

《物理与人工智能》

1. 人工智能概览

授课教师：马滢青

2025/09/08（第一周）

鸣谢：基于计算机学院《人工智能引论》课程组幻灯片



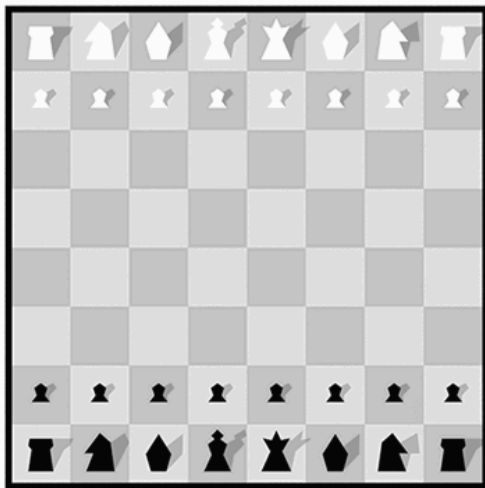
北京大学



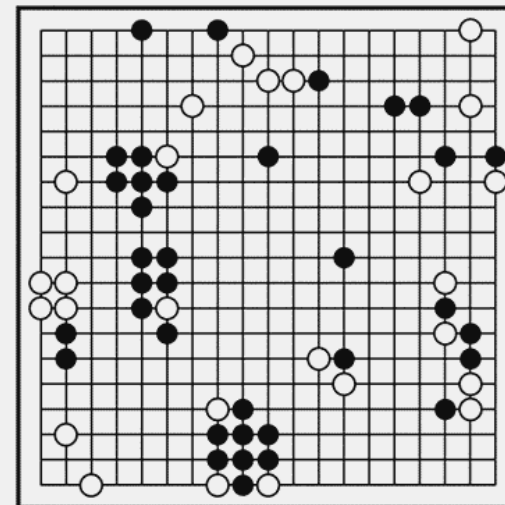
I. 疯狂的人工智能

II. 什么是人工智能

III. 人工智能的基础及发展



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




Human Actions IDLE

Agent Actions no_op

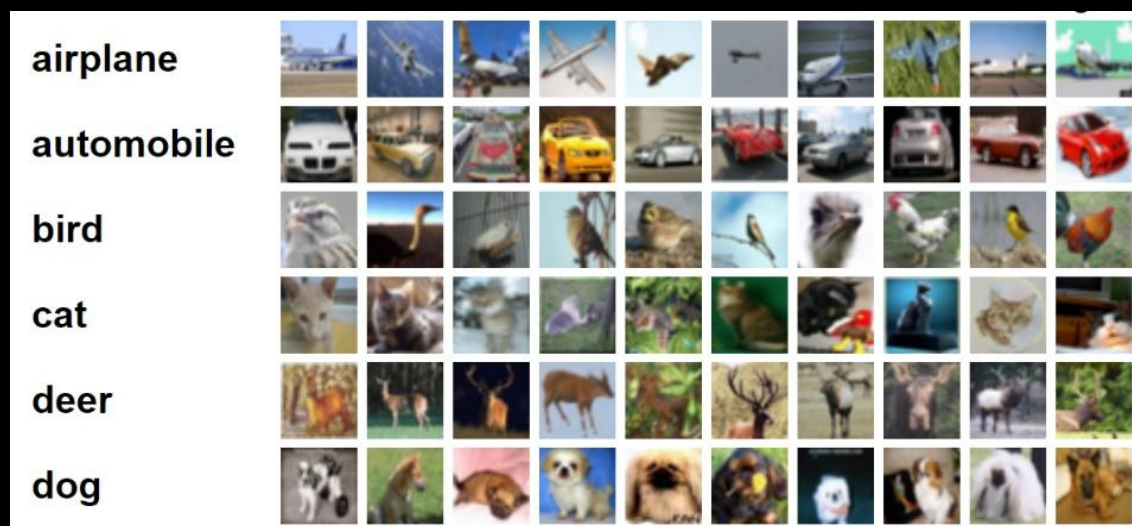
Available Actions

- Base action Point
- no_op
- rectangle select



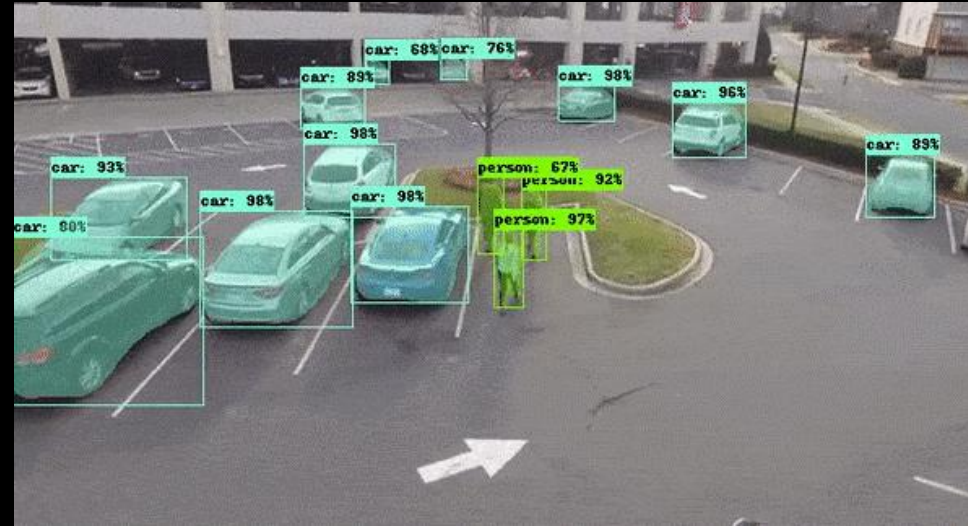
2019, AlphaStar人机大战以10:1战胜职业选手

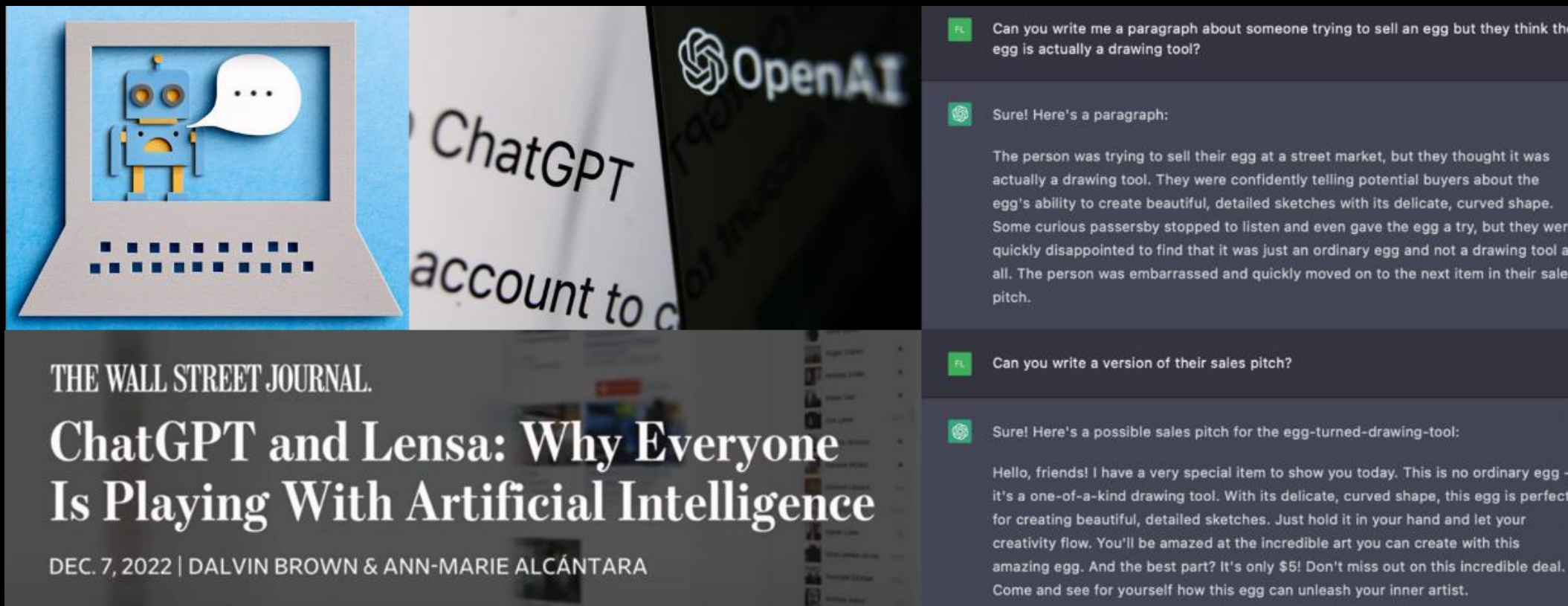
Grandmaster level in StarCraft II using multi-agent reinforcement learning, Nature 2019



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

人工智能——计算机视觉





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?

