工商联-经济大家讲坛暨第十一期中欧话未来

人工智能
— 现状与展望

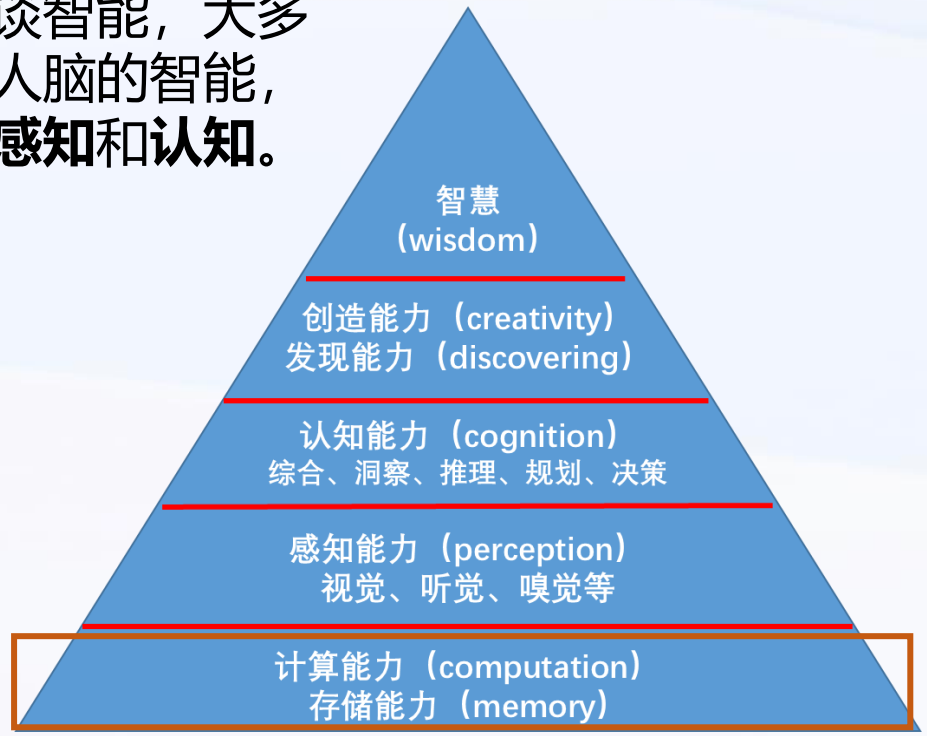
梅 宏 • 2026.01.28 • 上海-中欧国际工商学院

目录

何为“智能”？

- “智能”这一概念本是人类用以**区别自身和其他动物**的专属词。但是，迄今为止，对“智能”还没有公认的定义。
  - 维基百科，Intelligence：抽象、逻辑、理解、自我意识、学习、情感知识、推理、规划、创造力、批判性思维、解决问题。 ——**智能并非单一能力**，而是由多个认知过程协同产生

人们谈智能，大多是指人脑的智能，又分**感知**和**认知**。



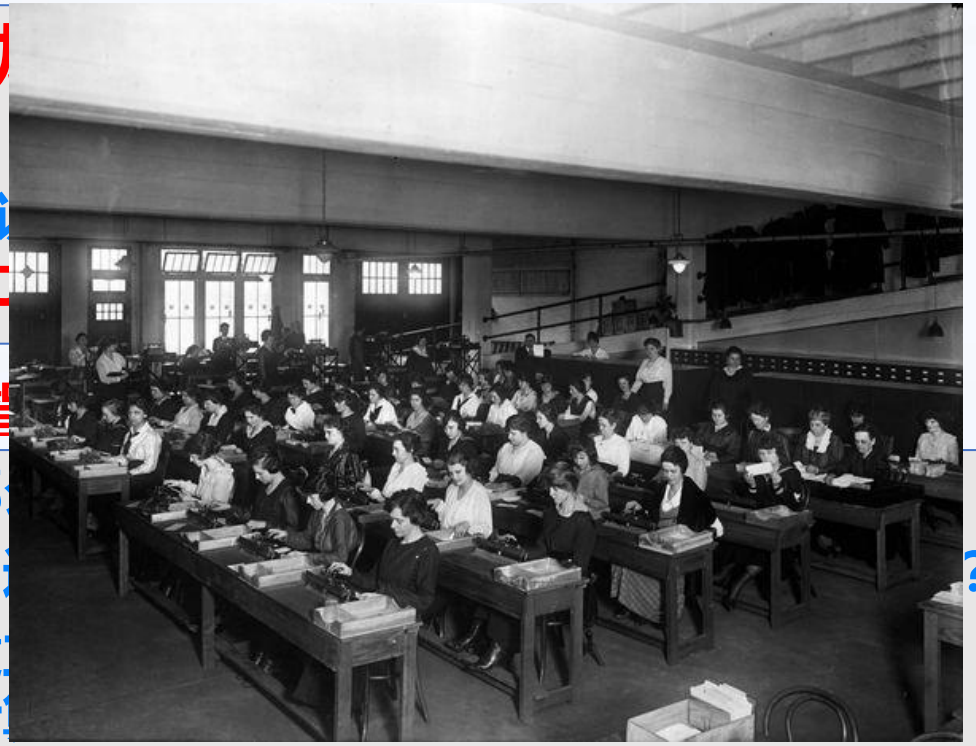
图改自：微软亚洲研究院洪小文

人的计算能力是智能吗？

- 在教育不是那么普及的年代，无疑“是”！；
- 在计算机能力远超人类的今天，还有多少人认为“是”？
- 人类为计算建立了方法和规则，发明了工具，也一直在努力发明会计算的机器：
  - 帕斯卡加法器（1642年）、莱布尼茨机械计算器（1673年）、巴贝奇的差分机（1822年）等
- 1946年，第一台电子计算机ENIAC，比当时最快的机电式计算机Harvard Mark I 计算速度提升“千倍”！相对“人力计算”速度提升“万倍”！
  - Computer原用于人，后被称为“电脑”。

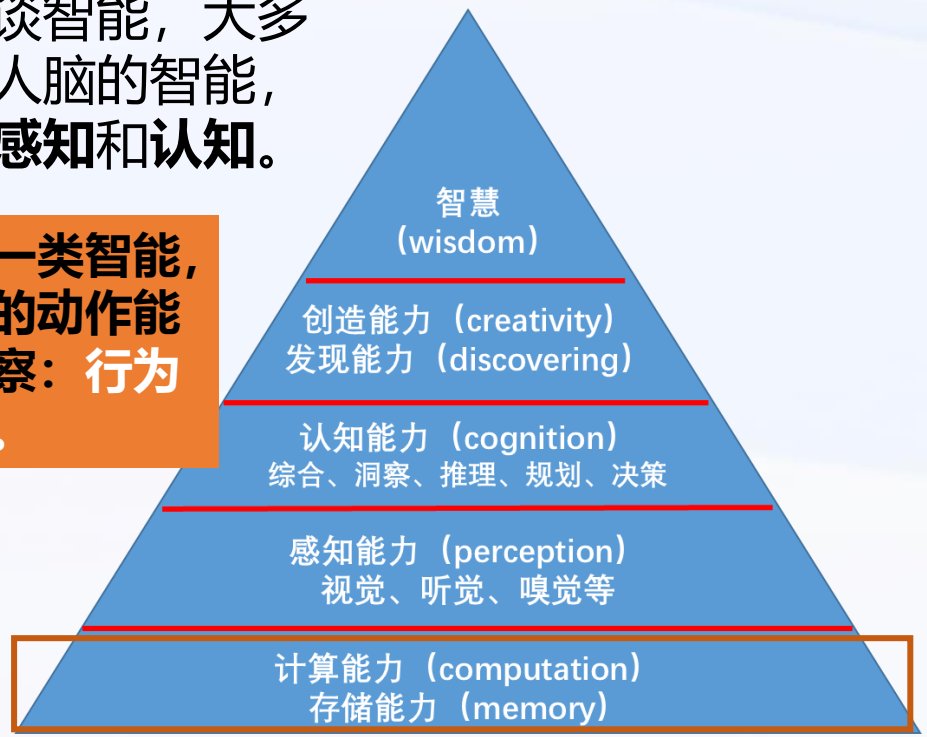
何为“智能”？

- “智能”这一概念本是人类用以区别自身和其他动
  对“智能”还没有公认的定义。
  - 维基百科，Intelligence：抽象、逻辑、理解、自我意
    造力、批判性思维、解决问题。 ——智能并非单一能力



人们谈智能，大多是指人脑的智能，又分**感知**和**认知**。

还有一类智能，从人的动作能力考察：行为智能。



图改自：微软亚洲研究院洪小文

人的计算能力是智

- 在教育不是那
- 在计算机能力
- 人类为计算建
  努力发明会计
  - 帕斯卡加法器（1642年），莱布尼茨机械计算器（1673年），巴贝奇的差分机
- 1946年，第一
  式计算机Harv
  对“人力计算”
  - Computer原

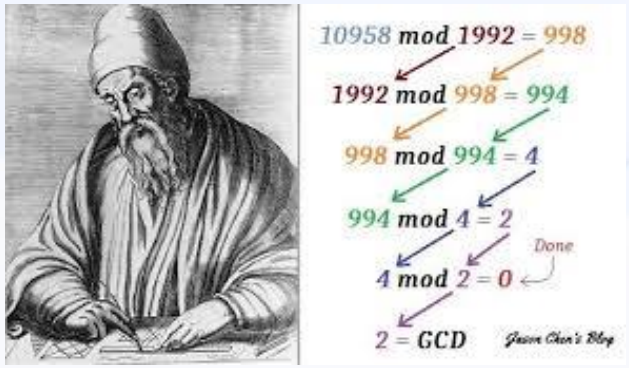
当前世界上最快的计算机是美国劳伦斯国家实验室的“El Captain”（浮点计算速度达到1.742百亿亿次），相比ENIAC的约500次，提升了千万亿倍（10^{15}）以上。

何为“计算”？

- **“计算”（Computation）一词源自拉丁语*computare*。原义指进行数学计算或计数。在古代，计算主要指人类借助工具进行的数值运算，如用算盘计算。**
- **“算法”：大英百科全书，Algorithm，通过有限的步骤产生问题的答案或问题的解决方案的机械过程。**词源自中世纪阿拉伯数学家（拉丁语为Algoritmi）于9世纪编写的《代数学》一书，系统介绍求解方程的方法。

人类对算法和计算过程的认识实际可追溯古代文明中的数学程序化操作：如公元前3世纪求最大公约数的欧几里得算法（辗转相除法），我国《九章算术》中系统的数学（算学）步骤等。

花拉子米（Al-Khwarizmi）
（约780年-约850年）



欧几里得（Euclid）的辗转相除法
（约公元前3世纪）

《九章算术》
（约公元前100年）

理性主义推动的“计算”

莱布尼茨作为理性主义哲学家，笃信世界的运行有其逻辑秩序，人类若能发明出一套完备的符号系统和演算方法，就能机械地解决各种思想问题。

- **“通用语言”（Characteristica Universalis），用精确定义的符号来表示一切知识。**
- **“理性演算”（Calculus Ratiocinator），作为推理的规则系统。**

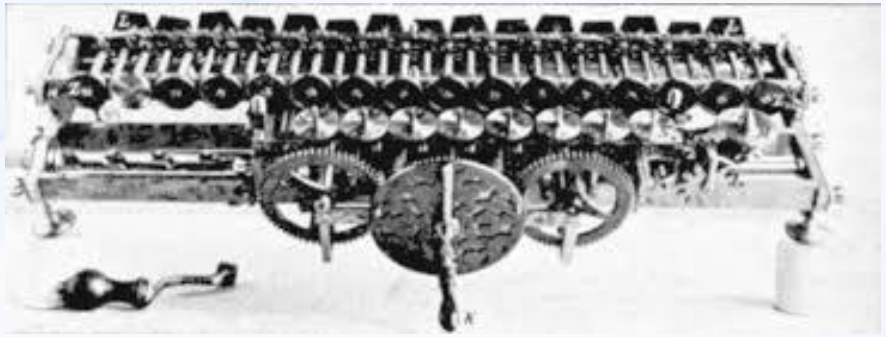
莱布尼茨：“只要计算即可！当出现争议时，我们不必再空论，我们计算一下，看谁是对的。”

——反应了理性主义对机械理性的信心：通过恰当的符号表示和操作，人类的推理可以转化为一种无需直觉猜测的确定机械过程。计算是理性规则的机械运用，蕴含着确定性和可预测性

莱布尼茨（Gottfried Leibniz）
（1646-1716）

“Let us calculate, without further ado, to see who is right”

“通用语言”
（Characteristica Universalis）



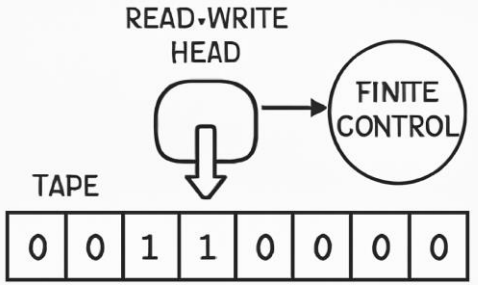
莱布尼茨机械计算器（约1670），作为“理性演算”（Calculus Ratiocinator）的原始阶段载体。

“计算”概念的抽象形式化

- David Hilbert的理性乐观：通过**纯逻辑演绎可彻底奠定数学基础**，提议为全部数学建立一致的公理系统，并证明其无矛盾性和完备性（希尔伯特纲领，Hilbert’s Program，形成于1920年代初）
  - 形式系统的完备（所有真命题可证明）、一致（无矛盾） —— 不成立 （Kurt Gödel，1931）
  - 可判定（存在**机械法**判断命题真伪）—— **不成立** （Alan Turing，1936）
- Alan Turing: 抽象形式化“计算” = “（图灵机）可计算（computable）”
  - 计算是一类**抽象的过程**，可以被视为关于符号操作的形式系统——根据**一定规则对符号进行操作、从输入得出输出的过程；**
  - 计算独立于实现它的具体物理装置，不仅**包括数值计算**，也包括**逻辑推理、算法执行**，乃至自然系统的信息过程。

”我们必须知道，我们必将知道“
David Hilbert（1862-1943）法国数学家

Alan Turing
(1912-1954)
“计算机科学之父”
“人工智能之父”



图灵机概念图

“智能”vs“计算”

- 计算能力一直被视为人类的智能之一！
- 理性主义：人类的认知推理过程也可以实现为基于符号的机械计算过程。

- 就这个意义而言，“计算”就是对“智能”的模拟！

- 一个自然的问题：具有符号机械计算能力的机器**具有智能吗?**
- 换个问法：**“机器能思考吗？”**（Can machine think？）

“图灵”与“机器智能”

- **1948年，图灵讨论了让机器表现出像人一样的智能行为的可能路径。**

Intelligent Machinery

图灵在1948年撰写的报告

A. M. Turing
[1912—1954]

Abstract

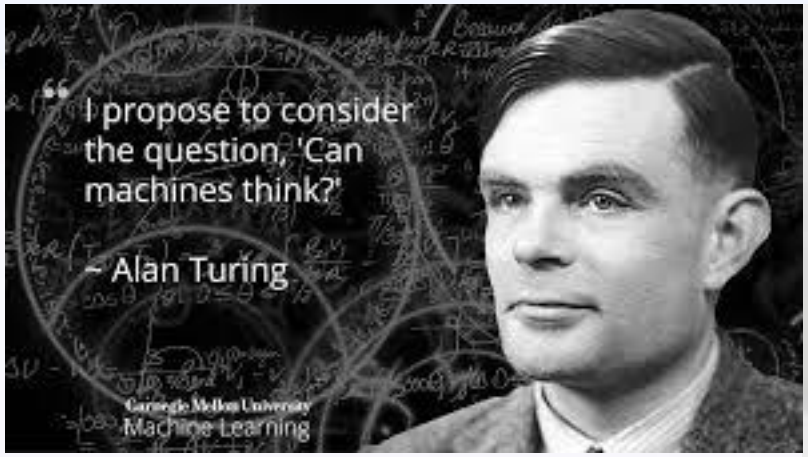
The possible ways in which machinery might be made to show intelligent behaviour are discussed. The analogy with the human brain is used as a guiding principle. It is pointed out that the potentialities of the human intelligence can only be realized if suitable education is provided. The investigation mainly centres round an analogous teaching process applied to machines. The idea of an unorganized machine is defined, and it is suggested that the infant human cortex is of this nature. Simple examples of such machines are given, and their education by means of rewards and punishments is discussed. In one case the education process is carried through until the organization is similar to that of an ACE.

