

人工智能概览

马滢青（北京大学）

yqma@pku.edu.cn

2025 年“微扰量子场论及其应用”前沿讲习班暨前沿研讨会，
山东大学，2025/07/09



北京大学



I. 疯狂的人工智能

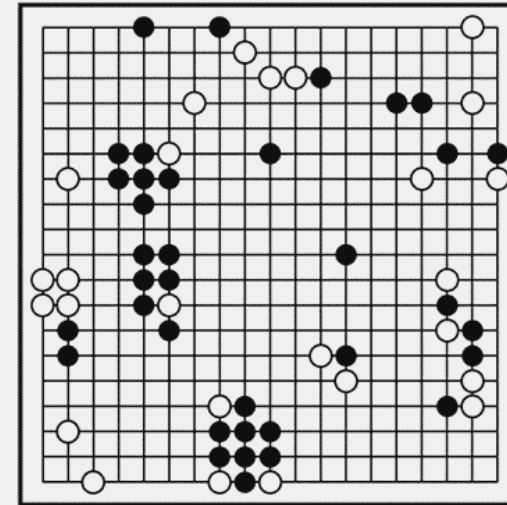
II. 什么是人工智能

III. 联结主义

IV. 展望



1997年， IBM深蓝战胜世界冠军



$2^{361} \approx 4 * 10^{108}$ ，巨大的搜索空间

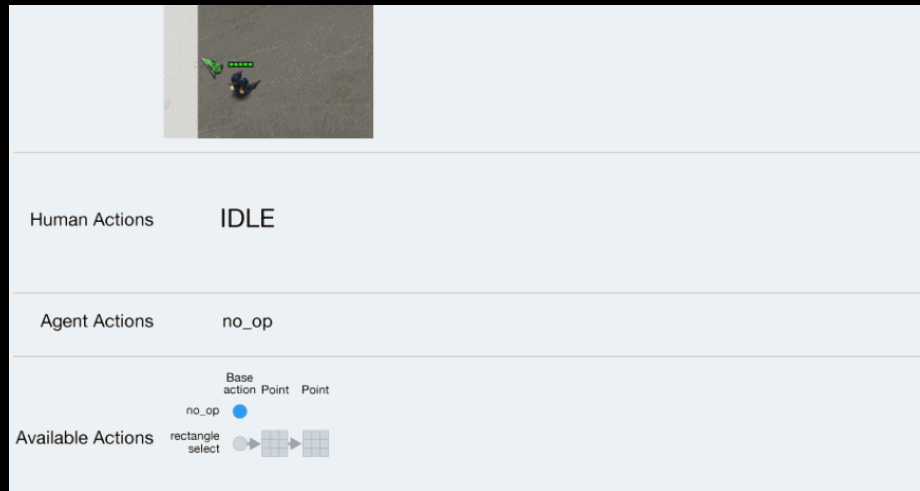


2016年，AlphaGo战胜世界冠军，蒙特卡洛树搜索+深度学习

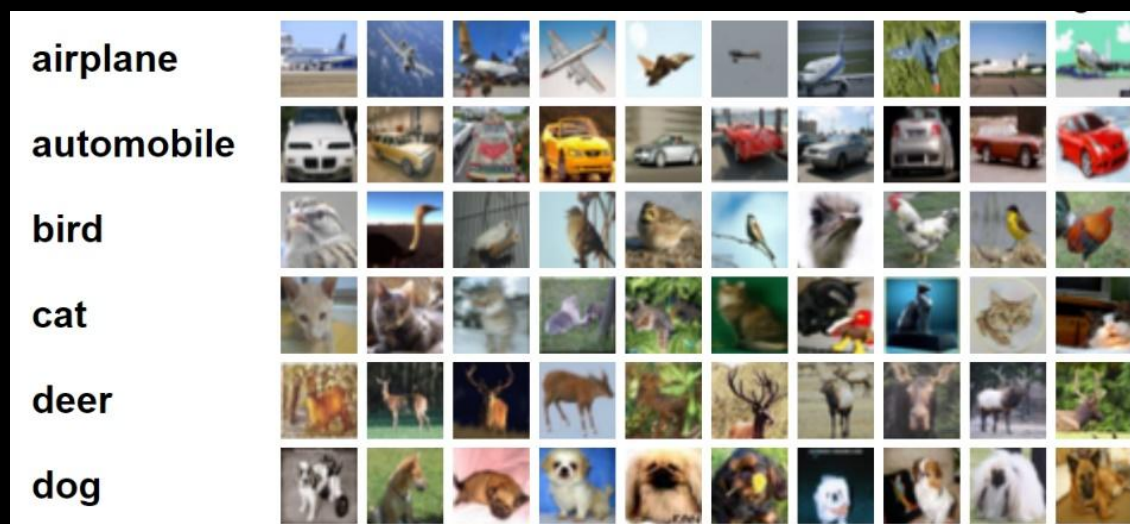
人工智能——多智能体



5/43



2019, AlphaStar人机大战以10:1战胜职业选手

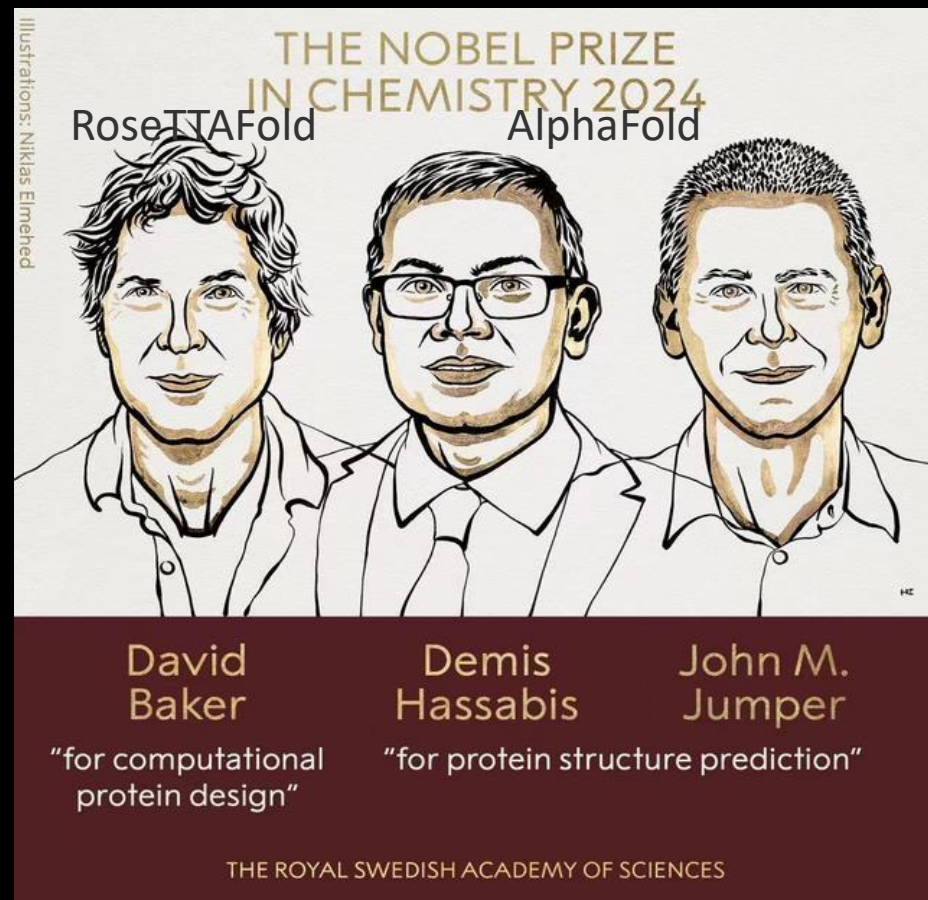


2009年, ImageNet数据集引发了计算机视觉领域的一场革命,
当前很多情况下AI的识别正确率超过人类

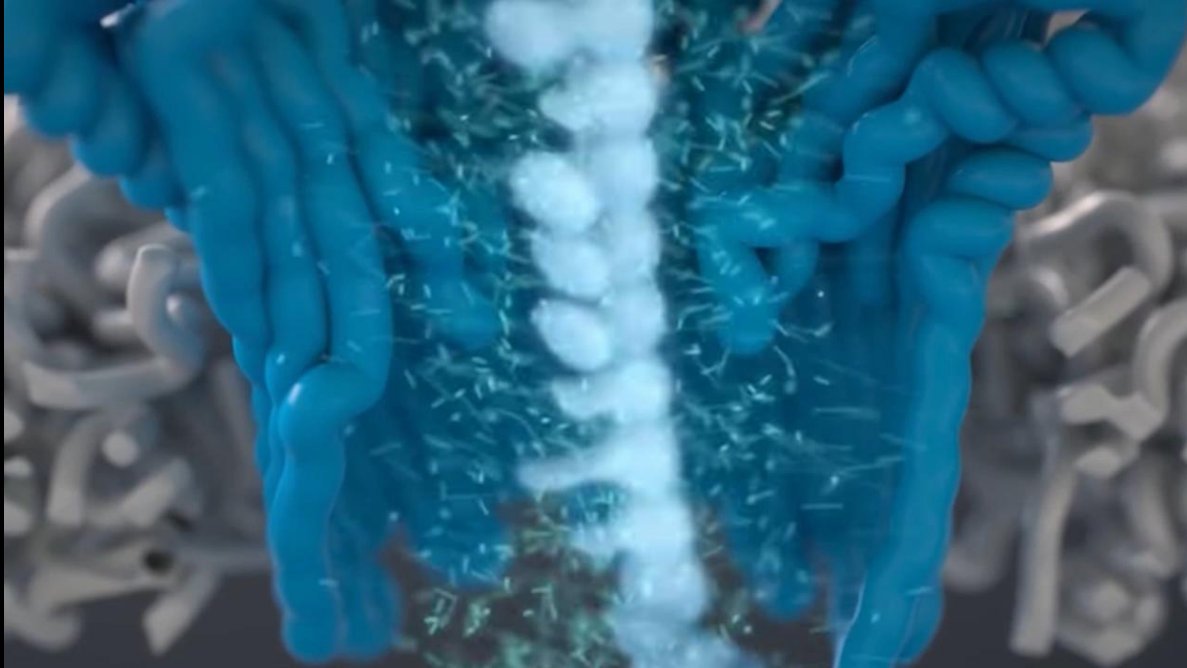


2020年11月30日，AlphaFold 2 蛋白质结构预测

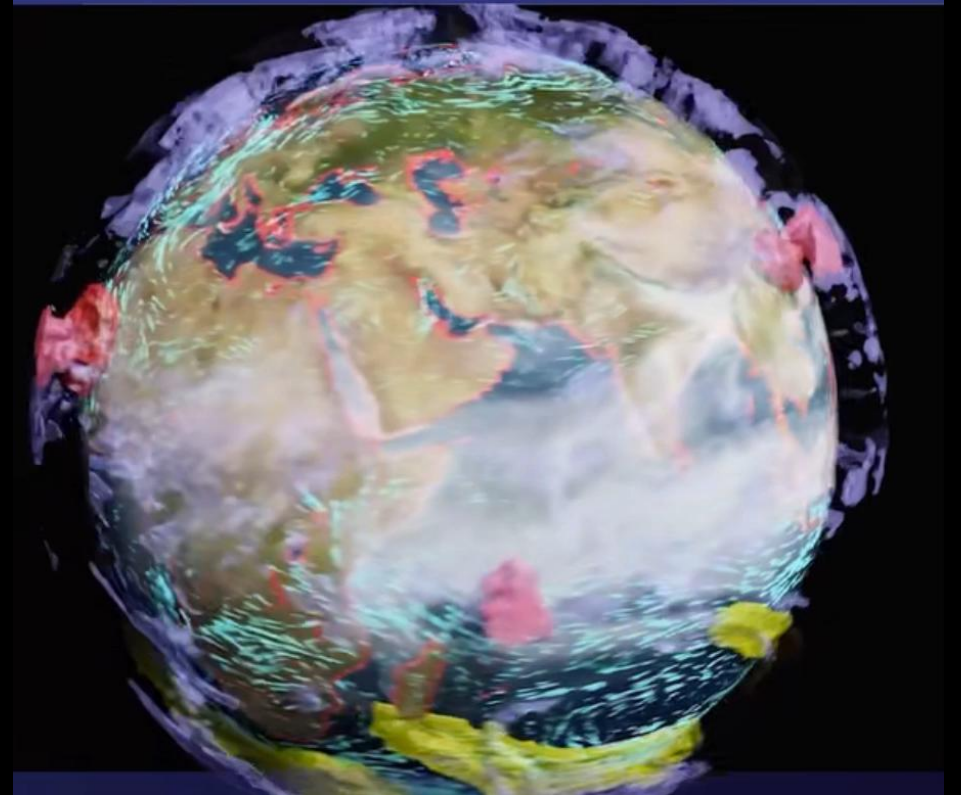
1.8亿个蛋白质结构：0.1% → 98.5%



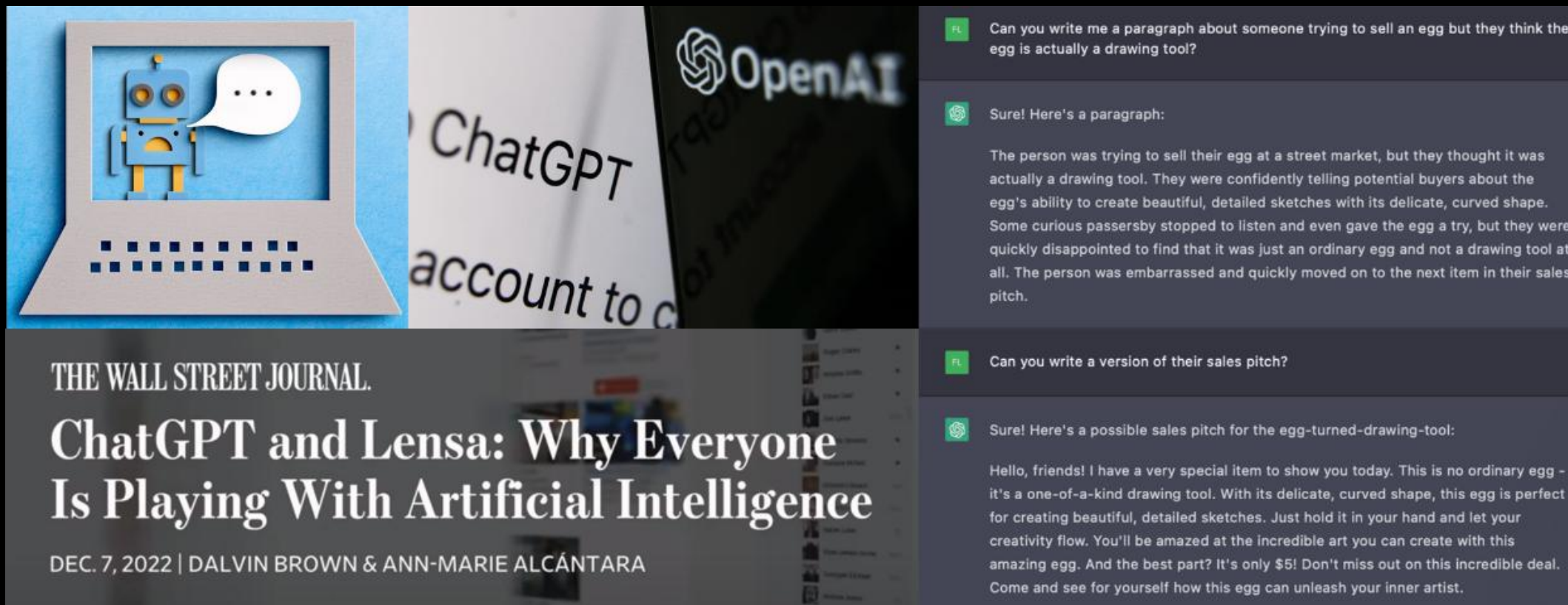
（基于人工智能）进行蛋白质设计和结构预言



2022年1月，DNA 测序的世界纪录：5小时2分钟



英伟达正在创建地球的数字孪生，Earth-2，
希望提前几十年来预测气候的变化



THE WALL STREET JOURNAL.

ChatGPT and Lensa: Why Everyone Is Playing With Artificial Intelligence

DEC. 7, 2022 | DALVIN BROWN & ANN-MARIE ALCÁNTARA

FL Can you write me a paragraph about someone trying to sell an egg but they think the egg is actually a drawing tool?

FL Sure! Here's a paragraph:

The person was trying to sell their egg at a street market, but they thought it was actually a drawing tool. They were confidently telling potential buyers about the egg's ability to create beautiful, detailed sketches with its delicate, curved shape. Some curious passersby stopped to listen and even gave the egg a try, but they were quickly disappointed to find that it was just an ordinary egg and not a drawing tool at all. The person was embarrassed and quickly moved on to the next item in their sales pitch.

FL Can you write a version of their sales pitch?

FL Sure! Here's a possible sales pitch for the egg-turned-drawing-tool:

Hello, friends! I have a very special item to show you today. This is no ordinary egg - it's a one-of-a-kind drawing tool. With its delicate, curved shape, this egg is perfect for creating beautiful, detailed sketches. Just hold it in your hand and let your creativity flow. You'll be amazed at the incredible art you can create with this amazing egg. And the best part? It's only \$5! Don't miss out on this incredible deal. Come and see for yourself how this egg can unleash your inner artist.

2022 ChatGPT

编程的终结：经典计算机科学的
末日即将来临

我们大多数人都是恐龙，
等待着流星的撞击。



攻城略地 所向披靡

10/43

5 : 0



K.O!