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?

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.

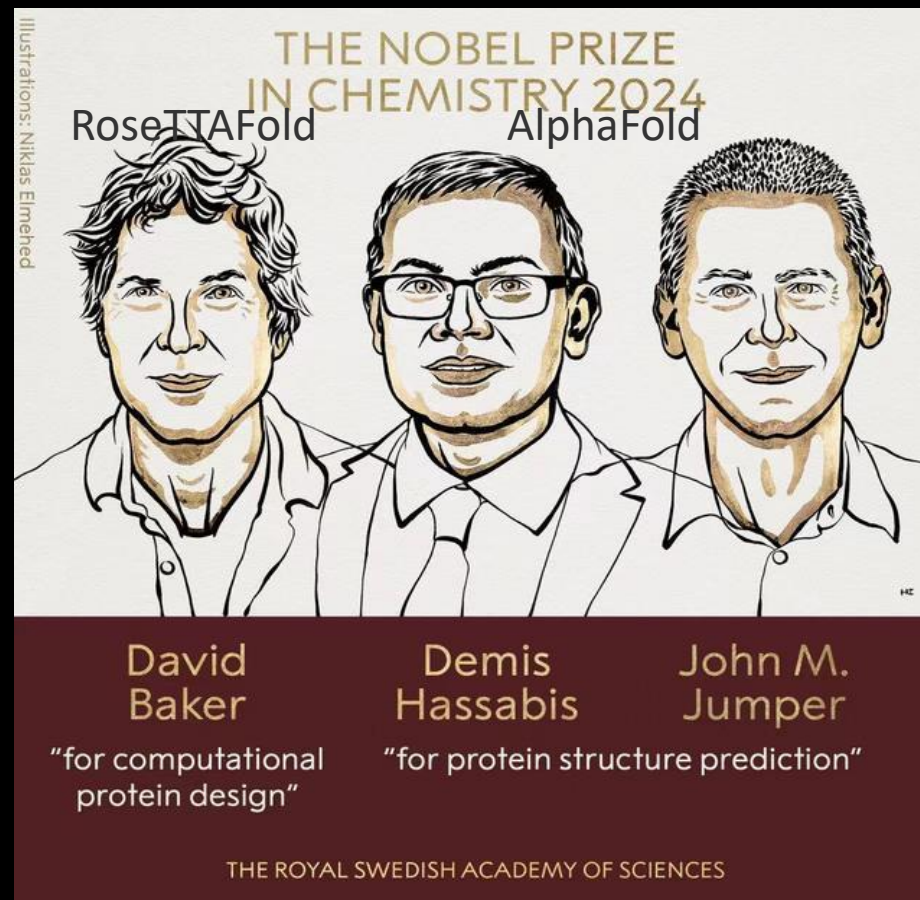
ChatGPT

Training language models to follow instructions with human feedback,
Arxiv 2022 <https://openai.com/blog/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。

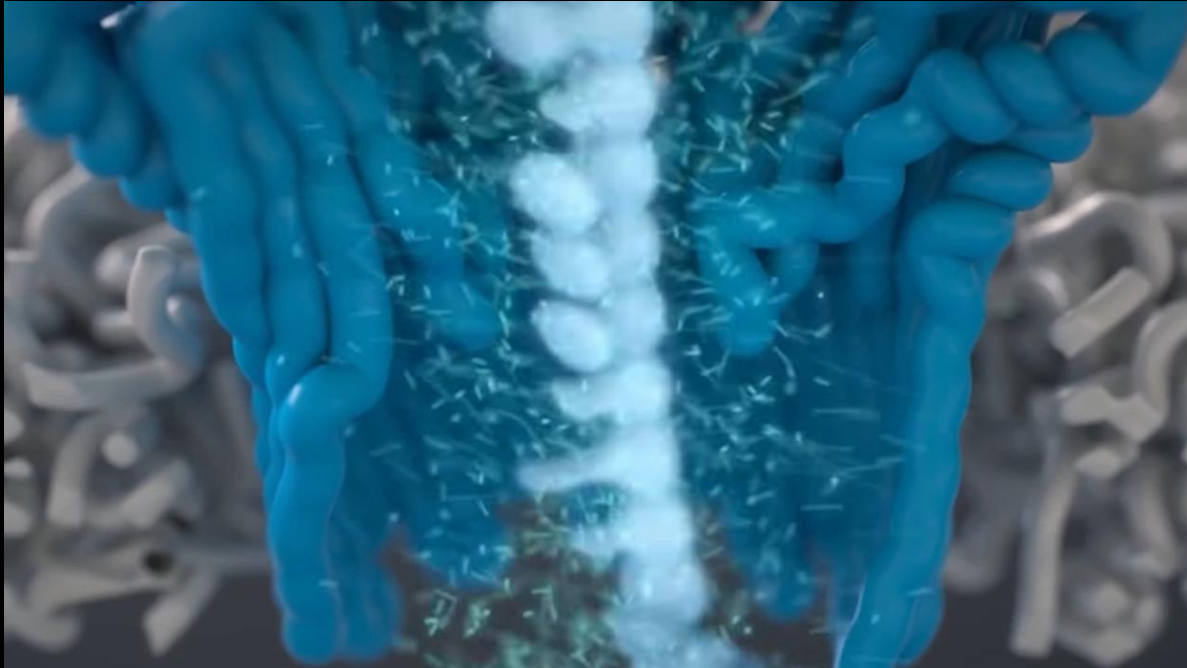




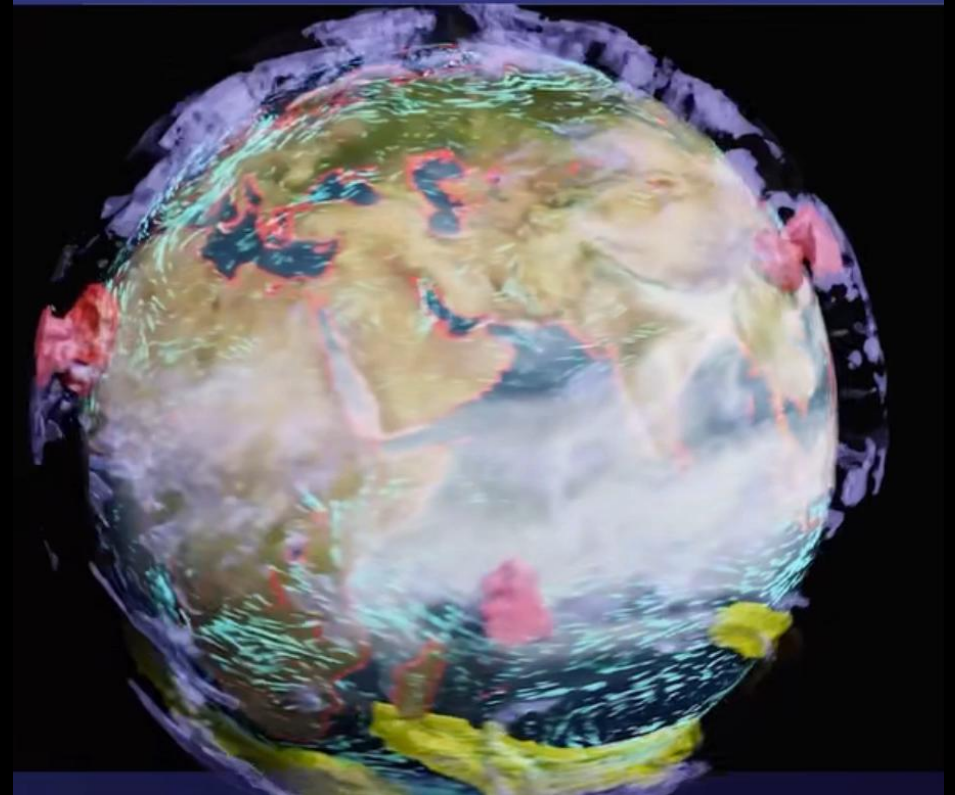
2020年11月30日，AlphaFold 2 蛋白质结构预测

(基于人工智能) 进行蛋白质设计和结构预言

1.8亿个蛋白质结构：0.1% -> 98.5%



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

攻城略地 所向披靡

5 : 0



K.O!

