When AI meets life sciences
当人工智能遇上生命科学

2025 清华大学 AI创新实践课程

张强锋，清华大学

Sep 15, 2025

Forever young, forever healthy

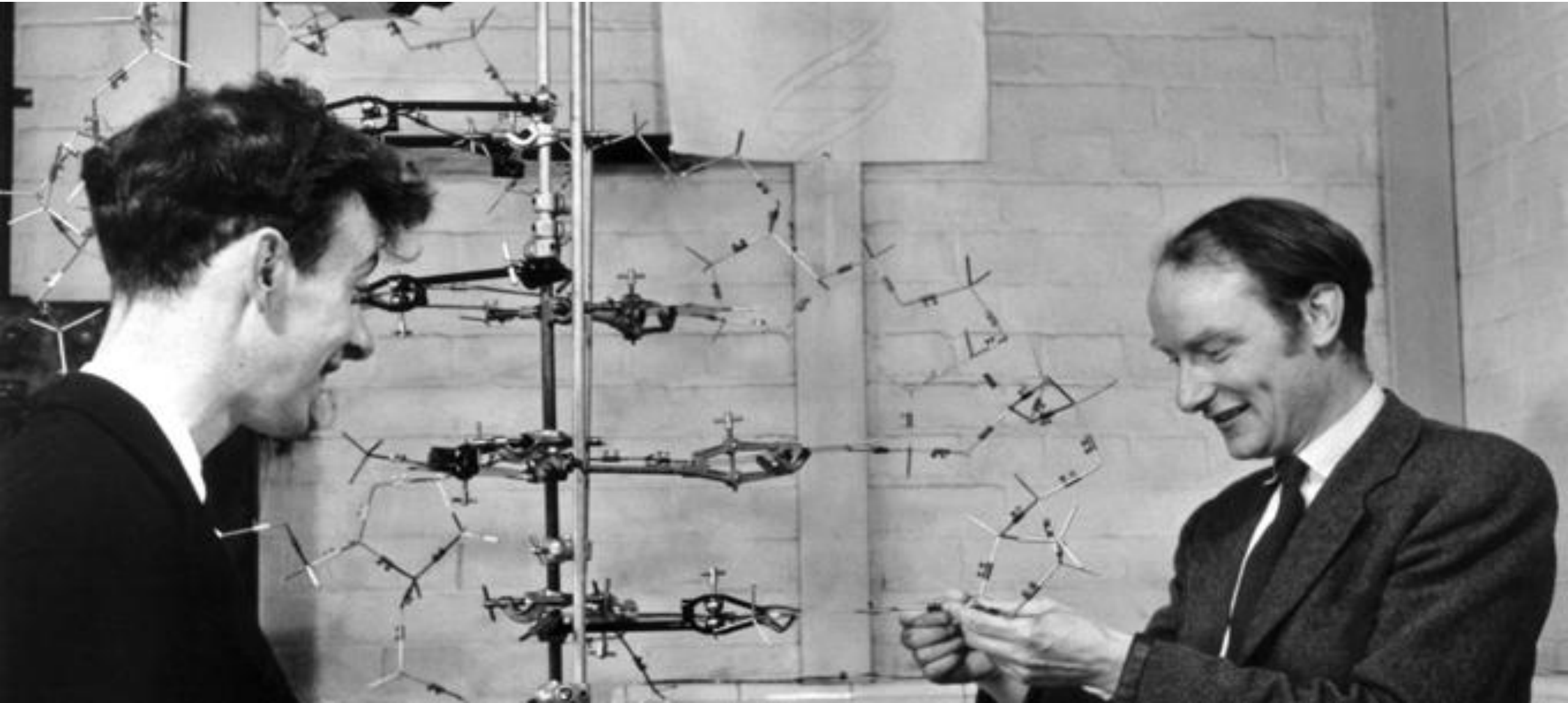


Do not go gentle into that good night. —— Dylan Thomas

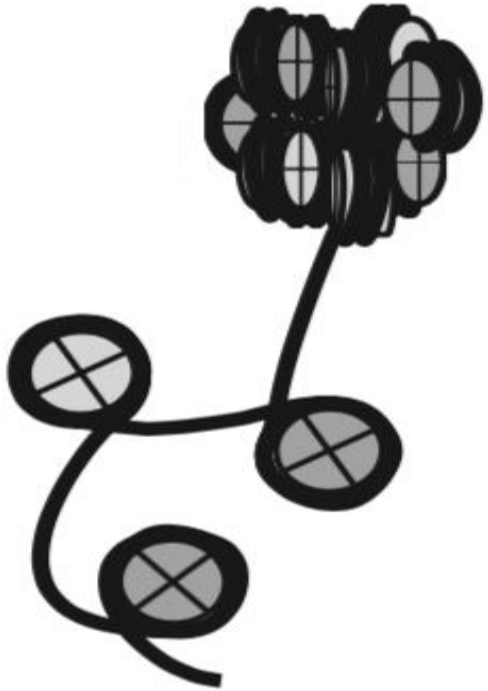
Life tree

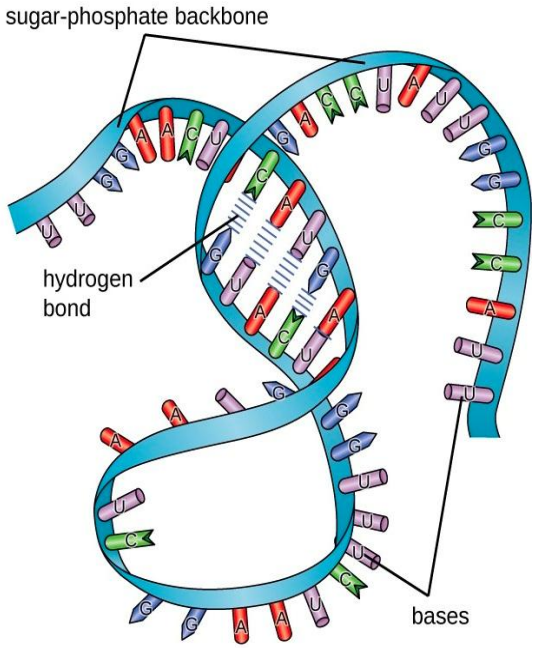