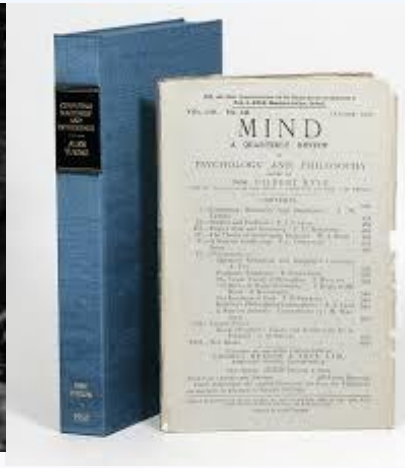
- 与人脑类比
- 人的智力只能通过教育被开发
- 对机器采用类人的教育过程
- 通过奖励和惩罚教育

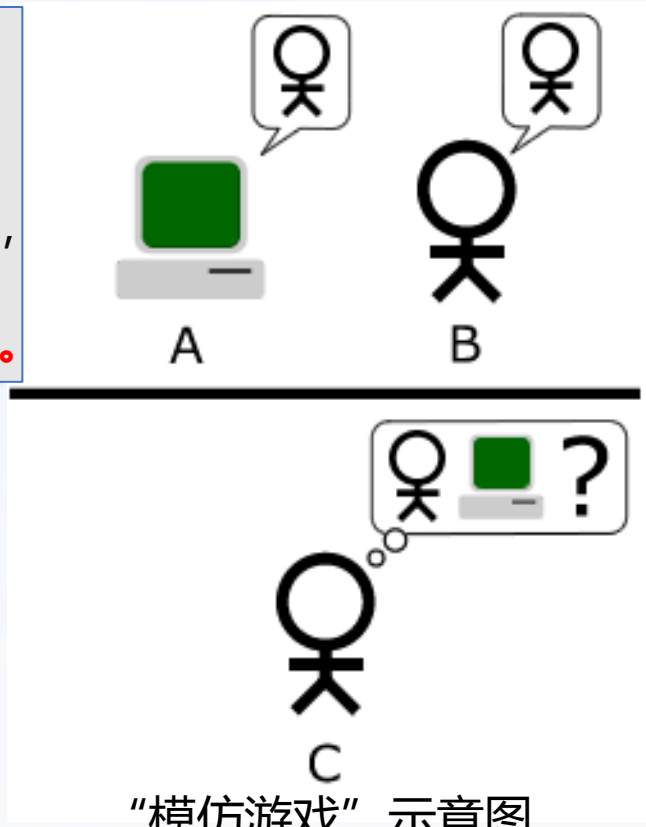
“图灵”与“机器智能”

- **1950年，图灵发表文章Computing Machinery and Intelligence（计算机器与智能），开篇首句：“机器能思考吗（Can machine think）？”**
  - 文中给出了机器智能的测试标准——图灵测试，改自“模仿游戏”。

➢ 模仿游戏（Imitation Game），性别判断游戏：询问者C通过文字对话（不可见对方）来辨别隐藏身份的男性A与女性B，其中A试图误导判断，B则协助正确识别。图灵将其改良为人工智能测试——当用计算机替代男性A的角色时，若询问者误判计算机为人类的概率与在原版游戏中混淆性别的概率相当，则视为机器通过了智能测试。**通过对话交互判断机器能否模拟人类思维的实验设计，成功将抽象的“机器能否思考”哲学问题转化为可操作的实证检验标准。**







“模仿游戏”示意图

“人工智能”一词的诞生

- “人类智能”：大英百科全书，Human intelligence，人的脑力素质，包括从经验中学习、适应新情况、理解和处理抽象概念、以及应用知识去操控所处环境的能力。
- “人工智能”：Artificial Intelligence，维基百科，指计算系统执行通常与人类智能相关的任务的能力，例如推理、解决问题、学习、感知和决策。

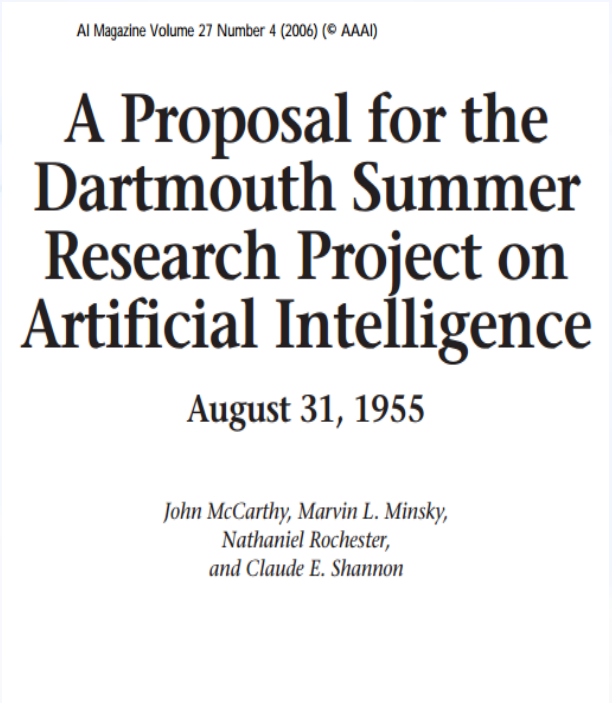
➢ **AI术语诞生：**

1956年达特茅斯会议

➢ 会议提案：“学习的每个方面或智力的任何其他特征都可以被精确地描述，以便机器可以模拟它。”（every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it）

——让机器模拟人类智能之研究领域

1956年达特茅斯会议主要参会人员


AI Magazine Volume 27 Number 4 (2006) (© AAAI)

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

August 31, 1955

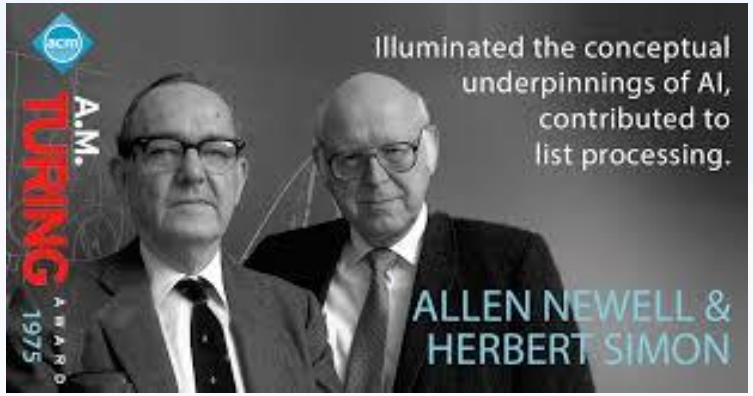
John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon

“符号主义” vs “连接主义”

- **如何模拟“智能”？——两个对立的观点：“符号主义”和“连接主义”**
  - **符号主义：与理性主义一脉相承，对智能来源的假设为“符号主义基本假设”**
  - **连接主义：与经验主义更相关，对智能来源的假设为“分布式并行处理”假设**

符号主义基本假设：物理符号系统假说（1976年正式提出，Physical Symbol System Hypothesis, PSSH）

- 核心主张：“物理符号系统具有**产生智能行为**的必要和充分条件。”——强调心智即计算



Allen Newell，Herbert Simon
共同获得1975年图灵奖

连接主义基本假设：分布式并行处理（1986年正式提出，Parallel Distributed Processing，PDP）

- 核心主张：“智能源于大量简单处理单元之间的相互作用，每个单元向其他单元发送兴奋和抑制信号。”——强调智能的涌现性和适应性

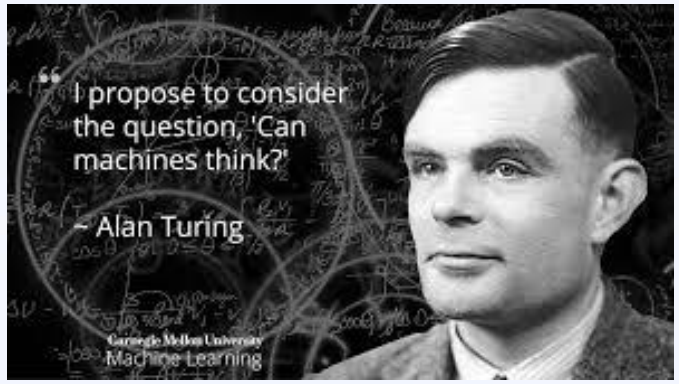
（解读：大量神经元通过加权连接及权重的调整形成的分布式并行计算可以产生智能）

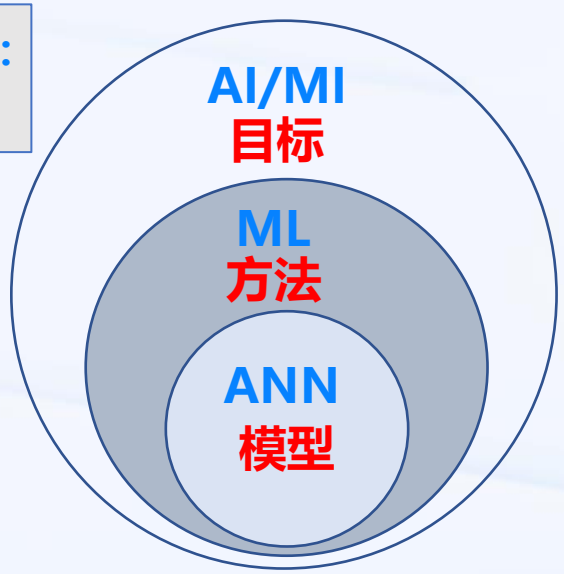
David Rumelhart（左）
美国心理学家

James McClelland（右）
美国心理学家

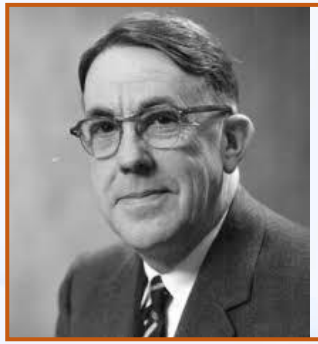
四个重要术语及其关系

- **机器智能（Machine Intelligence，MI）：指机器所展现的智能行为概念**





- **机器学习（Machine Learning，ML）：AI子领域，研究如何让机器通过数据/经验自动改进性能（智能的适应性）**

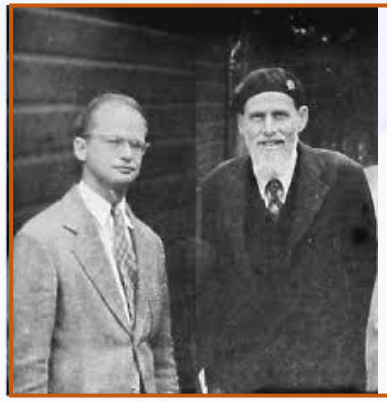


Arthur Samuel
“ML”一词提出者，1959年。
提出了“西洋跳棋”AI，→
AlphaGo AI （相同框架）。

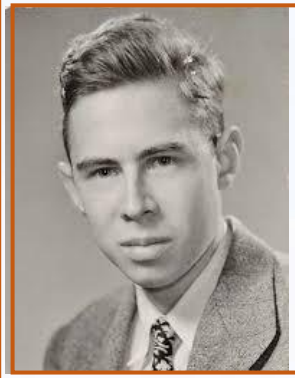
- **人工智能（Artificial Intelligence，AI）：让机器模拟人类智能之研究领域**

John McCarthy
“AI”一词提出者，1956年。
1971年图灵奖获得者。

- **人工神经网络（Artificial Neural Network，ANN）：受生物神经系统启发的计算模型，大量人工神经元相互连接组成**



Warren McCulloch
（右）和
Walter Pitts （左）
提出“人工神经元模型”，1943年



Frank
Rosenblatt
1957年提出第一个可学习的ANN“感知机”

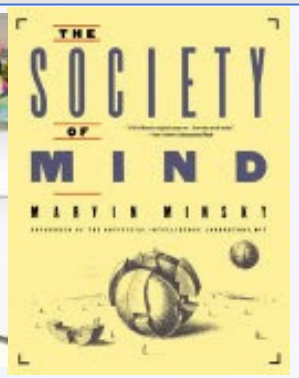
目录

从“知识”视角看待“人工智能”

知识：涉“认识论”（Epistemology）和“本体论”（Ontology）两个维度

- **认识论维度：知识如何获取？**
  - **理性主义**：倾向在模型中注入先验结构、规则、逻辑或公理体系，通过内在的推理或演绎来获得知识。
  - **经验主义**：倾向依赖外部数据、大规模统计、环境交互来“经验地”归纳学习。

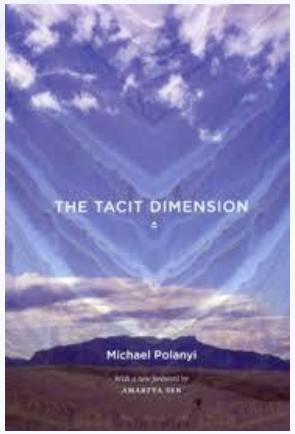
- **本体论维度：知识如何表示？**
  - **显式表示**：知识能被人类或符号系统直接解读、修改和推理，可读性和可解释性更强。
  - **隐式表示**：知识常以参数、权重或概率分布的形式存在，难以直接被人类理解或编辑。



Marvin Minsky
人工智能领域奠基者
1969年图灵奖

《The Society of Mind》
《心智社会》，1986

“我们最不了解自己的大脑最擅长做什么，人类最难逆向的技能是那些**低于意识水平**的技能”（The Society of Mind，1986）——**Marvin Minsky**



Michael Polanyi
匈牙利-英籍博学家
两个学生、一个儿子
获诺奖

《The Tacit Dimension》
《默会维度》，1966

人工智能技术：知识获取与表示

知识显式表示

归纳获得知识

象限(III)　经验主义+显式

1980s-2000s

归纳逻辑程序设计（ILP,1990s）

决策树 / 规则学习 (1980s-1990s)

部分专家系统（机器学习规则）

象限(I)　理性主义+显式

1950s-1980s

符号主义-基于逻辑推理的 AI (1950s)

专家系统　(1980s)

知识图谱 + 逻辑推理（基于先验构建）

象限(IV)　经验主义+隐式

现代

连接主义AI-神经网络 / 深度学习

（部分）强化学习（RL）　(现代)

演化算法（Evolutionary Algorithms）

群体智能（Swarm Intelligence）

象限(II)　理性主义+隐式

1980s-至今

带先验的贝叶斯模型 / 概率图模型（PGM）(1980s~至今)

在神经网络中嵌入先验结构的混合方法　(现代)

知识隐式表示

演绎获得知识

-1950s　AI思想孕育期

1960s　**符号主义兴起期**

1970s　AI第一次寒冬 神经网络低谷

1980s　专家系统热潮与寒冬 机器学习理论奠基 神经网络复苏

1990s　机器学习崛起期

2000s　数据与互联网时代

2010s　**深度学习革命**

2017-2020s　Transformer 与 生 成 式 AI

人工智能早期的乐观预期

HEURISTIC PROBLEM SOLVING: THE NEXT ADVANCE IN OPERATIONS RESEARCH*

Herbert A. Simon and Allen Newell
Carnegie Institute of Technology, Pittsburgh, Pennsylvania, and The Rand Corporation, Santa Monica, California

THE IDEA THAT the development of science and its application to human affairs often requires the cooperation of many disciplines and professions will not surprise the members of this audience. Operations research and management science are young professions that are only now beginning to develop their own programs of training; and they have meanwhile drawn their practitioners from the whole spectrum of intellectual disciplines. We are mathematicians, physical scientists, biologists, statisticians, economists, and political scientists.

Allen Newell和Herbert Simon 在1957年乐观预测

- **十年内**，数字计算机将成为世界国际象棋冠军，除非比赛规则禁止其参赛。 —1997年“深蓝”
- **十年内**，数字计算机将发现并证明一个重要的新数学定理。
- **十年内**，数字计算机将创作出被评论家认为具有相当审美价值的音乐作品。
- **十年内**，大多数心理学理论将以计算机程序的形式出现，或以关于计算机程序特性的定性陈述为基础。
- **可预见的未来内，机器能处理人类能处理的所有问题。**

— 除了国际象棋的预测，其余预测至今尚未完全实现

“在3到8年内，我们将拥有一台具有普通人一般智能的机器。我的意思是，一台能够阅读莎士比亚作品、给汽车换油、参与办公室政治、讲笑话、打架的机器。到那时，这台机器将开始以惊人的速度自我教育。几个月后，它将达到天才水平，再过几个月，它的力量将无法估量。”——**Marvin Minsky**

在1970年接受生活杂志采访时的乐观预期

“生活杂志”（LIFE）是美国一本具有重大影响力的新闻和摄影杂志，由美国时代出版公司出版，以高质量的摄影报道和深入的新闻分析著称。

媒体的乐观预期——对感知机的系列报道

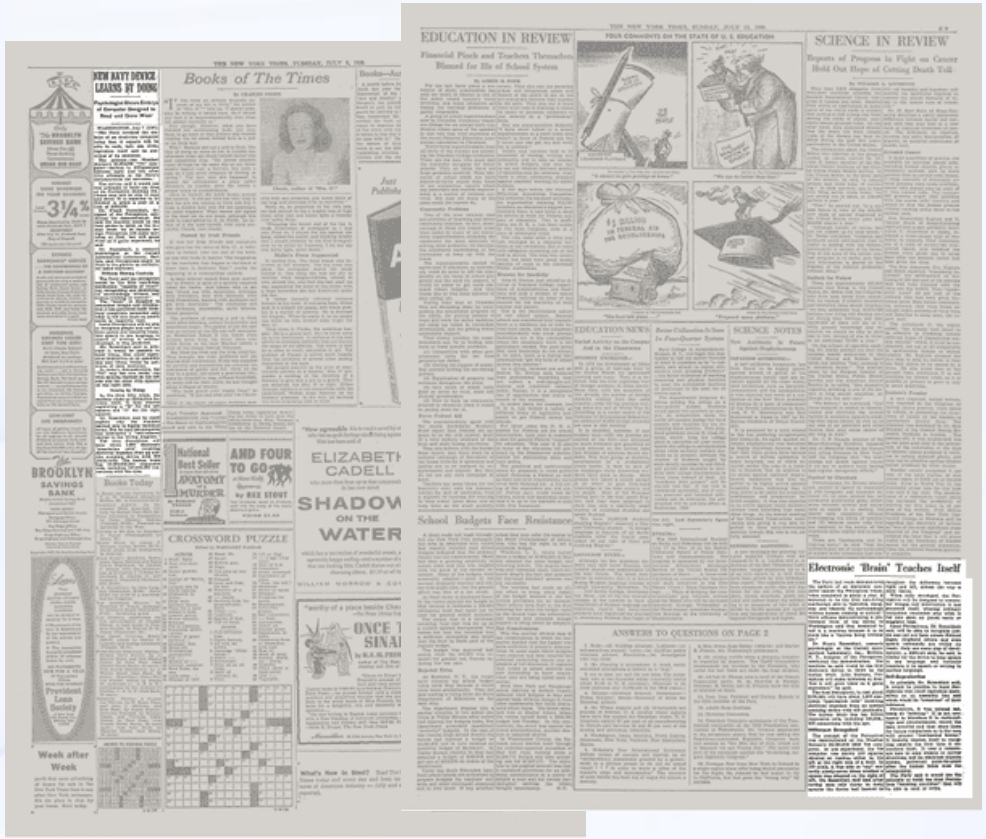
1957年7月8日和7月13日，美国纽约时报（New York Times）发表了两篇报道，表达了对基于“感知机”模型的电子计算机产生高级智能行为的乐观预期

The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

海军今天公布了一种电子计算机的胚胎，预计它将能够行走、说话、看、写、**自我复制并意识到自己的存在**。

The Navy last week demonstrated the embryo of an electronic com- puter named the Perceptron which, when completed in about a year, is expected to be the first non-living mechanism able to “perceive, recognize and identify its surroundings without human training or control.”

海军上周展示了一种名为感知器的电子计算机的胚胎，预计在大约**一年内完成**，这**将是第一个能够“在没有人类训练或控制的情况下感知、识别和识别周围环境”的非生命机制**



“纽约时报”创办于1851年，是美国历史最悠久的报纸之一。其影响力和读者群覆盖全球，在业内一直被视为美国全国性的“档案记录报”，是全美印刷发行量第二大的报纸。

媒体的乐观预期——出现担忧与恐慌

1957年6月9日，美国纽约时报（New York Times），发表了一篇报道，讨论了人工智能（电子“大脑”）可能的前景，并对可能带来的失业等社会问题提出了担忧。

Certainly the most dramatic thing about it ——and the most frightening thing from the standpoint of human society ——is the possibility of a completely automatic factory operating without human hands and governed primarily by electronic "brains". The key area of social change stimulated by automation is employment. Everywhere, one can find two things: a positive emphasis on opportunity and a keen sensitivity to change, often translated more concretely into fear.