I. 疯狂的人工智能

II. 什么是人工智能

III. 人工智能的基础及发展史



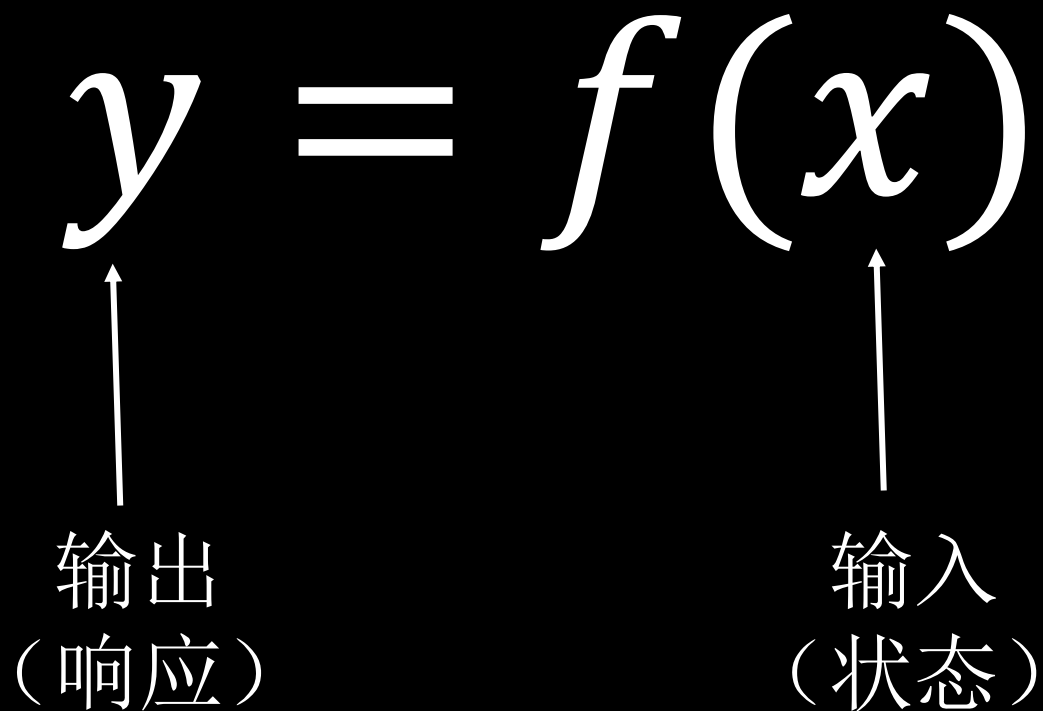
什么是智能？

- **智能** (intelligence)：是智力和能力的总称。从感觉到记忆到思维这一过程，称为“智慧”，智慧的结果就产生了行为和语言，将行为和语言的表达过程称为“能力”，两者合称“智能”
- **多元智能理论**：H. Gardner提出

