs) and can coordinate machine learning (ML) tools, experimental platforms, humans, or even combinations of them. Robotic agent, AI agent that operates robotic hardware for physical experiments; Database agent, AI agent that can information in databases via 'function calling' and APIs; Reasoning agent, AI agent capable of direct reasoning and reasoning with feedback; Hypothesis agent, AI agent that is creative and reflective when developing hypotheses, capable of characterizing its own uncertainty and using that as a driver to refine its scientific knowledge bases; Brainstorming agent, AI agent that generates a broad spectrum of research ideas; Search engine agent, AI agent that uses search engines as tools to rapidly gather information; Analysis agent, AI agent capable of analyzing experimental results to summarize findings and synthesize concepts; Experimental planning agent, AI agent that optimizes an experimental protocol for execution.







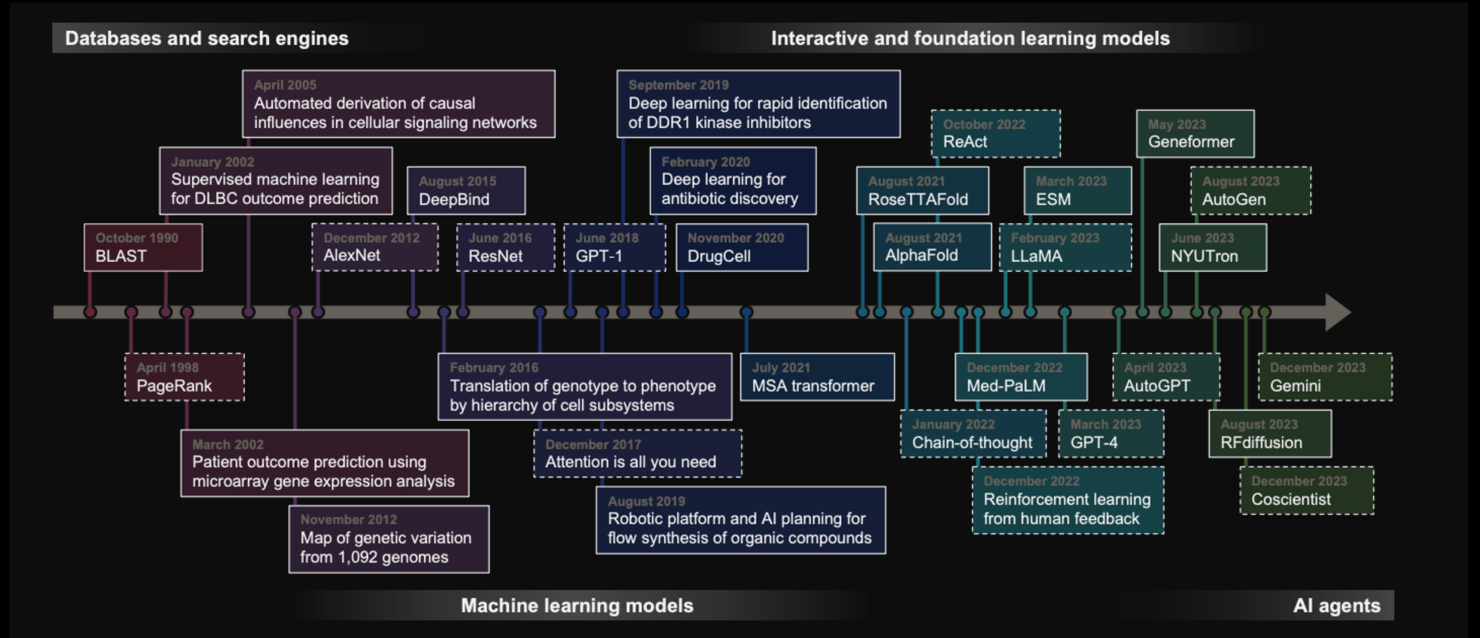


Figure 2: Evolving use of data-driven models in research. Data-driven approaches, from databases and search engines, machine learning, and interactive learning models to advanced agent systems (Section), have reshaped biomedical research throughout the last several decades. Dashed boxes represent studies focused predominantly on algorithmic machine learning innovation; solid-line boxes represent studies focused predominantly on biomedical discovery.

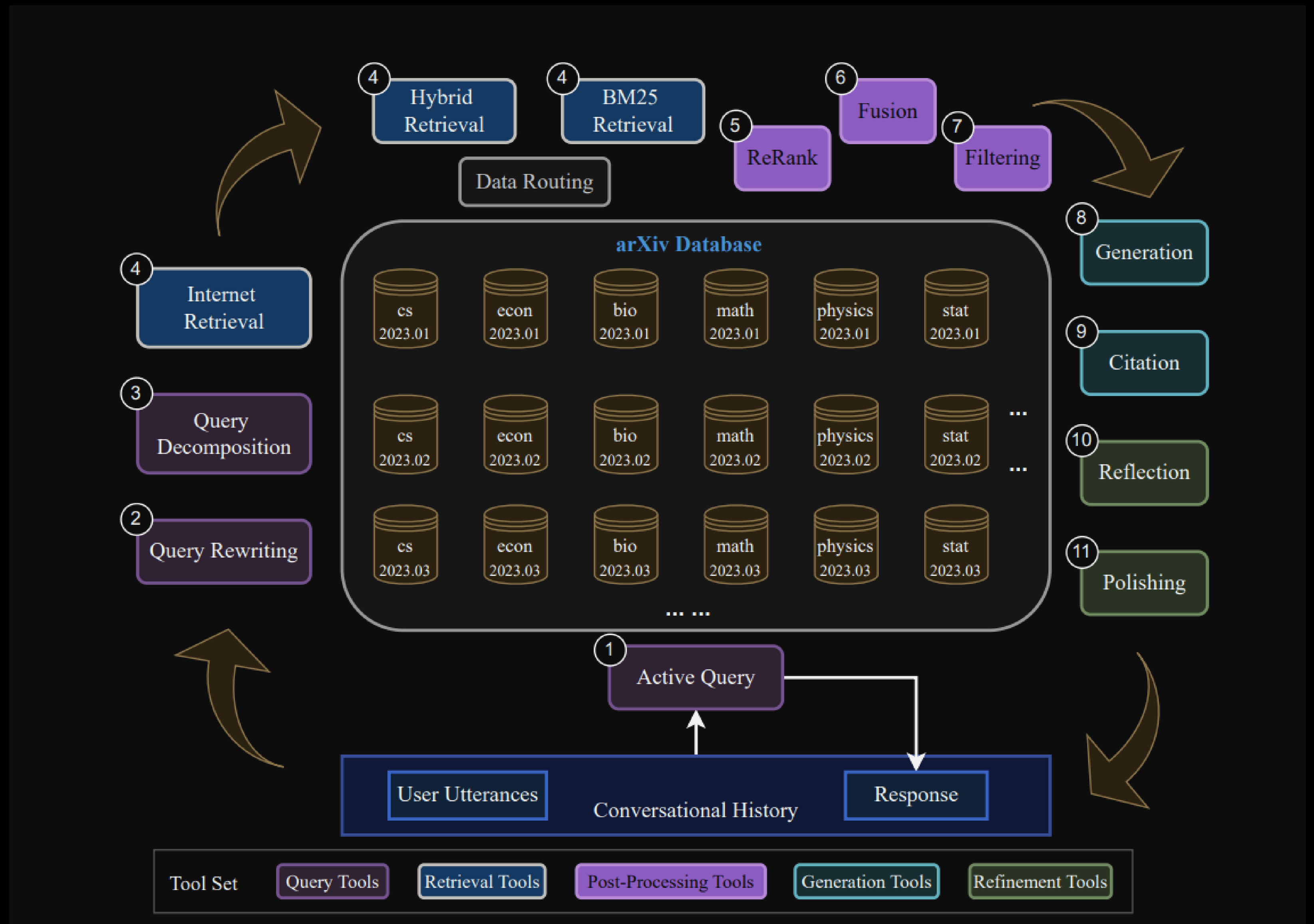
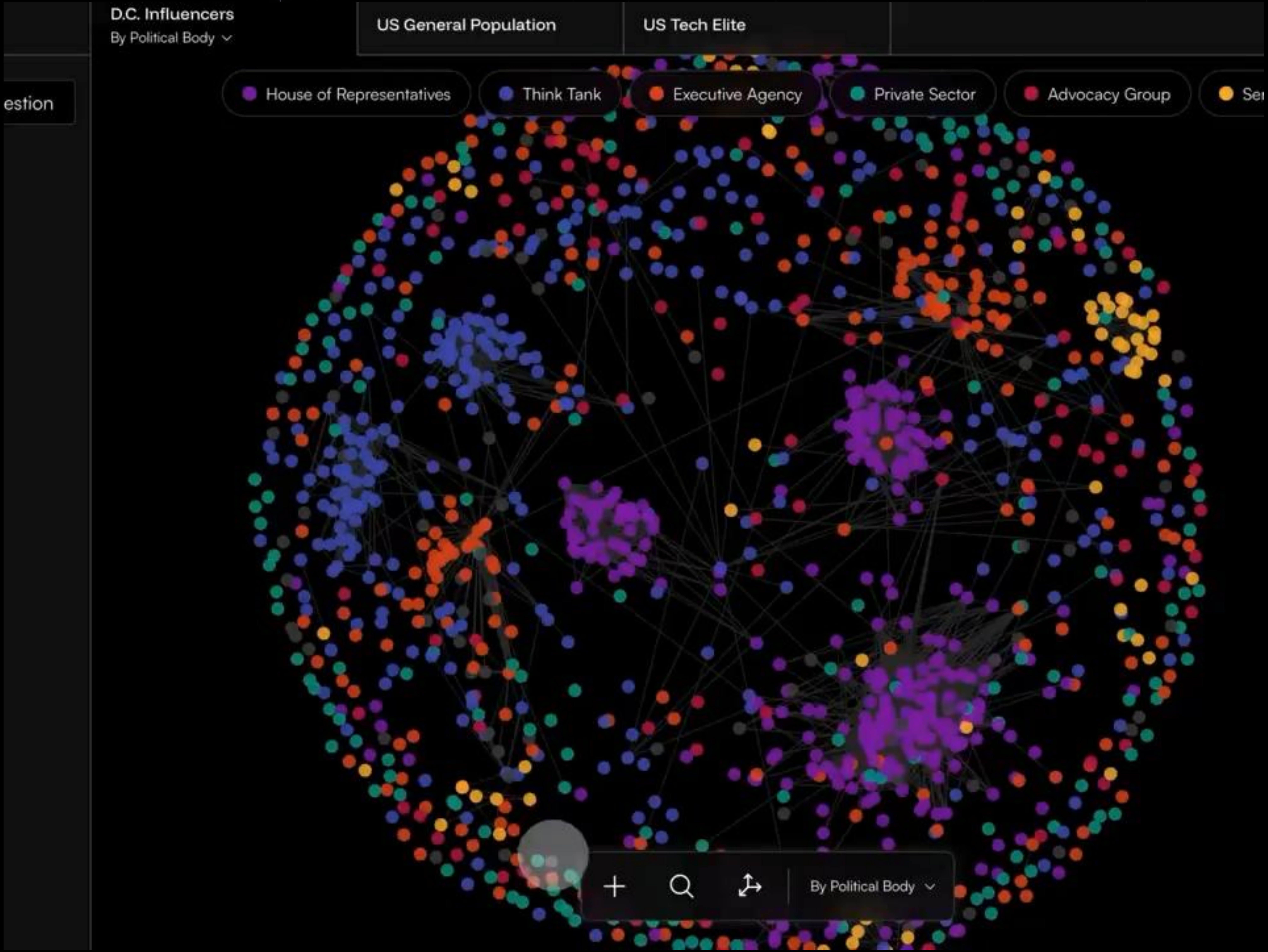


Figure 1: Main Workflow of OpenResearcher.