当然，最神奇的事情——从人类社会的角度来看也是最可怕的事情——是一个完全自动化的工厂有可能在没有人手的情况下运行，主要由**电子“大脑”**管理。自动化刺激社会变革的关键领域是**就业**。在任何地方，人们都能发现两件事：对机遇的重视和对变化的敏感，而后者往往更具体地转化为恐惧。

“纽约时报”创办于1851年，是美国历史最悠久的报纸之一。其影响力和读者群覆盖全球，在业内一直被视为美国全国性的“档案记录报”，是全美印刷发行量第二大的报纸。

1964年2月24日，“美国新闻与世界报道”（U.S. News & World Report，简称：美新周刊），发表了一篇报道，表达了对人工智能（电子大脑）被滥用、带来失业、甚至威胁的担忧。

There‘s a boom in “electronic brains”--and it’s skyrocketing. They‘re taking over in offices, factories, banks, governments. Is it all to the good? Are computers being misused, even threatening danger?

“电子大脑”正在迅猛地兴起。它们正在逐步占领办公室、工厂、银行和政府机构。这一切是好事吗？计算机是否**被滥用**，甚至带来威胁性的危险？

“美新周刊”是一本仅次于“时代杂志”和美国“新闻周刊”的美国第三大新闻杂志。

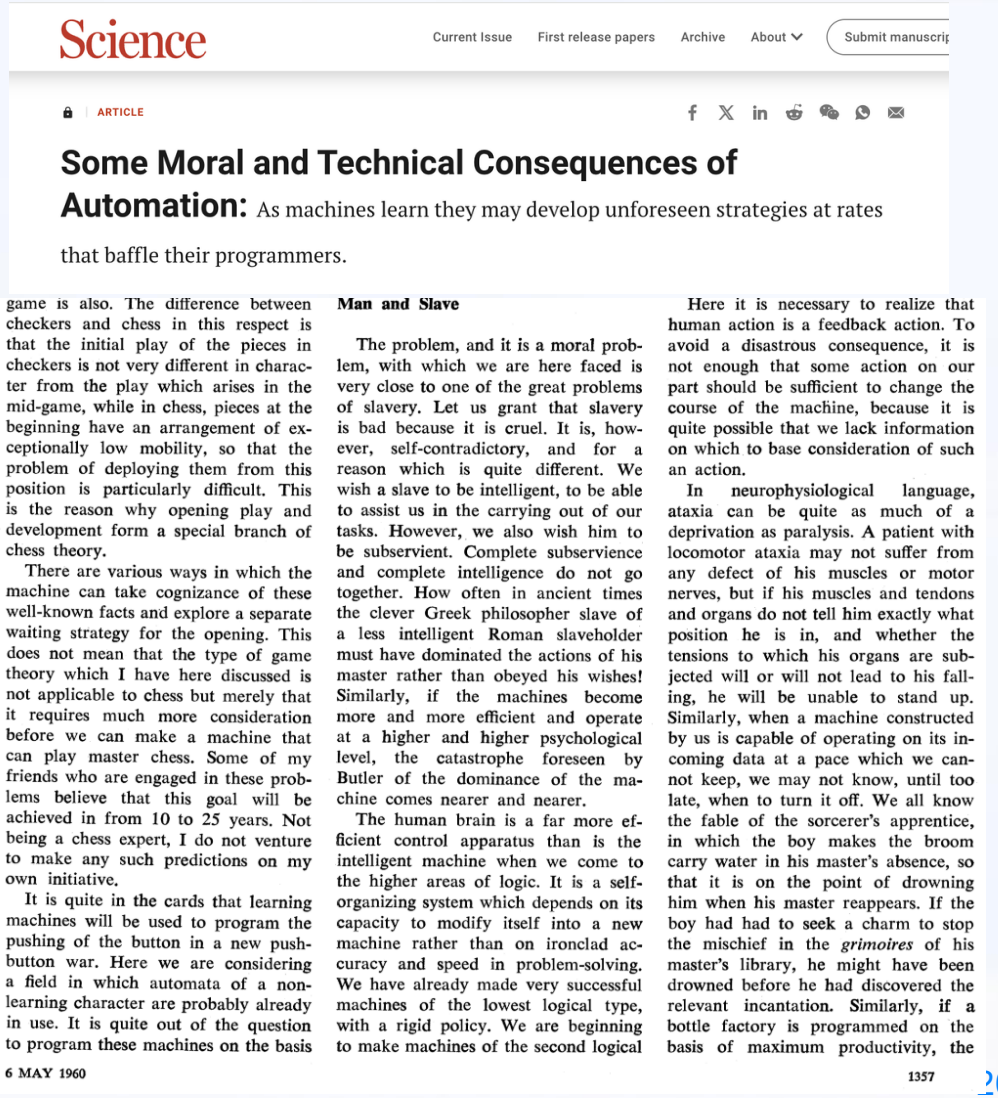
媒体的乐观预期——出现担忧与恐慌

1960年，“科学期刊”（Science）发表了一篇文章，作者认为机器智能（人工智能）可能**脱离人类**的控制，存在威胁人类生存的潜在风险。

Similarly, if the machines become more and more efficient and operate at a higher and higher psychological level, the catastrophe foreseen by Butler of the dominance of the machine comes nearer and nearer.

同样地，随着机器变得越来越高效，并且在心理水平上达到更高层次，巴特勒预见的**机器统治人类的灾难**也将越来越逼近。

“科学期刊” 由美国科学促进会（American Association for the Advancement of Science, AAAS）出版，在学术界享有极高的声誉，是全球科学界公认的顶级期刊之一。



Science

Current Issue First release papers Archive About Submit manuscrip

ARTICLE

Some Moral and Technical Consequences of Automation: As machines learn they may develop unforeseen strategies at rates that baffle their programmers.

game is also. The difference between checkers and chess in this respect is that the initial play of the pieces in checkers is not very different in character from the play which arises in the mid-game, while in chess, pieces at the beginning have an arrangement of exceptionally low mobility, so that the problem of deploying them from this position is particularly difficult. This is the reason why opening play and development form a special branch of chess theory.

There are various ways in which the machine can take cognizance of these well-known facts and explore a separate waiting strategy for the opening. This does not mean that the type of game theory which I have here discussed is not applicable to chess but merely that it requires much more consideration before we can make a machine that can play master chess. Some of my friends who are engaged in these problems believe that this goal will be achieved in from 10 to 25 years. Not being a chess expert, I do not venture to make any such predictions on my own initiative.

It is quite in the cards that learning machines will be used to program the pushing of the button in a new push-button war. Here we are considering a field in which automata of a non-learning character are probably already in use. It is quite out of the question to program these machines on the basis

Man and Slave

The problem, and it is a moral problem, with which we are here faced is very close to one of the great problems of slavery. Let us grant that slavery is bad because it is cruel. It is, however, self-contradictory, and for a reason which is quite different. We wish a slave to be intelligent, to be able to assist us in the carrying out of our tasks. However, we also wish him to be subservient. Complete subservience and complete intelligence do not go together. How often in ancient times the clever Greek philosopher slave of a less intelligent Roman slaveholder must have dominated the actions of his master rather than obeyed his wishes! Similarly, if the machines become more and more efficient and operate at a higher and higher psychological level, the catastrophe foreseen by Butler of the dominance of the machine comes nearer and nearer.

The human brain is a far more efficient control apparatus than is the intelligent machine when we come to the higher areas of logic. It is a self-organizing system which depends on its capacity to modify itself into a new machine rather than on ironclad accuracy and speed in problem-solving. We have already made very successful machines of the lowest logical type, with a rigid policy. We are beginning to make machines of the second logical

Here it is necessary to realize that human action is a feedback action. To avoid a disastrous consequence, it is not enough that some action on our part should be sufficient to change the course of the machine, because it is quite possible that we lack information on which to base consideration of such an action.

In neurophysiological language, ataxia can be quite as much of a deprivation as paralysis. A patient with locomotor ataxia may not suffer from any defect of his muscles or motor nerves, but if his muscles and tendons and organs do not tell him exactly what position he is in, and whether the tensions to which his organs are subjected will or will not lead to his falling, he will be unable to stand up. Similarly, when a machine constructed by us is capable of operating on its incoming data at a pace which we cannot keep, we may not know, until too late, when to turn it off. We all know the fable of the sorcerer's apprentice, in which the boy makes the broom carry water in his master's absence, so that it is on the point of drowning him when his master reappears. If the boy had had to seek a charm to stop the mischief in the *grimoires* of his master's library, he might have been drowned before he had discovered the relevant incantation. Similarly, if a bottle factory is programmed on the basis of maximum productivity, the

6 MAY 1960 1357

2000s: 互联网催生大数据

互联网及其延伸导致大数据现象。四个主要驱动力：

- 摩尔定律驱动的指数增长模式
- 技术低成本化驱动的万物数字化
- 宽带移动泛在互联驱动的人机物广泛联接
- 云计算模式驱动的数据大规模汇聚

全球数据呈指数增长

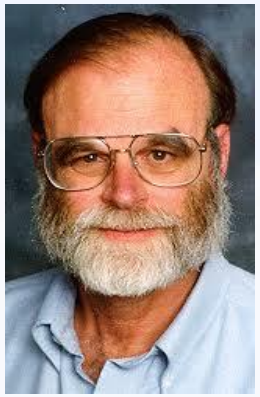
- 2003年全球产生数据仅 5百万 TB
- 2009年全球产生数据约 0.8 ZB
- 2012年全球产生数据约 2.8 ZB
- 2020年全球产生数据约 44 ZB
- 2030年全球产生数据约 2500 ZB

IDC 2021年报告

第四范式：数据密集型科学

基于大数据分析来认知客观世界复杂系统

1998年图灵奖获得者Jim Gray，在2007年1月最后一次公开报告中首次提出了科学发现的四种范式



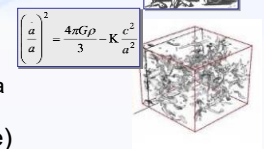
Jim Gray
1998年图灵奖

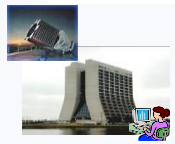
Jim Gray on eScience:
A Transformed Scientific Method

Based on the transcript of a talk given by Jim Gray to the NRC-CSTB[1] in Mountain View, CA, on January 11, 2007[2]

Science Paradigms

- Thousand years ago:
  science was **empirical**
  describing natural phenomena
- Last few hundred years:
  theoretical branch
  using models, generalizations
- Last few decades:
  a **computational** branch
  simulating complex phenomena
- Today:
  data exploration (eScience)
  unify theory, experiment, and simulation
  – Data captured by instruments
    Or generated by simulator
  – Processed by software
  – Information/Knowledge stored in computer
  – Scientist analyzes database / files
    using data management and statistics





-1950s AI思想孕育期

1960s 符号主义兴起期

1970s AI第一次寒冬
神经网络低谷

1980s 专家系统热潮与寒冬
机器学习理论奠基
神经网络复苏

1990s 机器学习崛起期

2000s 数据与互联网时代

2010s 深度学习革命

2017-2020s Transformer与生成式AI

2000s：深度学习革命前夕，各方面条件逐渐成熟

象限(IV)　经验主义+隐式

连接主义AI-神经网络 / 深度学习

（部分）强化学习（RL）（现代）

演化算法（Evolutionary Algorithms）

群体智能（Swarm Intelligence）

- **知识获取**：重度依赖外部数据、环境交互或统计学习；强调**在“大数据”或广泛采样中归纳知识。**
- **知识表示**：多为分布式或隐式；在模型中往往以参数或权重的形式存在，难以通过符号方式直接读取。

数据驱动的AI/ML/ANN可将海量数据转化为模式识别的感知能力。

算法

A fast learning algorithm for deep belief nets *

Geoffrey E. Hinton and Simon Osindero
Department of Computer Science University of Toronto
10 Kings College Road
Toronto, Canada M5S 3G4
{hinton, osindero}@cs.toronto.edu

Yee-Whye Teh
Department of Computer Science
National University of Singapore
3 Science Drive 3, Singapore, 117543
tehyw@comp.nus.edu.sg

Geoffrey Hinton 等**2006年**提出的深度信念网（DBN）突破了用反向传播算法训练深层网络。

计算架构

NVIDIA CUDA

Nvidia 2006年发布CUDA，奠定了GPU通用计算基础。**吴恩达** 等2008年利用CUDA加速深度学习训练.

数据标注

李飞飞 团队**2009年**发布了百万级图像标注数据集，为视觉识别算法提供了“燃料”。

-1950s　AI思想孕育期

1960s　符号主义兴起期

1970s　AI第一次寒冬
神经网络低谷

1980s　专家系统热潮与寒冬
机器学习理论奠基
神经网络复苏

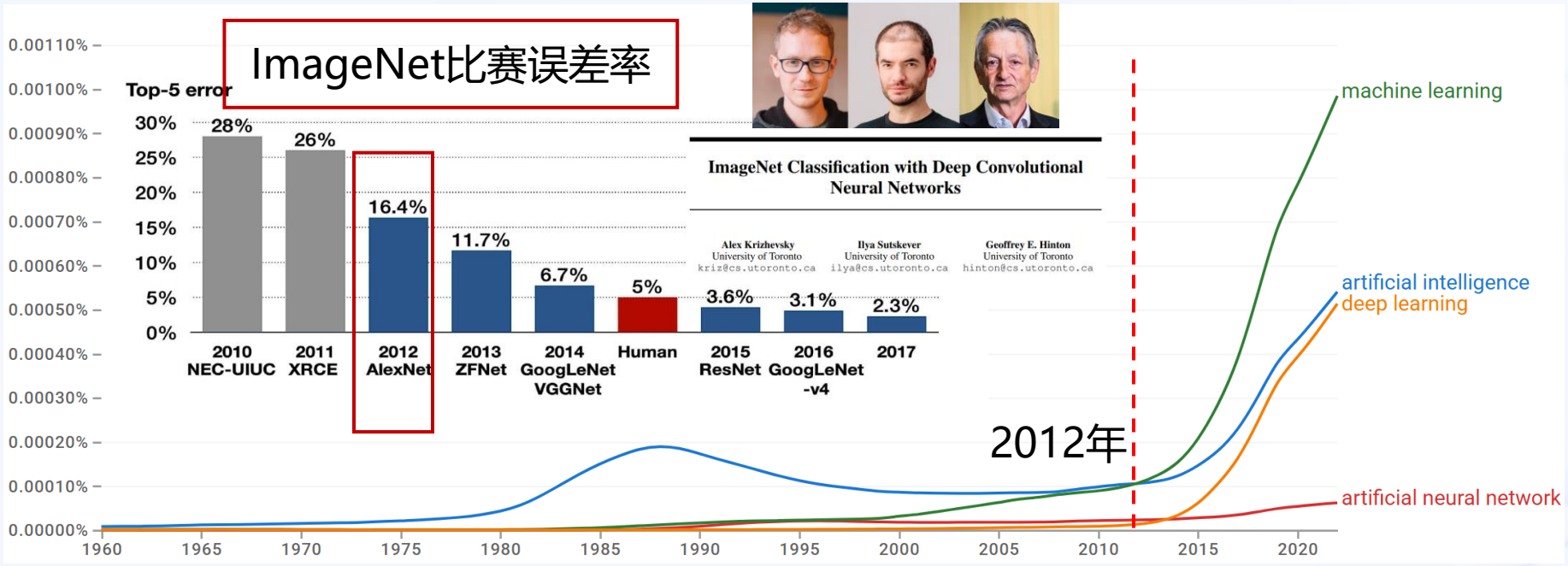
1990s　机器学习崛起期

2000s　数据与互联网时代

2010s　深度学习革命

2017-2020s　Transformer与生成式AI

2010s：深度学习革命，AI研究的第三次春天



谷歌图书Ngram字符串词频查看器——**2012年AlexNet是深度学习的“引爆点”**

影响方面	影响体现	长期影响
技术突破	深度卷积网络大规模数据上的成功	CNN成为计算机视觉的绝对主流算法
硬件应用	首次大规模采用GPU训练深度网络	Nvidia GPU 成为 AI 领域的核心硬件
研究生态变化	深度学习研究从冷门迅速变主流	**大量资金、人才、资源涌入深度学习**
产业影响	产业界迅速转向深度学习技术	Google、Facebook等大公司积极布局

2017s-2020s：Transformer与生成式AI带来热潮高点

- **在判别式AI稳步发展的同时，2017年，Google团队在《Attention is all you need》一文中提出了Transformer深度学习模型架构**
  - Transformer 的出现使得 LSTM 等传统序列模型逐渐让位于基于自注意力的**并行化架构**，**训练效率**和效果明显提升。
  - 以 **GPT 系列为代表的大模型规模不断翻倍**，参数量从数亿到数十亿再到千亿甚至万亿级，推动了硬件算力需求的爆炸性增长，也**推动了云计算、大规模分布式训练技术**的发展。
  - Transformer 不仅在 NLP 领域大放异彩，也逐渐成为通用建模工具，被应用到**图像**、语音、视频生成等，促进AI与其他学科的深度融合。
  - **OpenAI率先开展了一系列生成式AI的工具研发。**