The origins of modern biology



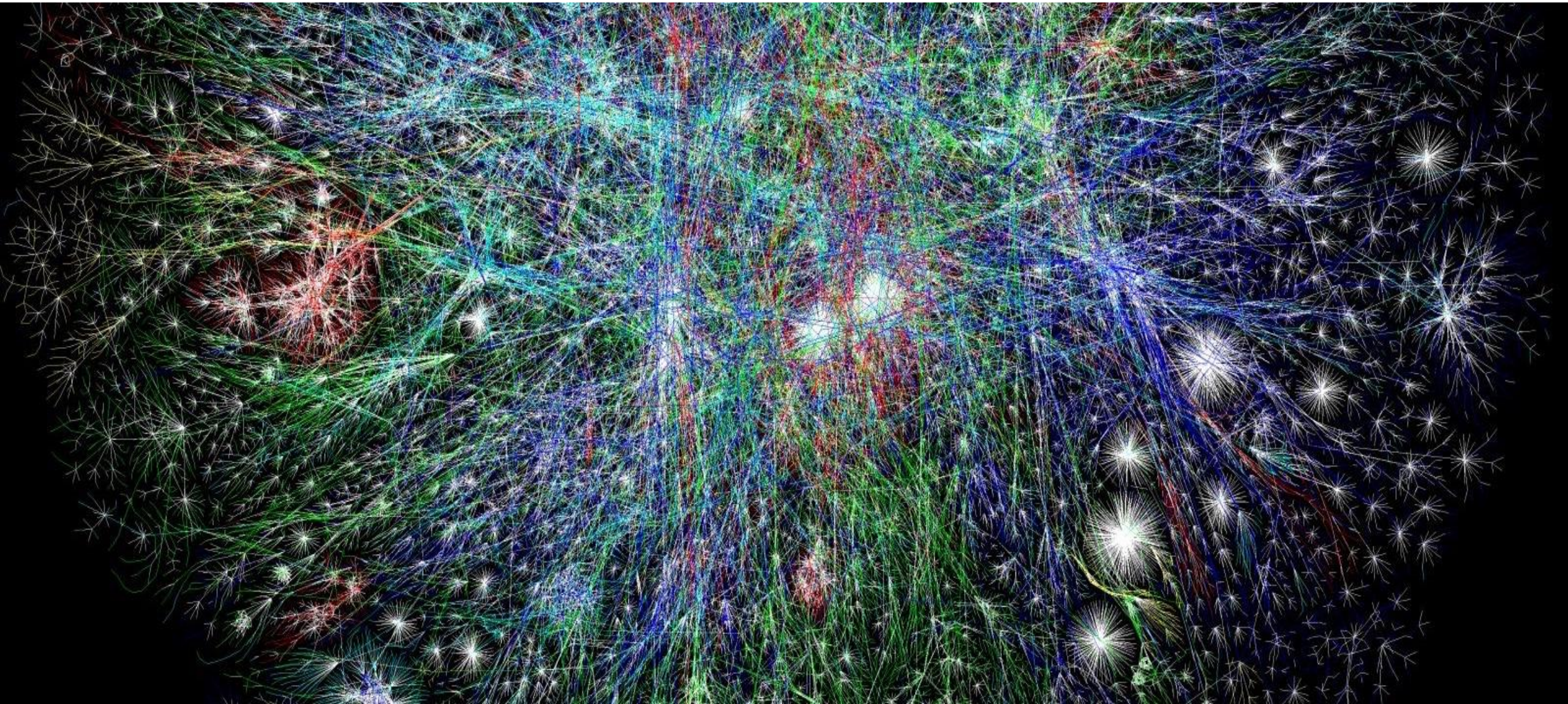
Genetic central dogma



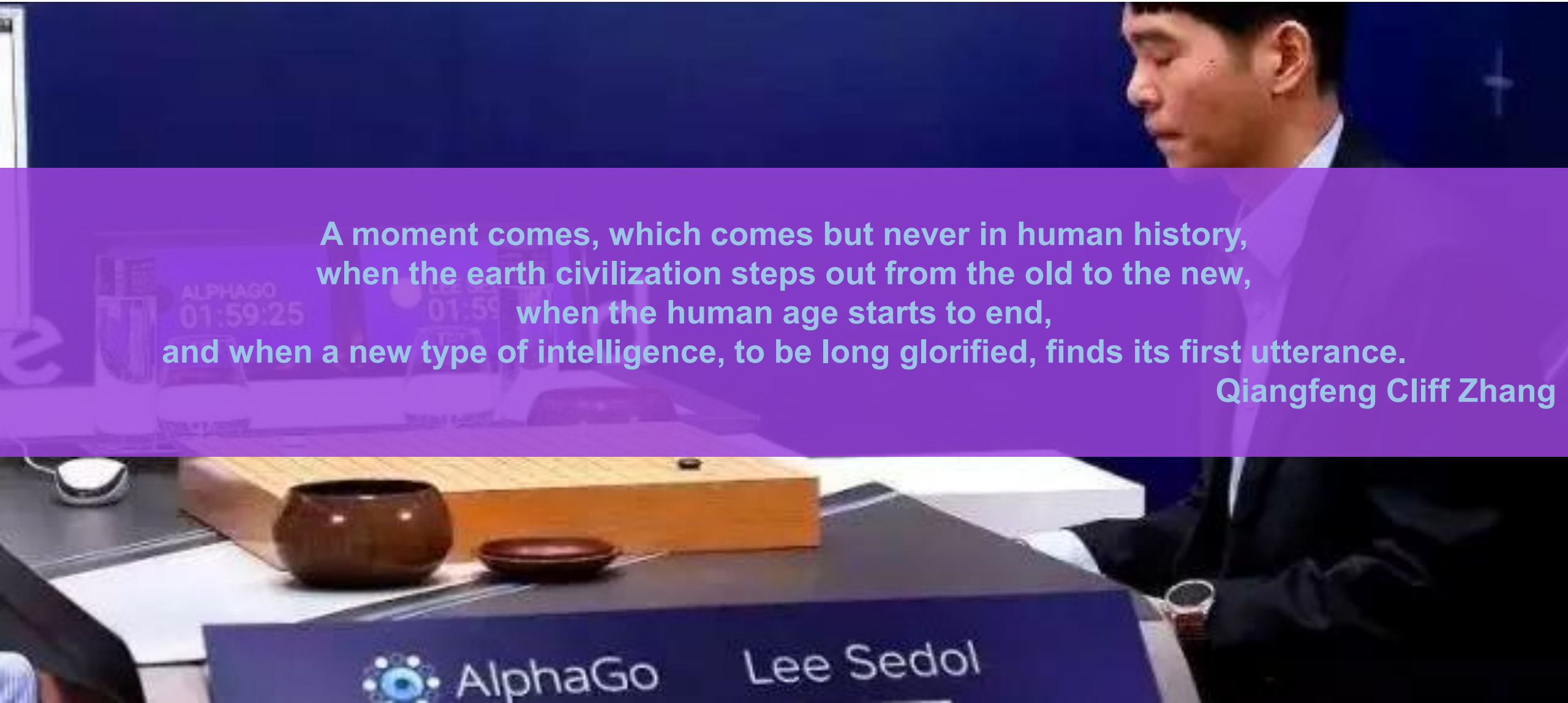




The complexity of modern life sciences



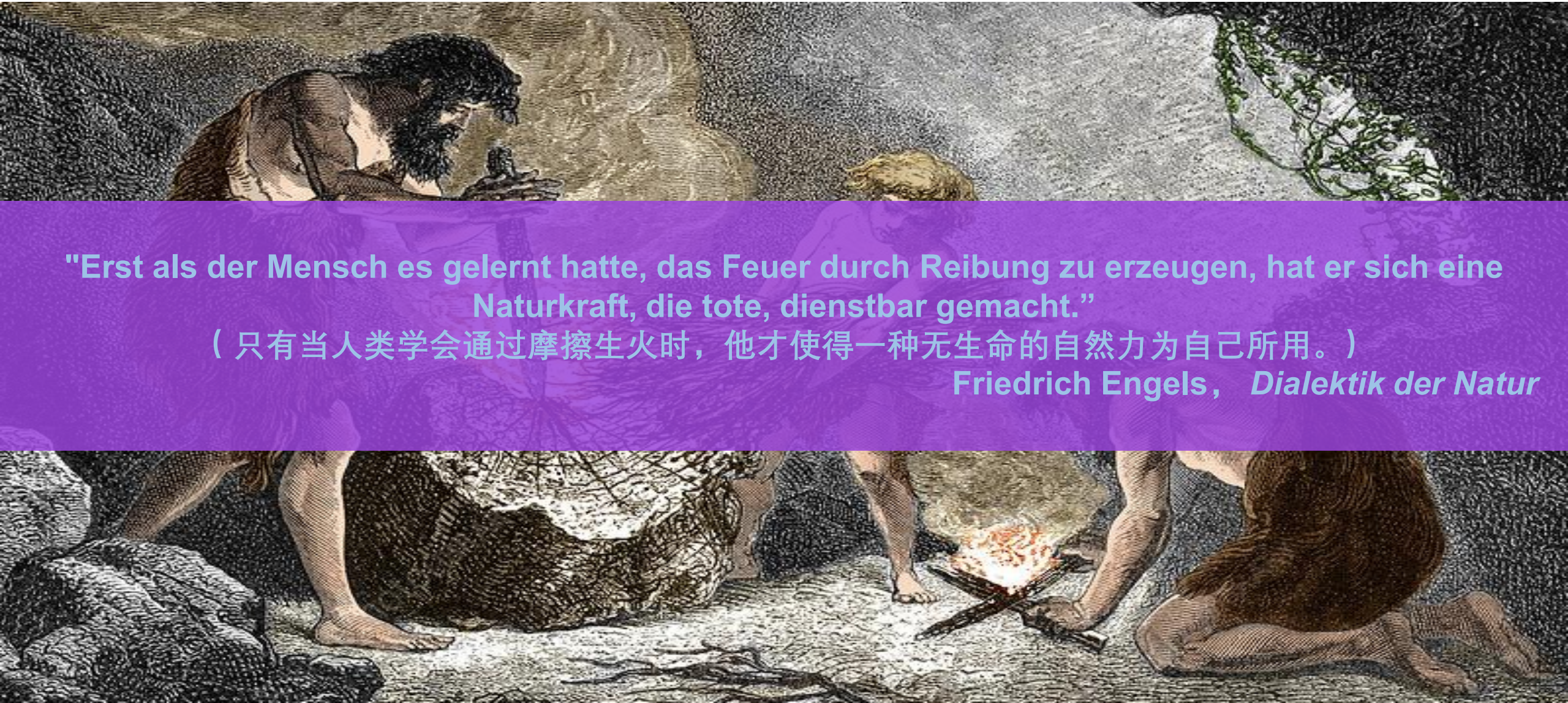
Will artificial intelligence be the last major invention and creation of humanity?