- ① 语言智能
- ② 逻辑智能
- ③ 空间智能
- ④ 肢体运动智能
- ⑤ 音乐智能
- ⑥ 人际智能
- ⑦ 自我认知智能
- ⑧ 自然探索智能

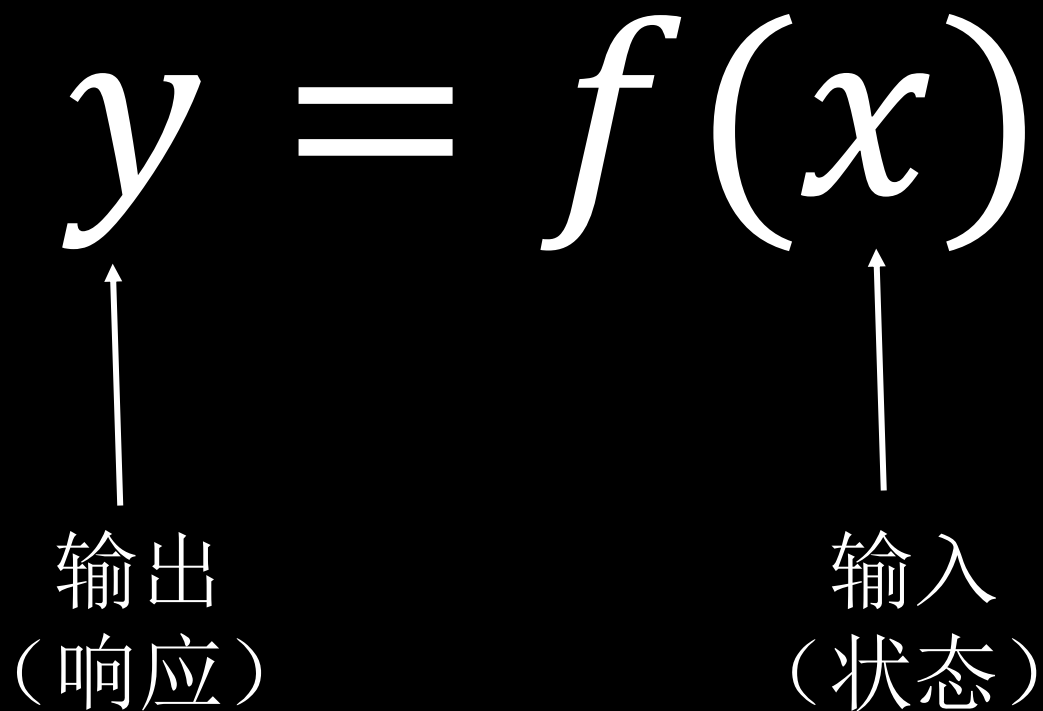


什么是智能？



- 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？



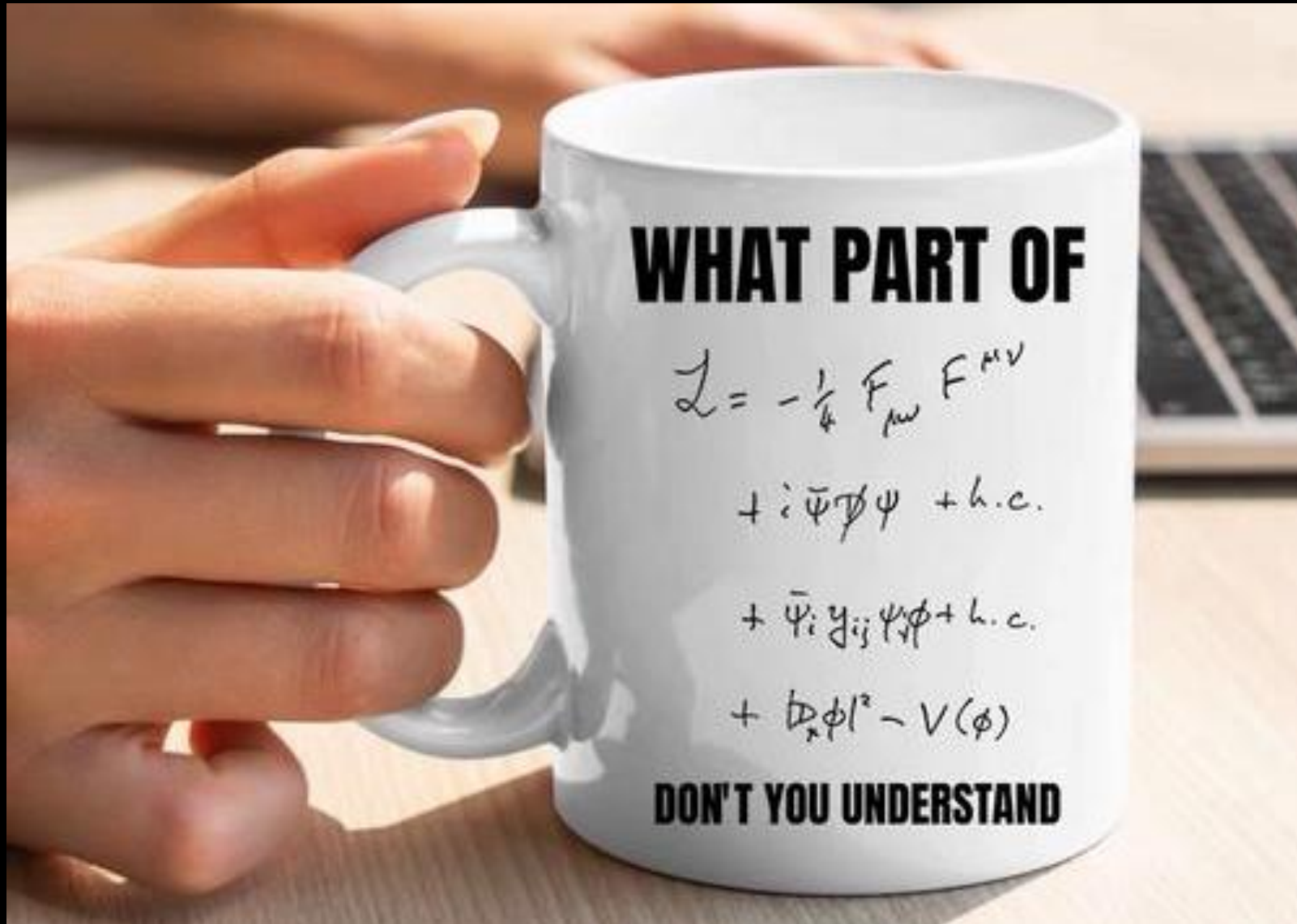
预言 (泛化) 能力?
不能简单地枚举

➤ 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
- $1+1=2$
-



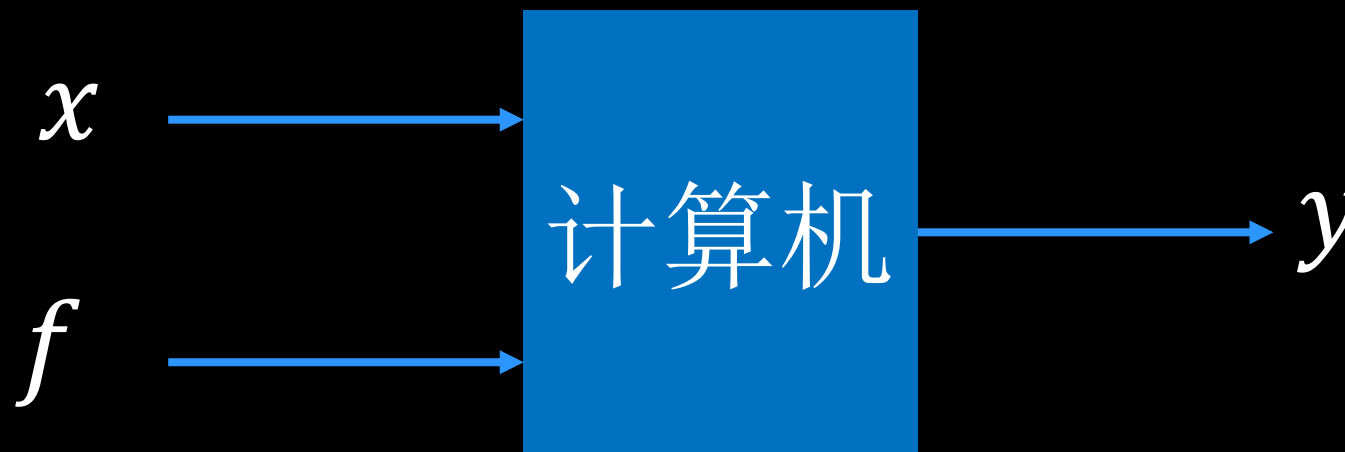
什么是人工智能？

- **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。

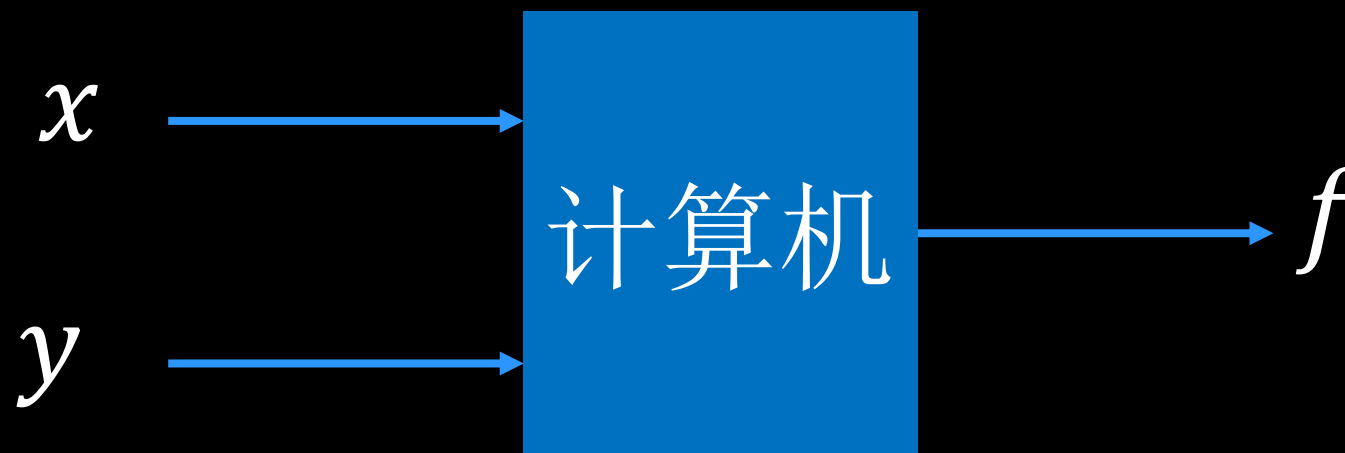


智能函数的来源

算法编程:



机器学习:



类比: 解析计算方法 vs 数值模拟方法

✓ 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

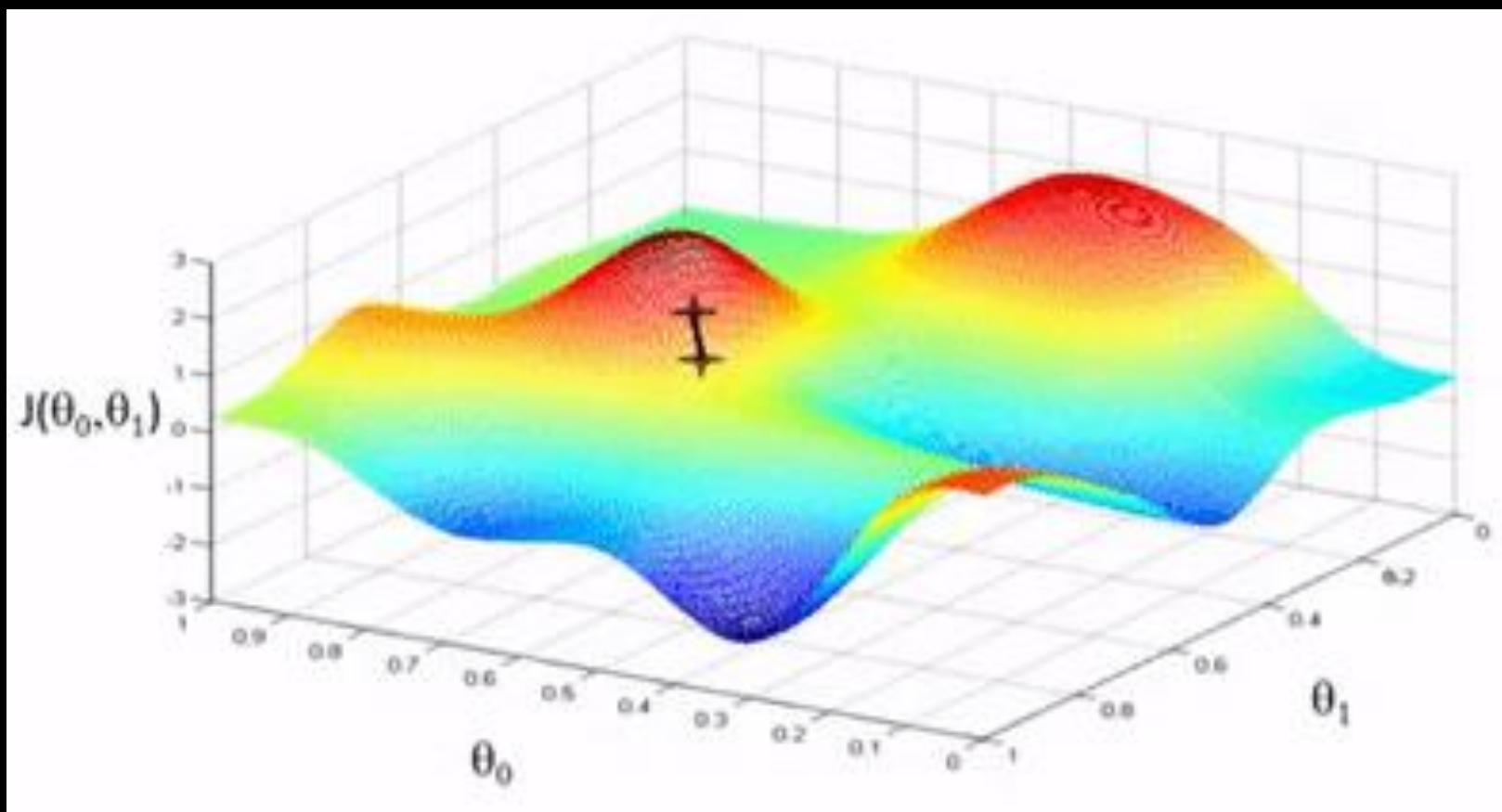
- 优化目标(Loss function), 如最小二乘法:

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

- 从而 $f_{\theta}(x) \approx f(x)$

梯度下降(GD)

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. 人工智能的基础及发展史

- **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
 - 人工神经元模型(1943)：联结主义的开端
 - 误差反向传播(BP)算法(1986)：联结主义的彻底复苏
 - 深度学习(2011，先语音识别后图像识别)：联结主义的迈向高潮

- 可以使用形式化规则得出有效结论吗？
 - 亚里士多德 (BC384–BC322)–三段论
 - 鲁尔 (1232–1315) 试图用一组旋转纸盘构成的机械来实现一种推理系统
- 思维如何从物质大脑中产生的？
 - 笛卡尔 (1596–1650) 二元论：人类思维一部分处于自然之外，不受物理定律约束
 - 唯物主义 (materialism)：大脑根据物理定律的运作构成了思维
- 知识从何而来？
 - 经验主义；逻辑实证主义；确证理论–将思维视为计算过程
- 知识如何导致行为？
 - 逻辑规划以实现确定目标；基于期望效用最大化的理性决策
 - 基于规则：“做正确的事情”不是由结果决定的，而由管理可行行为的普适社会法则决定

- 得出有效结论的形式化规则是什么？

- 形式化逻辑(formal logic)：古代中国、古希腊、古印度的哲学家
- 布尔(1815-1864)：命题和布尔逻辑；弗雷格(1848-1925)：一阶逻辑

- 什么可以被计算？

- 第一个非平凡的算法(algorithm)：欧几里得算法，计算最大公因数
- 哥德尔(1906-1978)不完全性定理：演绎的限制确实存在
- 图灵(1912-1954)：存在任何图灵机都无法计算的函数
- 可计算性(computability)；易处理性(tractability)；NP完全性(NP-completeness)理论

- 如何使用不确定的信息进行推理？

- 概率论(probability)：信息不确定下的广义逻辑；贝叶斯法则
- 统计学(statistics)：概率的形式化结合数据的可用性

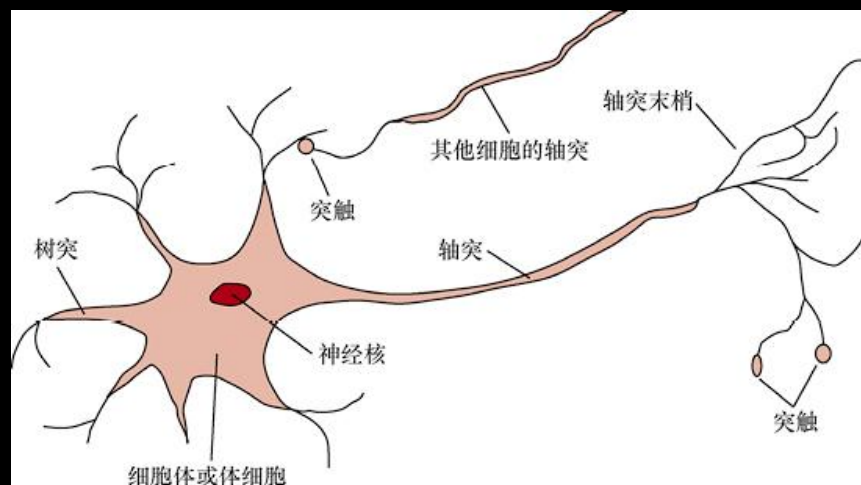


Herbert Simon
(1975年图灵奖)
(1978年诺贝尔经济学奖)

- 我们应该如何根据自己的偏好做出决定？
 - 经济学起源于1776年，亚当·斯密(1723-1790)《国富论》
 - 阿尔诺(1662)最大期望值公式；伯努利(1738)基于期望效用最大化准则，解释大众投资选择
 - 《博弈论与经济行为》(1944)：经济学不再是研究金钱的学科，而是对**欲望和偏好**的研究
- 当其他人可能不支持时，我们应该怎么做？
 - 决策论(decision theory)：概率论+效用理论，适用于“大型”经济体
 - 博弈论(game theory)：不对行为的选择提供明确的指示，适用于“小型”经济体
 - 多智能体系统(multiagent system)
- 当收益可能在很遥远的未来时，我们应该怎么做？
 - 运筹学(Operations research)：二战英国对雷达安装的优化
 - 强化学习(reinforcement learning)：马尔可夫决策过程，形式化贯序决策问题

• 大脑如何处理信息？

- 直到18世纪中叶，大脑才被广泛认为是意识所在地
- 布卡罗 (1861) 发现大脑左半球一个局部区域负责语音的产生，开启大脑功能组织的研究
- 高尔基，卡哈尔 (1906 诺奖) 提出“神经元学说”：神经元结构的电化学反应产生认知功能
- 伯杰 (1929) 脑电图仪 (EEG)；功能磁共振成像 (fMRI)；光遗传学 (optogenetics)
- 超级计算机 summit (2017) 在某些指标上已经与大脑相当，但需正确的理解智能的理论指导



	超级计算机	个人计算机	人类大脑
计算单元	10^6 个 GPU + CPU	8 个 CPU 内核	10^6 列
	10^{15} 个晶体管	10^{10} 个晶体管	10^{11} 个神经元
存储单元	10^{16} 字节 (10 PB) RAM	10^{10} 字节 (10 GB) RAM	10^{11} 个神经元
	10^{17} 字节 (100 PB) 磁盘	10^{12} 字节 (1 TB) 磁盘	10^{14} 个突触
周期时间	10^{-9} 秒	10^{-9} 秒	10^{-3} 秒
运算/秒	10^{18}	10^{10}	10^{17}

- 人类和动物是如何思考和行为的？
 - 科学心理学的起源：赫尔姆霍茨(1821-1894)和温特(1832-1920)，坚持严格控制的实验，进行感知/联想任务时，内省思维过程
 - 行为主义(behaviorism)：只研究施加动物的感知/刺激及其产生的行为/反应的客观度量
 - 认知心理学(cognitive psychology)：大脑是信息处理设备，感知涉及无意识的逻辑推断
 - 认知科学(cognitive science)：从信息处理的角度来描述认知功能的运作，1956年9月MIT的一个研讨会，3篇论文展示了如何用计算机模型处理记忆、语言和逻辑思维的心理问题
 - 人机交互(human-computer interaction, HCI)先驱恩格巴特倡导**智能增强**(intelligence augmentation, IA)的理念而非AI
 - IA强调人类控制，AI强调机器智能

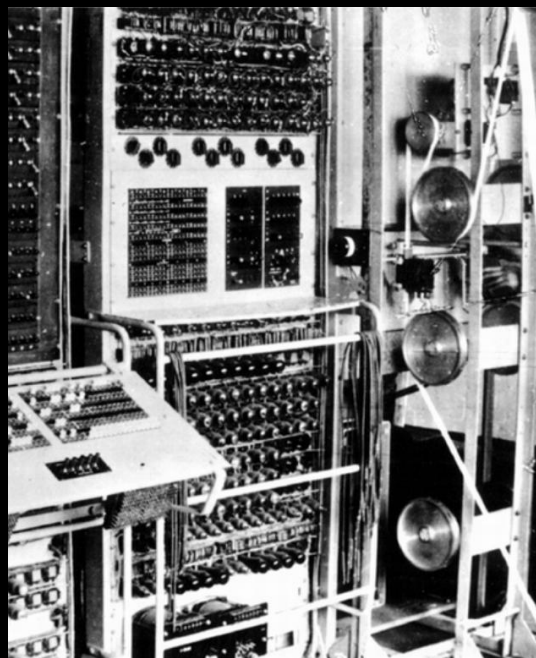
- 如何构建高效的计算机？

- 二战期间3个国家的科学家们独立且几乎同时发明了现代数字电子计算机



Heath Robinson

英国 (1943)