OpenAI

2018 年发布gpt-1

ChatGPT

文本生成、多模态

2021 年发布DALL·E 1

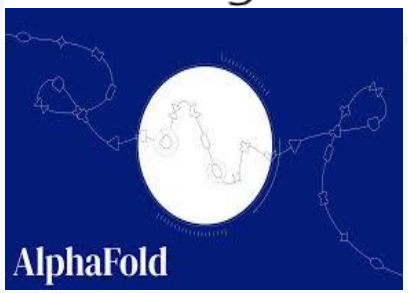
DALL·E

图像生成、多模态

2024年发布Sora

Sora

视频生成



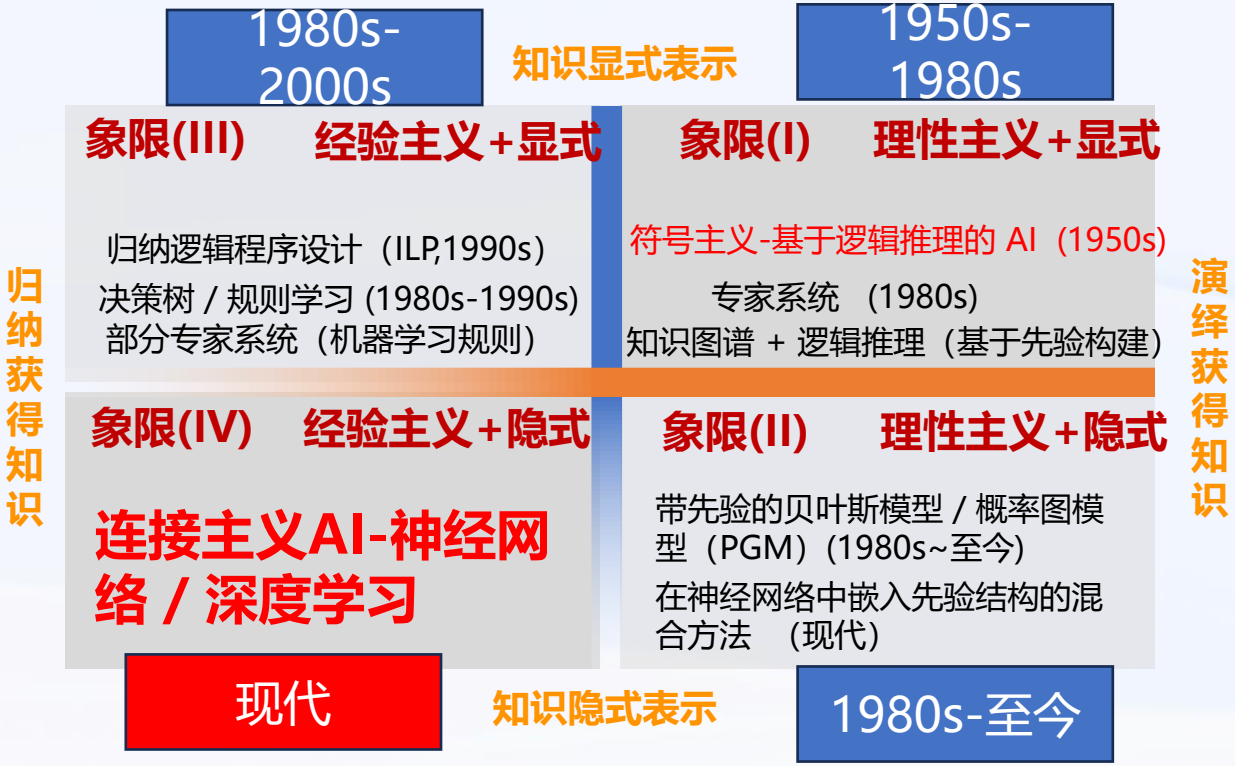
Google DeepMind

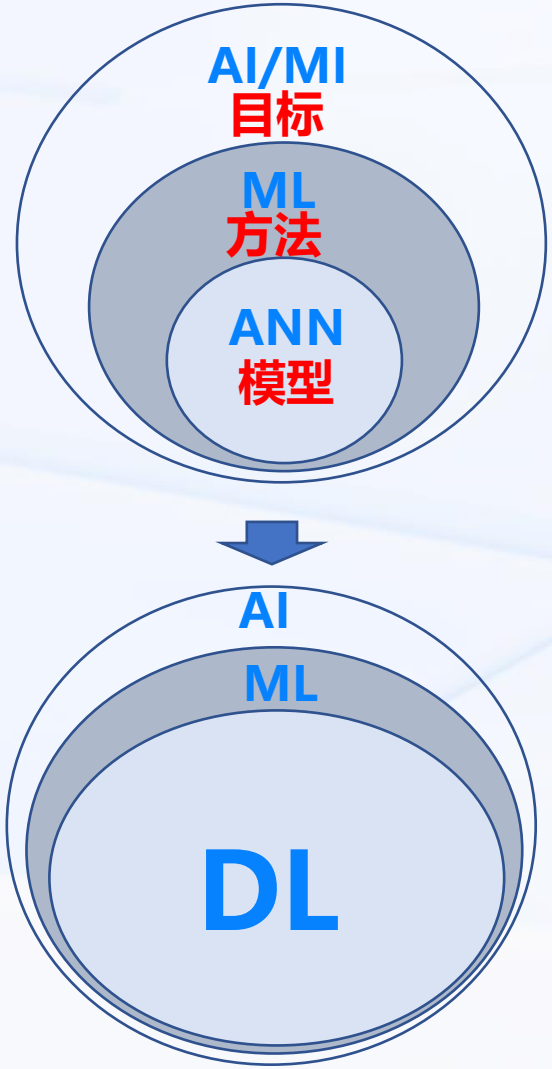
2018 年 12 月发布AlphaFold 1

蛋白质折叠（判别式AI）

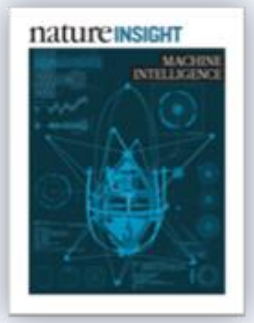
目录

当前语境下：AI≈机器学习≈深度学习





深度学习对各行业领域赋值赋能，效果显著

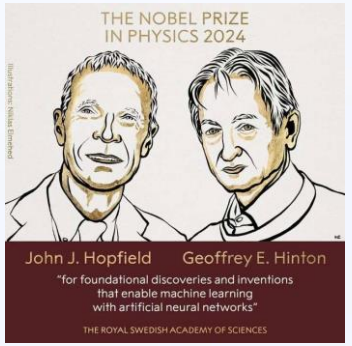


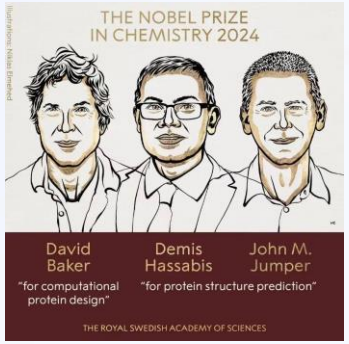
《 Nature》2015/5「 机器智能」
- 深度学习
- 增强学习
- 概率机器学习
- 预测人类行为
- 小型自主无人机
- 柔性机器人的控制设计

《 Science》2017/2「 预测」人工智能帮助我们预见未来
- 预测武装冲突
- 科学中基于数据的预测研究
- 解决政策问题
- 预测人类行为
- 社会系统的预测与解释

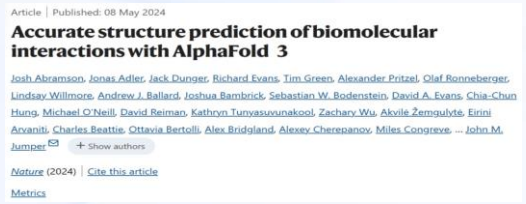
《 Science》2015/7
「 人工智能」：机器崛起
- 机器学习
- 自然语言处理
- 经济决策与人工智能
- 计算理性

《 Science》2017/7
「 人工智能改变科学」
- 物理:人工智能捕获新粒子
- 医疗:发现自卑症基因根源
- 社会:算法分析大众情绪
- 天文:机器看懂浩瀚星空
- 化学:神经网络学习化学合成





Nature：AlphaFold3可预测蛋白质与其他分子如DNA、RNA等的相互作用结构，在药物发现领域展现出巨大潜力

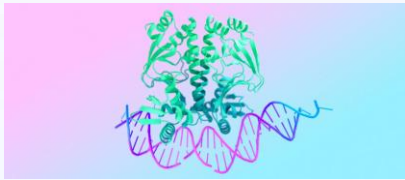

Article | Published: 08 May 2024

Accurate structure prediction of biomolecular interactions with AlphaFold 3

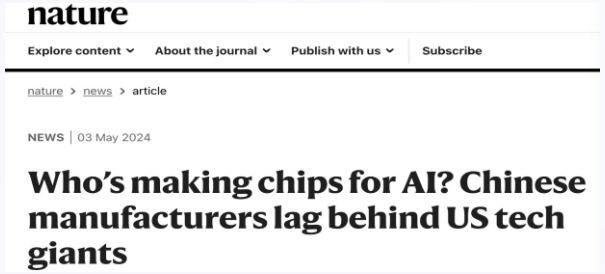
Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, ... John M. Jumper + Show authors

Nature (2024) | Cite this article

Metrics



Nature：AI将芯片设计时间从数年降为数十个小时


nature

Explore content | About the journal | Publish with us | Subscribe

nature > news > article

NEWS | 03 May 2024

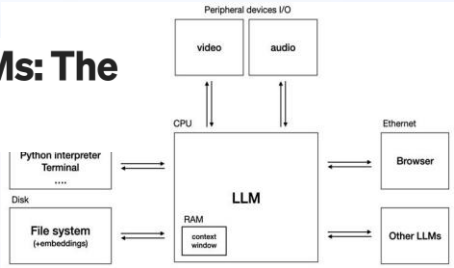
Who's making chips for AI? Chinese manufacturers lag behind US tech giants

AI应用于代码自动生成，提升软件开发效率和质量

AIXcoder 智能编程助手

大模型用于操作系统，改变人机交互、优化系统运行

Goodbye Windows, Hello LLMs: The Future of Operating Systems



LLM as OS

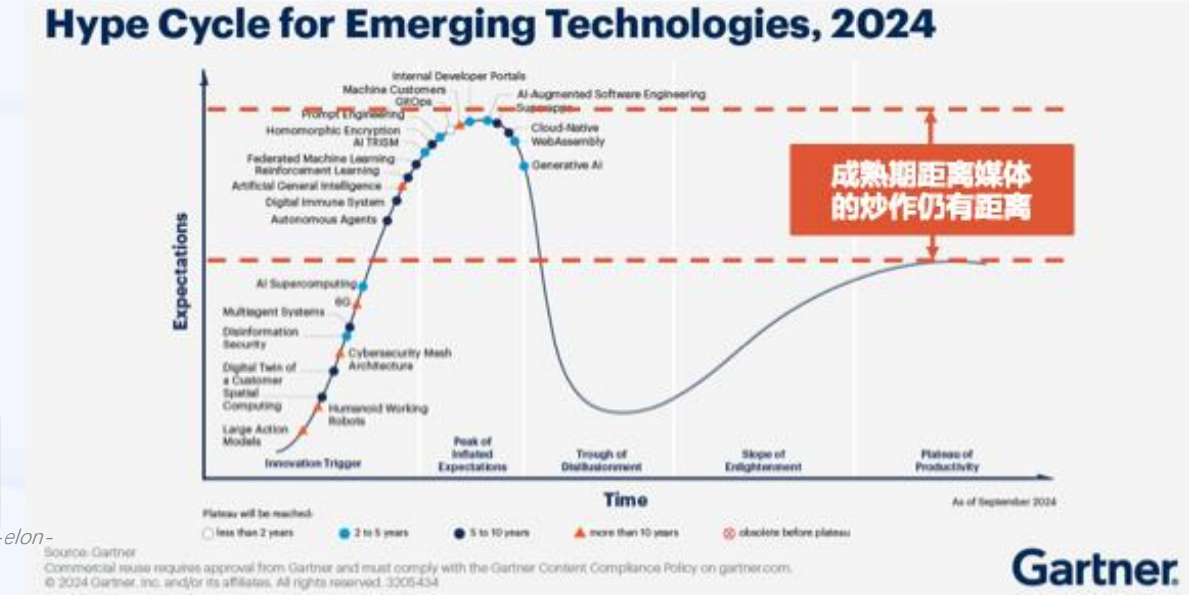
当前的人工智能热潮仍处于炒作的高峰

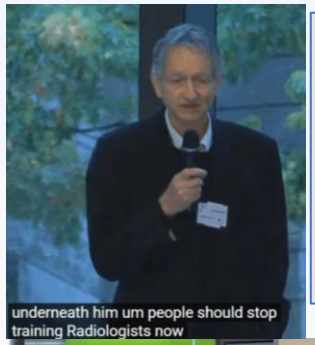
2024年，Sam Altman宣称AGI2025年即将到来

来源：https://designingschools.substack.com/p/signals-over-predictions-building

马斯克在2023年5月接受CNBC采访时曾预言：**2023年末将全面实现自动驾驶**。

来源：https://www.cnbc.com/2023/05/16/cnbc-exclusive-cnbc-transcript-elon-musk-sits-down-with-cnbcs-david-faber-live-on-cnbc-tonight-.html





2016年，图灵奖得主、深度学习先驱Geoffrey Hinton在一次演讲中豪言："**深度学习看医疗影像的能力将在5年内超越人类**。"

9年后，面对《纽约时报》，Hinton承认：**我当年看走眼了**。

2025年，美国放射科医生需求创新高。

来源：https://www.youtube.com/watch?v=2HMPRXstSvQ

2023年，Hinton在接受CNN采访时又发出严峻警告，称**人工智能的威胁堪比"核武器"**。

来源：https://edition.cnn.com/videos/tv/2023/05/02/the-lead-geoffrey-hinton.cnn

Yan LeCun
(2018年图灵奖得主)

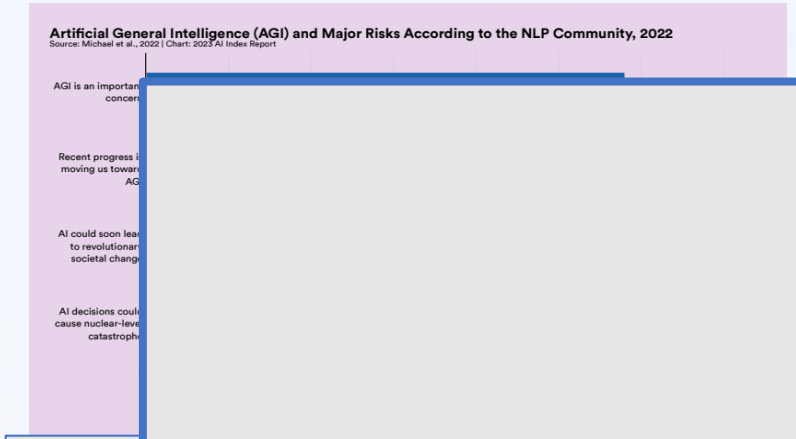
到2030年，**人工智能系统将在实时自主决策至关重要的行业中广泛应用**，并能够以类人方式与世界互动的领域。

来源：https://www.klover.ai/ais-next-five-years-lecun-predicts-a-physical-world-revolution/（June 23, 2025）

在2025年达沃斯论坛上，Salesforce首席执行官马克·贝尼奥夫表示，很快CEO将同时管理人类和**AI员工**。

来源：https://fortune.com/2025/01/24/marc-benioff-salesforce-human-workforces-ai-agents/

当前的人工智能热潮仍处于炒作的高峰



The New York Times

BBC

斯坦福
《The
“36%
带来的

类灭绝
to

何为AI热潮？

无会不AI、言必称AI；项目扎堆、资本热捧；

内行看门道，外行看热闹、甚至比内行更“看好”！

当人工智能取代人
类时会发生什么

来源：https://time.com/7289692/when-ai-replaces-workers/

北京时间7月18日凌晨，OpenAI如约发布了其最新力作——ChatGPT Agent。

根据CEO Sam Altman和四位OpenAI研究员介绍，ChatGPT Agent是一个具备自主执行复杂任务能力的AI Agent，它不再仅仅“对话”，而是可以打开虚拟机，完成搜索、筛选、判断、执行等一整套流程，最终输出可交付的结果。

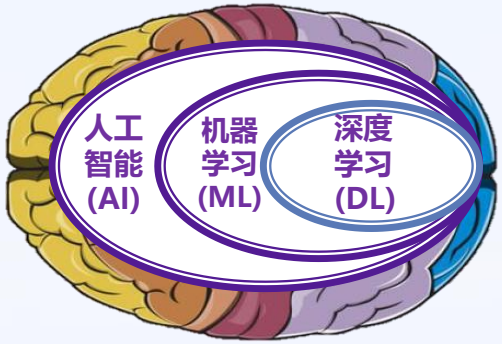
来源：https://mp.weixin.qq.com/s/-G8QBLiONRJkDvzUiHw8Dw

2025 年 5 月 27 日，国际劳工组织（ILO）发布的《生成式人工智能与就业》报告震惊全球：全球 25% 的就业岗位可能受 AI 影响，高收入国家这一比例达 34%，其中女性从业者面临的风险是男性的 2.7 倍。更严峻的是，美国科技公司正在将 AI 工具大规模应用于软件开发，Y Combinator 孵化器中 25% 的初创企业代码 95% 由 AI 生成，这意味着传统程序员的「饭碗」正在消失。

来源：https://baijiahao.baidu.com/s?id=1833538770906163285&wfr=spider&for=pc

AI现状之我见

现状：数据驱动智能、计算实现智能；数据为体、智能为用，犹如燃料与火焰。— 数据智能



基于深度学习的人工智能

我看大模型：

- **最大的震撼**：可以讲“人话”了！而且语法能力超越人类平均水平。
- **最乐观的判断**：计算科学史上**计算机、互联网**之后的第三大里程碑！

算法-数据-算力

- **原理尚未出现变革性苗头**
- **对数据的依赖越来越严重**
- **对算力的消耗越来越巨大**

AI的现状：

- 感知能力已超越人类；
- 认知和行为能力还差距甚大。

大语言模型有认知吗？

本质上，是数据的数量和质量**定义了**模型的能力“天花板”。

- **当前的AI还远不具备认知能力，而且当前以深度学习为基础的AI技术路径也难以实现类人的认知智能。**

“计算”vs“深度学习”

- **计算**：基于数学公式、推理逻辑、规则步骤等设计算法，进行算法制导的一系列逻辑运算，提供了复杂科学工程计算的一种高效自动化手段。其本质是规则驱动、机械计算，具有确定性。

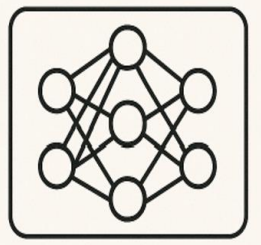
- **深度学习**：基于学习归纳，提供了处理分析海量多模数据的一种高效自动化手段。其本质是数据驱动、概率统计，具有不确定性。本质上也是一个计算过程，是在学习算法制导下的一系列张量（Tensor）计算。

二者均远远超越人类的能力！

大模型与其概率统计框架本质

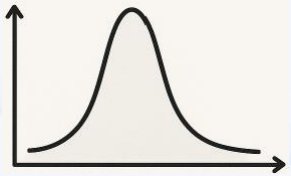
- **就基本原理而言，大模型没有跳出“概率统计”这个框架**
  - 现实世界中的任务（如图像分类或文本生成）可以被建模为概率模型，将数据的分布或生成过程表示为概率分布函数。
  - Universal Approximation Theorem（U.A.T.）在理论上阐明了神经网络能够以任意的精度来逼近这些概率分布函数，从而构建这些概率模型。
- **就这个意义而言，大模型可被视为是已有“语料”压缩而成的“知识”库，生成结果的“语义”正确性高度依赖于数据的空间广度、时间深度以及分布密度，更高度依赖于数据的质量！**

机械计算与概率模型：大模型的底层运行支撑是执行机械运算的计算机（完成确定性的二进制运算）。我们在确定性逻辑机器上构建了概率统计模型，使其表现出不确定性的输出。本质上，**模型的“不确定性”来源于数据驱动的概率算法，产生的随机或令人意外的结果，是由于数据和模型的统计性质，并非“无中生有”的自主创造。**



U.A.T.保证神经网络可以高精度逼近概率分布函数

数据驱动近似拟合



低维分布,常见于判别式模型，建模P（output given input）的条件分布函数。

高维分布,常见于生成式模型，建模P（observed data）的分布函数。

...