A moment comes, which comes but never in human history,
when the earth civilization steps out from the old to the new,
when the human age starts to end,
and when a new type of intelligence, to be long glorified, finds its first utterance.

Qiangfeng Cliff Zhang

Discovery of fire



"Erst als der Mensch es gelernt hatte, das Feuer durch Reibung zu erzeugen, hat er sich eine Naturkraft, die tote, dienstbar gemacht."

（只有当人类学会通过摩擦生火时，他才使得一种无生命的自然力为自己所用。）

Friedrich Engels， *Dialektik der Natur*

Breakthroughs in artificial intelligence (AI)

Unlike most technological innovations of the past, artificial intelligence represents one of the few truly transformative technological revolutions in human history.



Turing proposed **turing test** on BBC

1950



Dartmouth meeting, **definition of AI**

1956



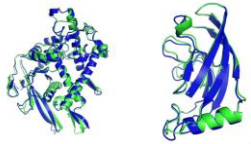
Deep Blue defeated Garry Kasparov

1997



Transformer, Revolutionary Natural Language Processing

2017



AI for science: **AlphaFold2** breakthrough on protein Structure Prediction

2020



Large language model **ChatGPT** has achieved human-like conversation capabilities

2022



Nobel Prize in Physics
Neural Network



Nobel Prize in Chemistry
protein Structure Prediction and Design

2024

2025

2012



AlexNet achieved 1st in ImageNet competition using CNN model

2016



AlphaGo defeated Lee Sedol

2019



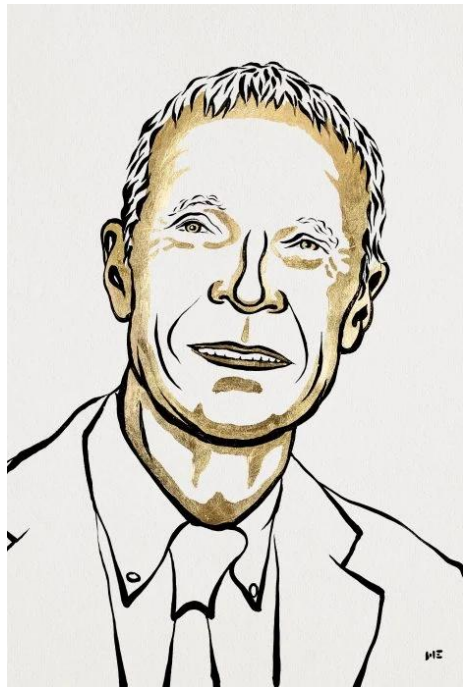
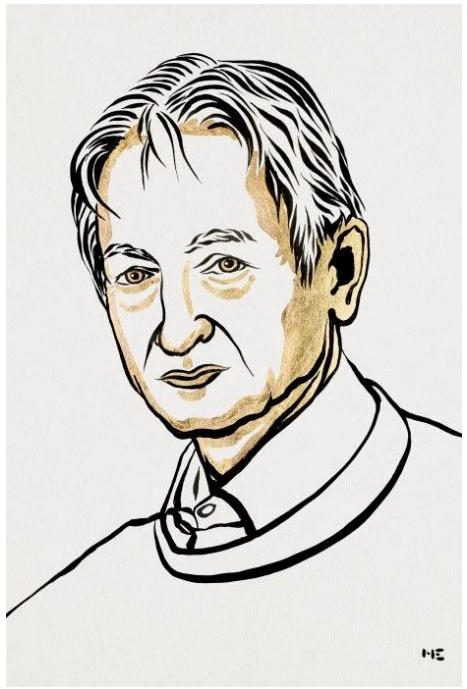
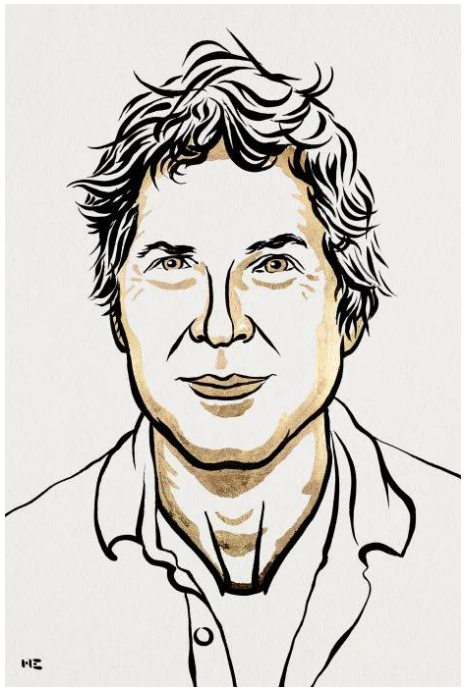
Hinton, LeCun and Bengio won The **Turing Award**

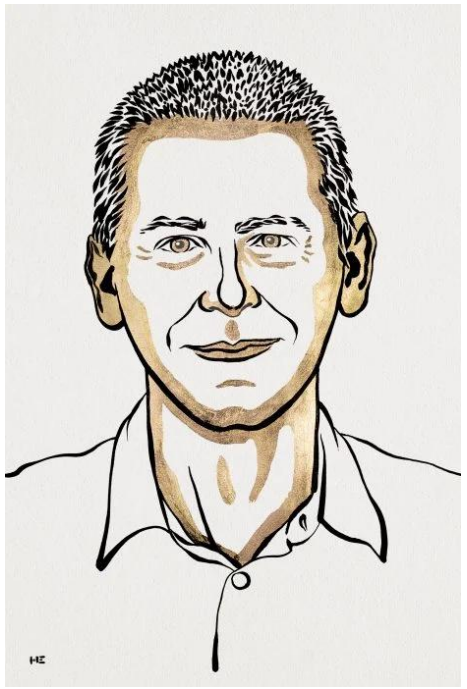
deepseek

domestic AI technology milestone

AI Moore's Law: Major breakthroughs in artificial intelligence occur approximately every two years.
—— Qiangfeng Cliff Zhang

The Nobel Prize's "year of AI"