D.C. Influencers
By Political Body
US General Population
US Tech Elite
estion
House of Representatives
Think Tank
Executive Agency
Private Sector
Advocacy Group
By Political Body

Me

The Denario project: Deep knowledge AI agents for scientific discovery

Francisco Villaescusa-Navarro[1,2,*], Boris Bolliet[3,4,*], Pablo Villanueva-Domingo[5,*], Adrian E. Bayer[1,2], Aidan Acquah[6], Chetana Amancharla[7], Almog Barzilay-Siegal[8], Pablo Bermejo[9,10,11], Camille Bilodeau[12], Pablo Cárdenas Ramírez[13,14,15], Miles Cranmer[16], Urbano L. França[17,18], ChangHoon Hahn[19,20], Yan-Fei Jiang[1], Raul Jimenez[21,22], Jun-Young Lee[1], Antonio Lerario[23], Osman Mamun[13], Thomas Meier[24], Anupam A. Ojha[25,26], Pavlos Protopapas[27], Shimanto Roy[12], David N. Spergel[1], Pedro Tarancón-Álvarez[21,28], Ujjwal Tiwari[7], Matteo Viel[23,29,30,31,32], Digvijay Wadekar[33,40], Chi Wang[34], Bonny Y. Wang[35], Licong Xu[36,4], Yossi Yovel[8,37], Shuwen Yue[13], Wen-Han Zhou[38], Qiyao Zhu[25], Jiajun Zou[39], Íñigo Zubeldia[16,36,4]

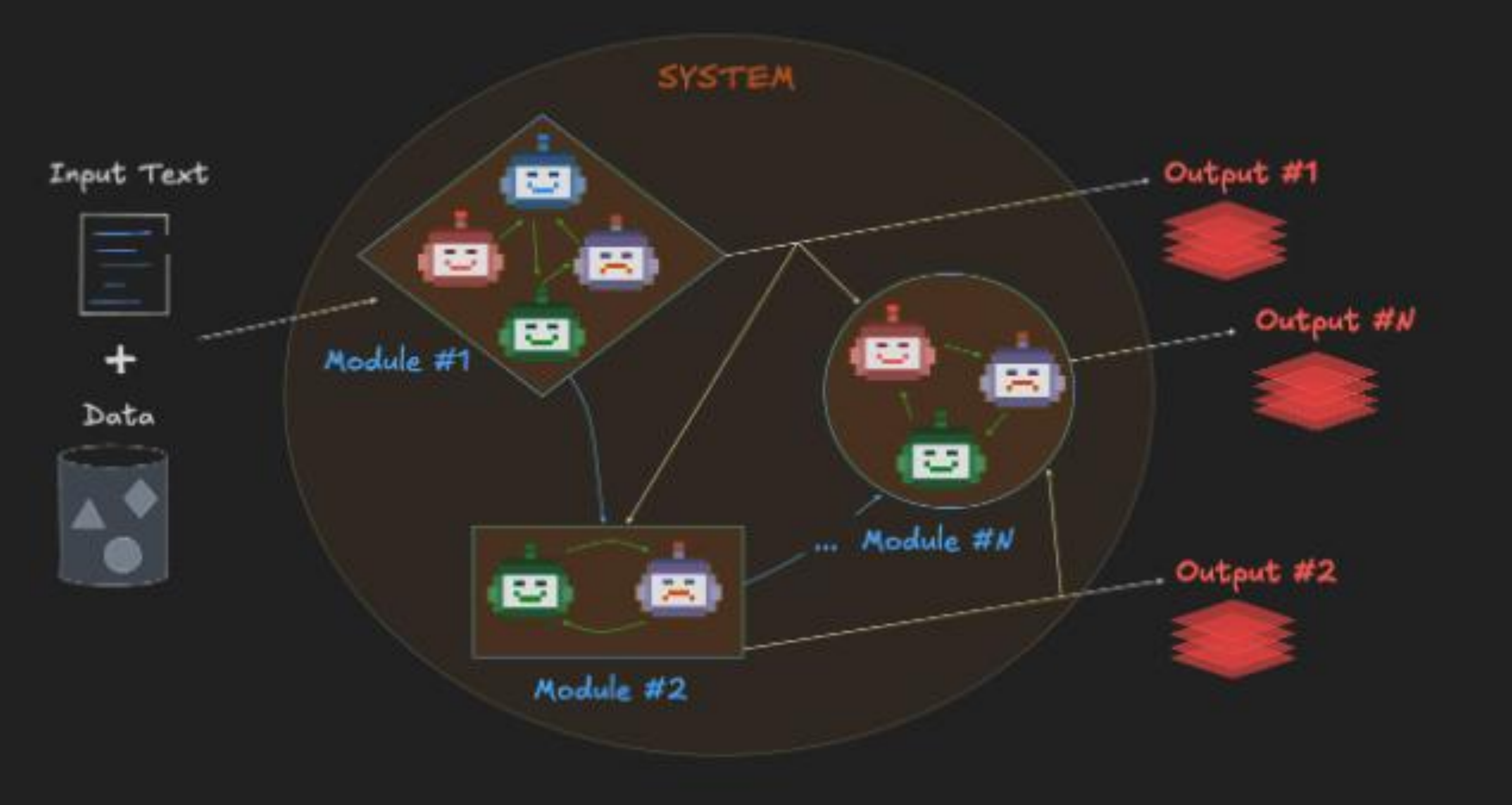
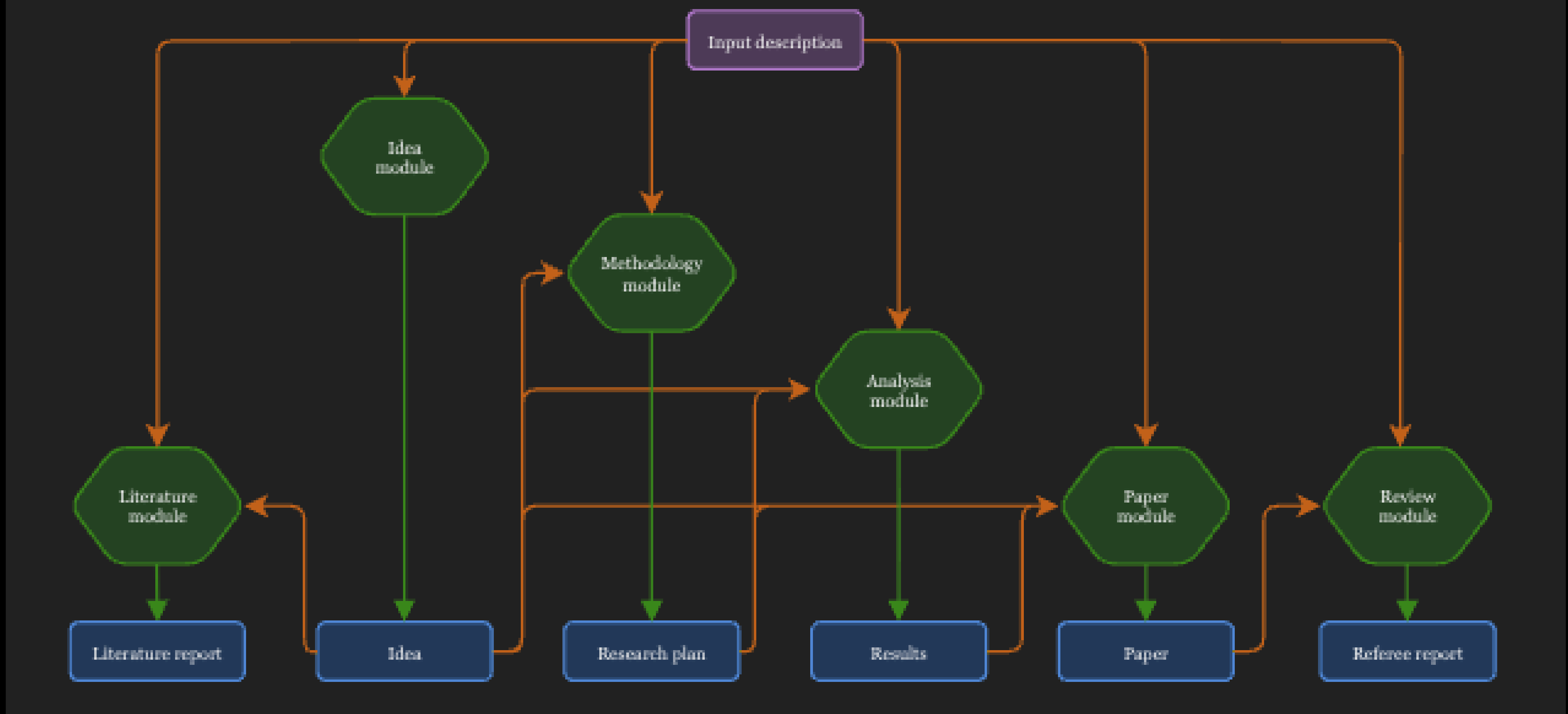


Figure 1: This cartoon shows the different components of Denario and their interplay. The system is shown as an orange circle, and it takes some input text and data (left side), and can generate one or several outputs (red icons on the right). The system is composed of different modules (yellow shapes), which can exchange messages among them (blue arrows outside modules). The modules are composed of multiple agents (bots icons), that can communicate with each other (blue arrows inside modules).



Module	Task	Input	Output
Idea	Generate project idea	input.md	idea.md
Literature	Determine if idea is new	input.md idea.md	literature.md
Methods	Develop project plan	input.md idea.md	methods.md
Analysis	Implement plan write and execute code make plots	input.md idea.md methods.md	results.md Plots
Paper	Write paper	input.md idea.md methods.md results.md	paper.pdf
Review	Review paper	(input.md) paper.pdf	referee.md

Table 1: This table shows the different modules contained in `Denario` along with their tasks, inputs, and outputs. `input.md` contains a description of the data or problem of interest and is provided by the human researcher in markdown format. `idea.md` contains the project idea, `literature.md` a report on whether the idea is new, `methods.md` the research methodology, `results.md` the research analysis, Plots is a folder containing the plots and `referee.md` contains a referee report. `paper.pdf` is the scientific paper in pdf format. Files between parentheses are optional.

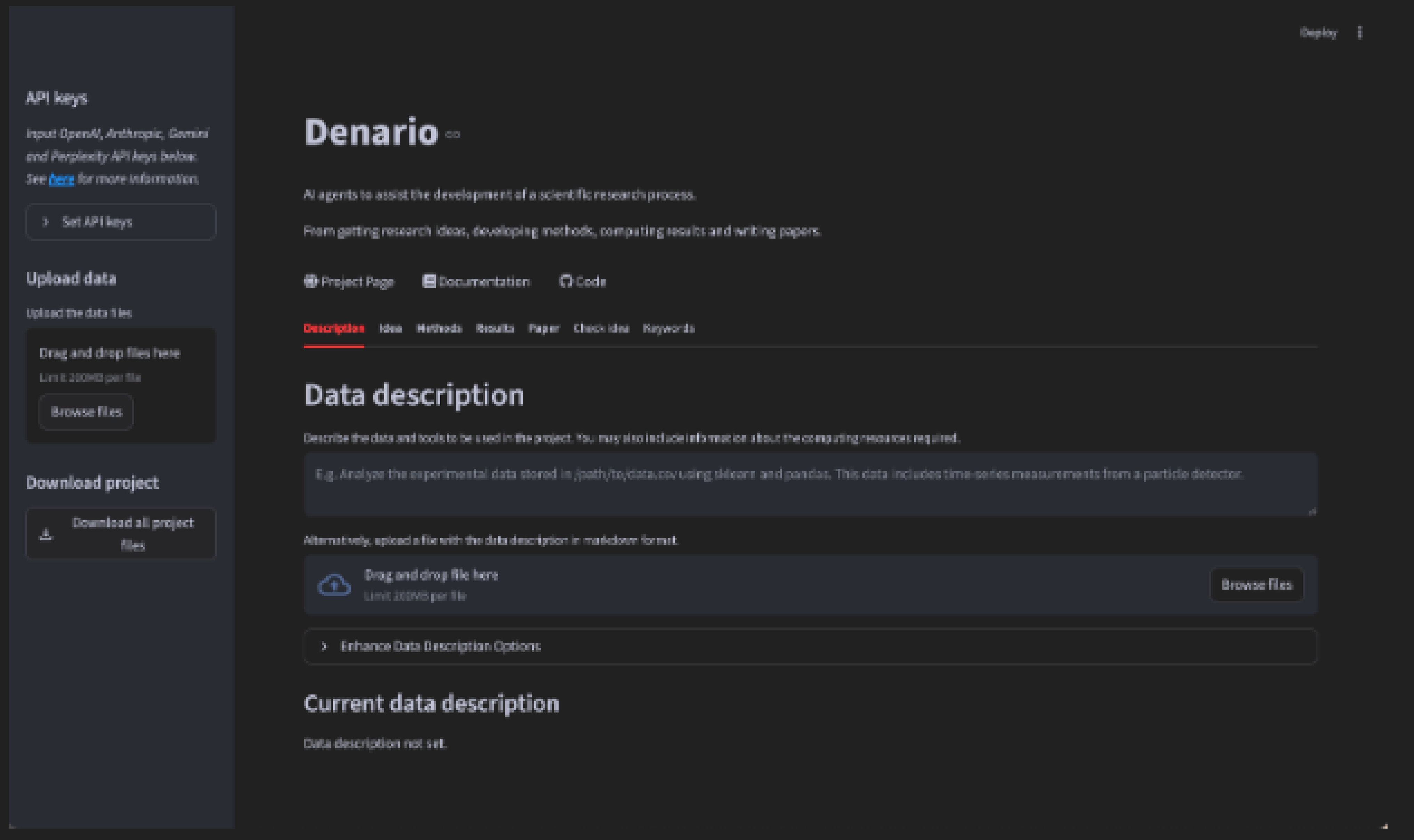


Figure 9: This figure shows the Graphical User Interface (GUI) of `Denario`. In the upper-left part, the user can set the API keys for the LLMs. Below that part, the user can upload data and download the files generated by `Denario`. In the central part, the user can choose which module to run and can tune the options available for it (e.g. LLM model). If the user wants to perform end-to-end research, she/he will have to run the different modules sequentially.