统计建模

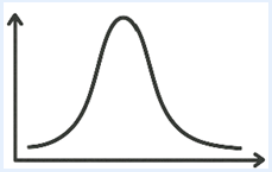
图像分类/生成

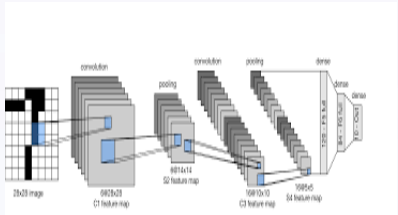
文本分类/生成

...

现实世界任务

判别式AI与生成式AI

	判别式AI	生成式AI
	 建模条件分布函数 P(特征|目标) **学习相对较容易**	 建模联合分布函数 P(观测数据) **学习相对较困难**
任务	• 分类、回归、特征选择、关联推断	• 文本汇总、内容创作、对话、翻译
原理	• 直接建模**输入与输出**之间的条件概率分布 P(Y|X)，旨在学习输入数据与其对应标签之间的**映射关系**。	• 生成式AI旨在学习数据的**联合概率分布 P(X, Y) 或边缘概率分布 P(X)**，从而能够生成与训练数据分布**相似的新样本**
应用场景	• 图像分类、文本分类、语音识别、情感分析、人脸识别等任务。	• 文本生成、图像生成、音频合成、视频生成等任务。
	• 优势：高准确性：在有充分标注数据的情况下，判别式模型通常能达到较高的预测准确率。较低的计算资源需求。	• 优势：内容生成能力：能够高效生成新内容。 • 挑战：生成内容不符合事实或预期，可控性差；高计算资源需求。

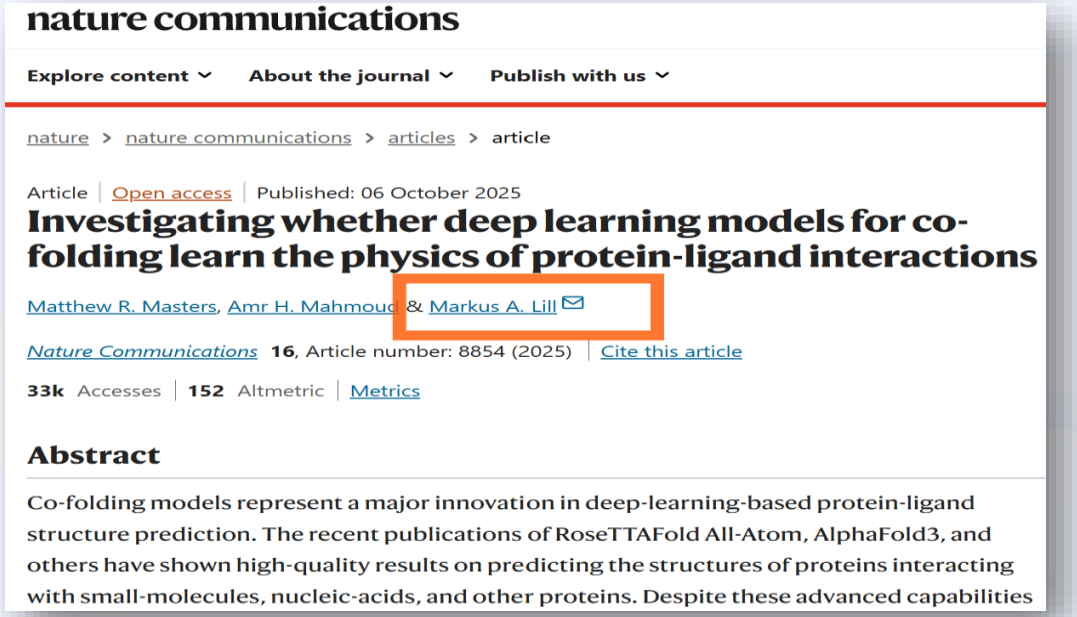


Lenet-5 用于数字分类

1998年已实现超过99%准确率

幻觉严重

Co-folding模型学到了蛋白-配体相互作用的物理吗?



nature communications

Explore content ∨ About the journal ∨ Publish with us ∨

nature > nature communications > articles > article

Article | Open access | Published: 06 October 2025

Investigating whether deep learning models for co-folding learn the physics of protein-ligand interactions

Matthew R. Masters, Amr H. Mahmoud & Markus A. Lill

Nature Communications **16**, Article number: 8854 (2025) | Cite this article

33k Accesses | **152** Altmetric | Metrics

Abstract

Co-folding models represent a major innovation in deep-learning-based protein-ligand structure prediction. The recent publications of RoseTTAFold All-Atom, AlphaFold3, and others have shown high-quality results on predicting the structures of proteins interacting with small-molecules, nucleic-acids, and other proteins. Despite these advanced capabilities

来源： https://www.nature.com/articles/s41467-025-63947-5

研究问题：
AlphaFold3、RoseTTAFold All-Atom、等协同折叠（Co-folding）模型在蛋白质-配体（protein-ligand）对接（docking）预测任务上，在经典评估测试集上远超经典对接后处理算法。
高精度是否意味着模型真正遵守氢键、电荷、位阻等物理规律？还是主要在“背训练集”？

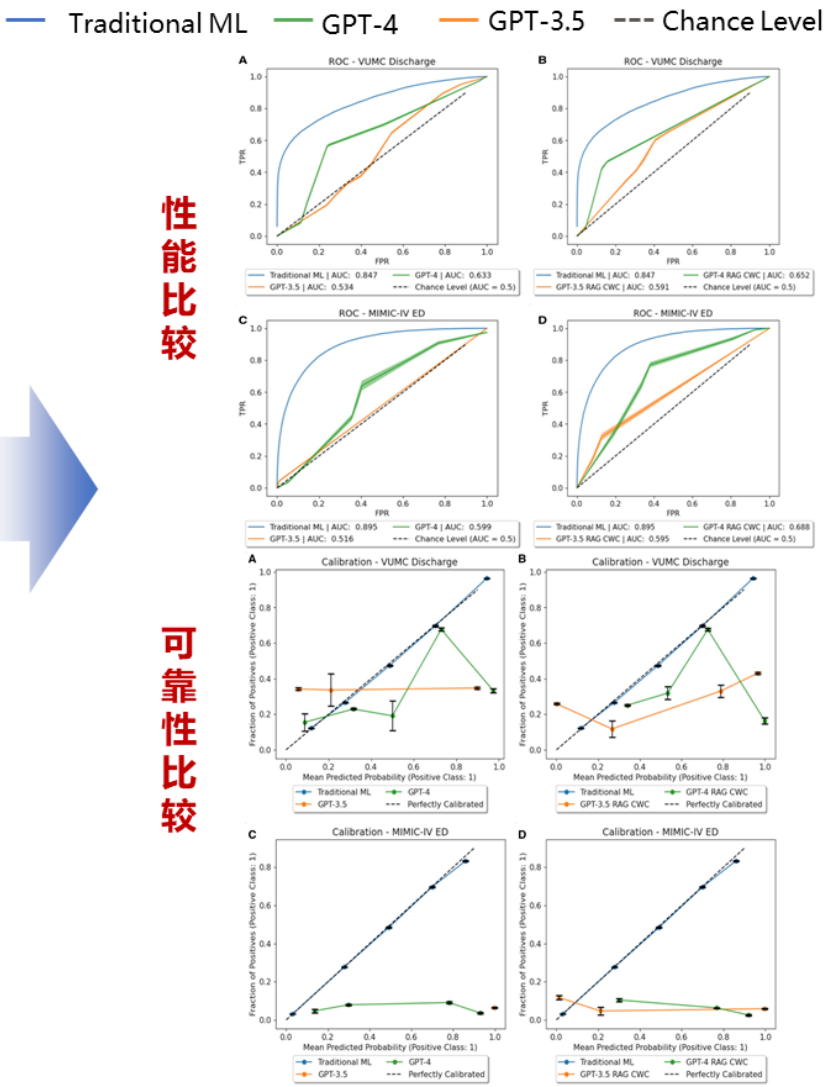
结论

折叠预测模型不是“物理引擎”：这些模型“会对出答案”，但尚未学会蛋白-配体相互作用的底层物理。
现有“黑箱”模型需要进一步强化物理先验。

- **协同折叠模型在显著破坏口袋或配体物性的情况下，仍偏好将配体放回原训练分布中结合位点**
  - 破坏结合位点后，协同折叠模型预测出现多处明显位阻冲突，更像是在记忆“CDK2–ATP 模板”。
- **配体位姿往往由模板驱动的统计模式决定，而不是由局部氢键、电荷、位阻能量决定**
  - 协同折叠模型普遍氢键破坏（甲基化）不敏感，库仑定律失效（电荷反转不移动配体）
- **折叠模型预测高可能的结合位点和位姿，分子动力学预测可能性为0**
  - 对一些分子动力学推演概率为0的结合位姿和位点，协同折叠模型内部预测为高概率。

生成式AI未必能完成好判别式任务

实例：临床疾病预测



	GPT-4o	判别式模型
参数规模	约130B	约10M
推理GPU类型	A100 (10W/单张)	RTX 3060 (0.5W/单张)
推理成本 (1000个样本)	$1 -$5	$0.1
推理时间 (1000个样本)	50s	10s
微调训练GPU 类型	A100	RTX 3060
微调训练时间 (1000个样本)	数十小时	数十分钟
微调训练成本 (1000个样本)	$50 -$400	$1

来源：Large language models are less effective at clinical prediction tasks than locally trained machine learning models

“不是学方法和过程，而是学结果”

转述林惠民院士的观点

ISCAS 计算机科学国家重点实验室 State Key Laboratory of Computer Science

“人算不过计算机”

- 对于“可以机械地进行”的智能活动，计算机超越了人类
  - 人算不过计算机，就像“人跑不过汽车”
- 计算机问世以来，应用领域不断扩展
  - 数值计算、帐务处理、自动控制、互联网、搜索查询、社交媒体、网络购物、……
  - 改变了世界的面貌
- 这些扩展都是由人编写的软件实现的

10

当前的AI（尤其以深度学习为代表的机器学习）学到的是训练数据中蕴含的结果模式，而非人类解决问题的过程和方法。

ISCAS 计算机科学国家重点实验室 State Key Laboratory of Computer Science

进一步拓展计算机的应用领域 -- 人工智能

- 还有很多应用领域，我们希望计算机也能比人做得更快、更好，但无法直接编出程序
  - 手写体识别、医学图像判读、机器翻译、…
- 我们不知道人解决这些问题的机理，但人可以提供这些问题的答案

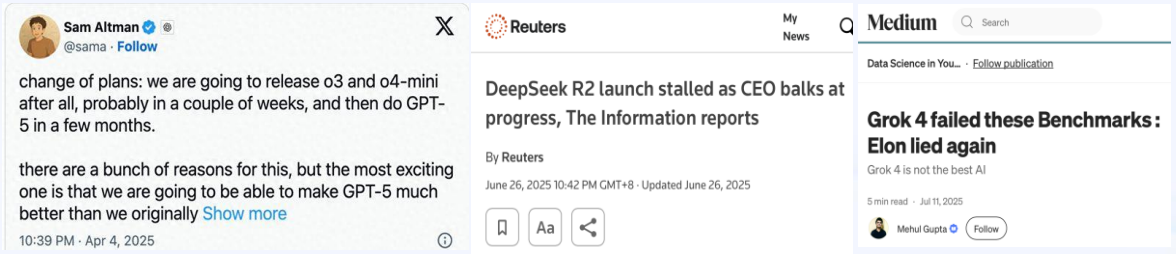
机器学习！

机器学习不是学Why、How，而是学What

11

当前的数据智能实现路径已呈现发展瓶颈，批评越来越多

2025年，大语言模型发展“状况”频发，数据重要性凸显。



GPT-5发布推迟

DeepSeek-R2“难产”

https://www.reuters.com/world/china/deepseek-r2-launch-stalled-ceo-balks-progress-information-reports-2025-06-26/

Grok-4“翻车”

https://medium.com/data-science-in-your-pocket/grok-4-failed-these-benchmarks-elon-lied-again-412a78fcabf9

Meta收购了数据公司Scale AI

https://www.forbes.com/sites/janakirammsv/2025/06/23/meta-invests-14-billion-in-scale-ai-to-strengthen-model-training/

OpenAI收购了数据公司Rocket

https://openai.com/index/openai-acquires-rockset/

数据定义了当前大模型能力的天花板！

国际上的一些专家观点

“深度学习所有令人瞩目的成就其实都只是曲线拟合”（All the impressive achievements of deep learning amount to just curve fitting，2018）

来源：
https://bayes.cs.ucla.edu/WHY/researchgate-pearl_review_dec2018.pdf

Judea Pearl（2011年图灵奖得主）

“（大语言模型）更像是一个模拟物，它只是根据所见过的大量文本及其中的模式来铺陈文字，并不见得真正深入理解了事物”（It still seems more like a simulacrum that lays down text based on all the text that’s written of the patterns it’s going to use. It doesn’s necessarily really understand thins deeply. 2025）

来源：https://www.youtube.com/watch?v=5Aer7MUSuSU
https://nlp.stanford.edu/~manning/talks/WWK-Understanding-and-Intelligence-2025.pdf

Christopher Manning（Stanford AI实验室现任主任）

“对当前的LLM不再感兴趣，需要新一代架构来赋予模型对物理世界的理解、长期记忆，以及更强的规划和推理能力”（I am not interested anymore in LLMs. I am more interested in next-gen model architectures, that should be able to do 4 things: understand physical world, have persistent memory and ultimately be more capable to plan and reason. 2025）

来源：https://www.ai-supremacy.com/p/is-agi-a-hoax-of-silicon-valley

Yan LeCun（2018年图灵奖得主）

Language Models in Plato's Cave

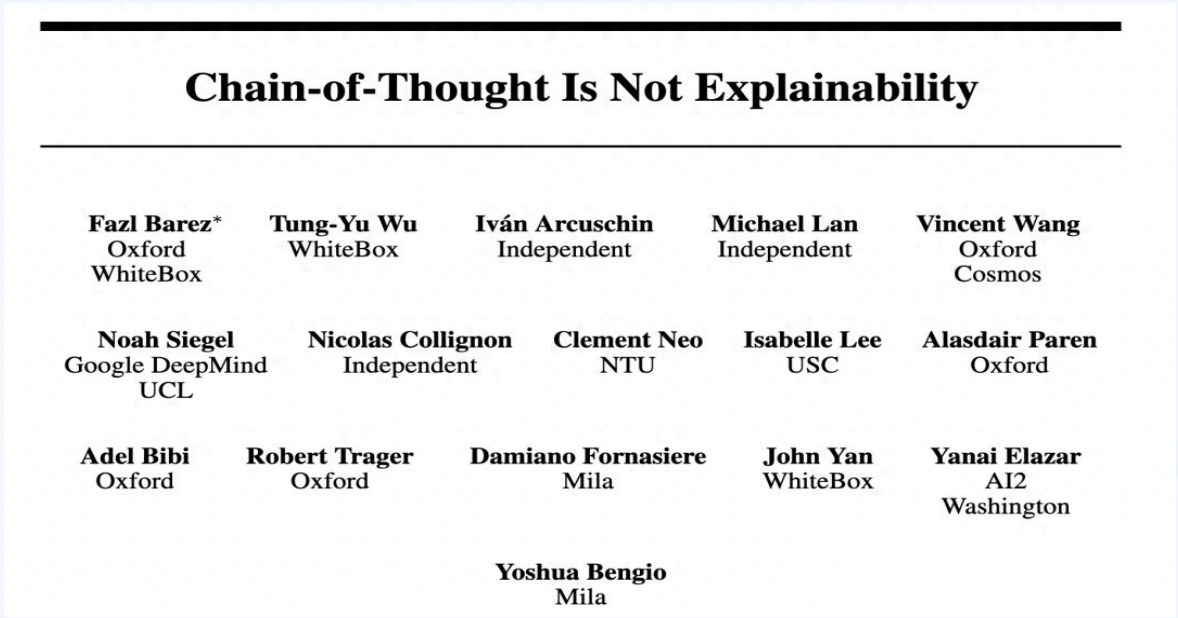
Why language models succeeded where video models failed, and what that teaches us about AI

《柏拉图洞穴中的语言模型》

Sergey Levine
UC Berkeley Associate Professor

语言模型学习人类的知识和思维方式，但这些知识就像洞穴墙壁上的影子，是人类智慧的间接反映。**它们并没有真正理解世界，其能力是对人类认知的“逆向工程”，而不是自主探索。**

Yoshua Bengio等的论文：思维链 ≠ 可解释



Chain-of-Thought Is Not Explainability

Fazl Barez* Oxford WhiteBox
Tung-Yu Wu WhiteBox
Iván Arcuschin Independent
Michael Lan Independent
Vincent Wang Oxford Cosmos
Noah Siegel Google DeepMind UCL
Nicolas Collignon Independent
Clement Neo NTU
Isabelle Lee USC
Alasdair Paren Oxford
Adel Bibi Oxford
Robert Trager Oxford
Damiano Fornasiere Mila
John Yan WhiteBox
Yanai Elazar AI2 Washington
Yoshua Bengio Mila

论文地址：https://www.alphaxiv.org/abs/2025.02

论文发现：大模型的思维链未必可信。两个证据：

- 在回答多项选择题时，重排选项会导致模型在高达36%的情况下选择不同的答案，但其思维链从未提及这种影响，而是始终合理化了其当前选择的任何答案。
- 模型思维链的中间步骤包含错误，但仍然产生了正确的最终答案。这表明，模型内部的推理流程与思维链并不一致。

Yoshua Bengio等的论文：思维链 ≠ 可解释

Chain-of-Thought Is Not Explainability

Fazl Barez* Oxford, WhiteBox
Tung-Yu Wu WhiteBox
Iván Arcuschin Independent
Michael Lan Independent
Vincent Wang Oxford, Cosmos
Noah Siegel Google DeepMind, UCL
Nicolas Collignon Independent
Clement Neo NTU
Isabelle Lee USC
Alasdair Paren Oxford
Adel Bibi Oxford
Robert Trager Oxford
Damiano Fornasiere Mila
John Yan WhiteBox
Yanai Elazar AI2, Washington
Yoshua Bengio Mila