I. 疯狂的人工智能

II. 什么是人工智能

III. 联结主义

IV. 展望

什么是智能？

12/43

$$y = f(x)$$

↑ ↑
输出 输入
(响应) (状态)

- 棋盘：下子位置
- 局势：兵种操作
- 氨基酸排序：蛋白质结构
- 当前天气：未来天气
- 已有语句：后续语句
-

什么是智能？

13/43

$$y = f(x)$$

↑ ↑

输出 输入
(响应) (状态)

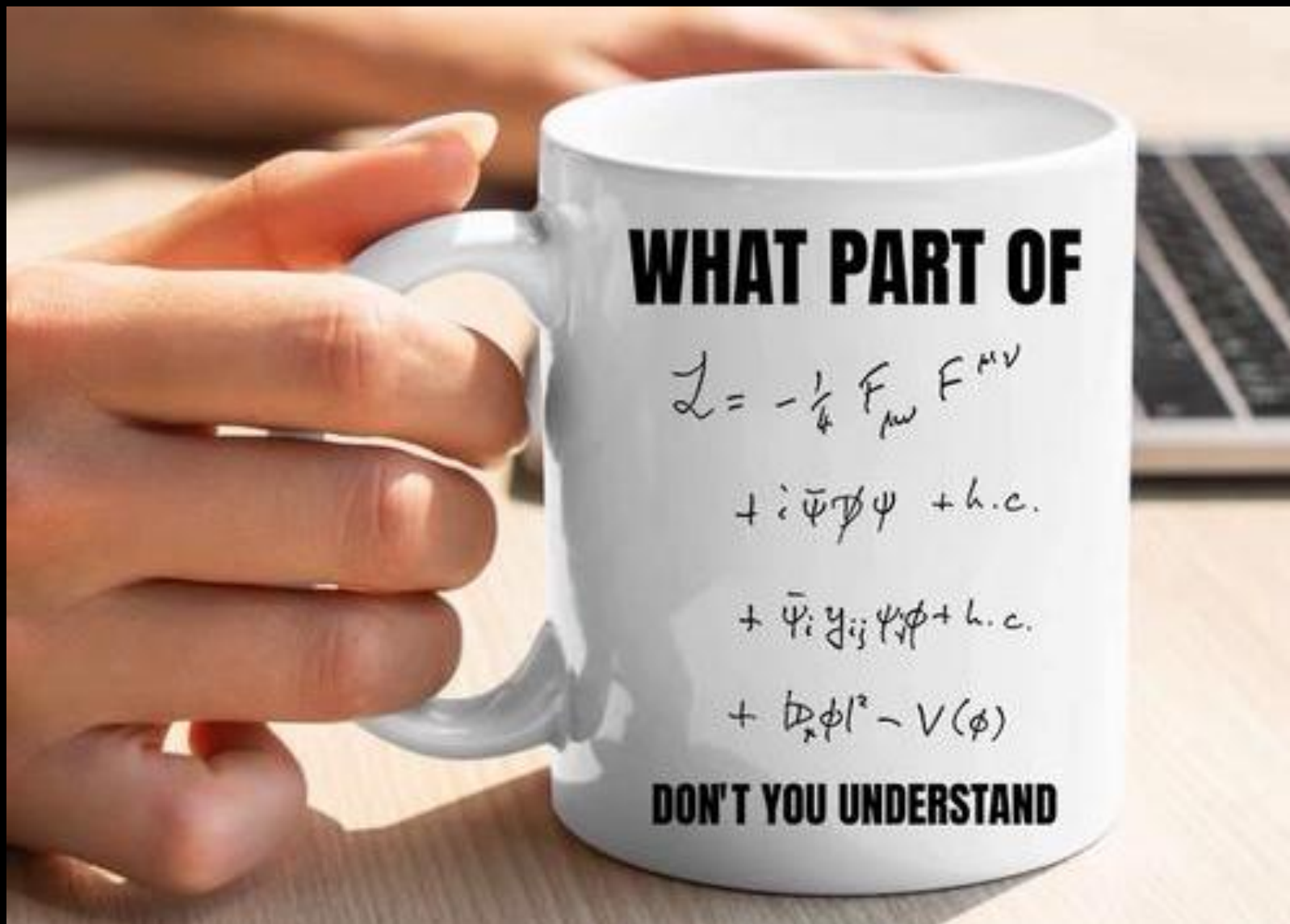
预言（泛化）能力？
不能简单地枚举

➤ 发现共性，产生规律

➤ 信息压缩 → 简单函数 → 智能

经验和规律：

- 见到乌云预测将下雨
- 见到台阶要抬脚
- 围棋定式
- $1+1=2$
-



什么是人工智能？

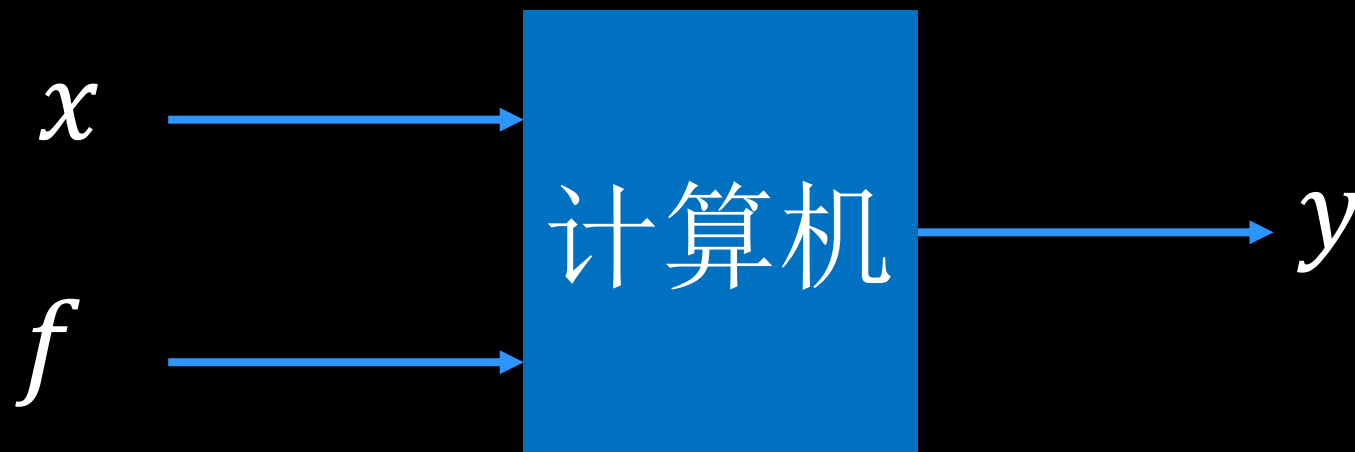
- **人工智能：** 是研究模拟、延伸和扩展人类智能的理论、方法、技术及应用系统的学科。



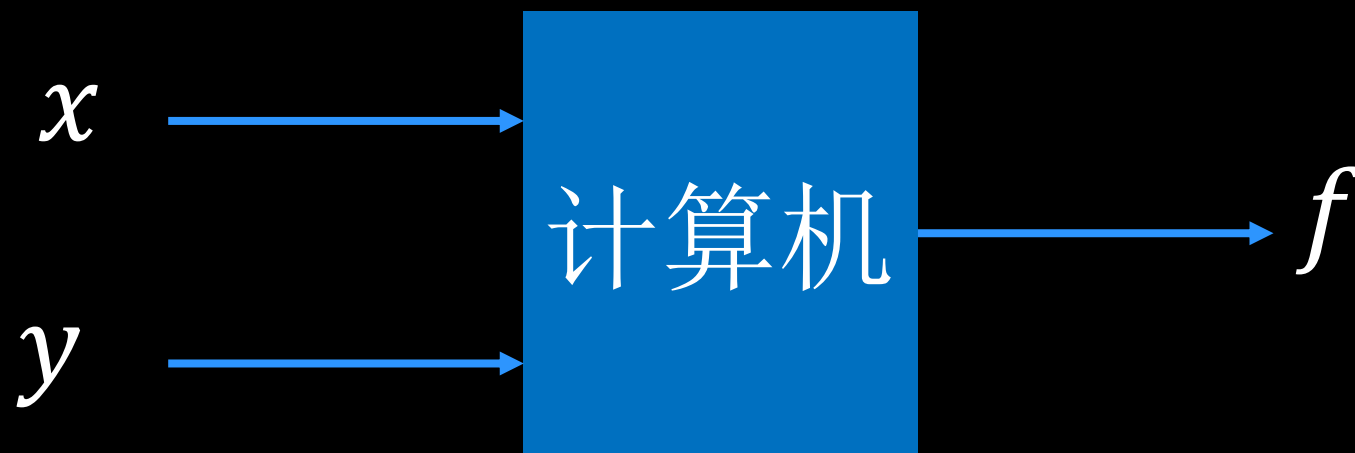
智能函数的来源

18/43

算法编程:



机器学习:



类比: 解析方法 vs 数值方法

✓ 准确性、可靠性

- × 使用局限性：需对问题有流程化的解决方案
- × 人力成本高：每个问题需单独编程

- ✓ 原则上可用于任意问题、人力成本低
- × 缺少解释性、可靠性差

- 优化目标(Loss function), 如最小二乘法:

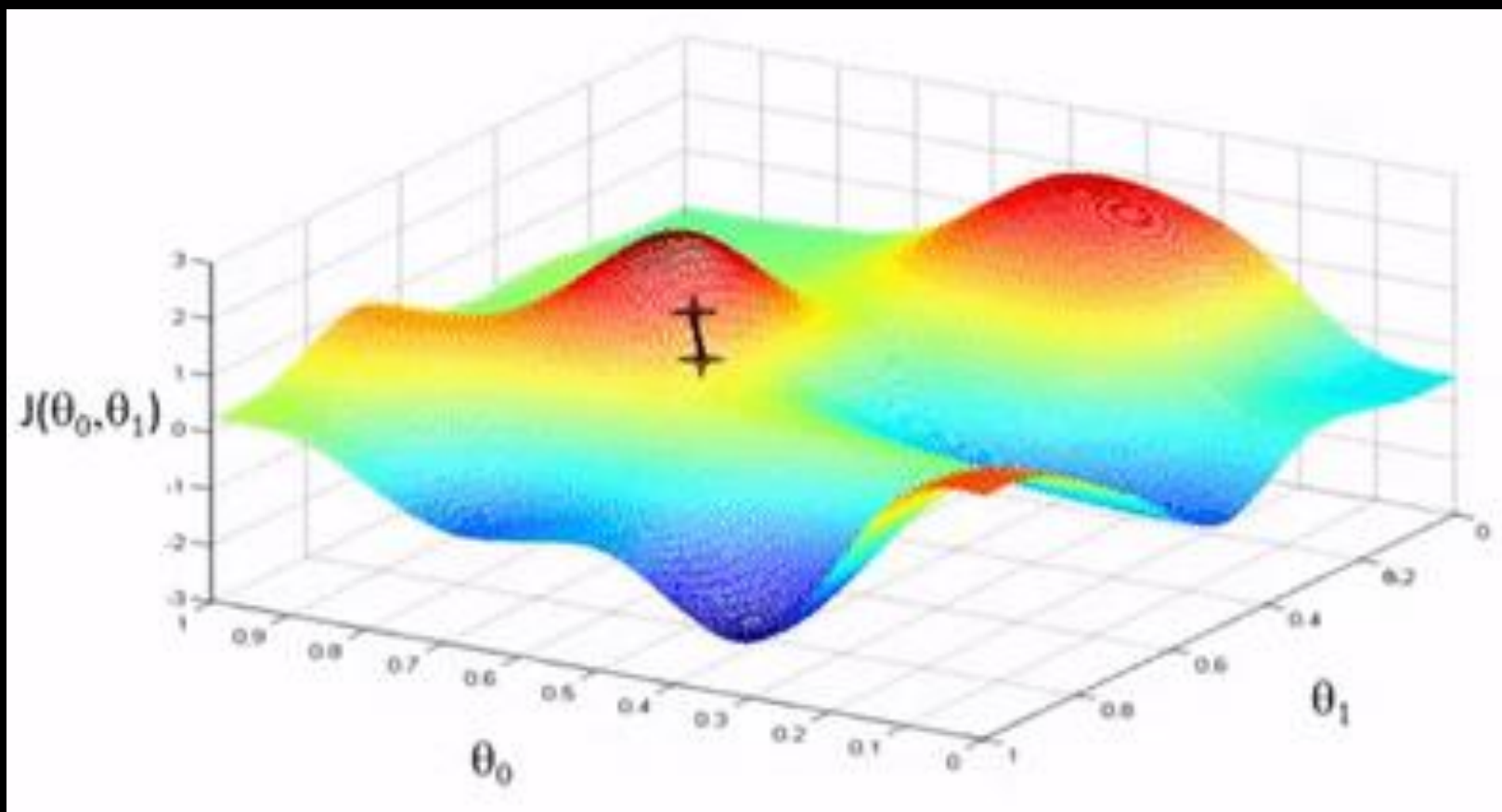
$$L(y, f_{\theta}(x)) = \frac{1}{n} \sum_{i=1}^n (y_i - f_{\theta}(x_i))^2$$

- 从而 $f_{\theta}(x) \approx f(x)$

梯度下降(GD)

22/43

每次沿梯度的相反方向走一个步长，直至目标函数收敛，或者达到预设的最大迭代次数。



□ 实际中一般在损失函数后加一项一起优化:

$$\min_f \frac{1}{n} \sum_{i \in [n]} L(f(x_i), y_i) + \lambda \cdot R(f)$$

- $\lambda \cdot R(f)$ 称为正则化项 (regularization term)，用于惩罚过于复杂的模型
- 模型越简单，发掘的共性越多，越有泛化能力
- 防止过拟合 (overfitting)，即测试误差远远高于训练误差