Input text

We have some data about asteroids. The data is located in /mnt/ceph/users/fvillaescusa/AstroPilot/Asteroids/data

Each asteroid is identified by a unique number. For example, 1 corresponds to the first discovered asteroid, Ceres.

All CSV files share a common structure:

- Column 1: Asteroid identification number
- Column 2: Corresponding physical or orbital property

These files are recommended to be opened with the Python package "pandas".
File Descriptions:

- asteroid_name.csv: number + name of the asteroid
- asteroid_diameter.csv: number + diameter of the asteroid in kilometers
- asteroid_semimajor_axis.csv: number + semimajor axis of the orbit in astronomical units (AU)
- asteroid_eccentricity.csv: number + eccentricity of the orbit
- asteroid_inclination.csv: number + inclination of the orbit in degrees
- asteroid_arg_peri.csv: number + argument of periapsis of the orbit in degrees
- asteroid_long_asc_node.csv: number + longitude of ascending node of the orbit in degrees
- asteroid_spin_period.csv: number + spin period of the asteroid in hours
- asteroid_obliquity.csv: number + obliquity of the asteroid in degrees (Obliquity is defined as the angle between the spin vector and the orbital angular momentum vector.)
- asteroid_type.csv: number + spectral type of the asteroid
- asteroid_family.csv: number + family name of the asteroid (The family name corresponds to the largest member of the family)
- asteroid_age.csv: number + age of the asteroid in gigayears (Gyr)

Please come up with an interesting project for a PhD thesis. Use state-of-the-art methods to analyze the data. For every step, make lots of plots and save the data you generate, as it may be used for other steps. When writing the code, write some lines to indicate whether the execution was successful or not. Join plots that are similar. Do not create dummy data. You have access to 128 cpus; for computationally heavy tasks, try to use all of them.

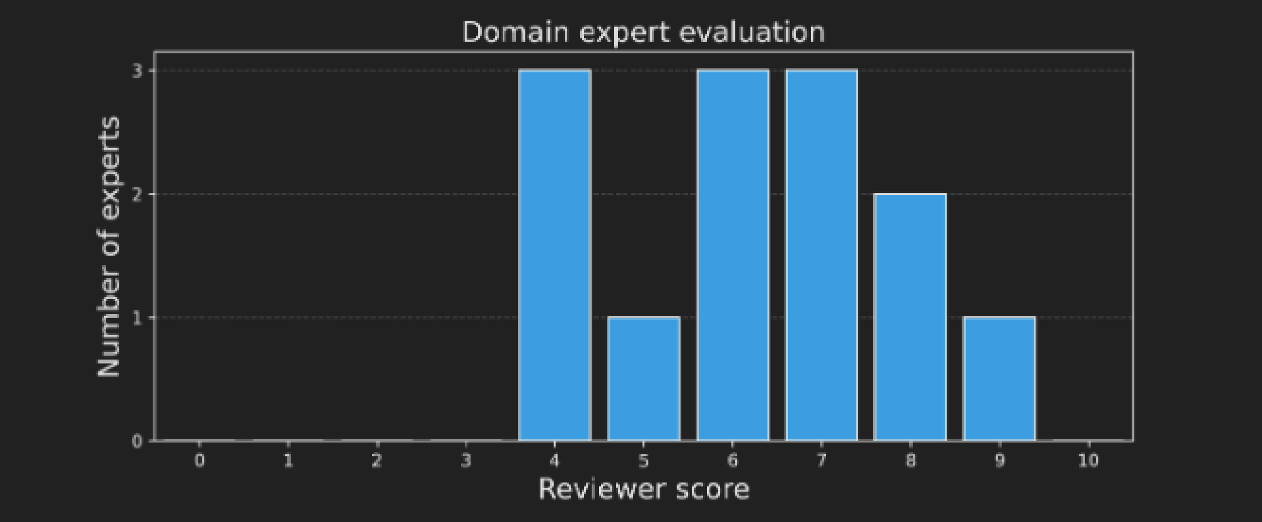
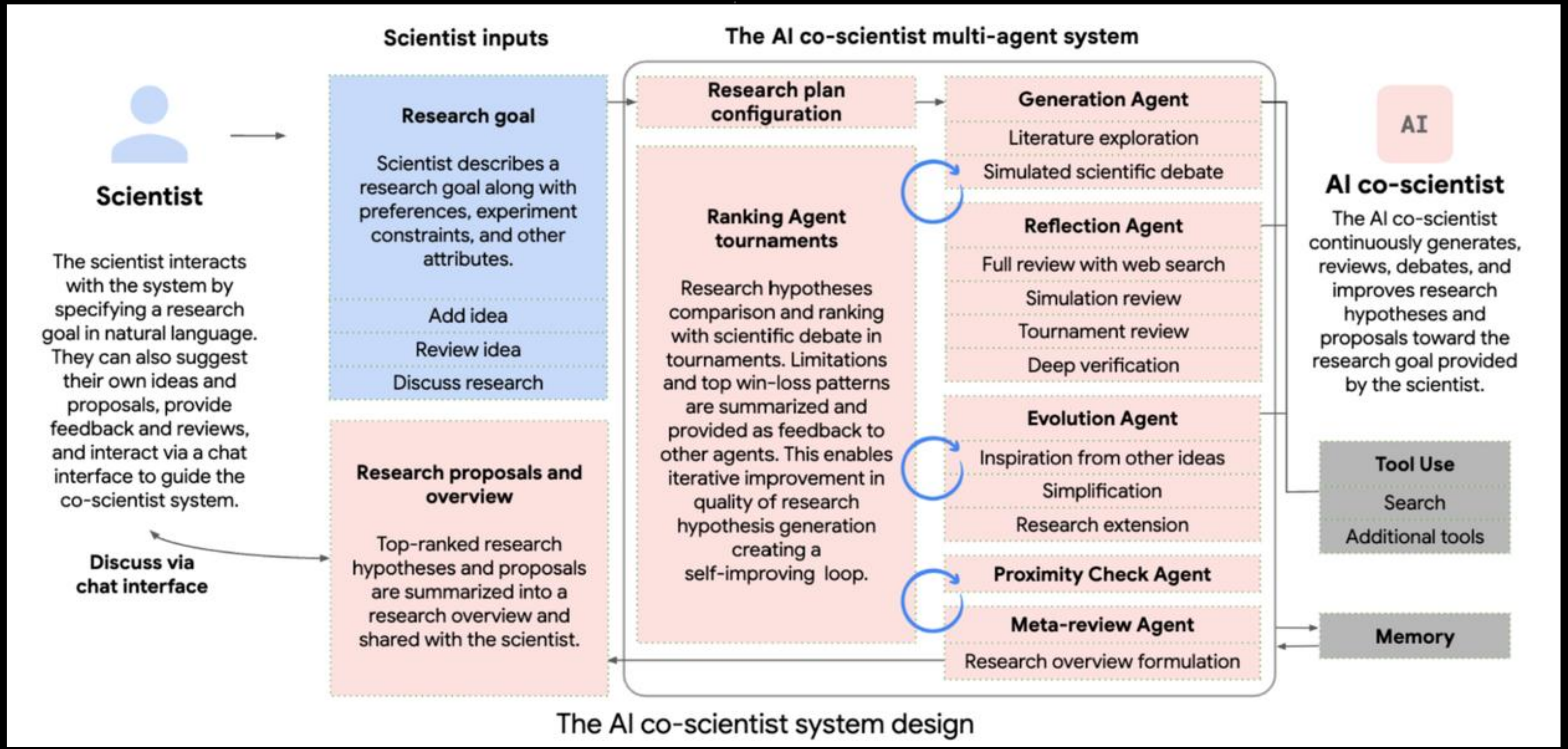


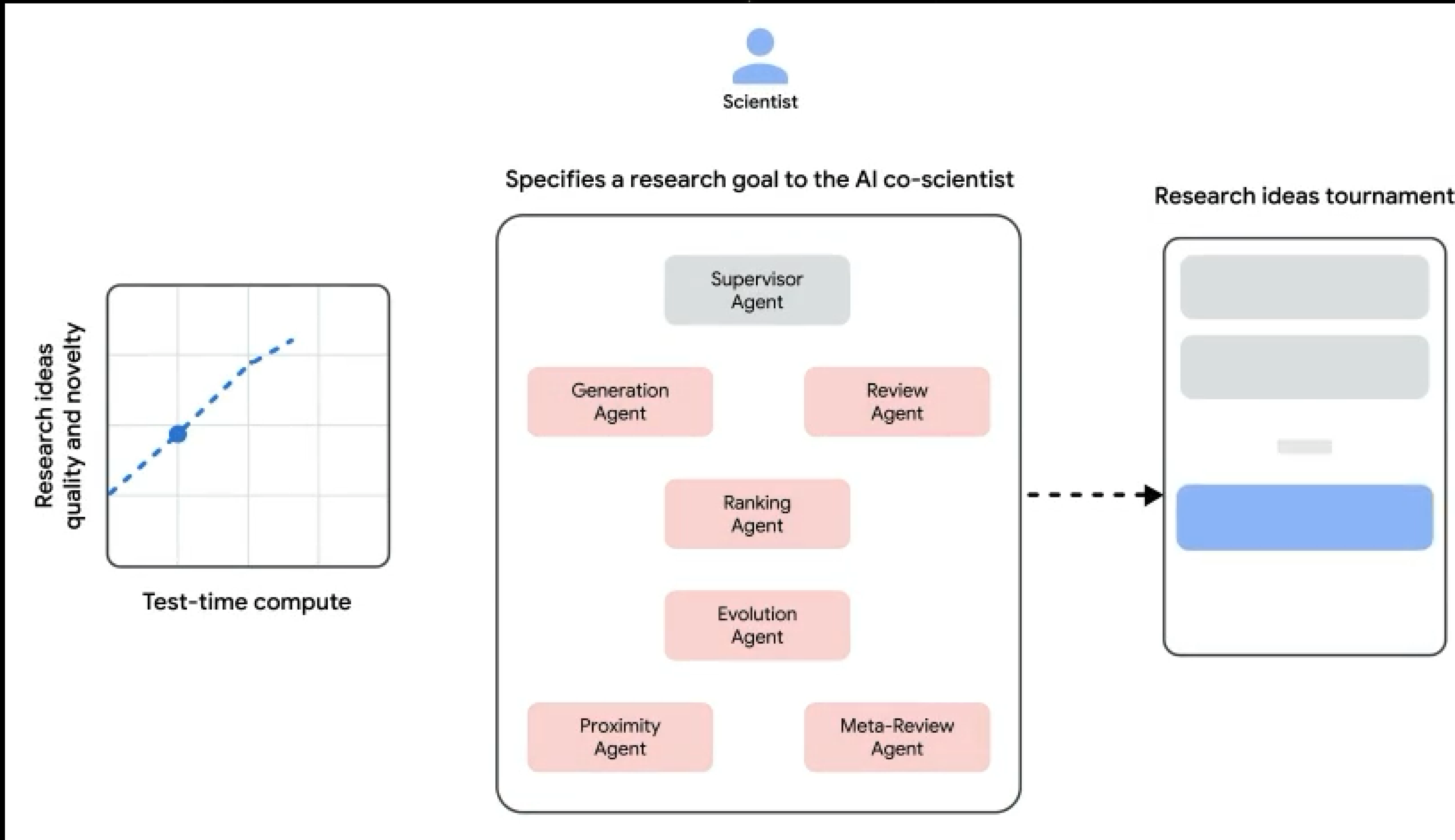
Figure 11: We have assigned each of the papers in the Appendix to a domain expert for evaluation. Each expert, knowing that these are AI-generated papers, was tasked with providing a numerical score to quantify the quality of the paper, from 0 (really bad paper) to 10 (really good paper). This graph shows the distribution of the scores. As can be seen, while some papers are ranked below average, most papers are above average, and some of them were highly ranked.