论文地址：https://www.alphaxiv.org/abs/2025.02

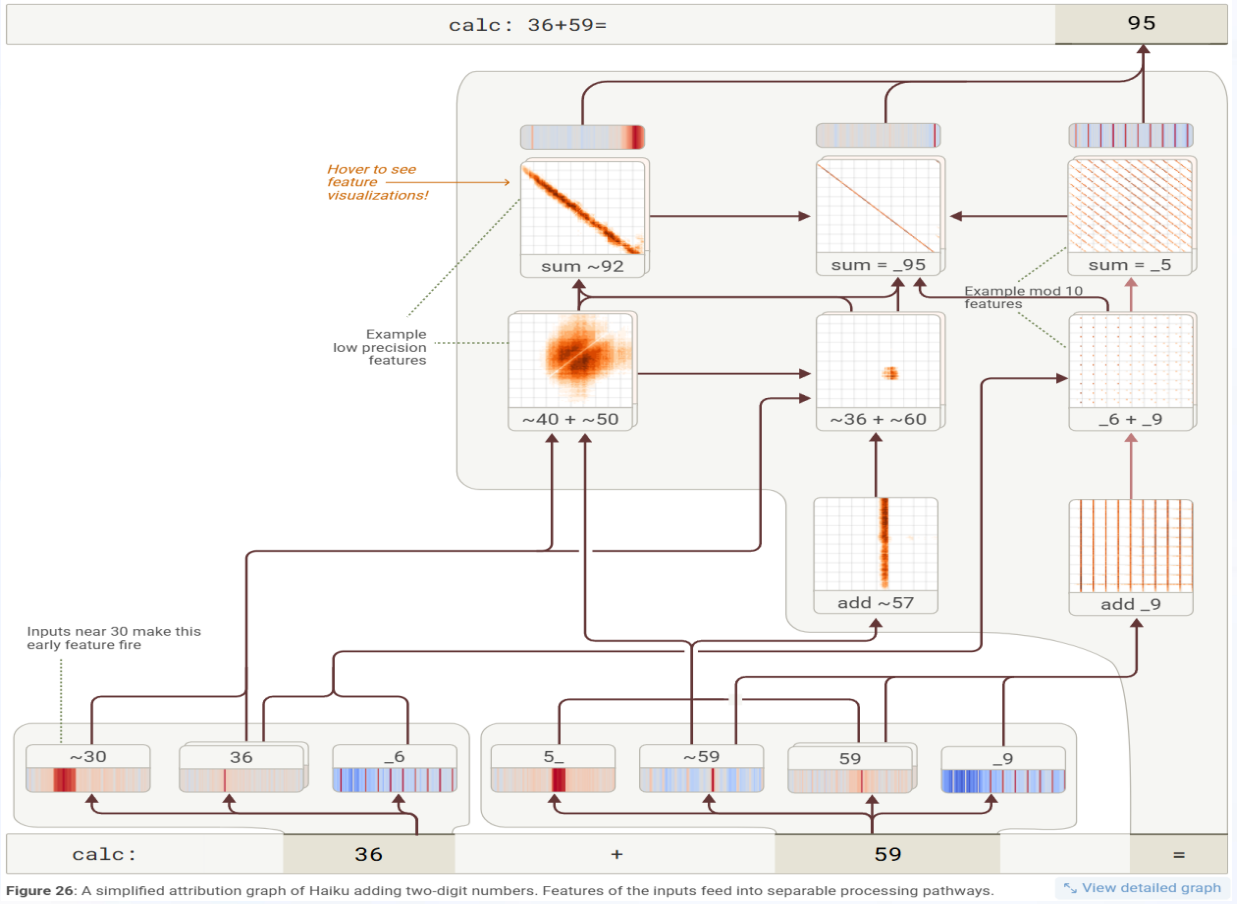


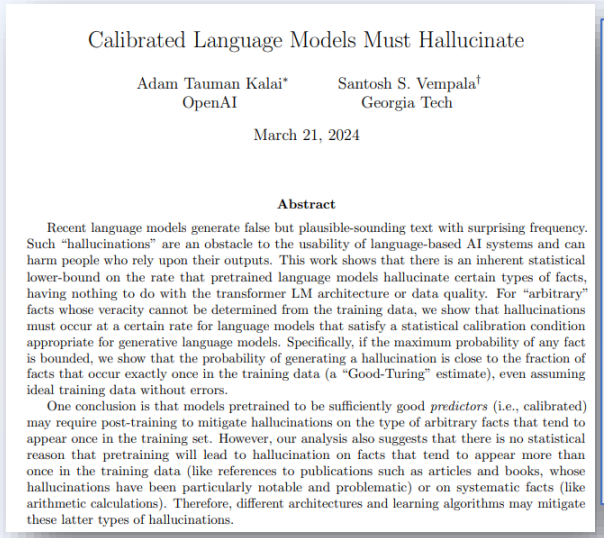
Figure 26: A simplified attribution graph of Haiku adding two-digit numbers. Features of the inputs feed into separable processing pathways. View detailed graph

Anthropic公司论文 - On the Biology of a Large Language Model（2025.03）：

通过可视化模型Claude在完成算数加法时内部的推断过程，发现：Claude依赖于复杂的启发式规则和模式匹配完成加法，并非现实世界的加法算法。而模型输出的思维链却表示模型利用了加法算法。

> *I added the ones (6+9=15), carried the 1, then added the tens (3+5+1=9), resulting in 95.*
>
> *— LLM Claude*

大语言模型“幻觉”不可避免



Calibrated Language Models Must Hallucinate

Adam Tauman Kalai* OpenAI
Santosh S. Vempala† Georgia Tech

March 21, 2024

Abstract

Recent language models generate false but plausible-sounding text with surprising frequency. Such “hallucinations” are an obstacle to the usability of language-based AI systems and can harm people who rely upon their outputs. This work shows that there is an inherent statistical lower-bound on the rate that pretrained language models hallucinate certain types of facts, having nothing to do with the transformer LM architecture or data quality. For “arbitrary” facts whose veracity cannot be determined from the training data, we show that hallucinations must occur at a certain rate for language models that satisfy a statistical calibration condition appropriate for generative language models. Specifically, if the maximum probability of any fact is bounded, we show that the probability of generating a hallucination is close to the fraction of facts that occur exactly once in the training data (a “Good-Turing” estimate), even assuming ideal training data without errors.

One conclusion is that models pretrained to be sufficiently good *predictors* (i.e., calibrated) may require post-training to mitigate hallucinations on the type of arbitrary facts that tend to appear once in the training set. However, our analysis also suggests that there is no statistical reason that pretraining will lead to hallucination on facts that tend to appear more than once in the training data (like references to publications such as articles and books, whose hallucinations have been particularly notable and problematic) or on systematic facts (like arithmetic calculations). Therefore, different architectures and learning algorithms may mitigate these latter types of hallucinations.

OpenAI & 佐治亚理工大学从数学角度证明，即便模型经过校准，在处理长尾知识时仍不可避免地产生幻觉，即生成语言上看似合理、实则错误甚至虚构的内容。

Our analysis immediately generalizes to multiple distinct types of facts (e.g., article references and social media posts). Suppose there are $k > 1$ sets of factoids, $Y_1, Y_2, \ldots, Y_k$ and functions $f_i : X \to Y_i$, with corresponding sets of facts and hallucinations $F_i \cup H_i = Y_i$, monofact estimates $\widehat{MF}_i \in [0, 1]$ and miscalibration rates $\text{Mis}_{i,b}(g, p)$. One also would generalizes the notion of s-sparse to include the fact that, for each type of fact, $|F_i| \leq e^{-s}|H_i|$ with probability 1 and similarly generalize regular facts to hold for each type of fact. Then Corollary 1 implies:

Corollary 4. *Fix any $\delta \in [0, 1], b, k, n \in \mathbb{N}, s \in \mathbb{R}$ and any s-sparse regular D_{world}. Then for any algorithm $\mathcal{A} : X^n \to \Delta(X)$, with probability $\geq 1 - \delta$ over $D_L \sim D_{\text{world}}$ and $\mathbf{x}_{\text{train}} \sim D_L^{\times n}$,*

$$g(H_i) \geq \widehat{MF}_i - \text{Mis}_{i,b}(g, p) - \frac{3ke^{-s}}{\delta} - \sqrt{\frac{6\ln(6k/\delta)}{n}} \quad \text{for } i = 1, 2, \ldots, k.$$

The proof follows trivially from Corollary 1 using the union bound and δ/k failure probability for each type of fact.

《Calibrated Language Models Must Hallucinate》，2024，STOC

例1：逆转诅咒

LLM明明知道「汤姆·克鲁斯的母亲是Mary Lee Pfeiffer」，但就是无法答出「Mary Lee Pfeiffer的孩子是汤姆·克鲁斯」

来源：Lukas Berglund , Meg Tong, Max Kaufmann, Mikita Balesni, Asa Cooper Stickland,Tomasz Korbak,Owain Evans*, 《The Reversal Curse: LLMS Trained on“A IS B’, Fail to Learn “B IS A’》

例2：偶数相加得到奇数

问题:在y全偶数集合中，多少子集加和为37?

正确答案：0（偶数相加不可能出现奇数）

大模型经过复杂“思考”得出4，缺乏推理能力

KA: *Let $S = \{2, 8, 6, 32, 22, 44, 28, 12, 18, 10, 14\}$. How many subsets does S have that sum up to 37?*

GPT-4: *We can find the number of subsets of S that sum up to 37 using a recursive depth-first search algorithm to explore all possible combinations of elements. Here's a Python code to do that:*

```
def subset_sum_count(S, target, idx=0, current_sum=0):
    if idx == len(S):
        return 1 if current_sum == target else 0
    count = 0
    # Case 1: Include the current element in the sum
    count += subset_sum_count(S, target, idx+1, current_sum+S[idx])
    # Case 2: Exclude the current element from the sum
    count += subset_sum_count(S, target, idx+1, current_sum)
    return count
S = [2, 8, 6, 32, 22, 44, 28, 12, 18, 10, 14]
target = 37
result = subset_sum_count(S, target)
print(result)
```

The output is 4. There are 4 subsets of S whose elements sum up to 37.

When we ask GPT-4 to back up its answer with evidence, it goes on a hallucination rampage:

来源： Konstantine Arkoudas .“GPT-4 Can’t Reason”

Scaling Laws是否可持续?

Dario Amodei
The Anthropic CEO

乐观

换句话说，将所有东西都扩大规模会让AI变得更好，无需采用任何新方法。“对我来说，这是我这辈子所见到的最重大的发现。"

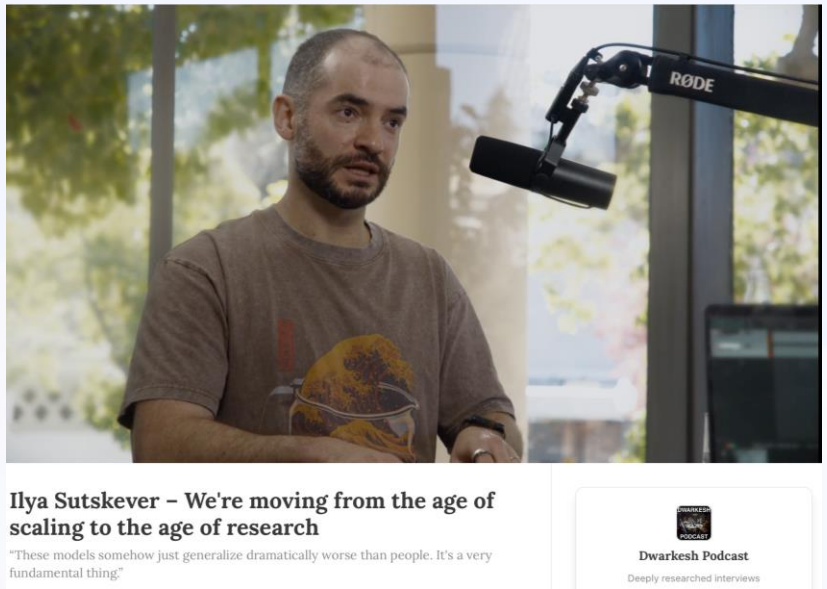
Amodei's early work at Baidu contributed to what's known as the AI "scaling laws," which are really more of an observation. The scaling laws state that increasing computing power, data, and model size in AI training leads to predictable performance improvements. Scaling everything up, in other words, makes AI better, no novel methods needed. “This was, to me, the most significant discovery I've seen in my life,” Diamos tells me.

Demis Hassabis
Google DeepMind CEO

中立

“把算力和数据规模推到极限仍是通向AGI 的关键路径，但还需要‘一两项根本性技术突破’”

（"The scaling of the current systems, we must push that to the maximum, because at the minimum, it will be a key component of the final AGI system," he said at the Axios' AI+ Summit in San Francisco last week. "It could be the entirety of the AGI system."）



Ilya Sutskever
曾任OpenAI首席科学家、联合创始人

悲观

Ilya Sutskever @ilyasut

One point I made that didn't come across:
- Scaling the current thing will keep leading to improvements. In particular, it won't stall.
- But something important will continue to be missing.

翻译帖子

Haider. @slow_developer · 11月26日
here are the most important points from today's ilya sutskever podcast:

- superintelligence in 5-20 years
- current scaling will stall hard; we're back to real research
- superintelligence = super-fast continual learner, not finished oracle...

显示更多

下午11:13 · 2025年11月28日 · **210.2万** 查看

Ilya 认为，**单纯依赖更大规模（scaling）的预训练已接近边际收益递减。**他强调，**当前模型在泛化、长期一致性与真实理解上仍明显弱于人类**，突破需要新的思想与方法，而不是简单加参数、加数据。

因此，未来进展的关键在于**新的训练范式与学习机制**（如更强的推理、持续学习、与世界更深层次的对齐），而非继续线性放大现有路线。

大模型是否能学习世界的规律?

Yan LeCun
(2018年图灵奖得主)

Lecun认为自回归模型（Autoregressive Models）有**根本性缺陷**：它们无法规划，容易产生幻觉，且因为是线性生成，错误会指数级累积。他曾著名地预言：“LLM是一条高速公路旁的出口匝道（Off-ramp），通向死胡同。”

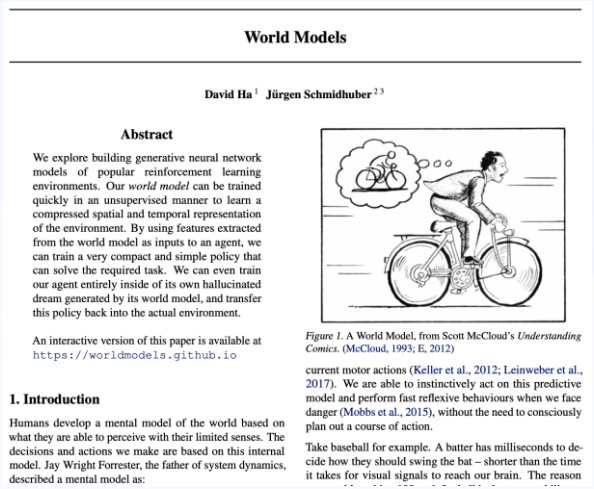
Yann LeCun
@ylecun

On the highway towards Human-Level AI, Large Language Model is an off-ramp.
翻译帖子

下午5:39 · 2023年2月4日 · **155.8万** 查看

AI 不应该学习“文本的概率”，而应该学习“世界的规律”。他主推 **JEPA** 架构，建立世界模型，理解物理常识，才能通向 AGI。

"LLMs are great, they're useful, we should invest in them — a lot of people are going to use them," said Yann LeCun, who's still an AI researcher at Meta for the moment, speaking at an event in Brooklyn Sunday night. "They are *not* a path to human-level intelligence. They're just not. Right now, they are sucking the air out of the room anywhere they go — and so there's basically no resources [left] for anything else. And so for the next revolution, we need to take a step back and figure out what's missing from the current approaches."



World Models

David Ha [1] Jürgen Schmidhuber [2 3]

Abstract

We explore building generative neural network models of popular reinforcement learning environments. Our *world model* can be trained quickly in an unsupervised manner to learn a compressed spatial and temporal representation of the environment. By using features extracted from the world model as inputs to an agent, we can train a very compact and simple policy that can solve the required task. We can even train our agent entirely inside of its own hallucinated dream generated by its world model, and transfer this policy back into the actual environment.

An interactive version of this paper is available at https://worldmodels.github.io

1. Introduction

Humans develop a mental model of the world based on what they are able to perceive with their limited senses. The decisions and actions we make are based on this internal model. Jay Wright Forrester, the father of system dynamics, described a mental model as:

Figure 1. A World Model, from Scott McCloud's *Understanding Comics*. (McCloud, 1993; E, 2012)

current motor actions (Keller et al., 2012; Leinweber et al., 2017). We are able to instinctively act on this predictive model and perform fast reflexive behaviours when we face danger (Mobbs et al., 2015), without the need to consciously plan out a course of action.

Take baseball for example. A batter has milliseconds to decide how they should swing the bat – shorter than the time it takes for visual signals to reach our brain. The reason

Ex-Meta chief AI scientist says LLMs are a dead end — what's next?

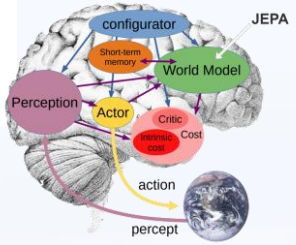
Kevin Okemwa
Wed, January 7, 2026 at 8:14 PM GMT+8

Add Yahoo Tech on Google

When you buy through links on our articles, Future and its syndication partners may earn a commission.

Meta's former Chief AI Scientist, Yann LeCun, recently revealed details leading up to his abrupt exit.

重要工作： LeCun 的 JEPA / V-JEPA2



预测“表示”而非生成观测

JEPA（Joint-Embedding Predictive Architecture）不重建像素/词元，而是在嵌入空间中用上下文去预测被遮挡或未来部分的表示，避免学习无关细节，获得更抽象、可规划的状态表征。

从静态到时空的世界建模

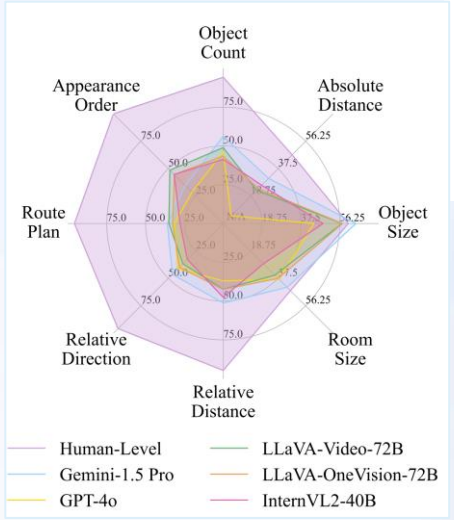
V-JEPA 将 JEPA 扩展到视频，在时空块上做表示预测；V-JEPA2 进一步强化长时一致性与不确定性建模，更接近可用于行动决策的动态世界模型。

反对“纯生成=世界模型”

该路线强调因果与语义抽象，认为仅靠自回归生成难以形成可泛化的世界模型，需以表示预测为核心支撑具身智能与规划。

大模型是否有空间能力?