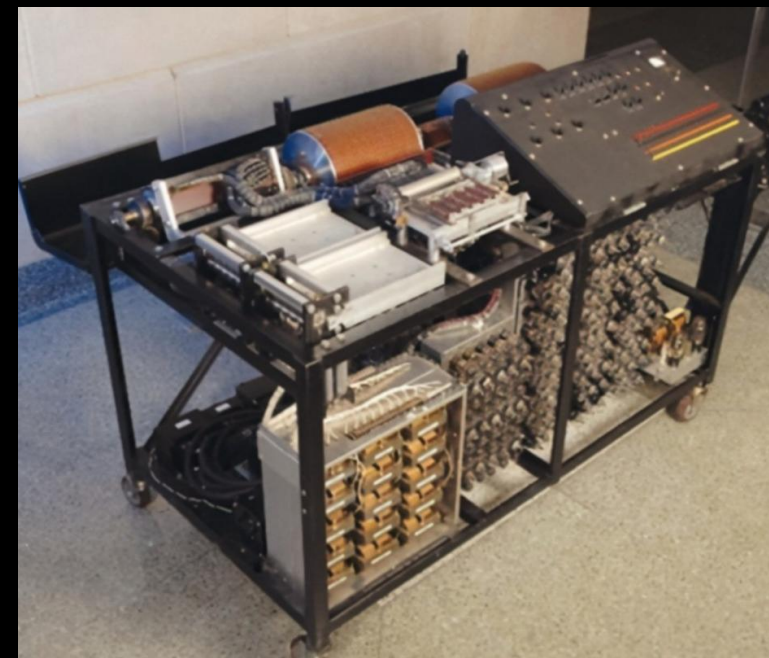


Colossus



Z-3

德国 (1941)

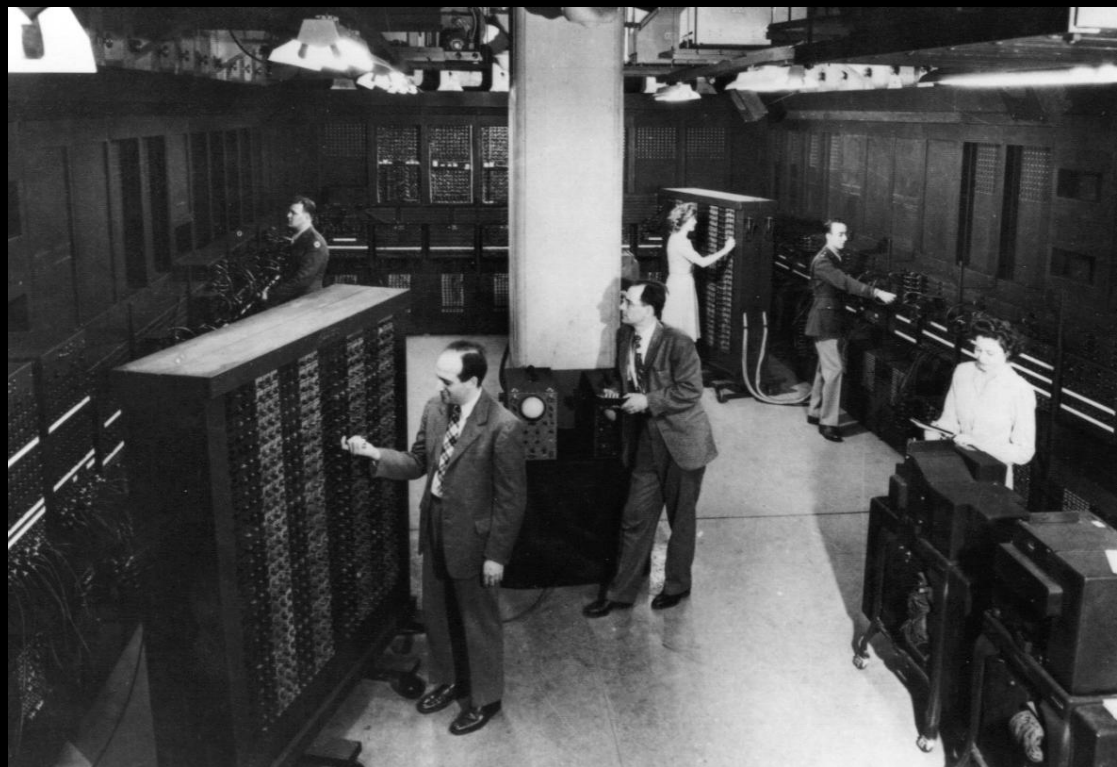


Atanasoff - Berry Computer, ABC

美国 (1940-1942)

• 如何构建高效的计算机？

■ 第一台通用电子计算机ENIAC (Electronic Numerical Integrator And Computer)

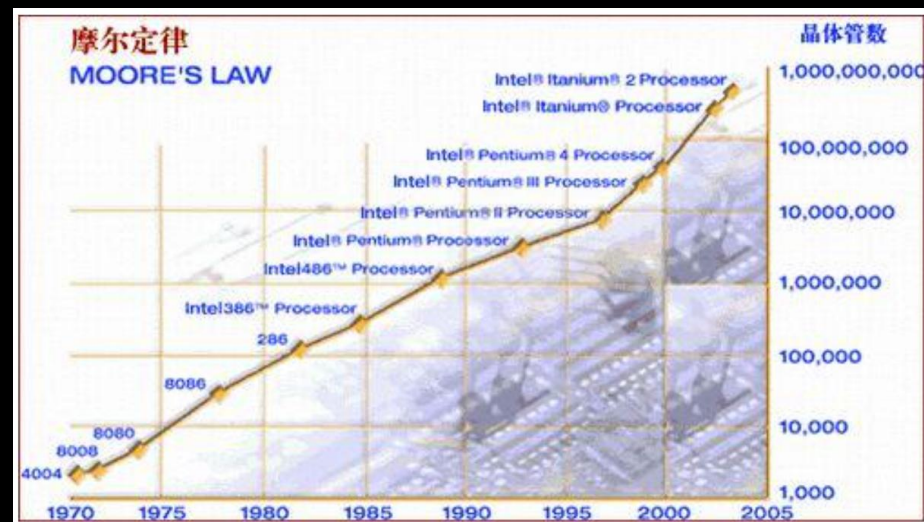
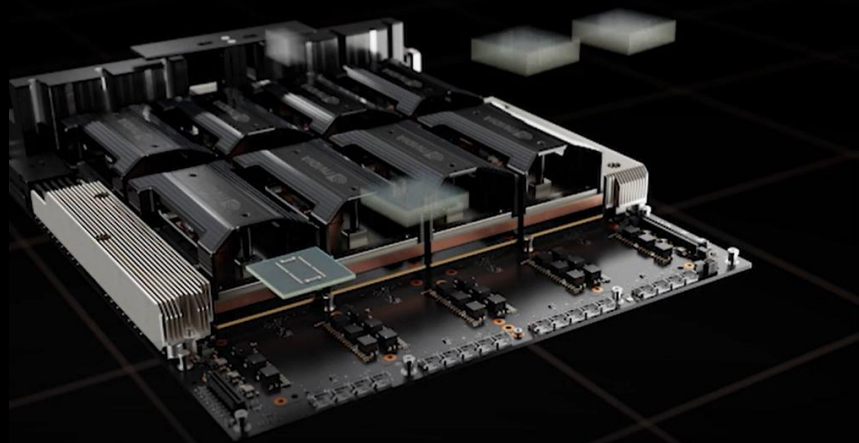


- ✓ 长30.48米，宽6米，高2.4米
- ✓ 占地面积约170平方米，30个操作台
- ✓ 重30英吨，耗电量150千瓦
- ✓ 包含17468根真空管（电子管）7200根水晶二极管，1500个中转，70000个电阻器，10000个电容器，1500个继电器，6000多个开关
- ✓ 计算速度是每秒5000次加法或400次乘法

美国宾夕法尼亚大学
(1946年2月14日)