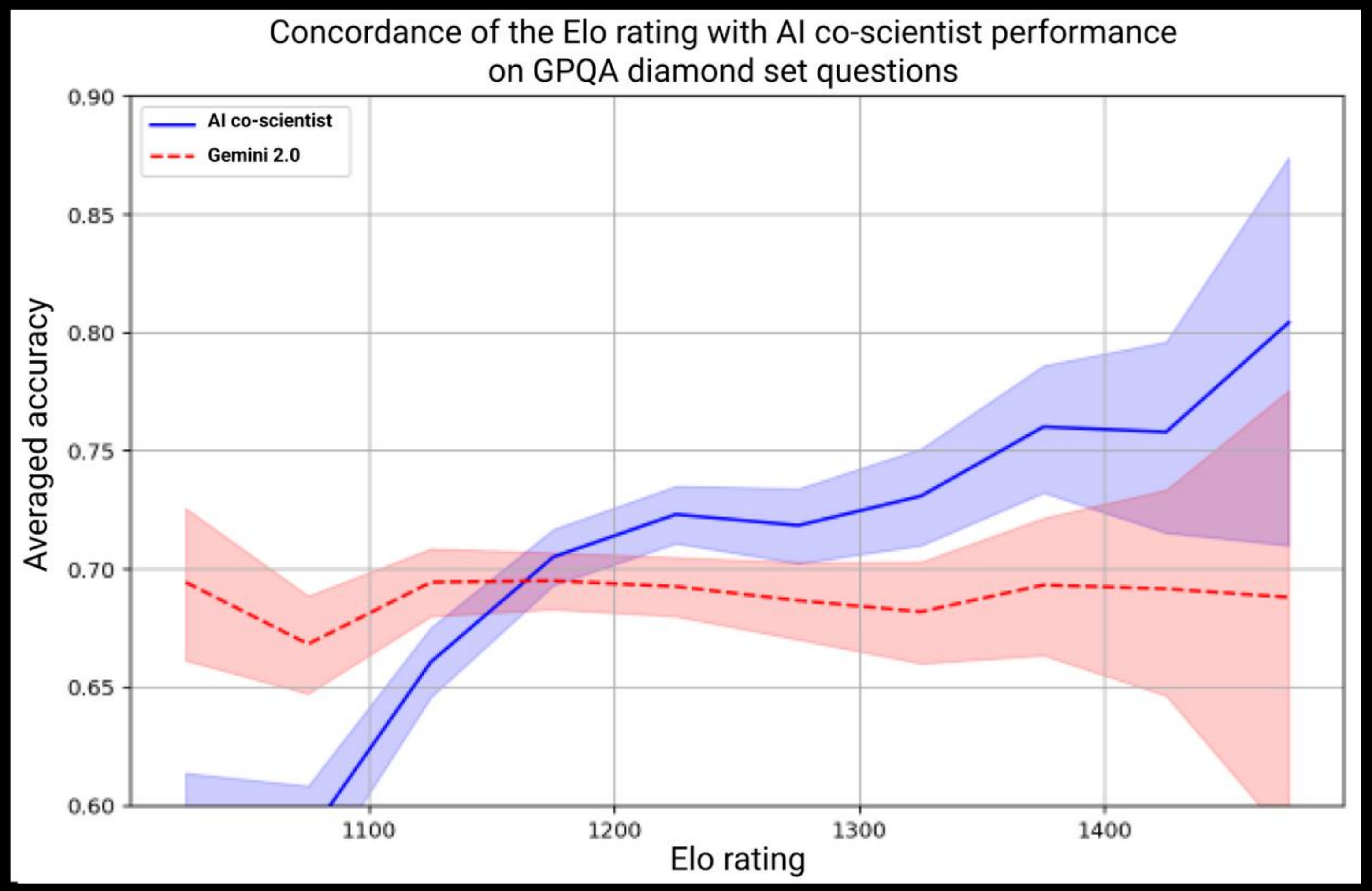
Accelerating scientific breakthroughs with an AI co-scientist

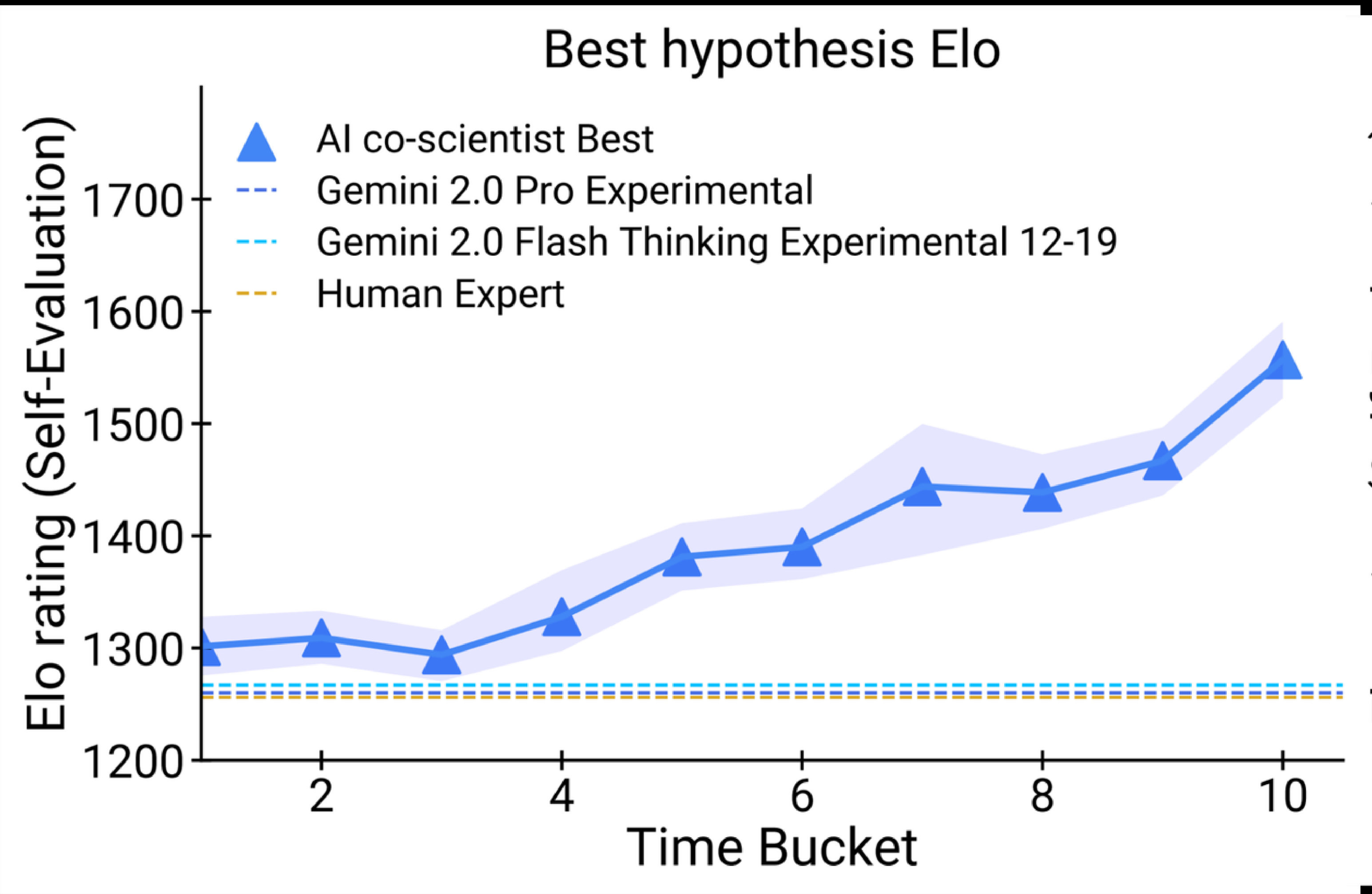
February 19, 2025 Juraj Gottweis, Google Fellow, and Vivek Natarajan, Research Lead