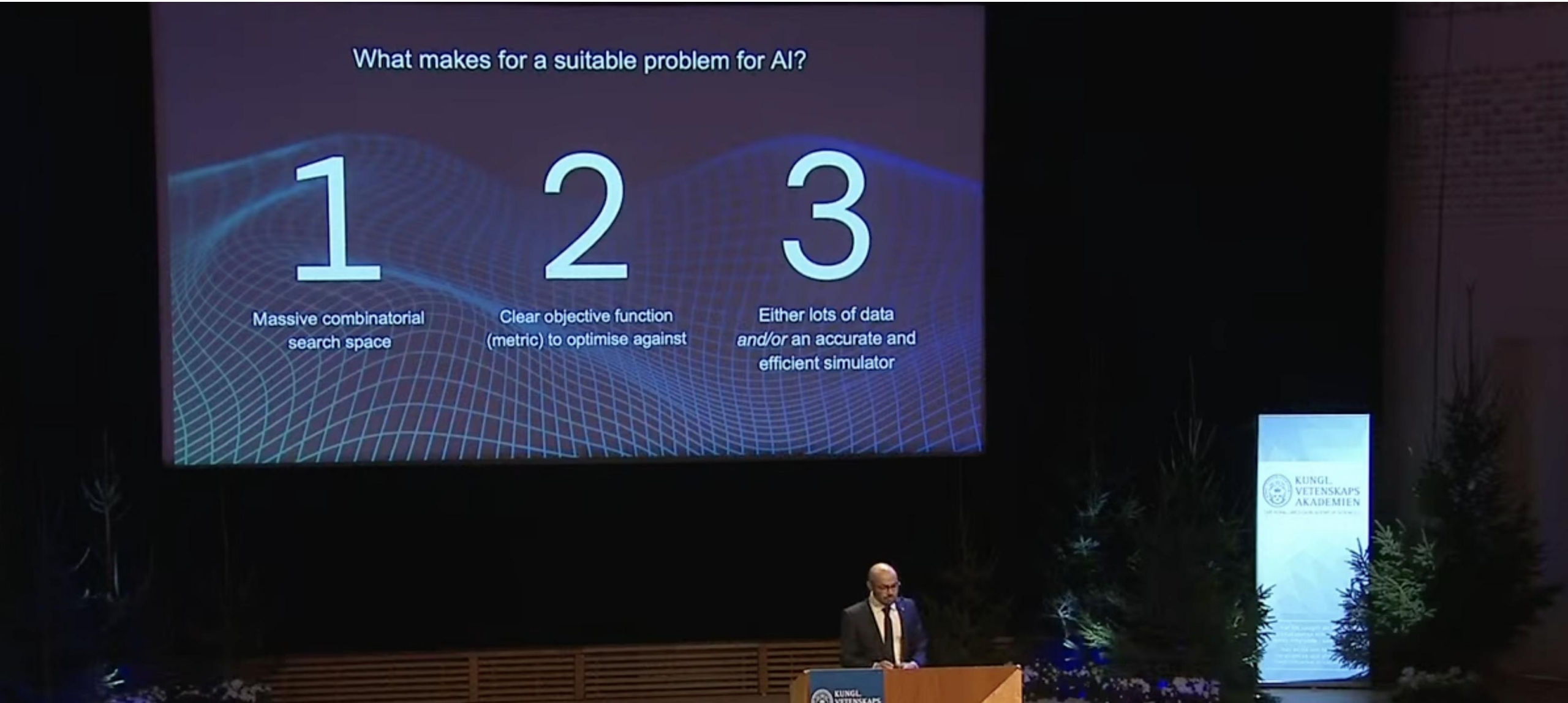






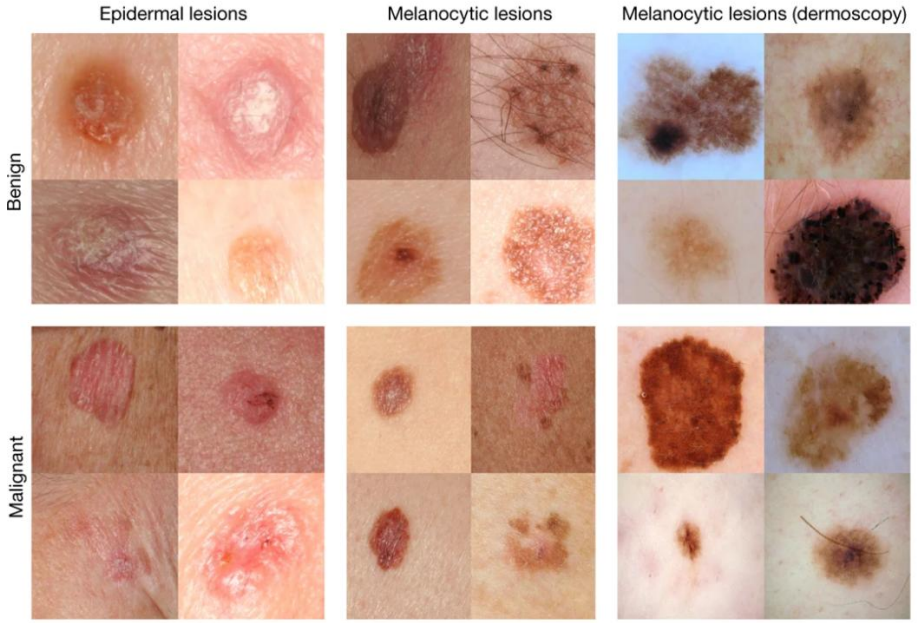
"For foundational discoveries and inventions that enable machine learning with artificial neural networks"

One half awarded "for computational protein design", the other half "for protein structure prediction"

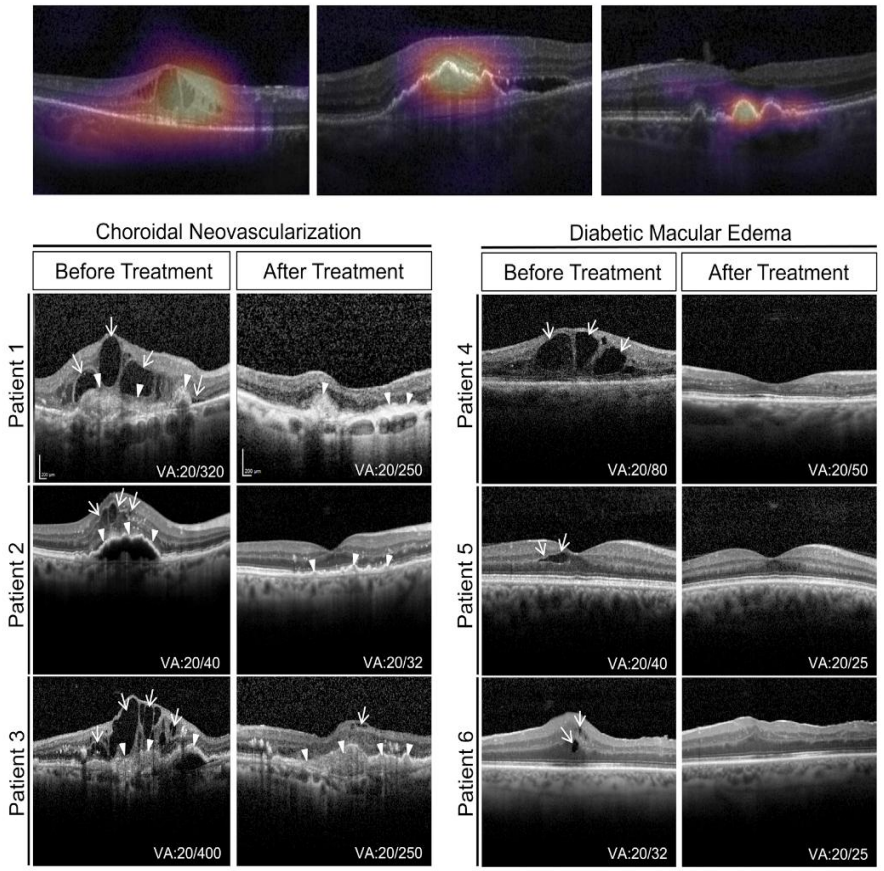
Why protein structure prediction?



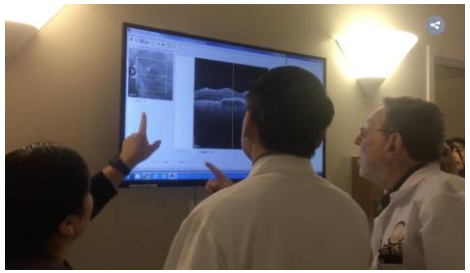
The earliest successful application of AI in biomedicine: medical image processing



Based on a dataset of 130,000 skin disease images, an AI model was trained to diagnose potential skin cancer, achieving expert-level accuracy.
Esteva *et al.*, *Nature* 2017



By leveraging transfer learning, AI models trained on real-world images can be adapted for medical imaging, achieving performance on par with human experts across multiple medical image diagnosis tasks.
Kermany *et al.*, *Cell* 2018



LLMs in 2024 Gaokao

GPT-4o ranked first with a score of 562. The domestic large-scale model Doubao performed exceptionally well, achieving a high score of 542.5 — surpassing the Henan liberal arts first-tier cutoff score by over 20 points.