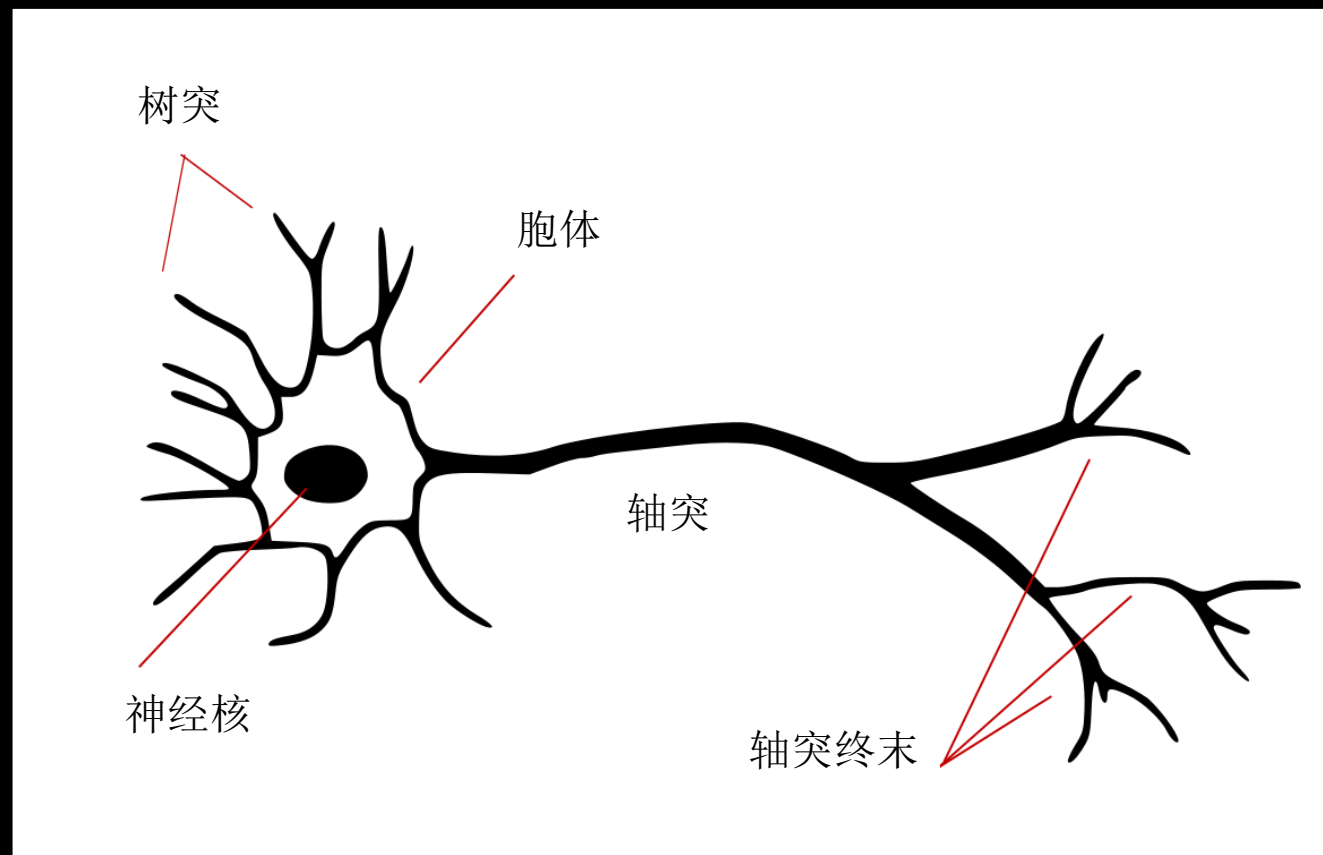
I. 疯狂的人工智能

II. 什么是人工智能

III. 联结主义

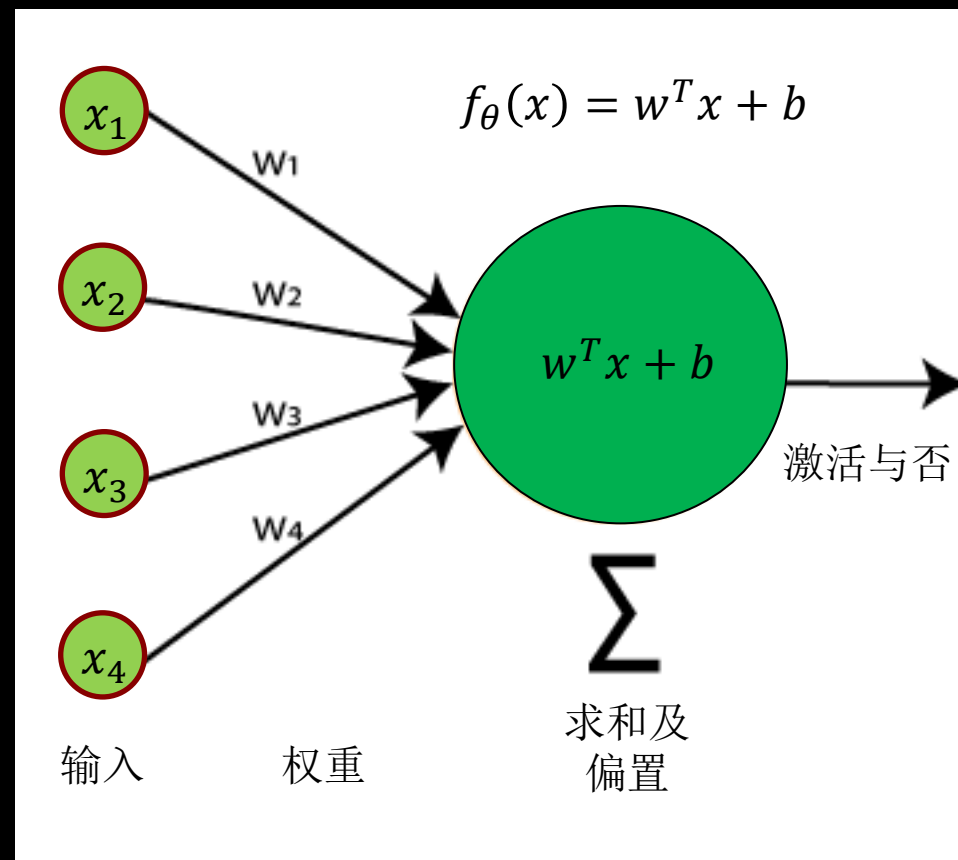
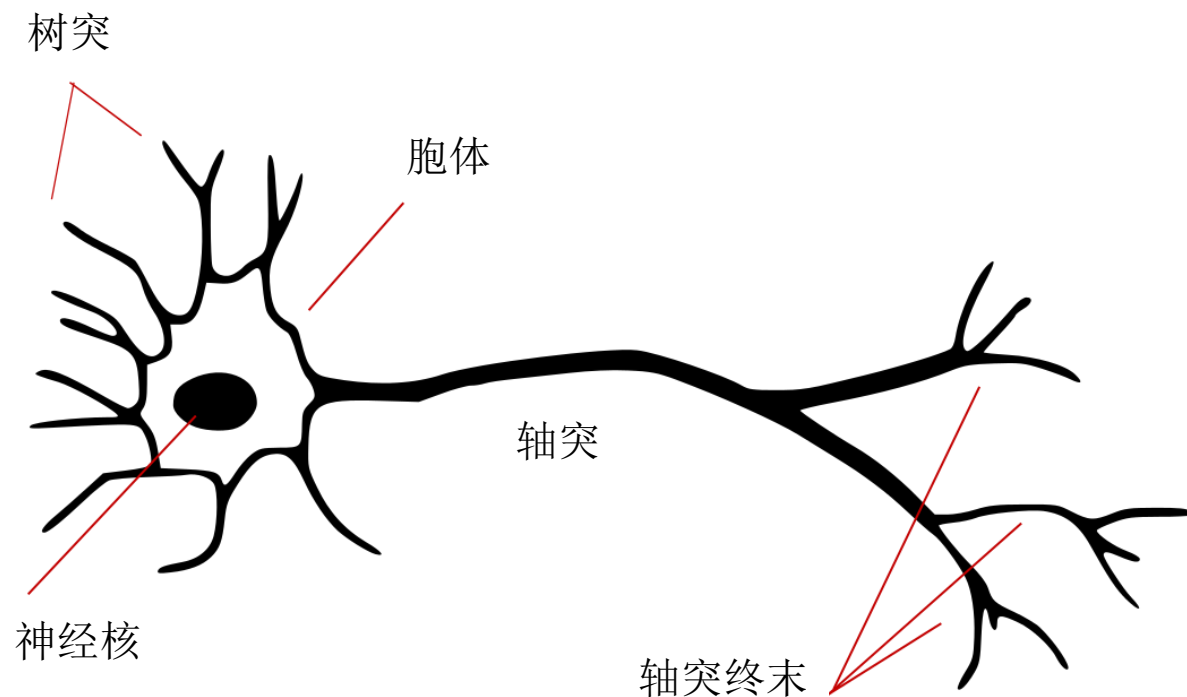
IV. 展望

神经元



人工神经元 (感知器)