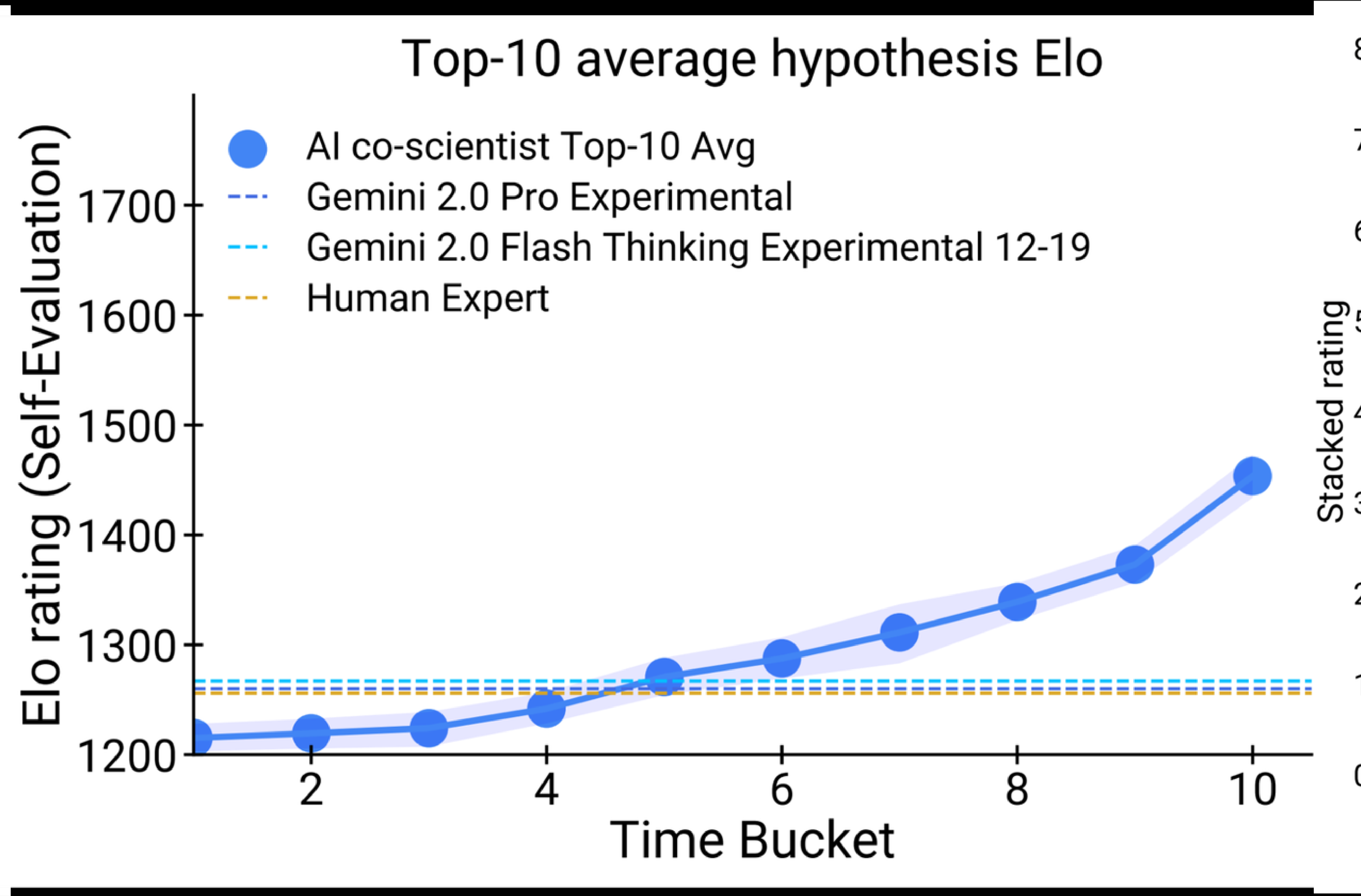


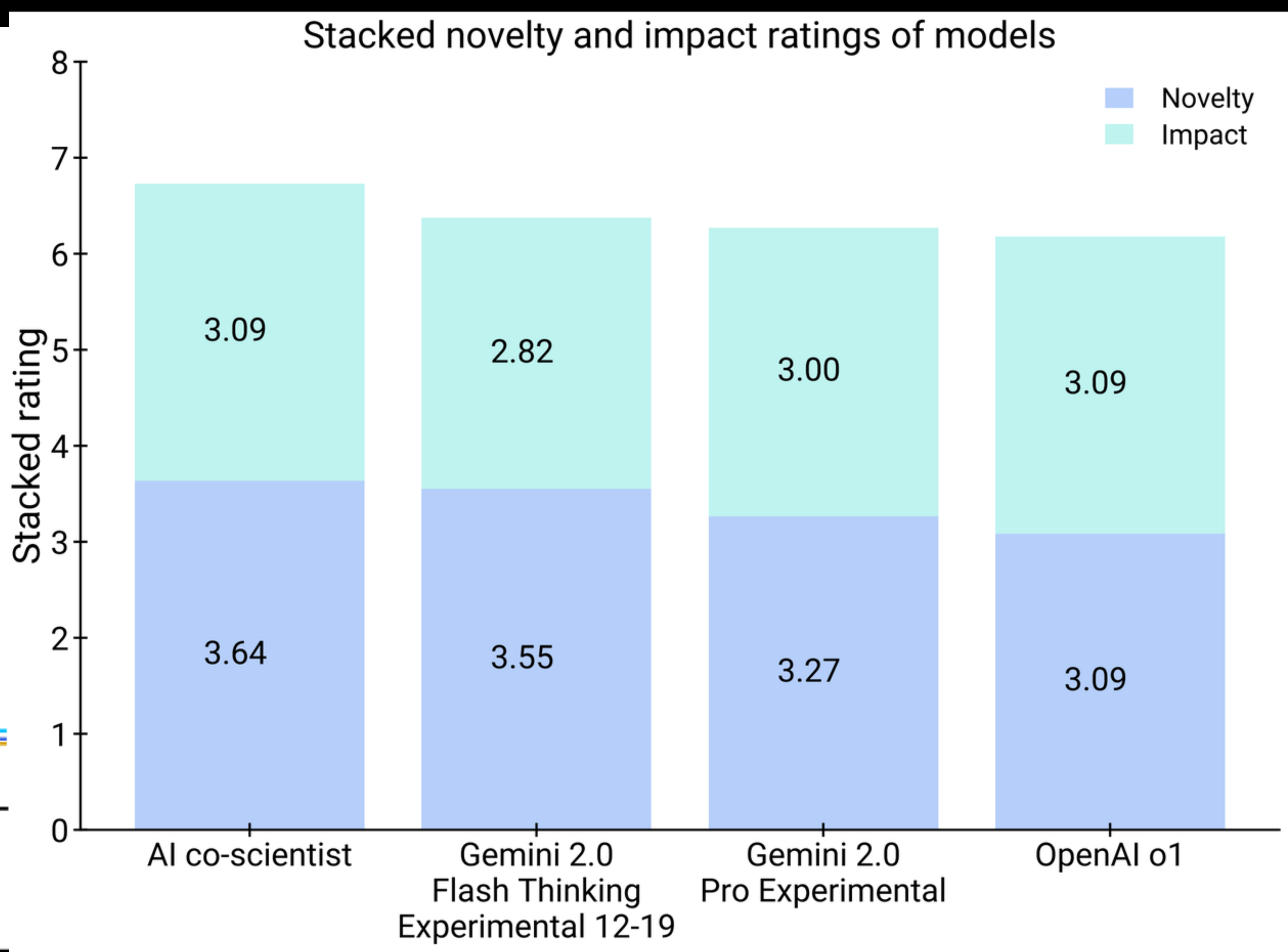
The AI co-scientist system design





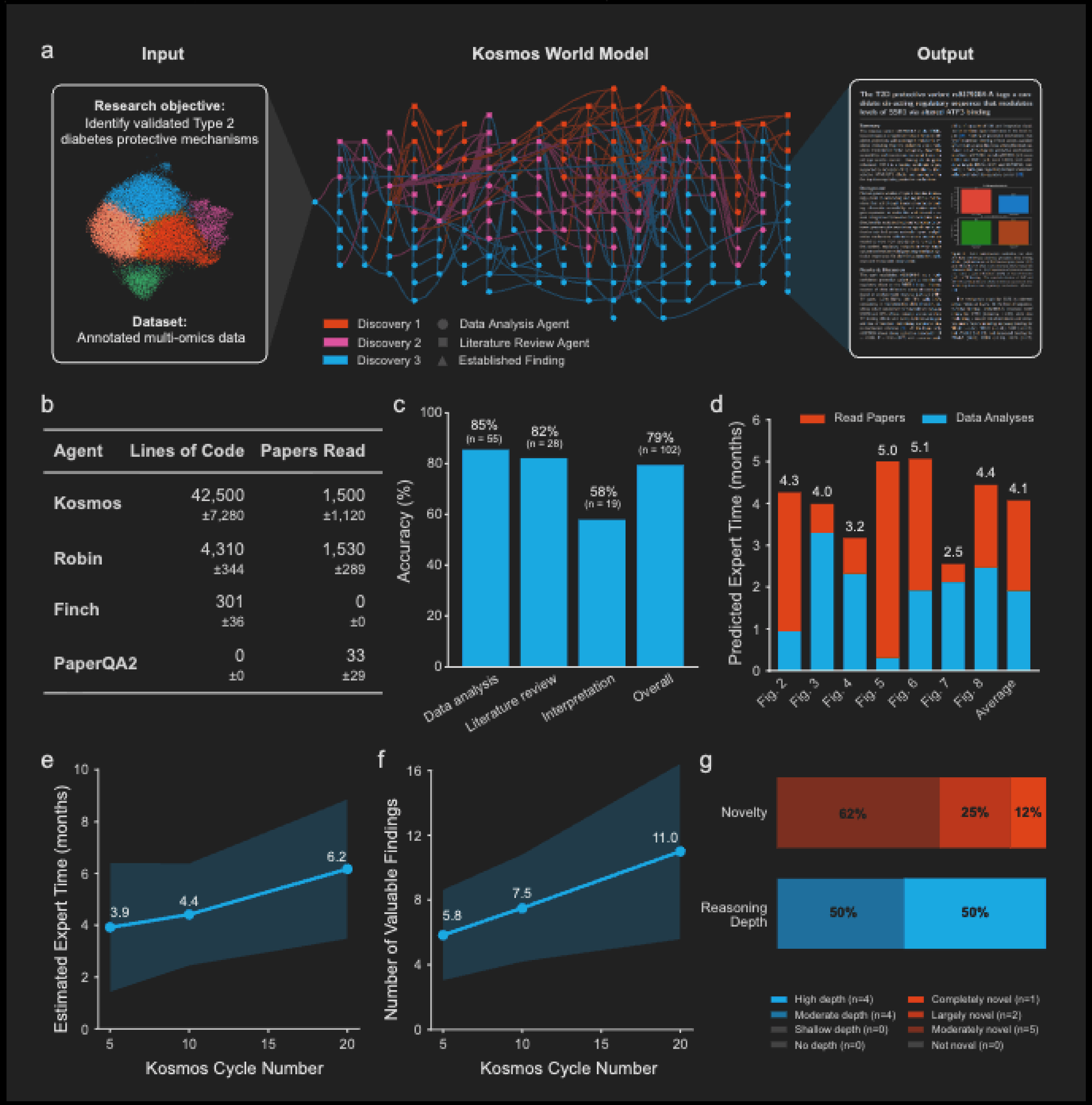


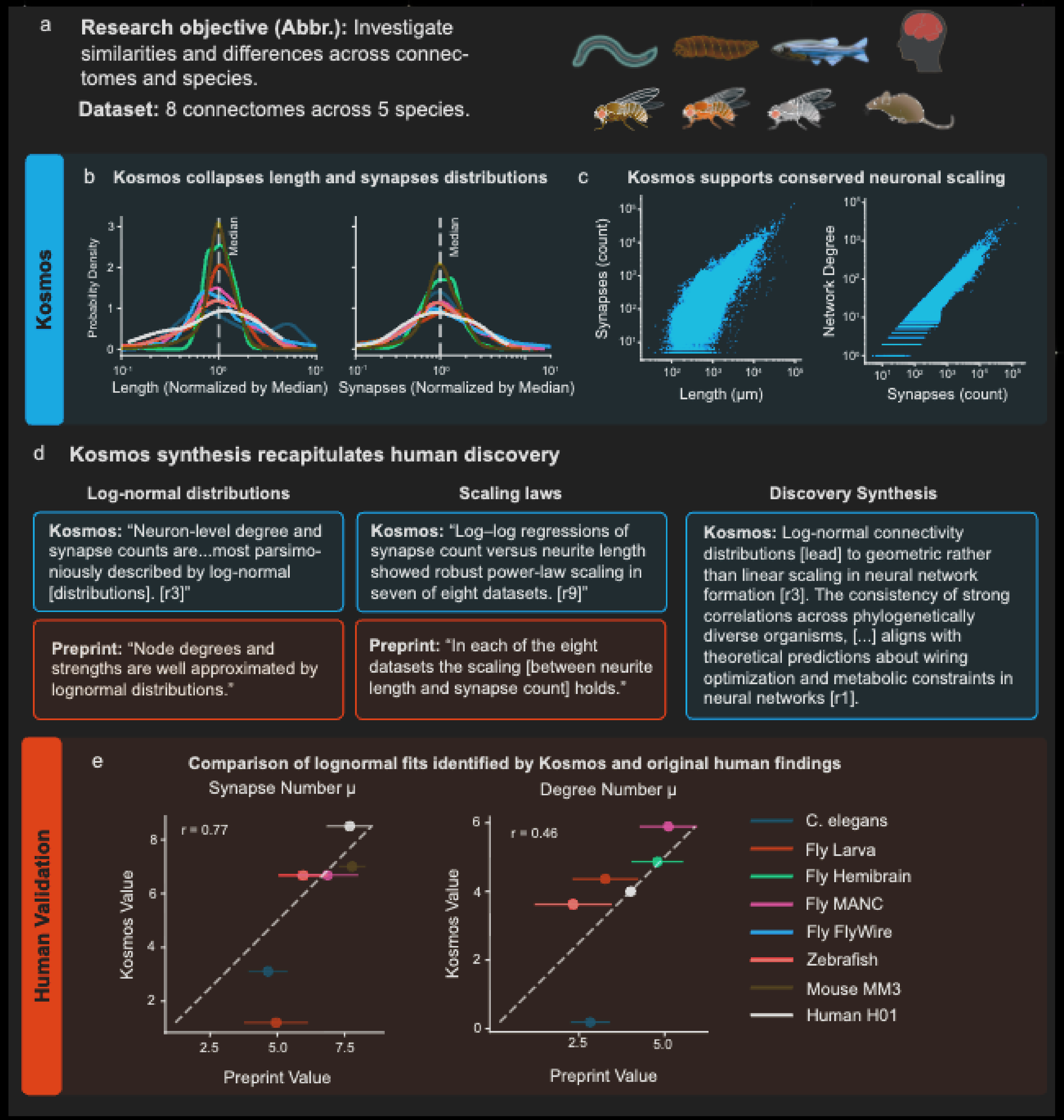


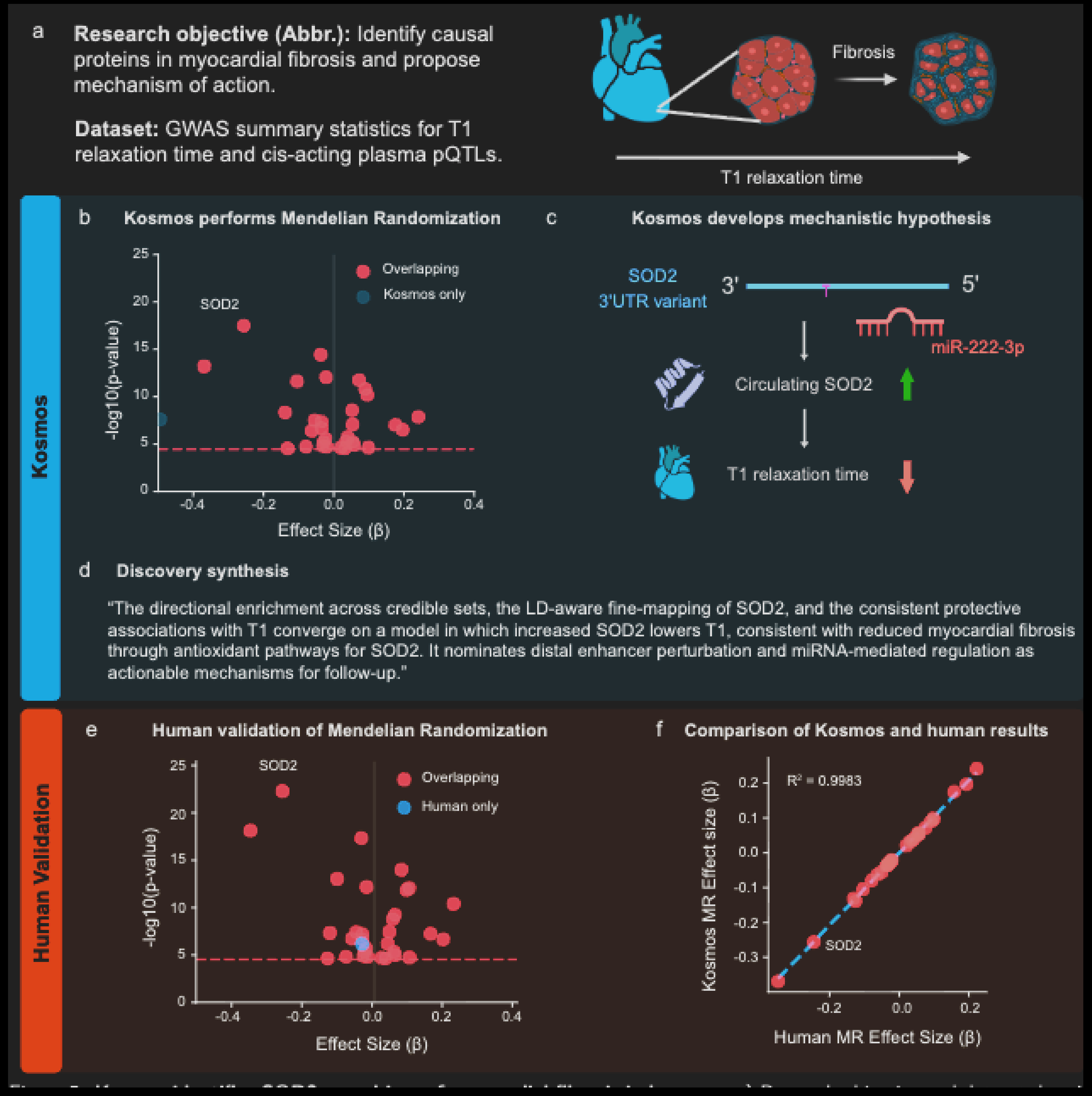


Kosmos: An AI Scientist for Autonomous Discovery

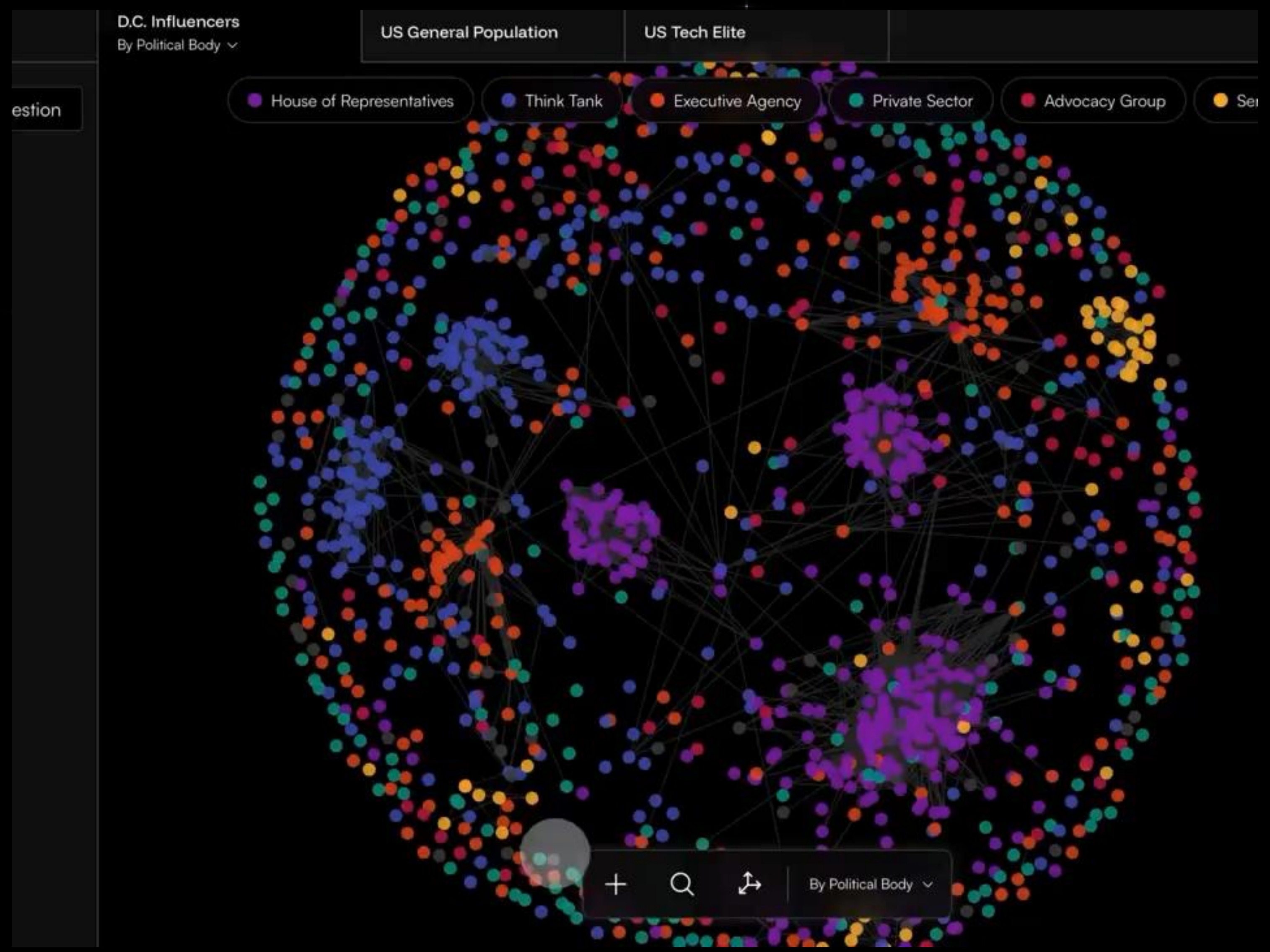
Ludovico Mitchener[*,1,†], Angela Yiu[*,1], Benjamin Chang[*,1,2], Mathieu Bourdenx[3,4,5], Tyler Nadolski[1], Arvis Sulovari[1], Eric C. Landsness[5,6], Dániel L. Barabási[7,8], Siddharth Narayanan[1], Nicky Evans[9], Shriya Reddy[10], Martha Foiani[3,4], Aizad Kamal[6], Leah P. Shriver[11,12,13], Fang Cao[10], Asmamaw T. Wassie[1], Jon M. Laurent[1], Edwin Melville-Green[1], Mayk Caldas[1], Albert Bou[1], Kaleigh F. Roberts[14], Sladjana Zagorac[15], Timothy C. Orr[6], Miranda E. Orr[6,16], Kevin J. Zwezdaryk[17,18,19], Ali E. Ghareeb[1], Laurie McCoy[1], Bruna Gomes[10], Euan A. Ashley[10], Karen E. Duff[3,4,5], Tonio Buonassisi[9,20], Tom Rainforth[2], Randall J. Bateman[5,6], Michael Skarlinski[1], Samuel G. Rodriques[1,7,‡], Michaela M. Hinks[1,†], Andrew D. White[1,7,‡]