2024 Gaokao New Curriculum I Roll + Shandong Roll Full Subject Test

Rank	LLMs	Math	Chinese	English	Physics	Chemistry	Biology	History	Geography	Politics	Science Overall	Arts Overall
01	GPT-4o	**66**	120	**139**	51	42	51.5	81	**68**	**88**	**469.5**	**562**
02	Doubao	61. 5	125.5	131	42.5	**49.5**	56. 5	**82.5**	62	80	**466.5**	**542. 5**
03	Wenxin Yiyan 4.0	62.5	119	137.5	**54.5**	40	**65**	78	61. 5	79	**478.5**	**537.5**
04	Baixiaoying	44	**128**	**139**	24.5	47.5	56	72	55	83	**439**	**521**
05	Qwen	35	111	131.5	18	37	62	82	44	75	**394.5**	**478.5**
06	Kimi	39	100	127	32	34	41	72.5	58.5	65	**373**	**462**
07	Tencent Yuanbao	39	120.5	118	27.5	36	47	73	39	70	**388**	**459.5**
08	Minimax	38.5	104.5	127	39	36.5	46.5	67.5	53.5	63	**392**	**454**
09	ChatGLM	37	102.5	134.5	13.5	24	50.5	60.5	39	64	**362**	**437.5**

521
the Henan liberal arts first-tier cutoff score

428
the Henan liberal arts second-tier cutoff score

From Natural Language to Life Language - Taking Proteins as an Example

Sequence

Primary

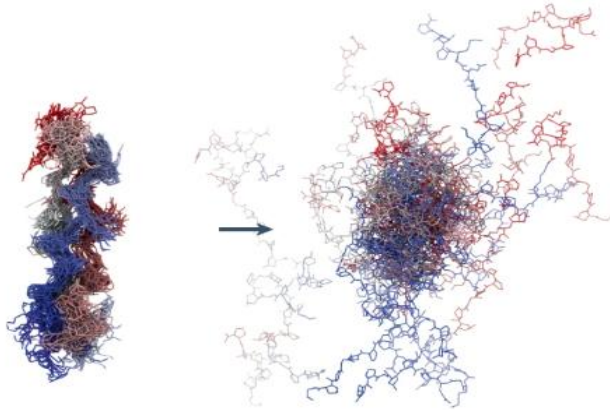
MKLPIYTRF
GVNCSWQA
AIPLMN…

Amino acids

ABCDEFGHIJKLMNO
PQRSTUVWXYZ

Alphabet

Your manuscript is nowt publishable



Secondary structure

Secondary

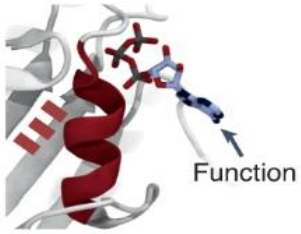


Protein

Jane bicycle ride

Tertiary structure

Tertiary



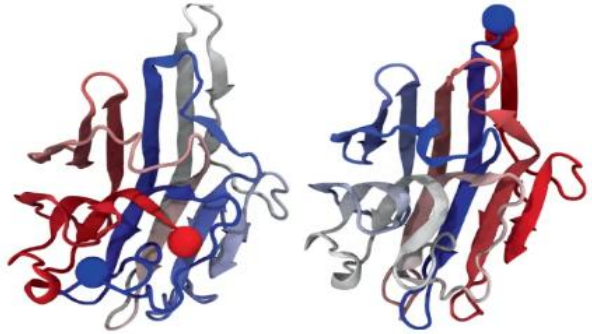
Function

Jane rides her bicycle



Meaning

I am happy when it rains. When it rains, I am happy.



Quaternary structure

Quarternary



Jane rides her bicycle to school. But always with a big bright light and a helmet …

Colourless green ideas sleep furiously.



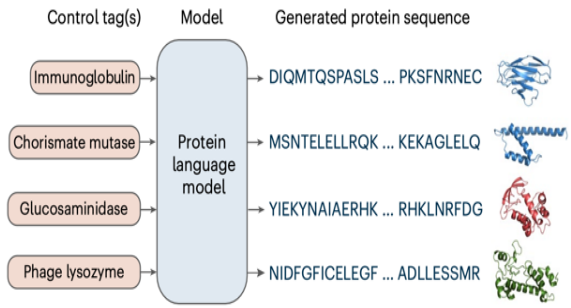
Amyloid fibril

Generative large language models enable the design of novel proteins

1 ProGen

Trained on 280 million protein sequences, the model learns to predict the next amino acid and has acquired the capability to design highly active novel proteins based on given labels.

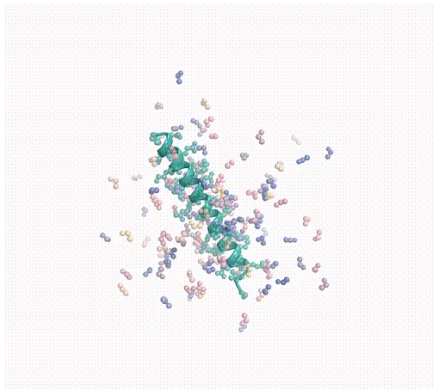
Madani *et al.*, *Nat. Biotechnol.* 2023



2 RFDiffusion

By integrating RoseTTAFold and diffusion models, we achieved de novo design of diverse functional proteins, including topological structures never observed in natural proteins.

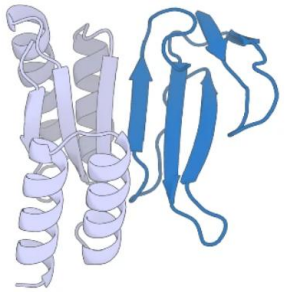
Watson *et al.*, *Nature* 2023



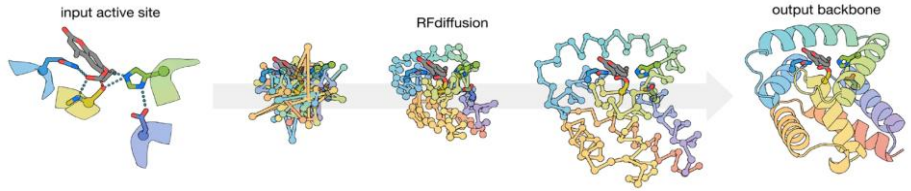
3 AI-Assisted Design of Highly Efficient Antivenom Proteins

Utilizing AI, α-neurotoxin binders were designed to prevent toxin binding to the neuromuscular nicotinic acetylcholine receptor through steric hindrance, achieving complete protective efficacy.