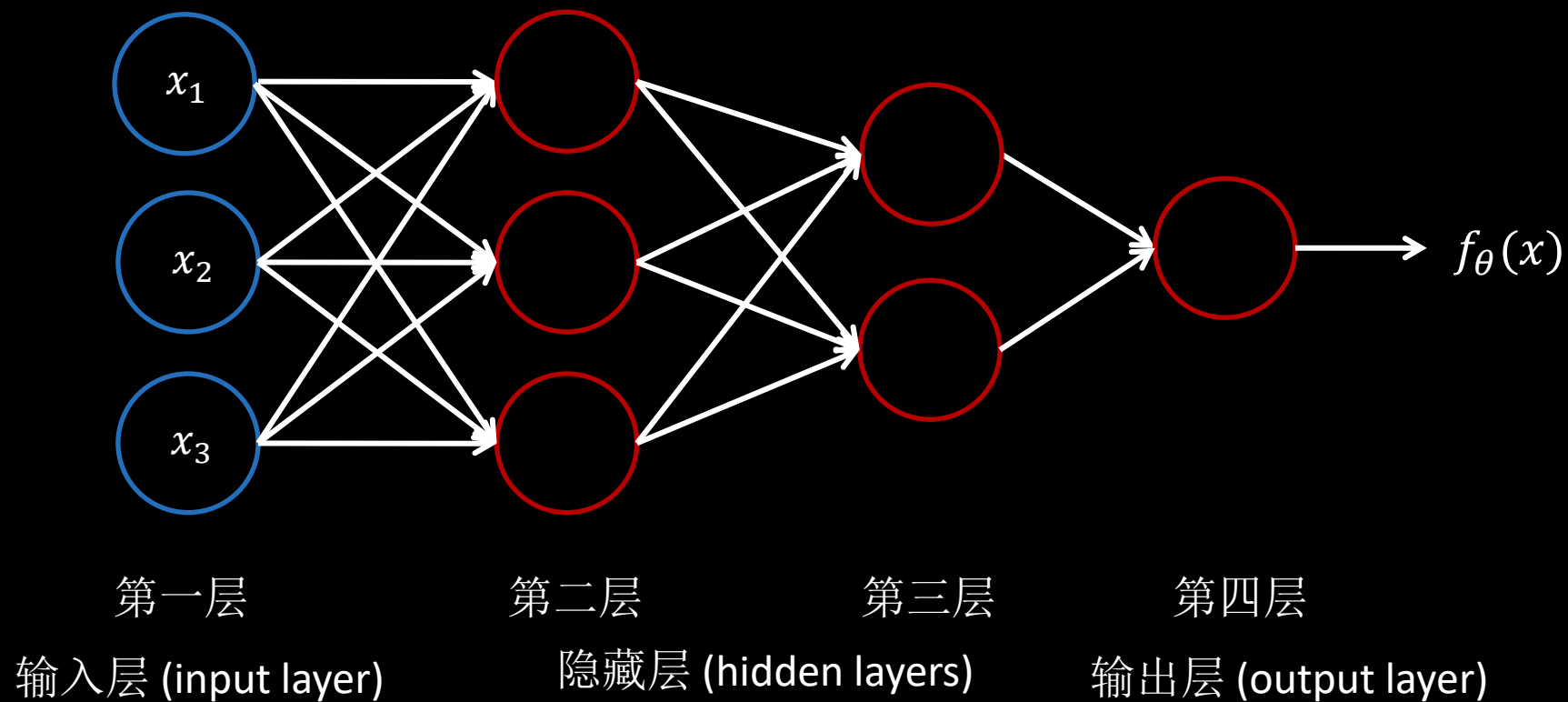
26/43



激活为非线性变换, 如ReLU: $\max(0, f(x))$

人工神经网络（多层感知器）

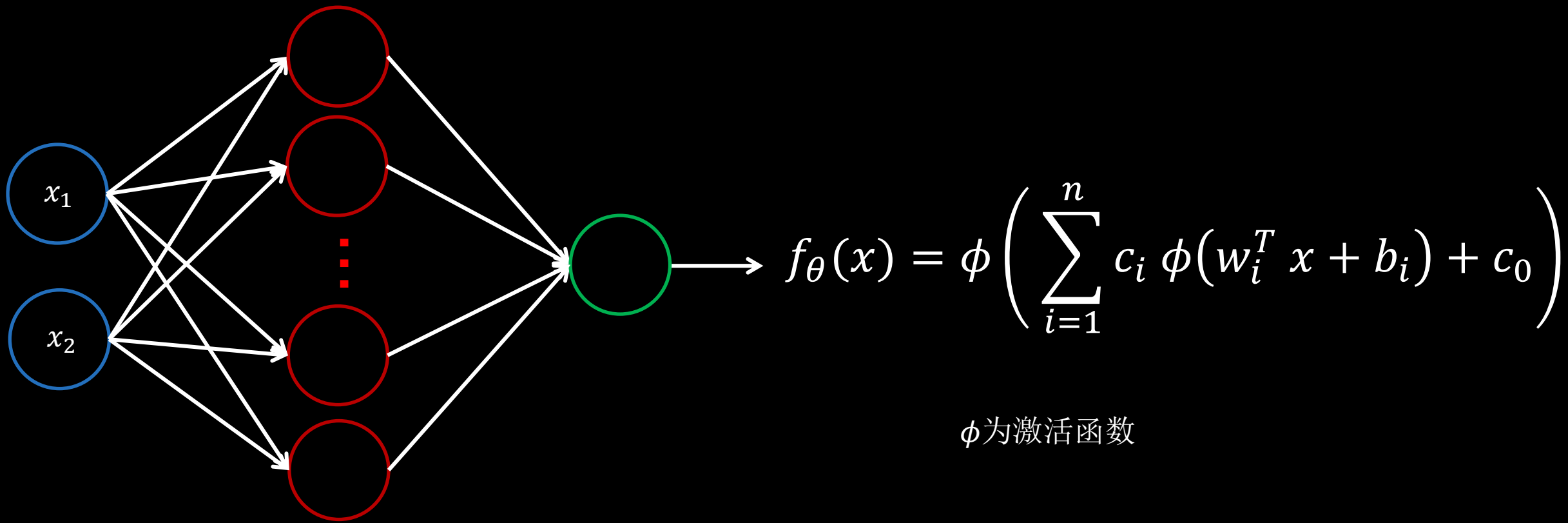
27/43



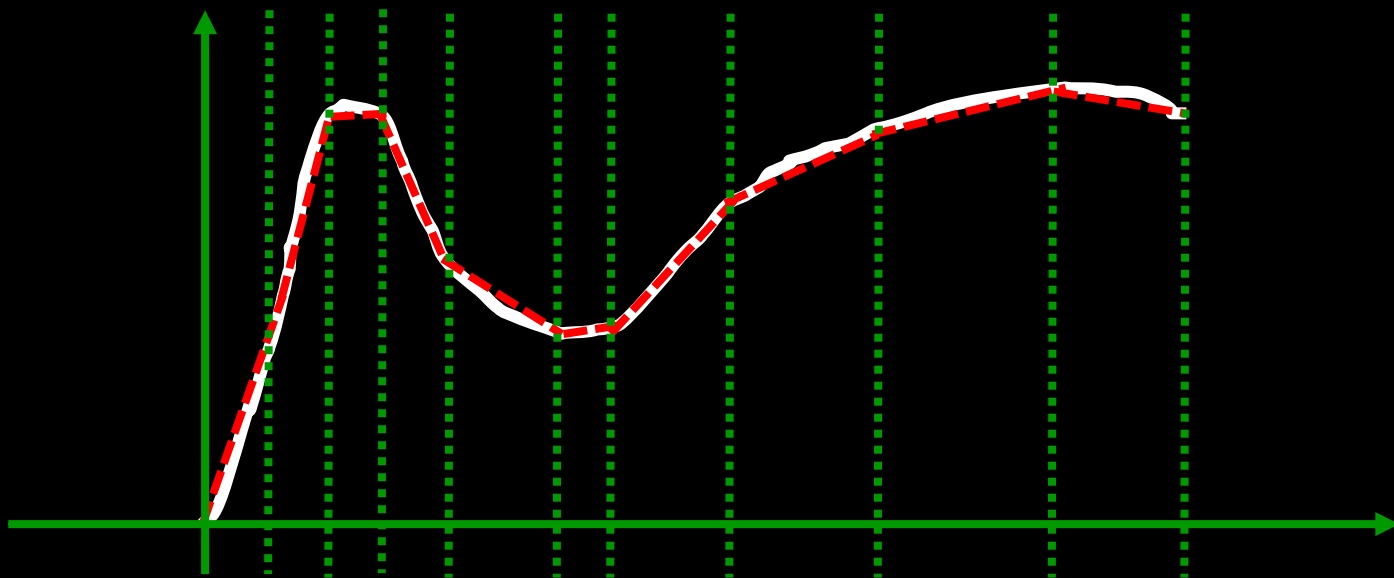
万能逼近定理

28/43

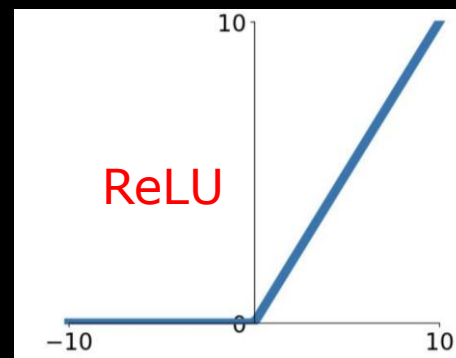
定理： 一个具有至少一个隐藏层的前馈神经网络，只要隐藏层中的神经元足够多，就可以逼近任何一个定义在有限维空间上的连续函数，精度可以任意接近。
这个结果适用于使用非线性激活函数（如Sigmoid或ReLU）。



举例：对于单变量



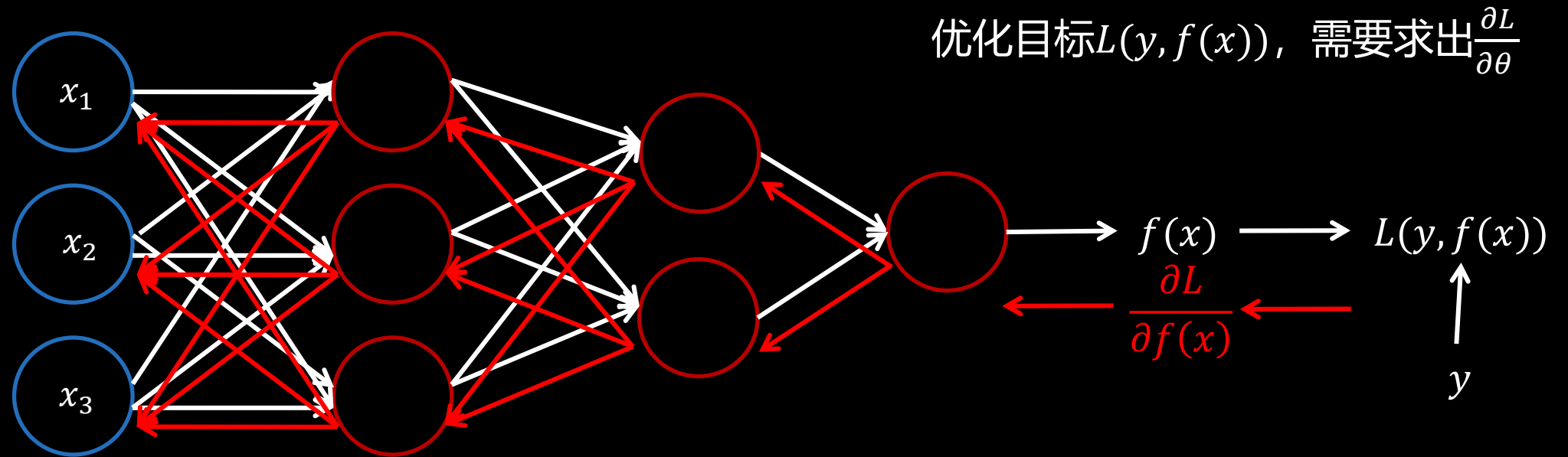
- 折线可以无限逼近任意函数
- ReLU激活函数的单隐藏层网络可描述任意连续折线： $\phi(c \phi(x) + \dots)$



- 其它激活函数可以近似ReLU

反向传播

30/43



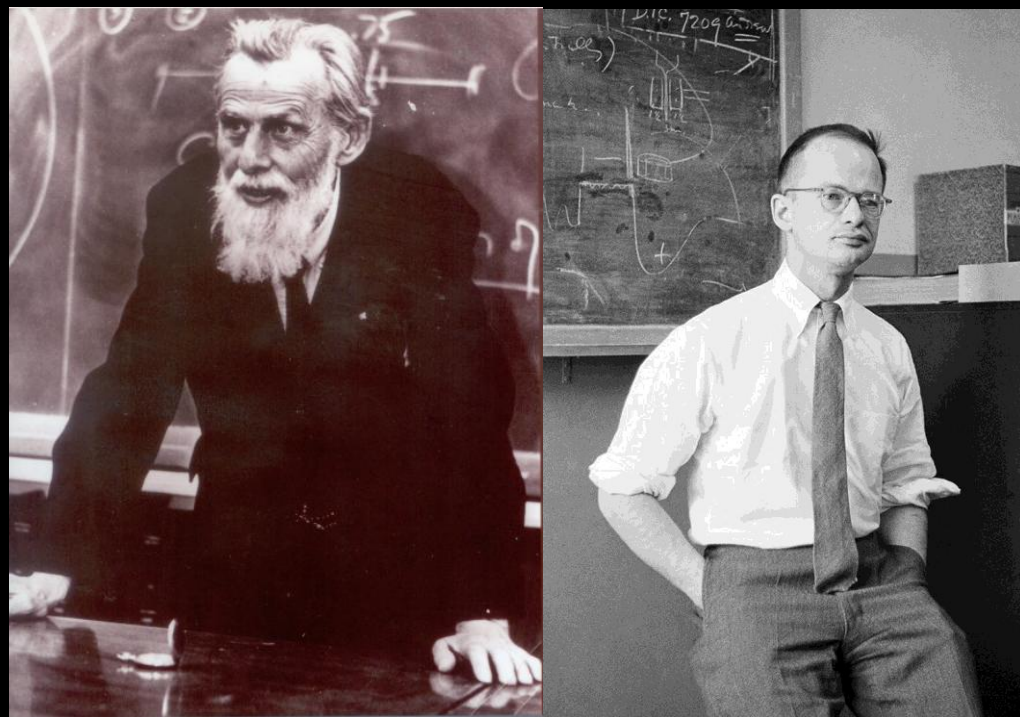
联结主义早期历史

31/43

- 无隐藏层+线性激活函数：
线性回归，高斯（1795）



- 1943：人工神经网络模型
McCulloch&Pitts
- 1940s-1960s：带有一个
自适应隐藏层的多层网络



符号主义高歌猛进

32/43

1965 西蒙 (Simon)