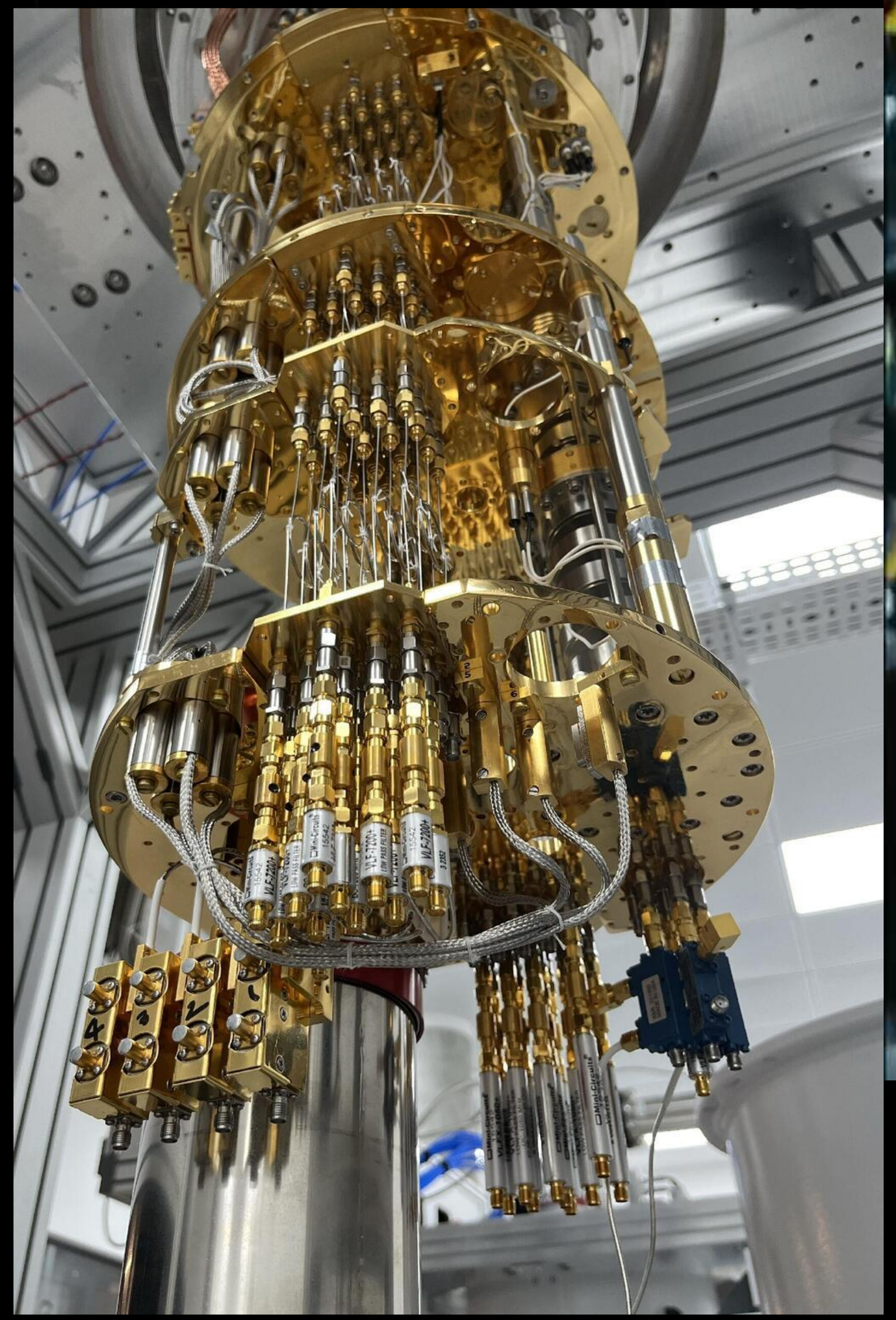


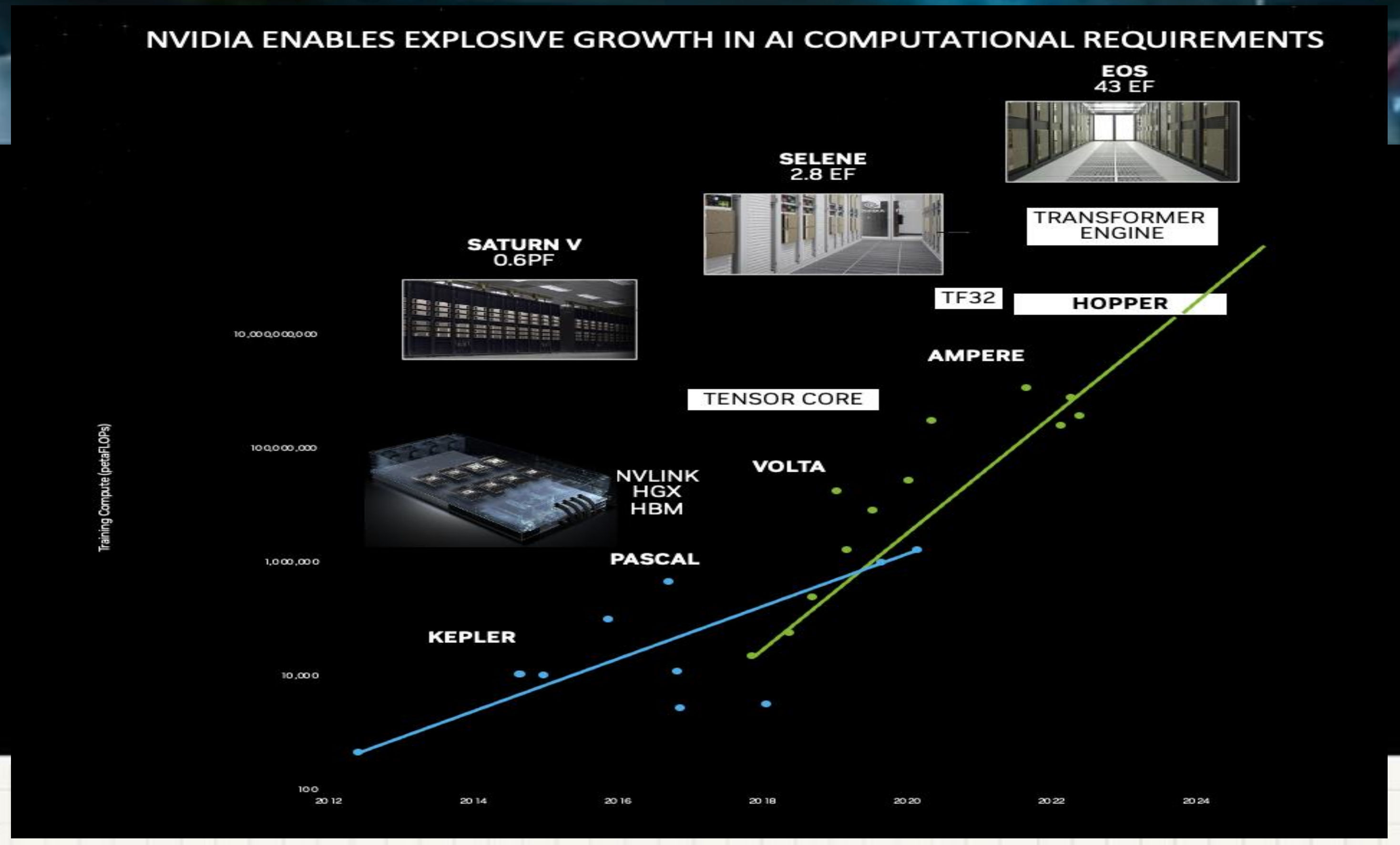


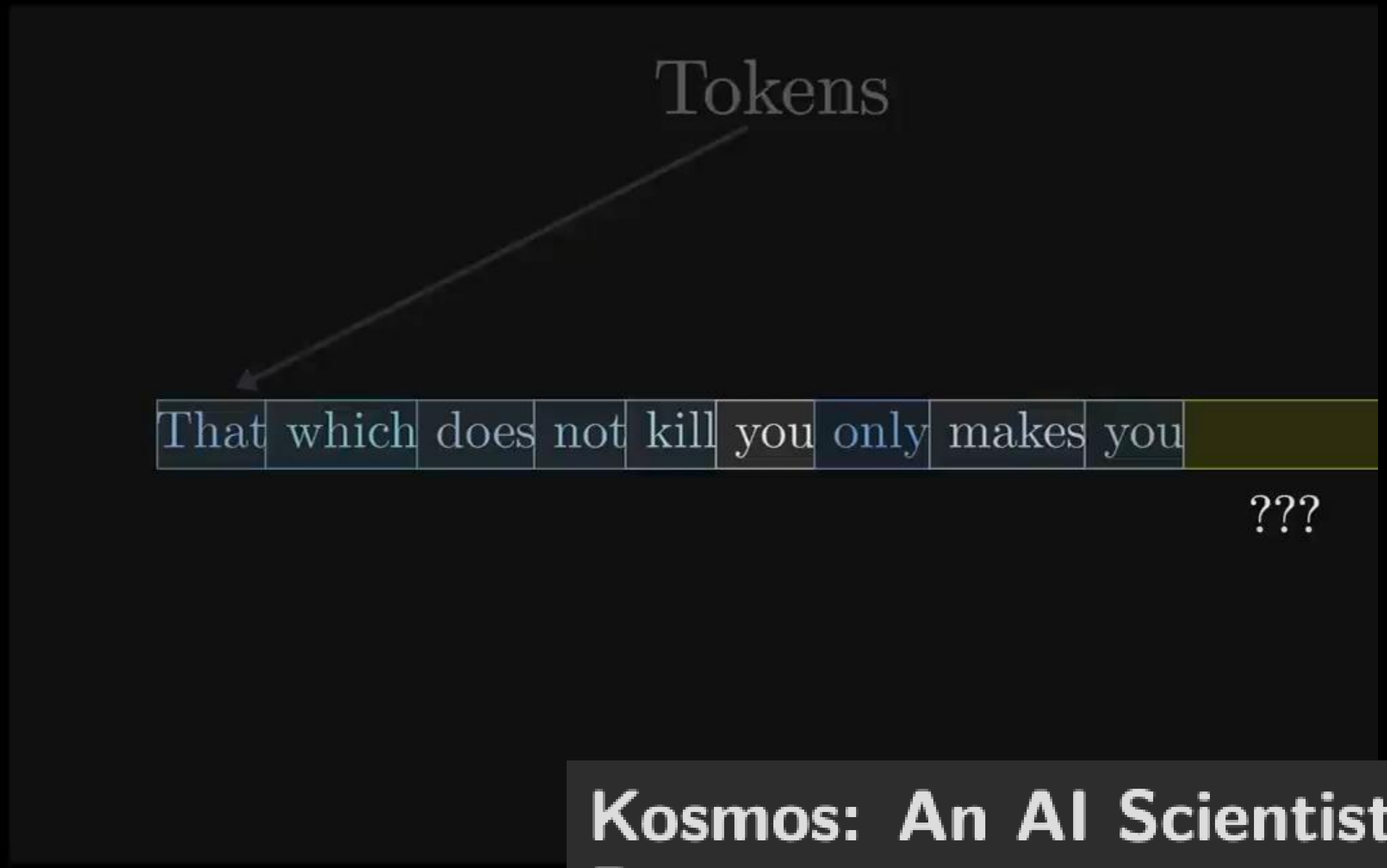


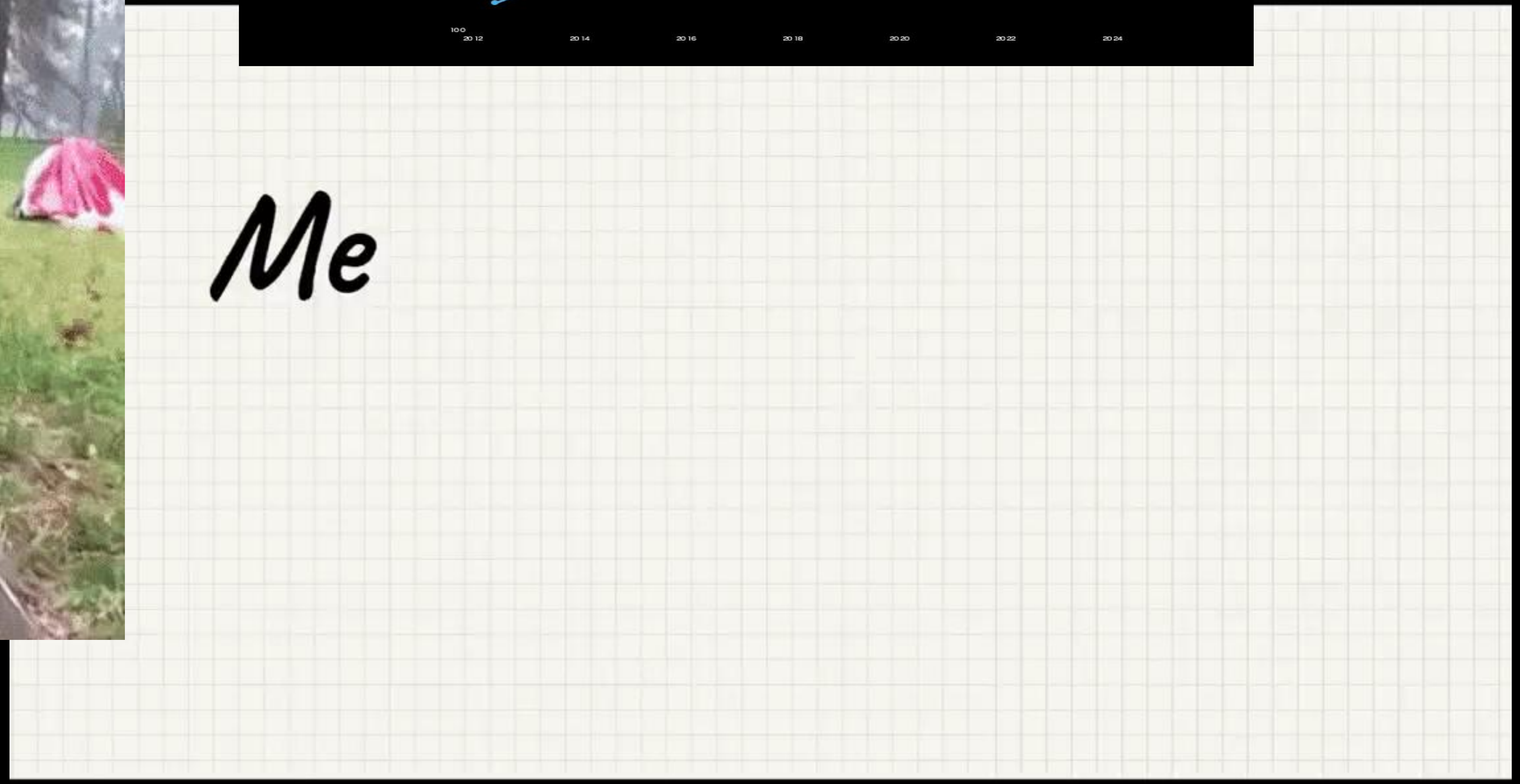

History
New chat
Strykr AI
Your financial market analysis assistant
Suggested questions
What's the forecast for Microsoft stock with AI developments?
How are interest rates affecting bank stocks like JPMorgan?
7/10
Message Strykr...
Strykr.ai may produce inaccurate information. Verify important information.