Torres *et al. Nature* 2025



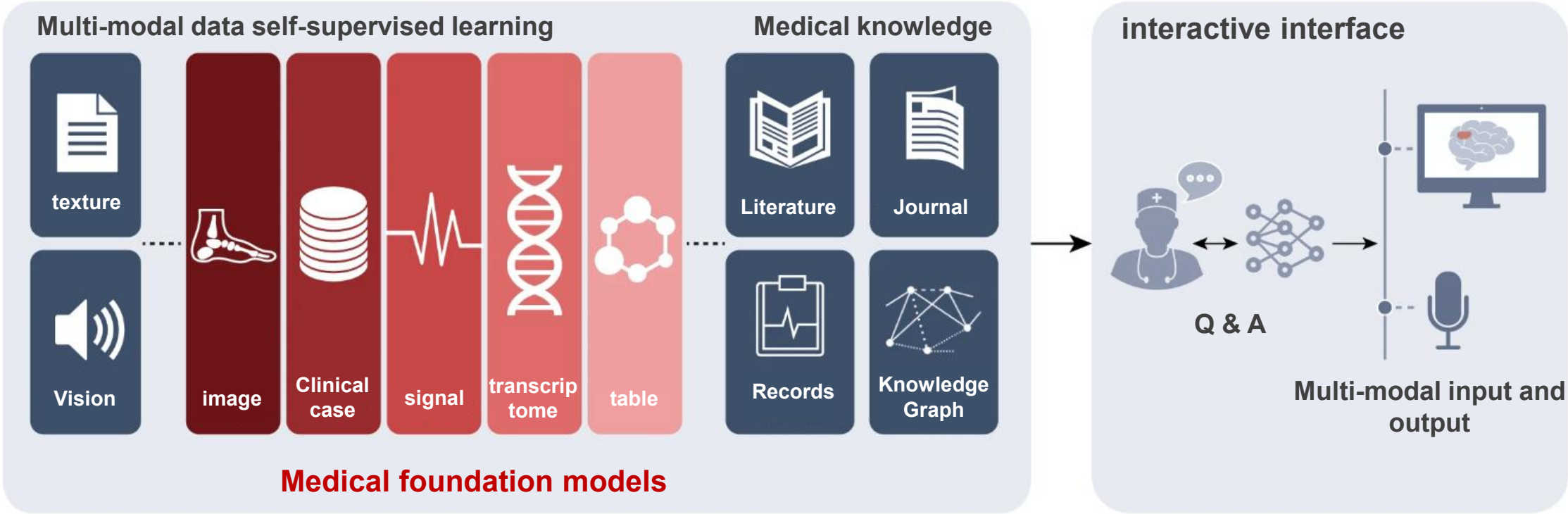
4 AI-Assisted Design of Highly Efficient Catalytic Enzymes



Leveraging AI and the geometric features of protein serine hydrolase active sites, a hydrolase was designed that not only catalyzes a four-step chemical reaction but also exhibits a catalytic efficiency 60,000 times higher than previous designs.

Lauko *et al. Science* 2025

Foundation models for medical AI are set to revolutionize medical practice



Applications

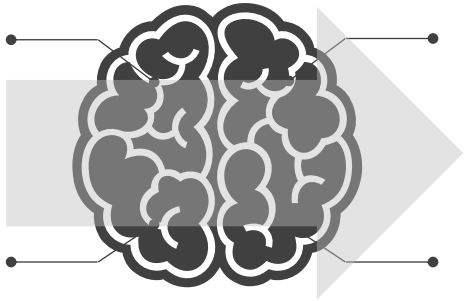
- Diagnose robot
- Interactive Notes
- Technical enhancement
- . . .
- Radiation report
- Text - protein generation
- clinical decision support

AI-powered innovative drug discovery: a long and challenging journey ahead

Disadvantage of traditional drug discovery:

- Long time (10-15 years)
- High cost (billions of dollars)
- Low success rate (<10%)

Scenarios where AI accelerates drug discovery



AI-powered drug discovery:

- Shortening development timelines
- Reducing costs
- Improving success rates

Insilico Medicine

Insilico Medicine designed **INS018_055** using AI, for the **treatment of idiopathic pulmonary fibrosis**. The entire process—from target discovery to preclinical trials—took only **18 months**, and significantly reduced drug development costs to **$2.6 million**。The drug has now entered Phase II clinical trials.

Exscientia

AI–driven drug discovery company Exscientia designed **EXS21546**, an A2A receptor antagonist, for **cancer immunotherapy.** The entire process from initial design to clinical trials took less than **12 months**。The drug has now entered Phase I clinical trials.

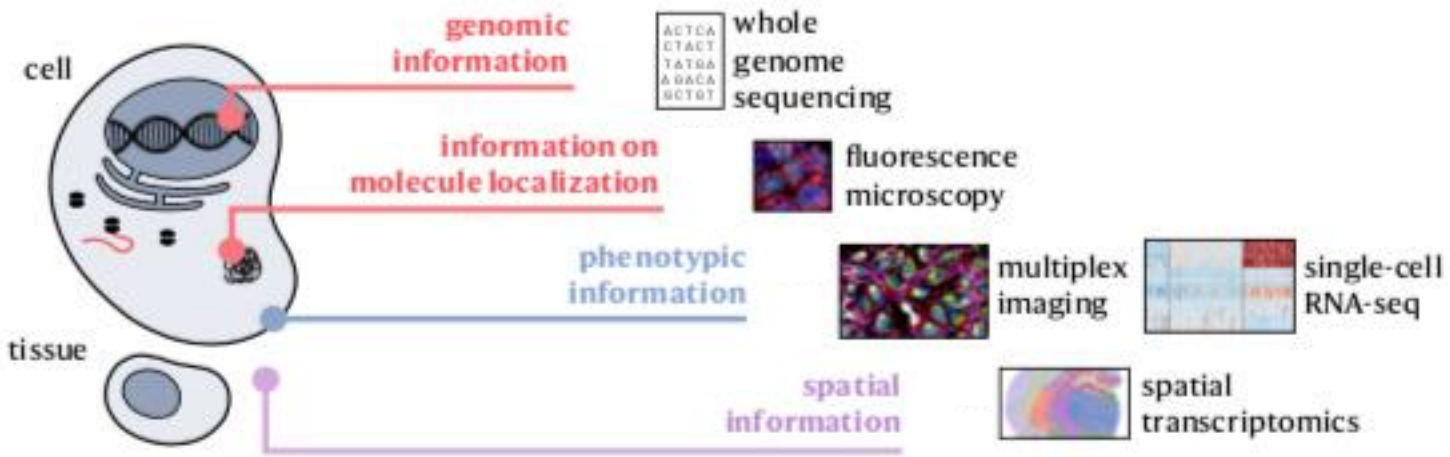
"Achieving AI-driven drug discovery (AIDD) would require **six AlphaFolds**."

Demis Hassabis
CEO and co-founder of DeepMind
Nobel Laureate 2024



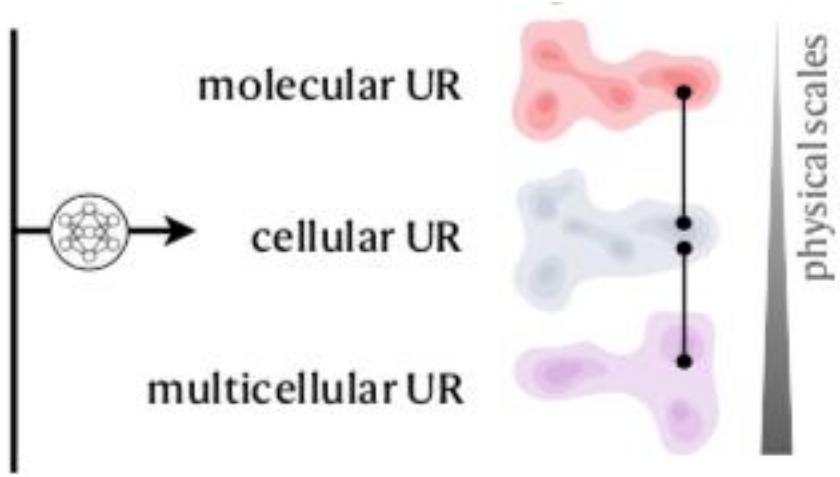
AI-based virtual cell

Multi-scale and multimodal data and knowledge from functional genome and cell biology experiments