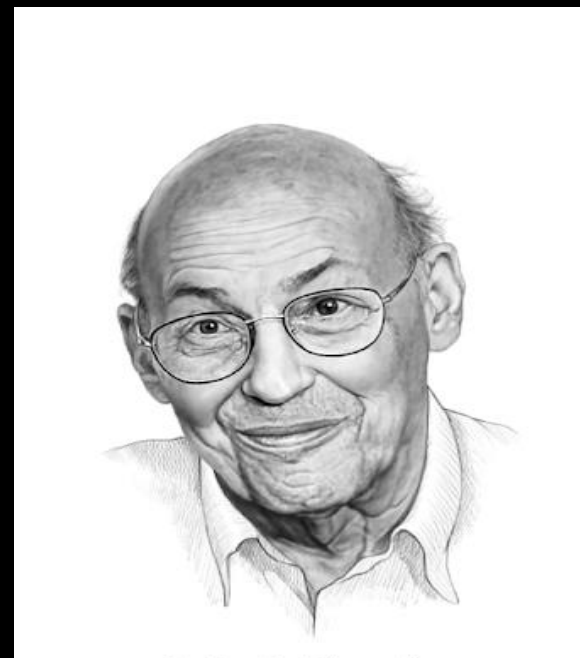
机器将在未来二十年内能够
做任何人类可以做到的工作



1975年图灵奖

1970 闵斯基 (Minsky)

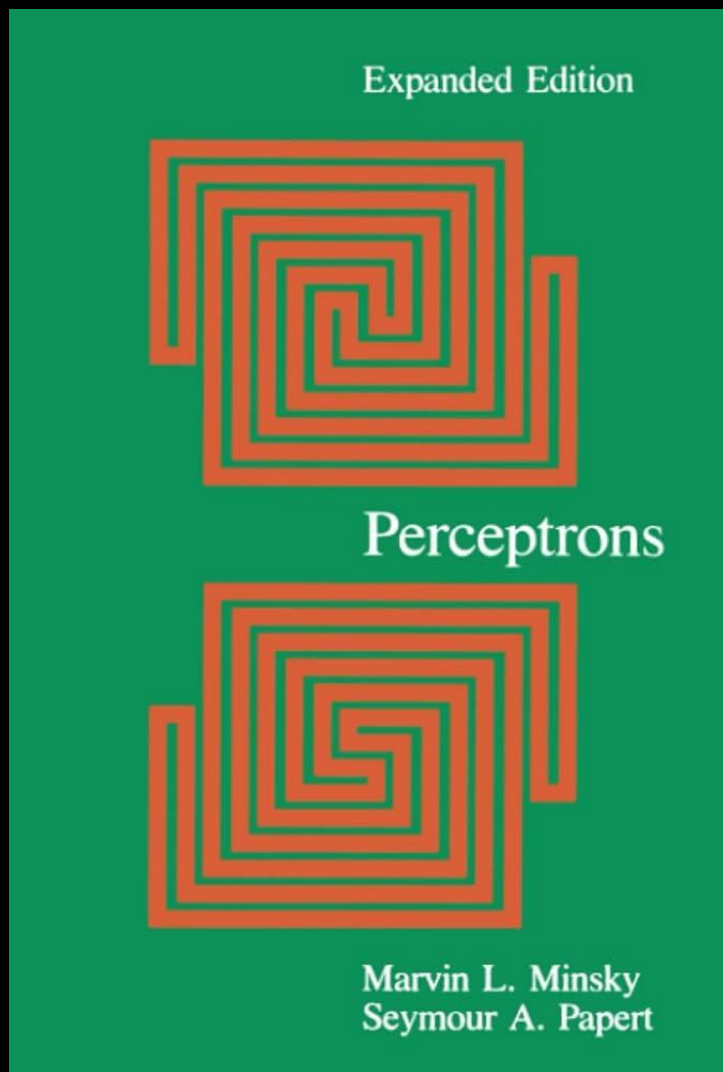
在3到8年内，我们将制造出具有
与普通人类一样智能的机器



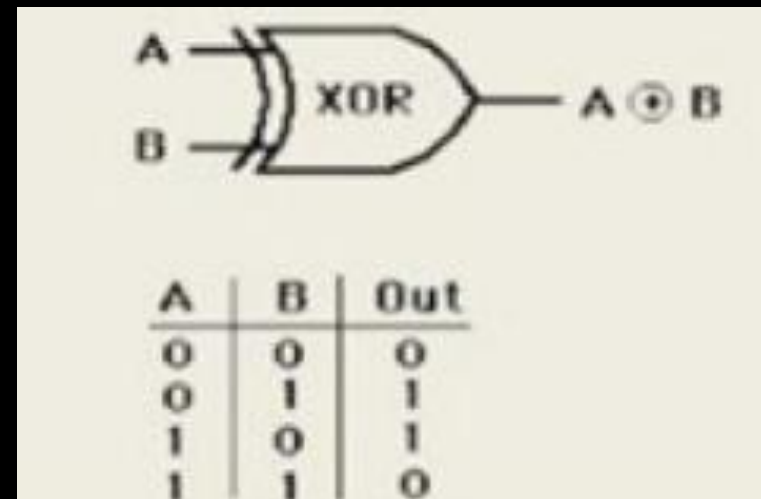
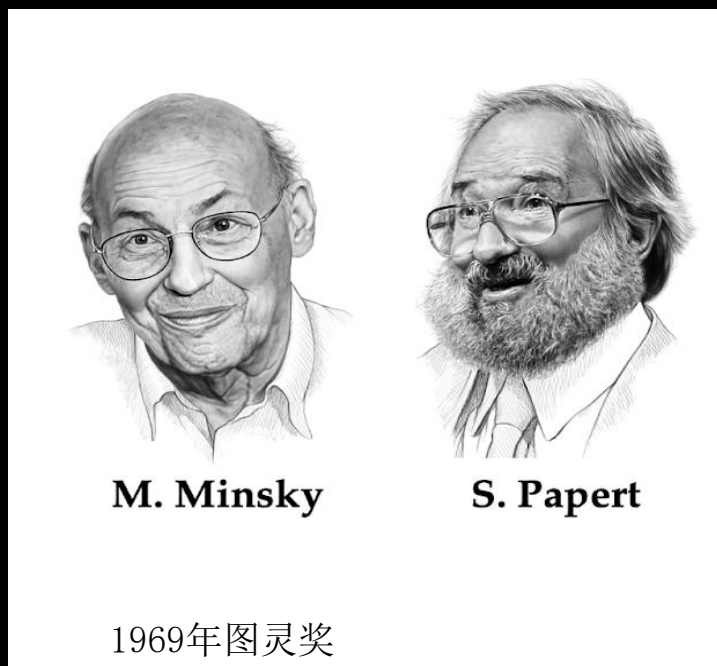
1969年图灵奖

联结主义的暴风雪

1969



- 指出单层感知器无法拟合“异或”问题
- 质疑多层感知器能够有效学习
- 泛化问题、理论挑战.....



THE NOBEL PRIZE
IN PHYSICS 2024

Illustrations: Niklas Elmehed



John J. Hopfield

Geoffrey E. Hinton

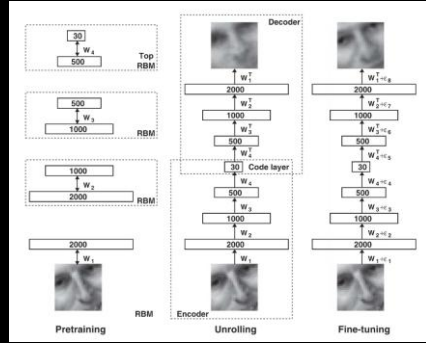
"for foundational discoveries and inventions
that enable machine learning
with artificial neural networks"

THE ROYAL SWEDISH ACADEMY OF SCIENCES

在人工神经网络机器学习方面的基础性贡献

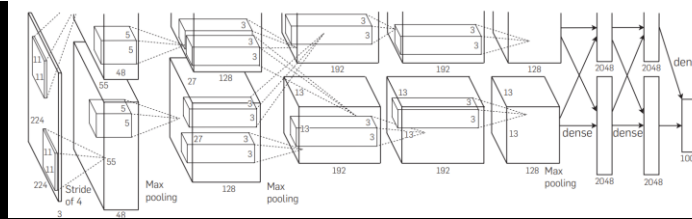
2018年图灵奖

人工智能的春天来临



AutoEncoder
Hinton et al. 2006

Model	Top-1 (%)	Top-5 (%)
Sparse coding ²	47.1	28.2
SIFT + FVs ²⁹	45.7	25.7
CNN	37.5	17.0



AlexNet
在ImageNet物体识别比赛中以绝对优势夺冠
Krizhevsky, Sutskever, Hinton. 2012

比尔·盖茨：ChatGPT表明人工智能历史意义不亚于“PC或互联网...”