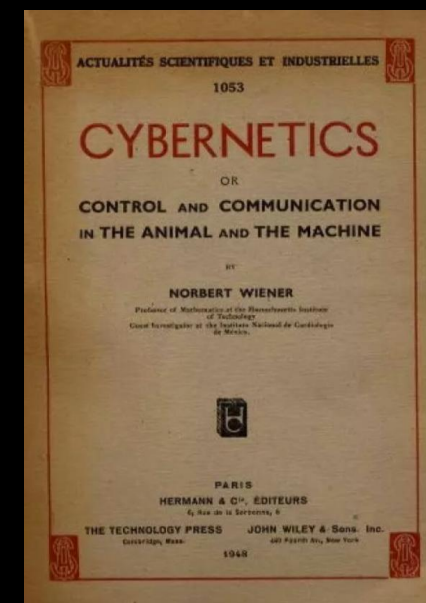
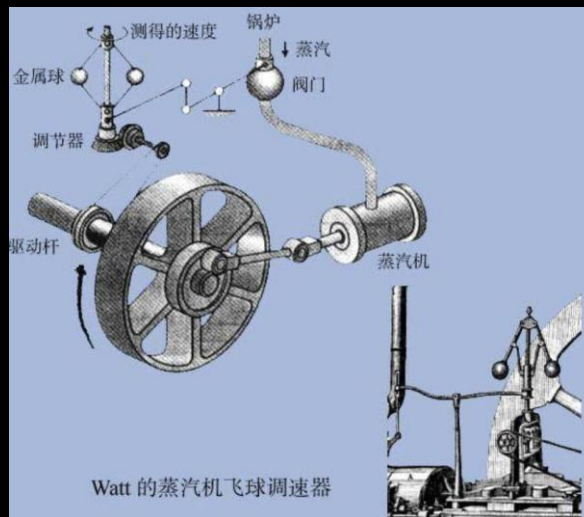
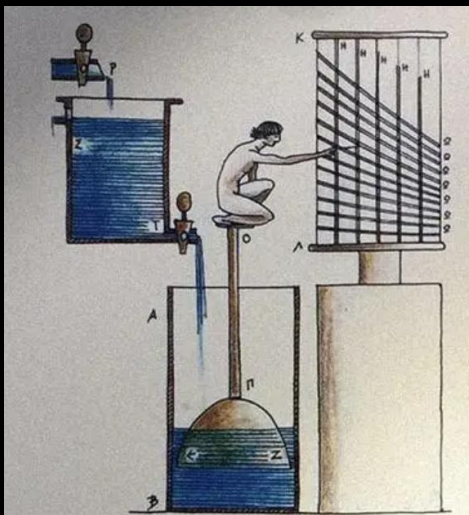
• 如何构建高效的计算机？

- 摩尔定律 (Moore's law): GPU性能18个月翻一番
- 2005年之后增加CPU核数而非提高CPU时钟频率，并行性对于性能提升更加重要
- 图形处理单元 (GPU)、张量处理单元 (TPU)、晶圆级引擎 (WSE)
- 训练机器学习应用的计算能力：2012年后，约100天翻一番
- AI还得益于计算机软件技术的发展：操作系统、编程语言和工具等



• 人造物如何在它们自己的控制下运行？

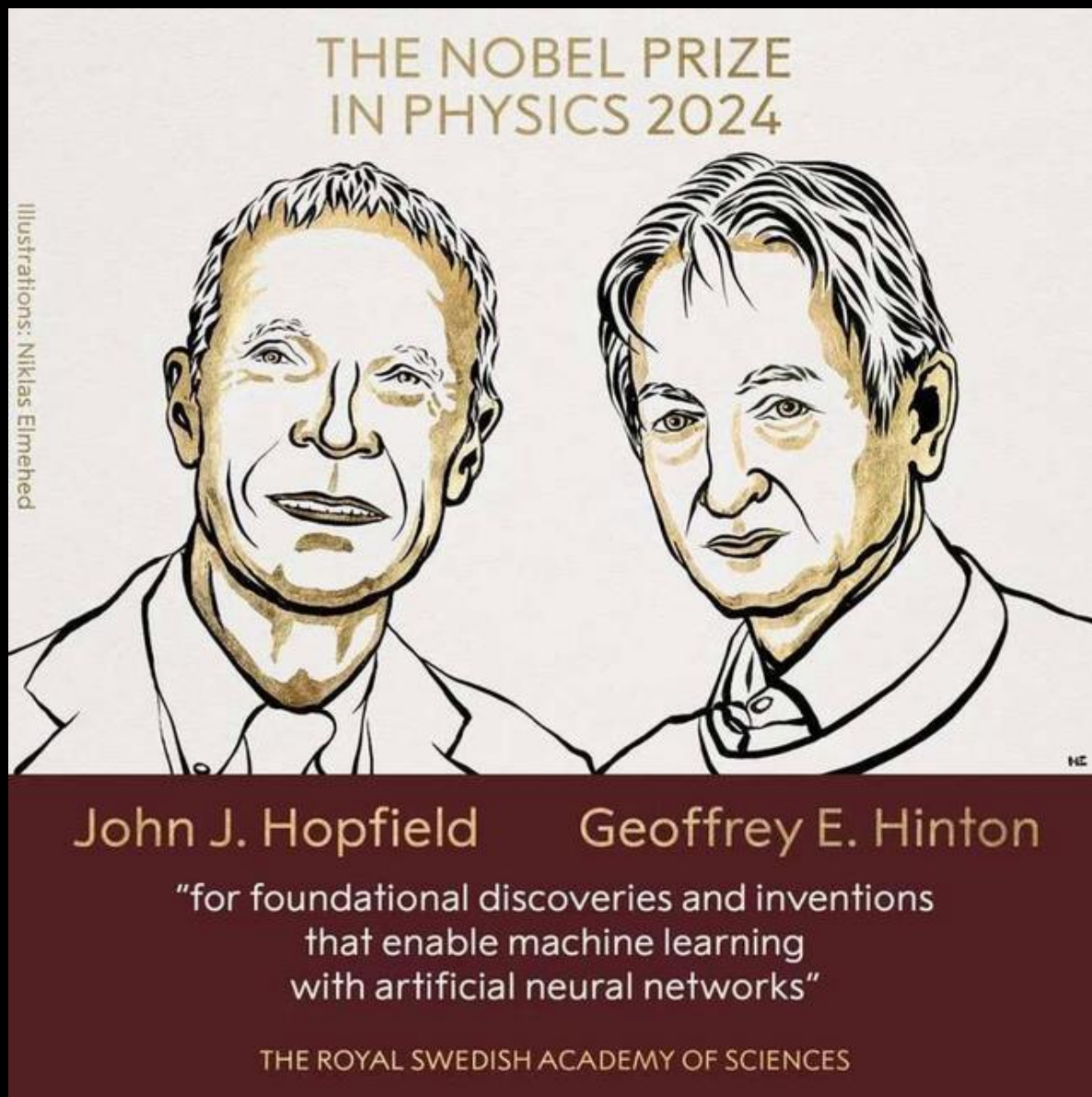
- 古希腊克特西比乌斯(公元前250)建造了第一个自我控制的机器：水钟；自调节反馈控制
- **控制理论**(control theory)：维纳《控制论》(1948)，有目的的行为源于试图最小化“错误”的调节机制，即当前状态和目标状态之间的差异
- 艾什比《大脑设计》：通过自我平衡(homeostatic)设备来实现智能，反馈回路
- **现代控制理论**：设计随时间最小化代价函数(cost function)的系统



- 语言是如何与思维联系的？

- 乔姆斯基《句法结构》(1957)：句法模型为基础的理论，可以解释语言创造力，足够形式化，原则上可被程序化
- 现代语言学
 - 计算语言学(computational linguistics)
 - 自然语言处理(natural language processing)
- 理解语言需要理解主题和上下文，而不仅仅是理解句子结构
- 知识表示(knowledge representation)：如何将知识转化为计算机可以推理的形式的研究

人工智能的基础——9、物理学???