AI-based model for virtual cell

Universal Representation



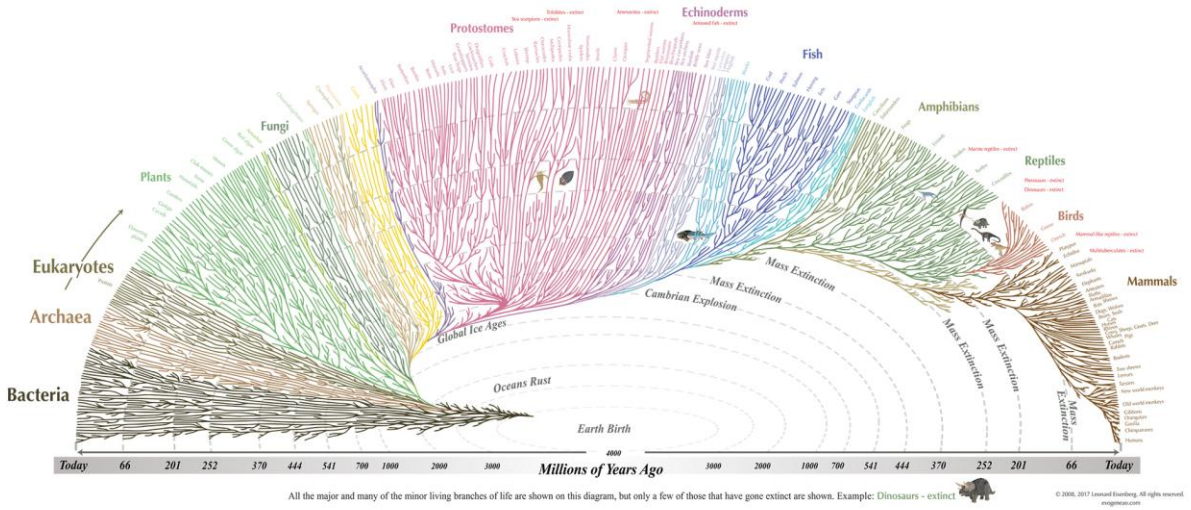
Based on **multi-scale, multimodal** functional genomic and cell biological experimental data and knowledge, **AI technology** is leverage to construct virtual cell models. This approach advances a comprehensive understanding of cellular function and disease mechanisms, accelerates drug development, and promotes personalized medicine.

Decoding the ultimate code of life — genomic foundation models

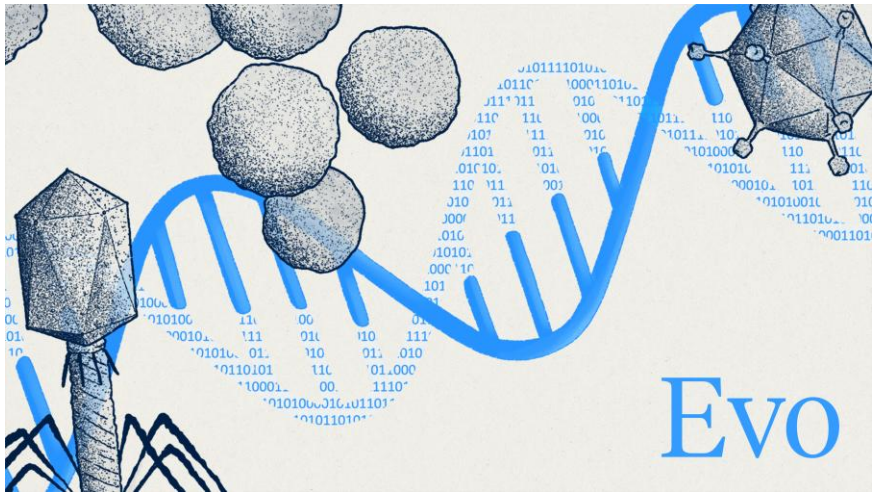
Science 125 questions: Exploration and discovery

1. Why are some genomes so big and others very small?
2. What genes make us uniquely human?
3. How many species are there on Earth?
4. How do organisms evolve?

Life tree



Genomic foundation models



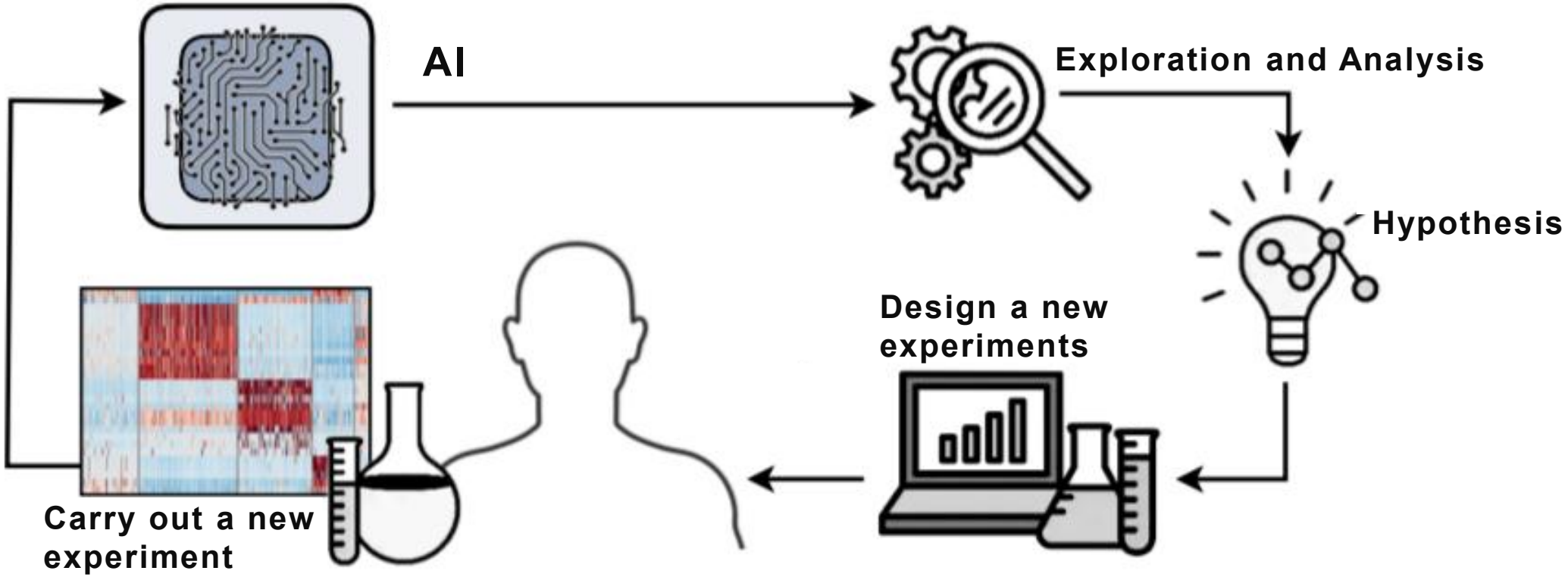
Based on billions of gene sequences, this approach infers the operational mechanisms of bacterial and viral genomes and leverages this information to design novel proteins - or even entire microbial genomes - demonstrating superior performance in generative gene editing systems.

Nguyen *et al*. *Science* 2024

"History began when humans invented gods—and will end when humans play gods… "