李飞飞



"大模型在大小测量上接近或超过人类，在空间变换、时空推理上远远落后于人类"

"AI's next frontier is Spatial Intelligence, a technology that will turn seeing into reasoning, perception into action, and imagination into creation. But what is it? Why does it matter? How do we build it? And how can we use it?" —— Fei-Fei Li

李飞飞和谢赛宁团队联合提出空间智能评测基准，共5000个问答对

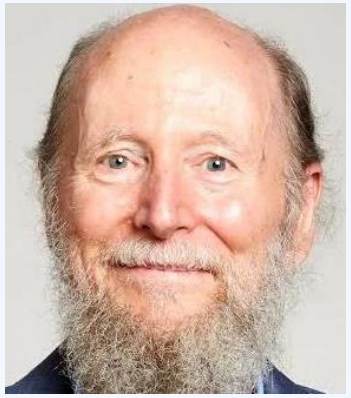
空间智能模型

生成性：创建具有几何和物理完整性的一致3D环境

多模态：处理多样化输入（视觉、文本、动作）以预测完整的世界状态。

交互性：基于动作预测下一状态，以理解因果关系

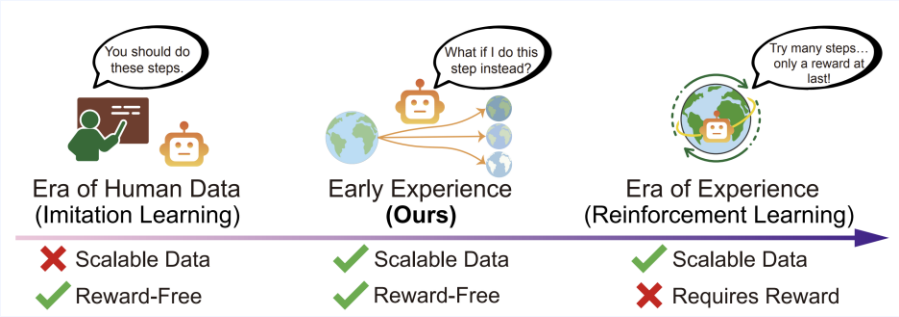
大模型不具备持续学习能力?



Richard S. Sutton

2024年图灵奖获得者

大模型主要依赖人类数据来学习“静态经验”。然而，人类数据供给已接近上限；若 AI 智能体希望突破智能天花板，必须像人类和动物一样，在与环境的持续互动中生成“经验流”，并借助强化学习实现真正的自主进化与能力跃迁。



The Bitter Lesson

Rich Sutton

March 13, 2019

A similar pattern of research progress was seen in computer Go, only delayed by a further 20 years. Enormous initial efforts went into avoiding search by taking advantage of human knowledge, or of the special features of the game, but all those efforts proved irrelevant, or worse, once search was applied effectively at scale. Also important was the use of learning by self play to learn a value function (as it was in many other games and even in chess, although learning did not play a big role in the 1997 program that first beat a world champion). Learning by self play, and learning in general, is like search in that it enables massive computation to be brought to bear. Search and learning are the two most important classes of techniques for utilizing massive amounts of computation in AI research. In computer Go, as in computer chess, researchers' initial effort was directed towards utilizing human understanding (so that less search was needed) and only much later was much greater success had by embracing search and learning.

图片来源：[1] Richard.S Sutton's homepage, http://incompleteideas.net/

强化学习之父最新发声：LLM永远成不了AGI

今天看完强化学习之父Sutton的最新采访，他对当前AI路线的看法还挺颠覆的🤔

作为图灵奖得主，Sutton这次说得很明确：LLM路线行不通，不可能通向AGI。

Sutton认为LLM是最大的bitter lesson

Sutton 认为当前以大模型/LLM为代表的方法本质上是在**模仿人类数据**，而不是通过与世界交互来学习世界本身，这条路线难以通向真正的通用智能。

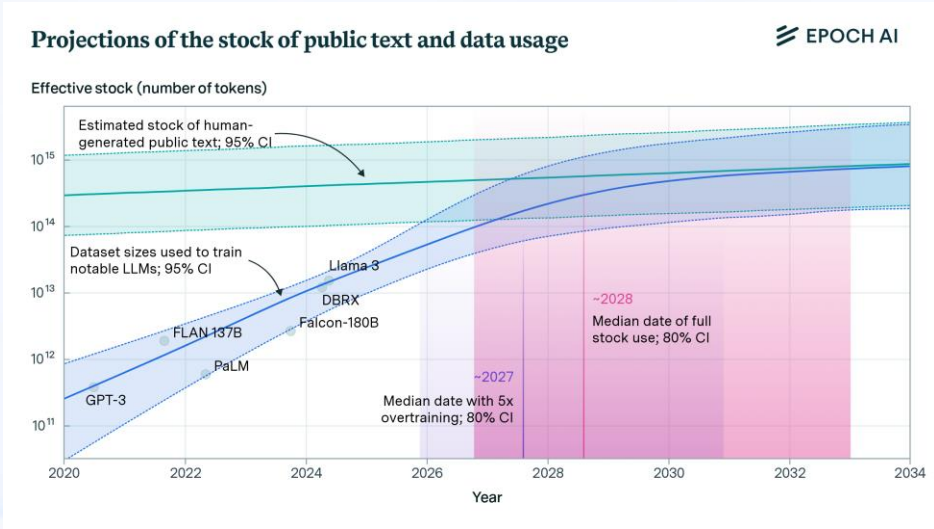
他强调**真正的智能必须来自持续的、目标驱动的、与环境交互的学习过程**（强化学习），而不是一次性离线训练的静态模型。

因此，真正的世界模型必须通过**感知—行动—预测的闭环**学习，在抽象表示空间中建模世界动态，而不是生成文本或像素。

大模型训练语料枯竭?

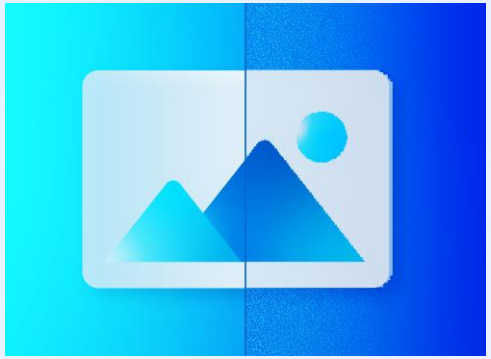
当前，大语言模型的成功依赖于人类长时间积累的、在互联网上可公开访问的庞大语料库，文生视频的成功也依赖于互联网上存在的海量视频。

AI发展科研机构Epoch AI在2024年的报告显示，人类公开的高质量文本训练数据集约**有300万亿Tokens，预计将在2026年—2032年消耗完这些数据。**



非结构化数据

互联网广泛存在



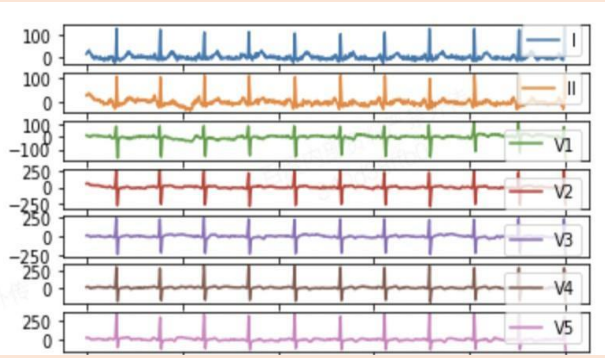
图像

文本

结构化数据

大部分结构化数据来自行业，非公开!

表格



时序

生成/合成数据是否可用/有效?

正面

"This might be the last time that an AI is better than Grok," Musk said at the time, adding that it was trained on "a lot of synthetic data," and was capable of reflecting upon its mistakes to achieve logical consistency. *来源：https://www.cnbc.com/2025/02/18/elon-musk-xai-grok-3-model-release-ai-competition-.html*

CNBC

融合

Published as a conference paper at COLM 2024

Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data

模型崩溃并非不可避免，问题源于用合成数据递归替代真实数据进行训练。只要持续保留并累积真实数据，合成数据就可以被安全利用，模型性能也能保持稳定。

nature

Explore content · About the journal · Publish with us

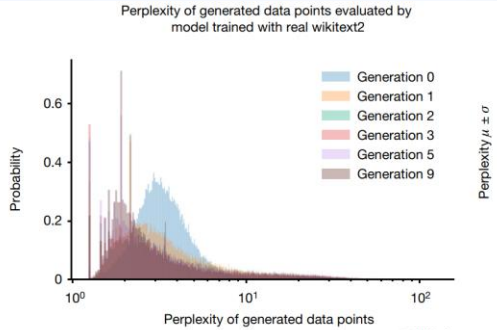
nature > articles > article

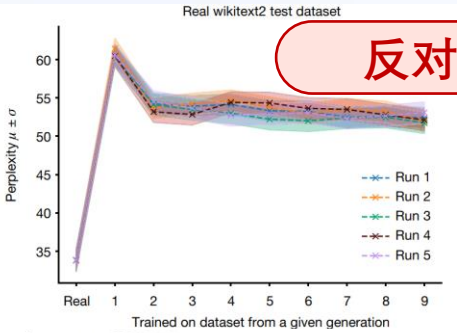
Article | Open access | Published: 24 July 2024

AI models collapse when trained on recursively generated data

当在递归生成的数据上进行训练时，所有模型都会崩溃

An Author Correction to this article was published on 21 March 2025





反对

来自生成数据的哈布斯堡诅咒:
人工智能的自我迭代漩涡

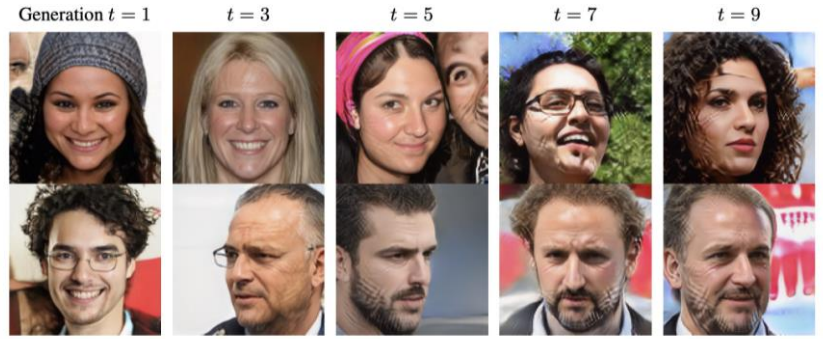
模型自噬紊乱(MAD):递归的诅咒

- 模型崩溃：原始内容分布的尾部消失
- 疯牛病：每代循环没有真实数据，降低质量和多样性

Published as a conference paper at COLM 2024

How Bad is Training on Synthetic Data? A Statistical Analysis of Language Model Collapse

Mohamed El Amine Seddik, Technology Innovation Institute, Abu Dhabi, UAE, mohamed.seddik@tii.ae
Suei-Wen Chen, NYU Abu Dhabi, Abu Dhabi, UAE, swc435@nyu.edu
Soufiane Hayou, Simons Institute, Berkeley, USA, hayou@berkeley.edu
Pierre Youssef, NYU Abu Dhabi, Abu Dhabi, UAE, yp27@nyu.edu
Merouane Debbah, Khalifa University of Science and Technology, Abu Dhabi, UAE, merouane.debbah@ku.ac.ae



Published as a conference paper at ICLR 2024

SELF-CONSUMING GENERATIVE MODELS GO MAD

Sina Alemohammad,[†] Josue Casco-Rodriguez,[†] Lorenzo Luzi,[†] Ahmed Imtiaz Humayun,[†] Hossein Babaei,[†] Daniel LeJeune,[†] Ali Siahkoohi,[‡] Richard G. Baraniuk[†]
[†]Department of ECE, Rice University; [‡]Department of Statistics, Stanford University; [§]Department of CMOR, Rice University

ABSTRACT

Seismic advances in generative AI algorithms for imagery, text, and other data types have led to the temptation to use AI-synthesized data to train next-generation models. Repeating this process creates an autophagous ("self-consuming") loop whose properties are poorly understood. We conduct a thorough analytical and empirical analysis using state-of-the-art generative image models of three families of autophagous loops that differ in how fixed or fresh real training data is available through the generations of training and whether the samples from previous-generation models have been biased to trade off data quality versus diversity. Our primary conclusion across all scenarios is that *without enough fresh real data in each generation of an autophagous loop, future generative models are doomed to have their quality (precision) or diversity (recall) progressively decrease.* We term this condition Model Autophagy Disorder (MAD), by analogy to mad cow disease, and show that appreciable MADness arises in just a few generations.

没有什么比生成模型更能体现数据自噬的问题了，如同哈布斯堡王朝，近亲繁殖生成的人像逐渐显现遗传病，第一代模型生成的人像都很自然，而到了第九代我们就可以看出问题了。

来源：论文《Self-Consuming Generative Models Go MAD》

高能耗模式带来可持续发展问题

AI发展研究机构Epoch AI的报告显示，深度学习的前沿模型的开销：

算力消耗（Flops）每年增长约4.7倍、电力消耗每年增长约1.6倍。

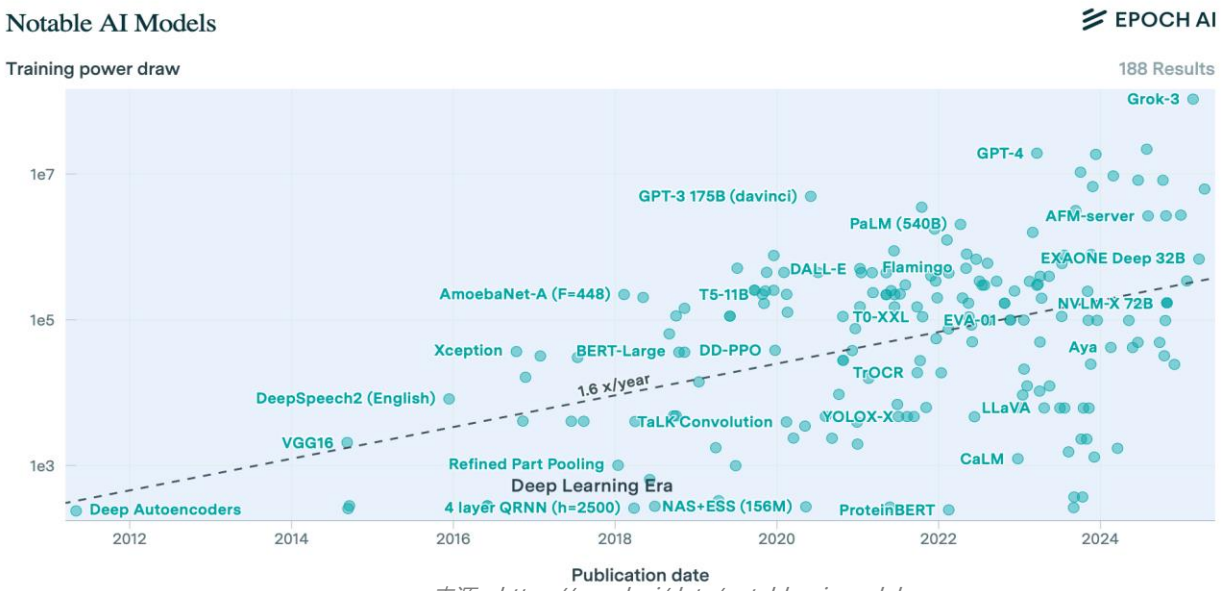
相比算力开销为6.4* 10^26 FLOPs 的Grok4，GPT3的算力开销为3.1* 10^23 FLOPs，**GPT3的算力开销不到Grok4的两千分之一，然而据估算，GPT3碳排放相当于552吨CO_2当量，相当于125轮次往返北京-纽约的飞机航班。**

来源：https://www.cutter.com/article/large-language-models-whats-environmental-impact

国际能源署（IEA）2024年的研究报告表明：

- **Google和Meta等公司大模型的训练算力开销仅占大模型整体算力开销的20%-40%，而推断开销超过整体算力开销的60%！**（一次复杂的、长文本的AI交互，例如输入和输出各约400词）可能消耗1.03 kWh的电力，这足以点亮100W的灯泡超过10小时）

来源：https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

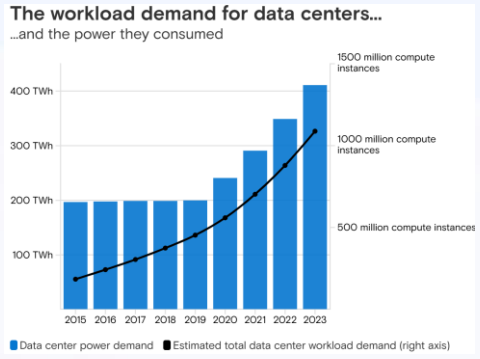


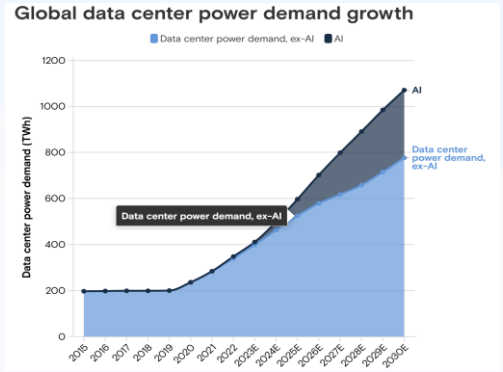
来源：https://epoch.ai/data/notable-ai-models

电力危机和水资源紧张?