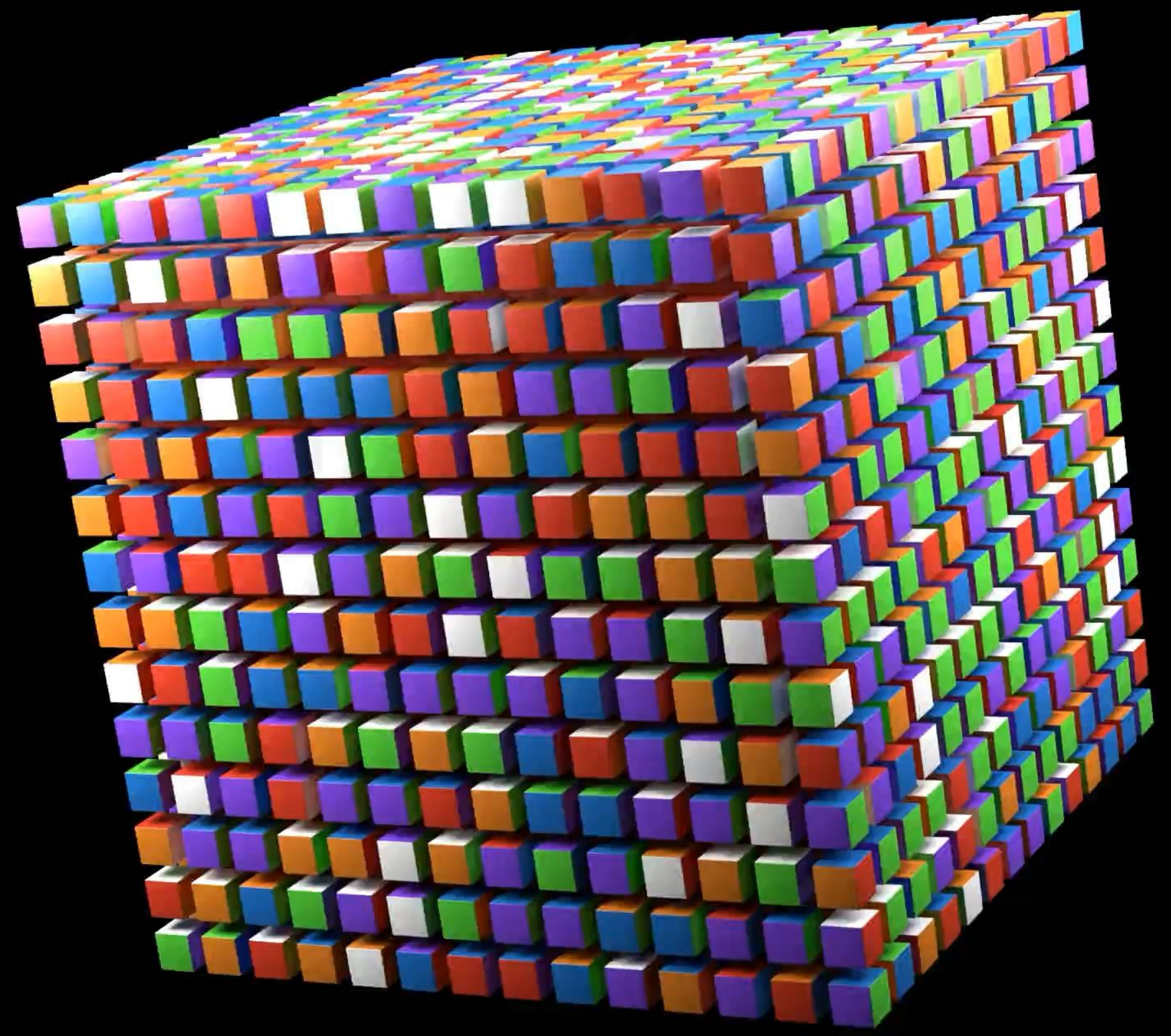


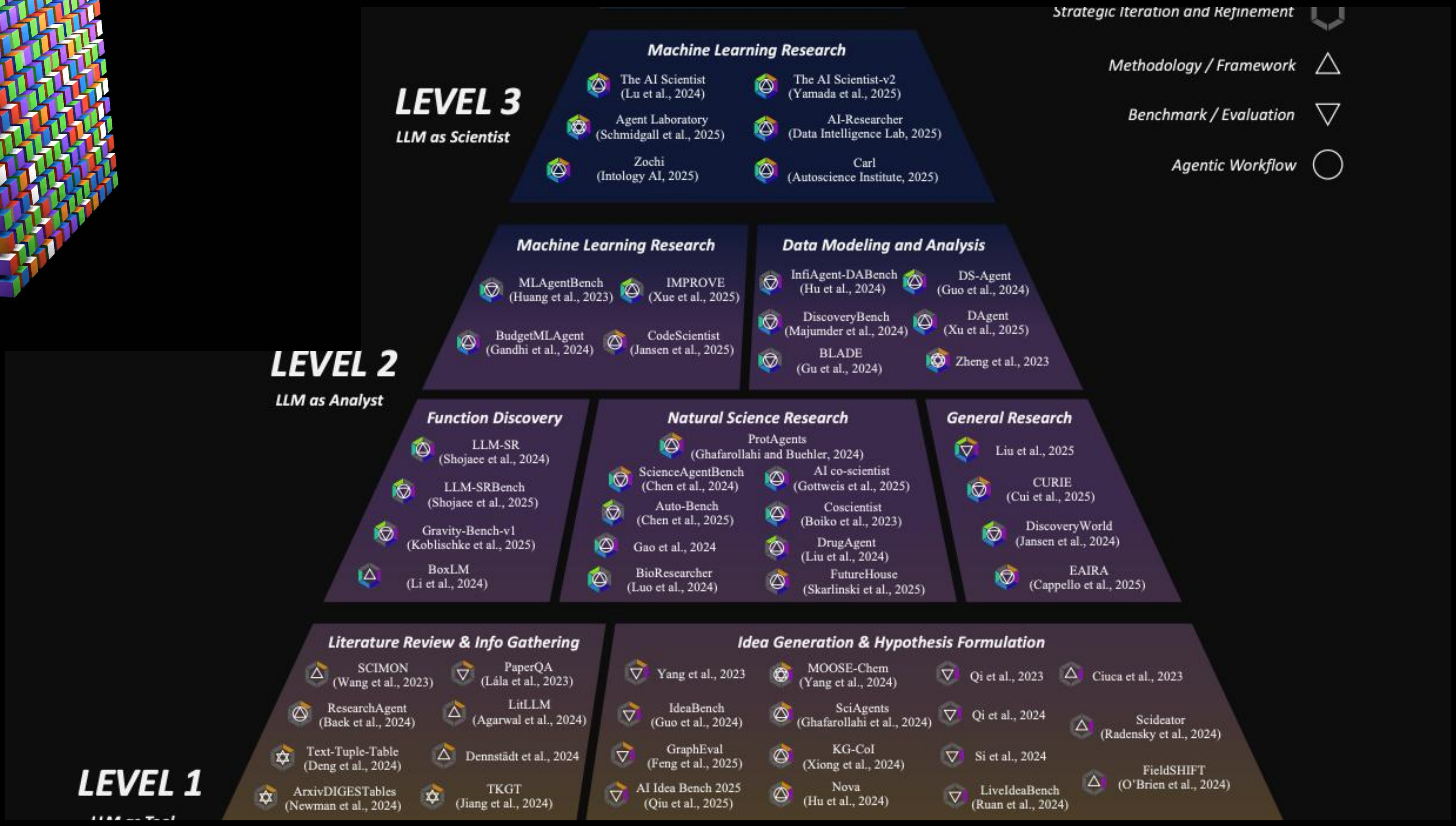




Kosmos: An AI Scientist for Autonomous Discovery

Ludovico Mitchener[*,1,†], Angela Yiu[*,1], Benjamin Chang[*,1,2], Mathieu Bourdenx[3,4,5], Tyler Nadolski[1], Arvis Sulovari[1], Eric C. Landsness[5,6], Dániel L. Barabási[7,8], Siddharth Narayanan[1], Nicky Evans[9], Shriya Reddy[10], Martha Foiani[3,4], Aizad Kamal[6], Leah P. Shriver[11,12,13], Fang Cao[10], Asmamaw T. Wassie[1], Jon M. Laurent[1], Edwin Melville-Green[1], Mayk Caldas[1], Albert Bou[1], Kaleigh F. Roberts[14], Sladjana Zagorac[15], Timothy C. Orr[6], Miranda E. Orr[6,16], Kevin J. Zwezdaryk[17,18,19], Ali E. Ghareeb[1], Laurie McCoy[1], Bruna Gomes[10], Euan A. Ashley[10], Karen E. Duff[3,4,5], Tonio Buonassisi[9,20], Tom Rainforth[2], Randall J. Bateman[5,6], Michael Skarlinski[1], Samuel G. Rodriques[1,7,‡], Michaela M. Hinks[1,†], Andrew D. White[1,7,‡]




Strategic Iteration and Refinement
Methodology / Framework
Benchmark / Evaluation
Agentic Workflow
LEVEL 3
LLM as Scientist
Machine Learning Research
The AI Scientist (Lu et al., 2024)
The AI Scientist-v2 (Yamada et al., 2025)
Agent Laboratory (Schmidgall et al., 2025)
AI-Researcher (Data Intelligence Lab, 2025)
Zochi (Intology AI, 2025)
Carl (Autoscience Institute, 2025)
LEVEL 2
LLM as Analyst
Machine Learning Research
MLAgentBench (Huang et al., 2023)
IMPROVE (Xue et al., 2025)
BudgetMLAgent (Gandhi et al., 2024)
CodeScientist (Jansen et al., 2025)
Data Modeling and Analysis
InfiAgent-DABench (Hu et al., 2024)
DS-Agent (Guo et al., 2024)
DiscoveryBench (Majumder et al., 2024)
DAgent (Xu et al., 2025)
BLADE (Gu et al., 2024)
Zheng et al., 2023
Function Discovery
LLM-SR (Shojaee et al., 2024)
LLM-SRBench (Shojaee et al., 2025)
Gravity-Bench-v1 (Koblischke et al., 2025)
BoxLM (Li et al., 2024)
Natural Science Research
ProtAgents (Ghafarollahi and Buehler, 2024)
ScienceAgentBench (Chen et al., 2024)
AI co-scientist (Gottweis et al., 2025)
Auto-Bench (Chen et al., 2025)
Coscientist (Boiko et al., 2023)
Gao et al., 2024
DrugAgent (Liu et al., 2024)
BioResearcher (Luo et al., 2024)
FutureHouse (Skarlinski et al., 2025)
General Research
Liu et al., 2025
CURIE (Cui et al., 2025)
DiscoveryWorld (Jansen et al., 2024)
EAIRA (Cappello et al., 2025)
LEVEL 1
Literature Review & Info Gathering
SCIMON (Wang et al., 2023)
PaperQA (Lála et al., 2023)
ResearchAgent (Baek et al., 2024)
LitLLM (Agarwal et al., 2024)
Text-Tuple-Table (Deng et al., 2024)
Dennstädt et al., 2024
ArxivDIGESTables (Newman et al., 2024)
TKGT (Jiang et al., 2024)
Idea Generation & Hypothesis Formulation
Yang et al., 2023
MOOSE-Chem (Yang et al., 2024)
Qi et al., 2023
Ciuca et al., 2023
IdeaBench (Guo et al., 2024)
SciAgents (Ghafarollahi et al., 2024)
Qi et al., 2024
Scideator (Radensky et al., 2024)
GraphEval (Feng et al., 2025)
KG-CoI (Xiong et al., 2024)
Si et al., 2024
FieldSHIFT (O'Brien et al., 2024)
AI Idea Bench 2025 (Qiu et al., 2025)
Nova (Hu et al., 2024)
LiveIdeaBench (Ruan et al., 2024)

AI ASSISTANT FOR DISCOVERY

Prof (Dr) Simon See

Prof (Adj), Univ of Newcastle, Coventry, Mahindra, UI, BUPT, SJTU, NTU

Distinguish Fellow, Fudan Uni

And

Head of AI Technology Center, NVIDIA