ChatGPT

1995

2006

2011

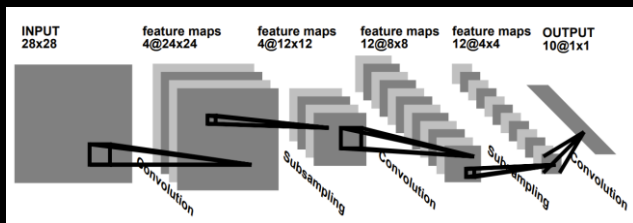
2012

2016

2018

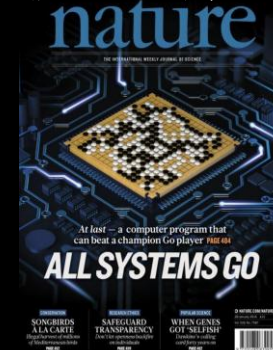
2022

卷积神经网络**CNN**
手写数字识别**LeNet**
LeCun et al. 1995



深度学习开始流行
从语音识别开始
到视觉物体识别

AlphaGo
深度网络评估盘面



深度学习三巨头
获图灵奖



I. 疯狂的人工智能

II. 什么是人工智能

III. 联结主义

IV. 展望

人工智能的历史——AI图灵奖得主

37/43

提出关于问题求解和
人类认知的符号模型

Allen Newell Herbert Simon
(1975年获奖) (1975年获奖)



提出处理不确定性的
概率因果推理技术

Judea Pearl
(2011年获奖)



Marvin Minsky
(1969年获奖)



John McCarthy
(1971年获奖)

定义了基于表示和推理
的领域基础



Ed Feigenbaum
(1994年获奖)



Raj Reddy
(1994年获奖)

开发对知识编码
来解决真实问题的专家系统



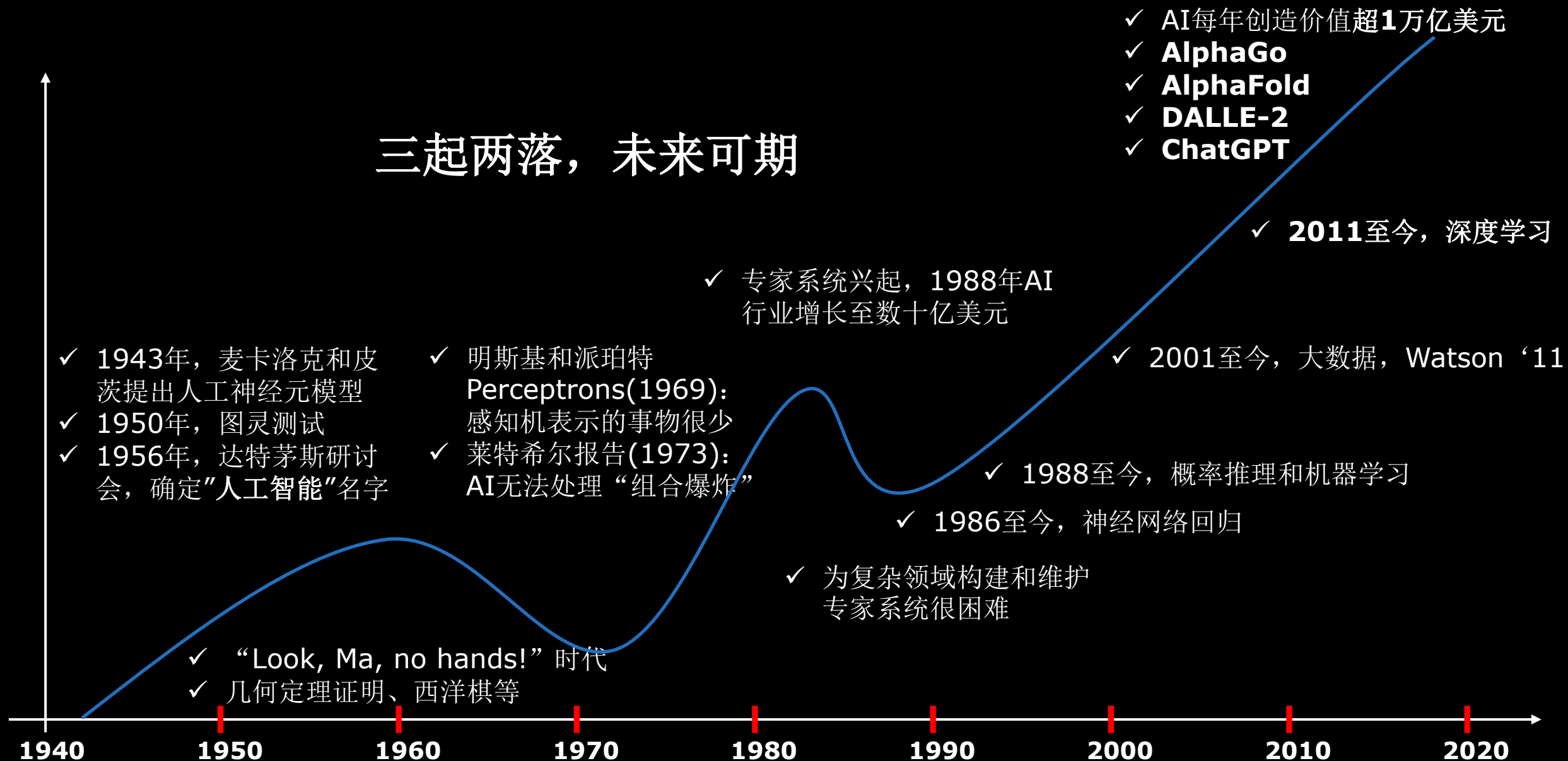
Y. LeCun G. Hinton Y. Bengio
(2018年获奖)

深度神经网络
深度学习

人工智能的历史

38/43

三起两落，未来可期

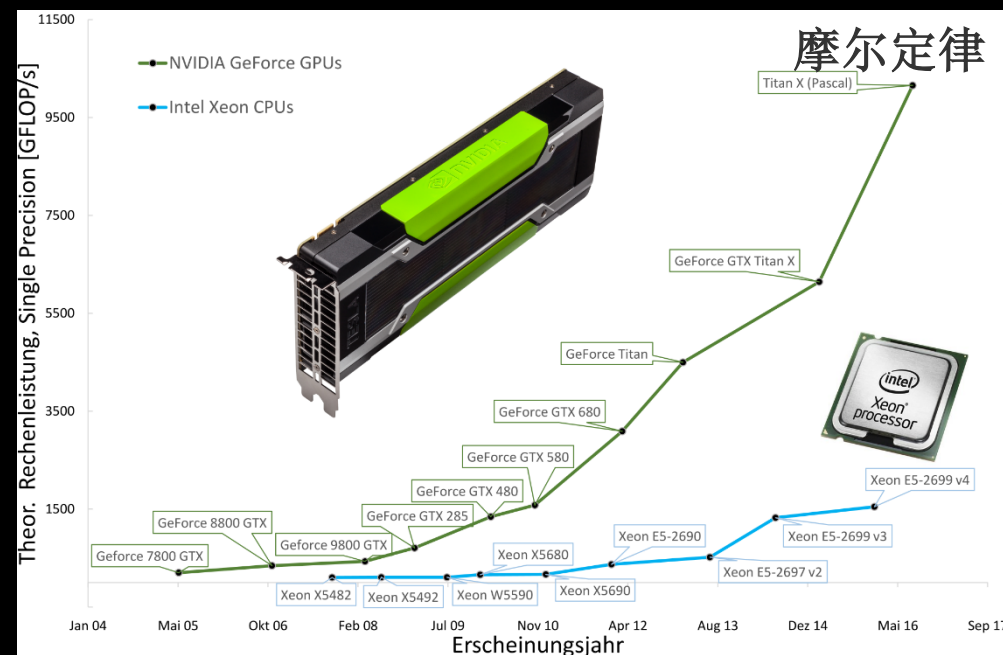


早期智人复原图（约20万年前）
智人脑容量已有1300-1600毫升





2009年, ImageNet数据集引发了计算机视觉领域的一场革命



CPU: $10^9 \sim 10^{10}$ 次运算/s
GPU/TPU上的DL: $10^{14} \sim 10^{17}$ 次运算/s

基于深度学习的困境

41/43

- 大数据、大计算量：模型参数指数增长，数据支撑和资源支持困难
- 小数据问题：科学发现时的常态
- 可解释性：黑匣子，对其预言的不安
- 因果、约束：难以施加，容易产生低级错误
-

- **符号主义(symbolism)**（逻辑主义或计算机学派）：认为人工智能源于数理逻辑
 - 将符号作为人工智能的基本元素，认为人工智能的运行建立在由符号构成的数理逻辑之上
- **行为主义(behaviorism)**（进化主义或控制论学派）：认为人工智能源于控制论
 - 专注于主体与环境的相互作用，并将这种相互作用看作是智能行为
- **联结主义(connectionism)**（仿生主义或生理学派）：认为人工智能源于仿生学
 - 试图使机器模拟大脑，通过建立一个类似于人脑中神经元的模拟节点网络来处理信号