在人工神经网络机器学习方面的基础性贡献

人工智能的历史——AI图灵奖得主

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

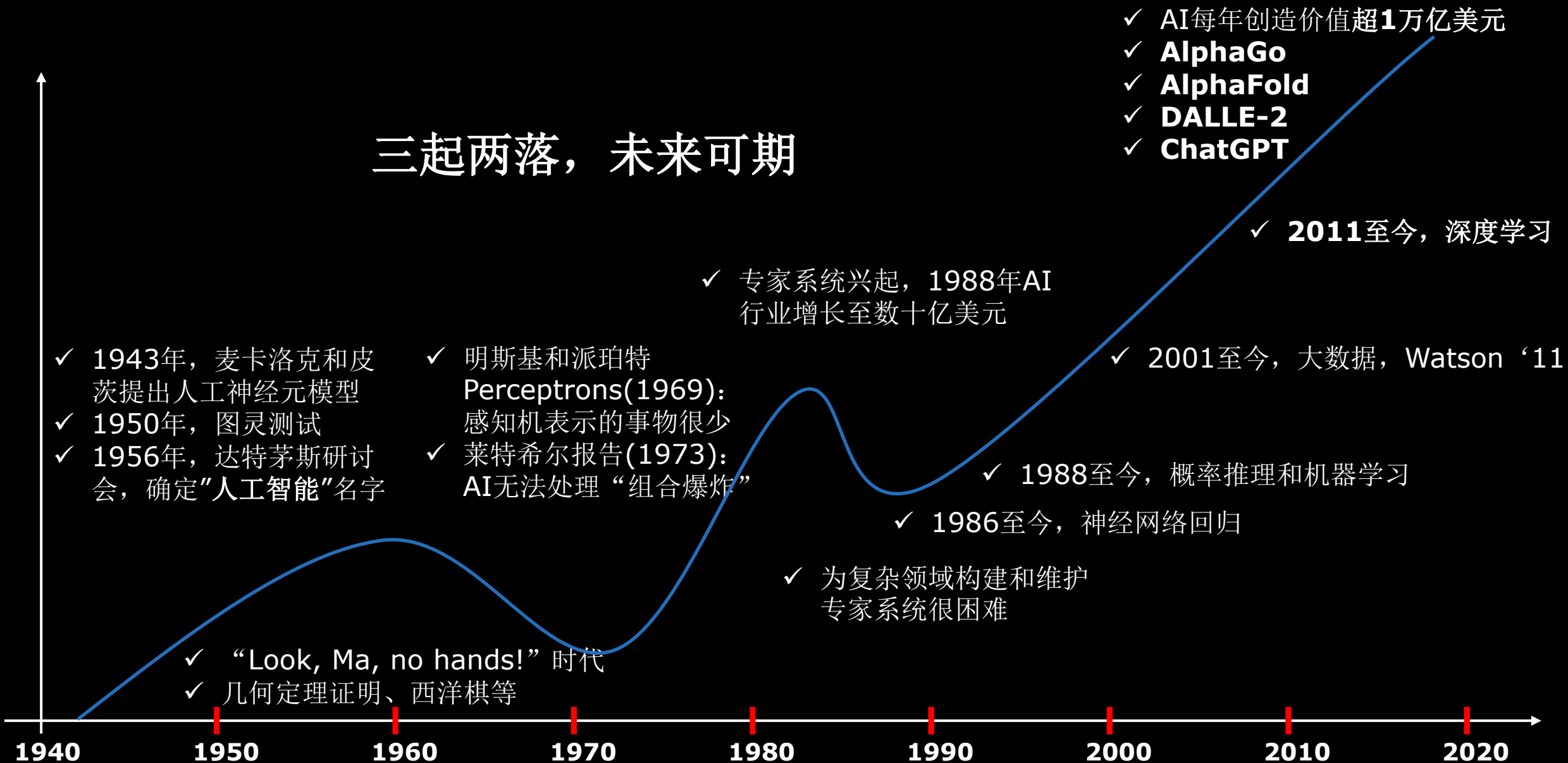


Y. LeCun G. Hinton Y. Bengio
(2018年获奖)

深度神经网络
深度学习

人工智能的历史

三起两落，未来可期

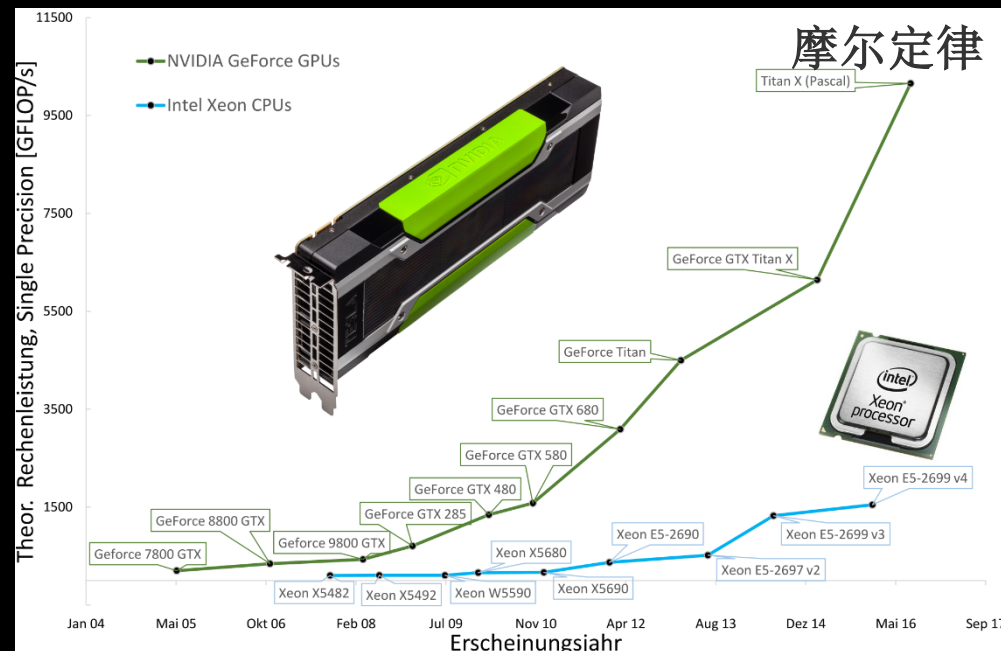


早期智人复原图（约20万年前）
智人脑容量已有1300-1600毫升





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

未来的AI发展更可能是一条“波浪式上升、螺旋式前进”的道路，而非简单的“起落”循环。

- **不会全面寒冬**：因为AI已经创造了不可逆的商业价值，并成为了像“电力”一样的基础设施。
- **但会经历“冷静期”**：当资本市场意识到某些领域的AI投资回报不及预期时，该领域的投资会降温。当技术遇到明显瓶颈时，进展会放缓。这会感觉像是一个“局部寒冬”或“预期低谷”。
- **未来属于融合与落地**：纯粹的模型规模竞赛会放缓，重点将转向：
 - 如何更高效（用更少的数据和算力做更多的事）。
 - 如何更可靠、更安全（解决幻觉、偏见问题）。
 - 如何与具体行业深度结合（在医疗、教育、制造、科研等领域找到刚需场景并创造价值）。
 - 如何与人类协同（AI作为增强人类能力的工具，而非替代）。

总之，AI不再是一个等待破灭的“泡沫”，而是一个正在经历“挤掉水分、夯实价值”过程的成熟领域。它可能会遇到挫折和调整，但整体向上发展的趋势是不可逆转的。下一次的“低落”更可能是一次健康的盘整，淘汰掉浮夸和泡沫，让资源更集中地流向真正有价值的方向，为下一次更具质量的飞跃积蓄力量。

- 课程概括

- 研本课程：3学分
- 考核含：AI基础的平时作业（40%）+专题部分的考核（40%，每个专题5分，任课老师可采用签到、随堂练习、课后作业等形式）+期末大作业（20%）

- 课程内容

- 人工智能基础
- 物理与人工智能的关联（物理中的应用及物理对其的推动）

课程安排

第1周（9月8日） AI概览、 LLM概览、API配置 和Python开发实践	第2周（9月15日） LLM的局限性、提示 工程、Tool calling、 RAG	第3周（9月22日） 逻辑推理、Agent工 作流、基于Agent的 开发实践	第4周（9月29日） 机器学习基础、决 策树与随机森林、 神经网络与优化	第5周（10月13日） CNN、RNN、 Transformer等架构
第6周（10月20日） 模型训练、规律、评 估及实践	第7周（10月27日） 盲目搜索、启发式搜 索、对抗搜索、遗传 算法	第8周（11月3日） AI与凝聚态物理-陈 基	第9周（11月10日） AI 代理模型与天体 物理-王力乐	第10周（11月17日） AI与核技术-林晨
第11周（11月24日） AI与量子多体物理-张 亿	第12周（12月1日） 朱华星	第13周（12月8日） 宋慧超	第14周（12月15日） AI与大规模数值模 拟-冯旭	第15周（12月22日） AI for discovery: 进展、困难及未来