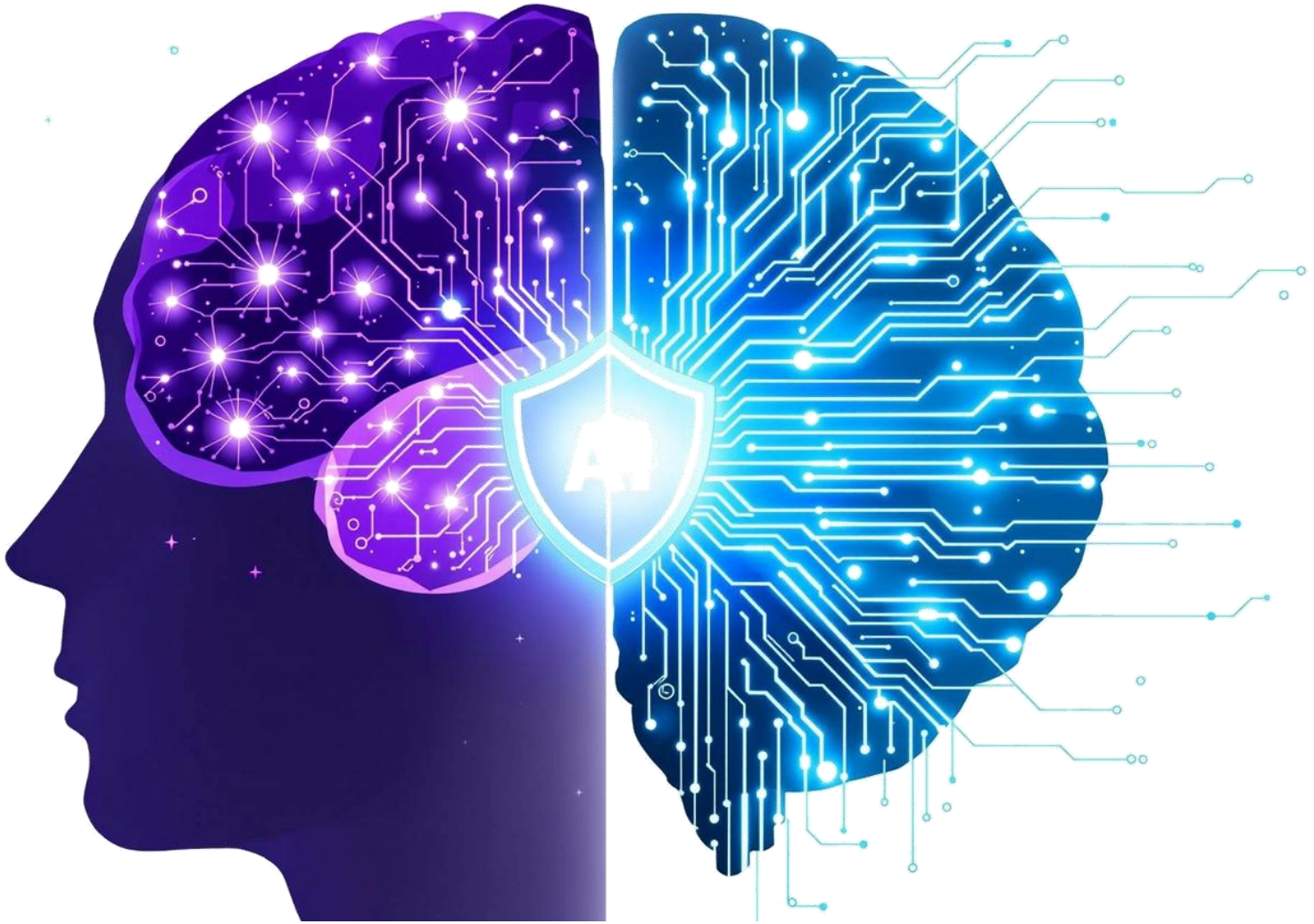
Yuval Noah Harari, *A brief history of humankind*

AI is revolutionizing the paradigm of scientific research



Using AI to build a hypothesis generation framework for scientific research, exploring a vast number of possible hypotheses.

Future?



Artificial Intelligence-Driven Life Science Research

Brain-Inspired High-Efficiency, Explainable, and General Artificial Intelligence

Course outline

- **Foundation Module: Fundamentals of Artificial Intelligence and Life Sciences**
  - Artificial intelligence introduction (week 1-4, team up)
  - Review major fields in life sciences (week 5)
  - Intersection logic of AI and biological (week 6)

- **Application Module: The AI-Driven Revolution in Life Sciences**
  - Gene and Genome Scale (week 7-9, midterm)
  - Structural Biology and Drug Design (week 10-12)
  - Cellular and Medical Applications (week 13-14)

- **Practical Module: From Data to Solutions (week 15-16, final)**

Instructor & TA

Instructor



Qiangfeng Cliff Zhang (张强锋)
Principal Investigator
qczhang@tsinghua.edu.cn
Lab website: https://zhanglab.net/

Teaching assistants



Jiasheng Zhang (张佳胜)
PhD Student
zjs21@mails.tsinghua.edu.cn



Kunting Mu(牟坤汀)
PhD Student
mkt22@mails.tsinghua.edu.cn

Assignments and project

- **Grading: attendance (20%) + assignments (50% + 10%) + final project (30%)**
- **Assignment:**

  Assign homework at the end of each chapter (maybe N=5).

  Combine with the final hands-on exercise in class to answer questions or write reports. (coding would get more score)

- **Final project:**

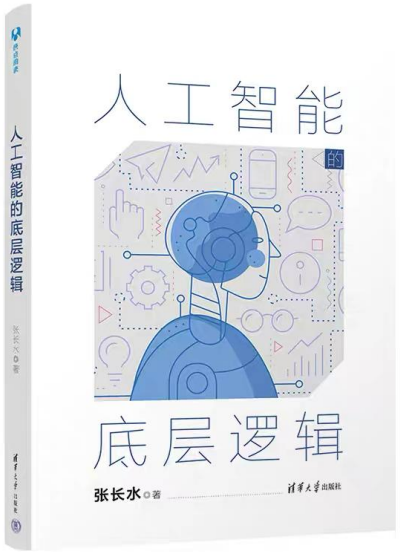
  2~5 students / group

  Work together to finish the mini project, and apply the AI-related knowledge learned in class to solve a biological problem.

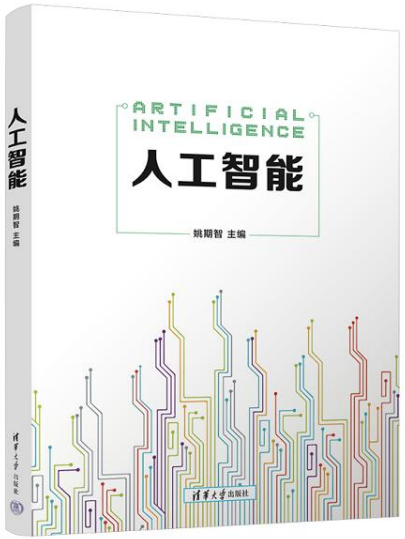
  - **Literature reproduction**
  - **Kaggle or other competitions**
  - **Custom mini project (welcome to discuss with TAs)**

  presentation is on the last two week

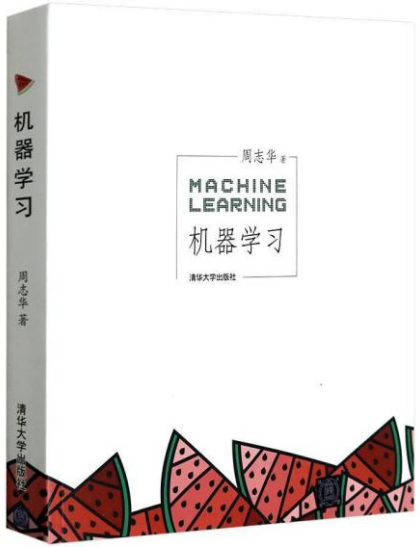
Reference books



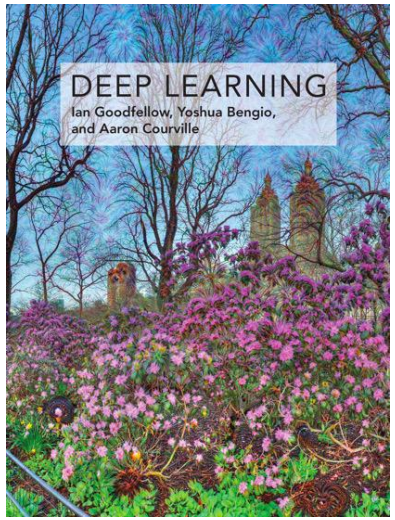
Artificial Intelligence: Underlying Logic



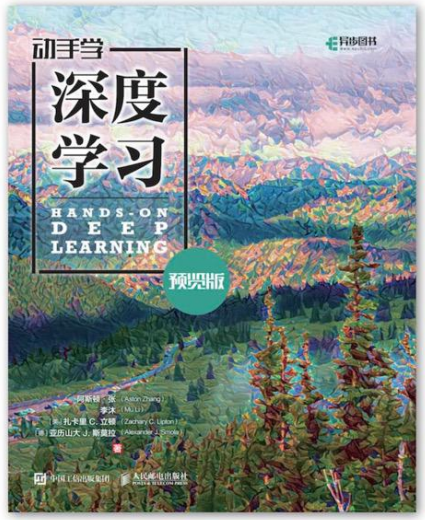
Artificial Intelligence



Machine Learning



Deep Learning



Dive into Deep Learning
https://zh.d2l.ai/