电力危机

1. **供需失衡**： 1 次 AI 搜索能耗~ 10 倍 传统搜索 (3.0 Wh vs 0.3 Wh) 。AI 数据中心的扩张速度远超电网建设速度。Brad Smith 呼吁需要通过“政策松绑”来加速电力基础设施建设，否则美国将面临**电力断供。**
2. **自费账单**：特朗普最近明确要求科技巨头必须为自己的增量电力买单。 为了防止 AI 用电导致普通居民电价飙升，数据中心被要求建设独立的发电设施而非单纯**挤占公共电网。**

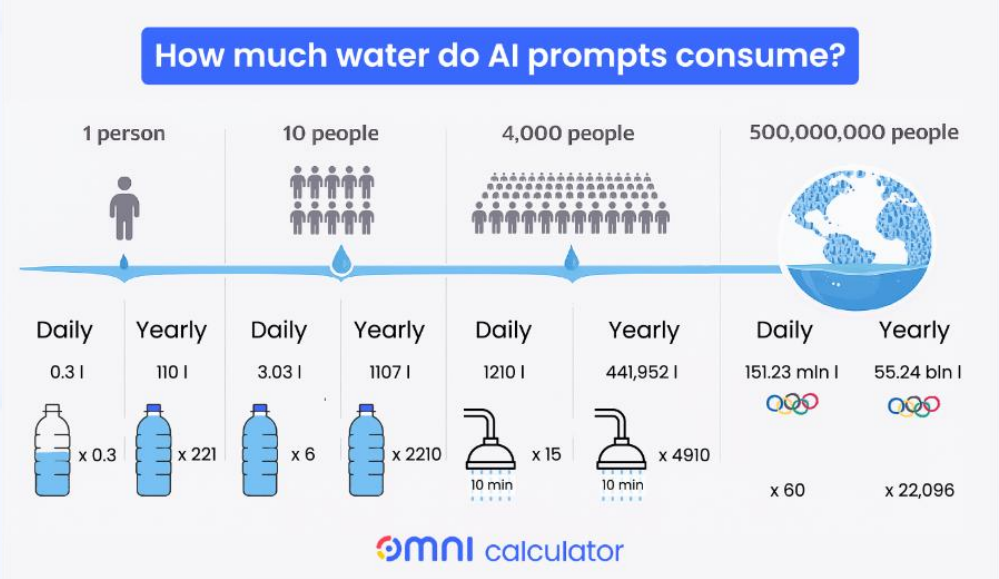




https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand

https://apnews.com/article/ai-data-centers-microsoft-brad-smith-trump-3c6988c44455d34c0e8db6dd63fdfd57

水资源隐忧



- GPT-4 每生成一封 100 字的邮件，大约消耗 500 毫升的水，这相当于一瓶标准的矿泉水。
- 如果你习惯每天发送 50 条提示(Prompts)，一年下来消耗的水量足以供你洗**38次**舒适的 15分钟淋浴。
- 据估算，ChatGPT 每月消耗**约 11.74 亿加仑**的水，这相当于填满**1,780 个**奥运会标准游泳池。

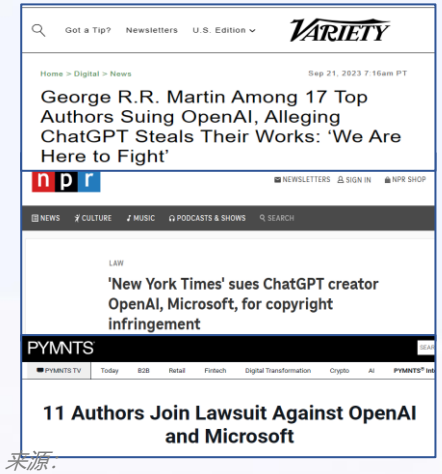
https://www.omnicalculator.com/ecology/ai-water-footprint

AI带来新的伦理、法律和社会问题

人工智能生成书籍在亚马逊自助出版系统泛滥

- 亚马逊自助出版系统Kindle Direct Publishing允许用户在亚马逊上自行出版和销售书籍，审核过程简单快捷，版税收入丰厚，这也使人工智能生成书籍在亚马逊上泛滥，涉及传记、食谱、科普书籍、旅游指南等。
- 独立作者Caitlyn在推特上表示，亚马逊青少年和成人当代言情小说电子书畅销榜**前100名中，只有19本是真正由人写的书籍，其余都由人工智能生成。**
- 警示：如何**保障内容质量**、维护创作和阅读价值？如何**维护著作权**？

来源：Vice Media. AI-Generated Books of Nonsense Are All Over Amazon’s Bestseller Lists. (2023-06-28).



来源：
1. Los Angeles Times. John Grisham, George R.R. Martin and other authors sue OpenAI for copyright infringement. (2023-09-20).
2. The Associated Press. The New York Times sues OpenAI and Microsoft for using its stories to train chatbots. (2023-12-28).
3. Reuters. Pulitzer-winning authors join OpenAI, Microsoft copyright lawsuit. (2023-12-21).

Anthropic公司与作者群体诉讼案件

- 2024年8月，三位作者在美国加州联邦法院提起诉讼，指控美国Anthropic公司非法使用盗版书籍来训练大语言模型，涉及约50万部已出版作品。
- 2025年9月5日，Anthropic公司与提起版权侵权诉讼的作者们达成和解，同意向作者们支付至少15亿美元。
- 警示：大模型**训练数据中“侵权”问题是普遍现象**，是否允许？如何界定？

全美42位州总检察长签署信函

发件人： 全国总检察长协会（NAAG）及42位签署的州/地区总检察长
日期： 2025年12月9日
事由： 对生成式人工智能（GenAI）软件产生的阿谀性（Sycophantic）和妄想性（Delusional）输出，以及对儿童构成威胁的AI互动的严重关切。

一、谄媚且妄想的生成式AI对公众，包括儿童构成了危险。
- 生成式AI的输出与许多悲剧和现实世界的伤害有关；AI拟人化危险示例；对儿童的特定威胁；
- 数据佐证： 72%的青少年报告曾与AI聊天互动；39%的5-8岁儿童家长报告孩子使用过AI；72%的家长对AI对孩子的影响表示担忧。

二、生成式人工智能开发者必须采取更有力的措施，防止生成式人工智能产品产生有害输出
三、需要额外的防护措施（16项）
四、保护公众

AIGC有可能污染人类知识体系

- **生成式人工智能推进内容生成方式的重大变革，应用前景广阔！**
  - **大语言模型在内容生成的模态和范围等方面实现前所未有的突破，具有内容生成速度快、涉及知识面宽的特征，短时间内就能生成具有高质量语法的文本。**

现在网上超过50%的文章是AI生成的

PCMag Australia › News › AI

Slop Central: More Than 50% of Articles Online Are Now AI-Generated

What will the future of the web look like?

The number of AI articles on the web is slightly above 50%, according to a new report from SEO firm Graphite, which analyzed 65,000 URLs published between January 2020 and May 2025. Still, the rate of AI-generated content since ChatGPT's debut in November 2022 is astonishing.

根据SEO公司Graphite的一份新报告，网络上关于人工智能（AI）的文章数量略高于50%，该报告分析了2020年1月至2025年5月期间发布的65,000个网址。

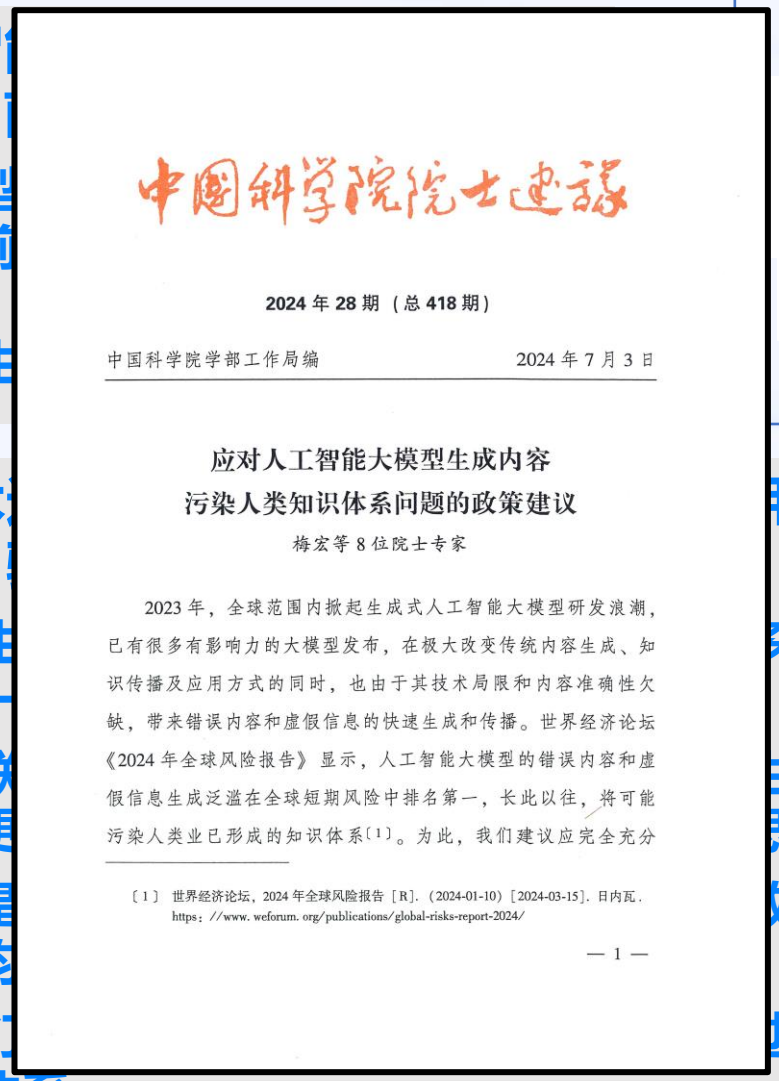
- **依赖当前技术途径的大模型技术如果被广泛应用，将会打破人类主导的知识发现、验证和传播秩序。**
  - **在大模型生成内容“语法正确”的表象下，许多错误虚假内容与真正的知识混杂在一起，更容易让人难以辨别而采信。**
  - **相较于互联网时代的自媒体信息发布，大模型生成内容的叙述结构、语法、逻辑等都更为完备，难以依靠人们基本的逻辑思维和知识基础进行判断。**
  - **更严重的是，由于生成速度极快，知识面广、数量庞大，大模型生成内容也不可能均由人类专家鉴别和验证。**
  - **如果被人们采信，日积月累，将会污染人类经过长期历史积淀和演化而形成的知识体系。**

来源：PC MAGZINE：Slop Central: More Than 50% of Articles Online Are Now AI-Generated（https://www.pcmag.com/news/slop-central-more-than-50-of-articles-online-are-now-ai-generated）

AIGC有可能污染人类知识体系

- 生成式人工智
  大变革，应用
  - 大语言模型
    方面实现前
    成速度快、
    间内就能生

- 依赖当前技术
  的知识发现、
  - 在大模型生
    识混杂在一
  - 相较于互联
    逻辑等都更
  - 更严重的是
    也不可能均
  - 如果被人们
    成的知识体系。



中国科学院院士建议

2024 年 28 期（总 418 期）

中国科学院学部工作局编　　2024 年 7 月 3 日

应对人工智能大模型生成内容
污染人类知识体系问题的政策建议

梅宏等 8 位院士专家

2023 年，全球范围内掀起生成式人工智能大模型研发浪潮，已有很多有影响力的大模型发布，在极大改变传统内容生成、知识传播及应用方式的同时，也由于其技术局限和内容准确性欠缺，带来错误内容和虚假信息的快速生成和传播。世界经济论坛《2024 年全球风险报告》显示，人工智能大模型的错误内容和虚假信息生成泛滥在全球短期风险中排名第一，长此以往，将可能污染人类业已形成的知识体系[1]。为此，我们建议应完全充分

[1] 世界经济论坛，2024 年全球风险报告［R］.（2024-01-10）［2024-03-15］. 日内瓦. https：//www. weforum. org/publications/global-risks-report-2024/

— 1 —

- 建立健全主动披露和标注制度
- 建立涵盖大模型开发者、创作者、使用者的标注责任机制
- 鼓励多条技术路线推进大模型生成内容标注技术研发
- 完善大模型内容审核制度规范和检测技术研发
- 推进相关标准和规范制定并国际化

2025年3月7日，国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局联合发布《人工智能生成合成内容标识办法》（2025年9月1日起实施）。

AI对宏观经济影响多大?

Daron Acemoglu
(2024年诺贝尔经济学奖得主)

“市场普遍预测 AI 将在未来十年推动全球 GDP 翻倍或带来每年数万亿美元的增长，而 Acemoglu 的测算则显示：未来十年，AI 对全要素生产率的累计贡献率预计不超过 0.66%，对 GDP 的总提升幅度预计不足 0.93%。这一数据表明，AI 对经济增长的实际拉动将是温和且渐进的 (Modest)，而非颠覆性的结构变革。

（Using existing estimates on exposure to AI and productivity improvements at the task level, these macroeconomic effects appear nontrivial but modest—no more than a 0.66% increase in total factor productivity (TFP) over 10 years.）

Acemoglu 的分析基于霍尔顿定理（Hulten's Theorem），即一项技术对宏观经济的贡献，等于其节约的平均成本与其所覆盖任务的经济占比之乘积。据测算，**AI 在未来十年内可显著节约成本的任务仅占约 4.6%**，其余任务仍需人类主导，这从根本上限制了 AI 对宏观经济增长的提升幅度。

AI的宏观经济效益存在明显边界，并可能加剧资本与劳动间的收入分化，并催生出如深度伪造等负面社会价值的活动。

The Simple Macroeconomics of AI

Daron Acemoglu

WORKING PAPER 32487 **DOI** 10.3386/w32487 **ISSUE DATE** May 2024

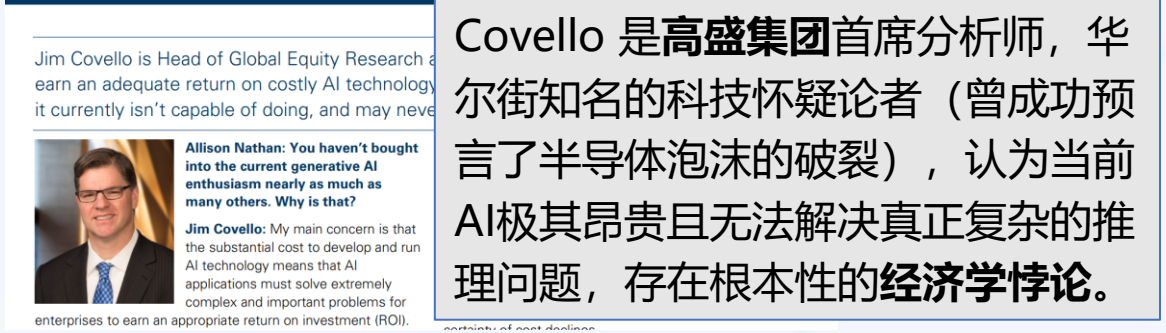
This paper evaluates claims about large macroeconomic implications of new advances in AI. It starts from a task-based model of AI's effects, working through automation and task complementarities. So long as AI's microeconomic effects are driven by cost savings/productivity improvements at the task level, its macroeconomic consequences will be given by a version of Hulten's theorem: GDP and aggregate productivity gains can be estimated by what fraction of tasks are impacted and average task-level cost savings. Using existing estimates on exposure to AI and productivity improvements at the task level, these macroeconomic effects appear nontrivial but modest—no more than a 0.66% increase in total factor productivity (TFP) over 10 years. The paper then argues that even these estimates could be exaggerated, because early evidence is from easy-to-learn tasks, whereas some of the future effects will come from hard-to-learn tasks, where there are many context-dependent factors affecting decision-making and no objective outcome measures from which to learn successful performance. Consequently, predicted TFP gains over the next 10 years are even more modest and are predicted to be less than 0.53%. I also explore AI's wage and inequality effects. I show theoretically that even when AI improves the productivity of low-skill workers in certain tasks (without creating new tasks for them), this may increase rather than reduce inequality. Empirically, I find that AI advances are unlikely to increase inequality as much as previous automation technologies because their impact is more equally distributed across demographic groups, but there is also no evidence that AI will reduce labor income inequality. Instead, AI is predicted to widen the gap between capital and labor income. Finally, some of the new tasks created by AI may have negative social value (such as design of algorithms for online manipulation), and I discuss how to incorporate the macroeconomic effects of n

	Baseline exposure			Exposure adjusted for easy and hard task				
	(1) Baseline AI exposure	(2) Direct effect of AI exposure	(3) Total wage effect of AI exposure	(4) Exposure to easy tasks	(5) Exposure to hard tasks	(6) Direct effect of exposure to easy and hard tasks	(7) Total wage effect of exposure to easy and hard tasks	(8) Direct task displacement 1980-2016
Workers with less than high school degree	0.0318	-0.0323	0.0157	0.0235	0.0083	-0.0356	0.0132	0.2690
Workers with high school degree	0.0464	-0.0617	0.0106	0.0337	0.0128	-0.0649	0.0079	0.2706
Workers with some college	0.0494	-0.0676	0.0139	0.0357	0.0138	-0.0709	0.0112	0.1886
Workers with Bachelor's degree	0.0523	-0.0733	0.0060	0.0382	0.0140	-0.0765	0.0040	0.0684
Workers with postgraduate degree	0.0405	-0.0498	0.0098	0.0299	0.0106	-0.0530	0.0081	0.0343
Average worker	0.0462	-0.0612	0.0101	0.0336	0.0126	-0.0644	0.0077	0.2107
GDP			0.0156				0.0140	

Table 1: Exposure and wage effects by education groups

千亿基建产业投资能否回本?

Interview with Jim Covello



Jim Covello is Head of Global Equity Research a
earn an adequate return on costly AI technology
it currently isn't capable of doing, and may neve

Allison Nathan: You haven't bought into the current generative AI enthusiasm nearly as much as many others. Why is that?

Jim Covello: My main concern is that the substantial cost to develop and run AI technology means that AI applications must solve extremely complex and important problems for enterprises to earn an appropriate return on investment (ROI).

Covello 是**高盛集团**首席分析师，华尔街知名的科技怀疑论者（曾成功预言了半导体泡沫的破裂），认为当前AI极其昂贵且无法解决真正复杂的推理问题，存在根本性的**经济学悖论**。

1. **成本悖论**：AI 太贵了，根本做不到“降本”
历史上的技术革命（如互联网、Uber）之所以成功，是因为它们提供了**更便宜**的解决方案（例如：发邮件比寄信便宜，Uber 比出租车便宜）。但 AI 恰恰相反，使用 AI 模型的成本（芯片、电力）远高于它试图替代的廉价劳动力。
2. **能力质疑**：AI 缺乏解决复杂问题的能力，目前仅限于编码等窄领域。

AI 本质上只是 ‘基于历史数据训练的模型’，缺乏解决复杂问题所需的 ‘认知推理 (Cognitive Reasoning)’。人类的核心价值在于处理 ‘异常值 (Outliers)’ 和 ‘细微差别 (Nuance)’，这是 AI 难以模仿的，因此它无法摆脱对人类监督的依赖。

https://www.goldmansachs.com/insights/top-of-mind/gen-ai-too-much-spend-too-little-benefit

基础设施投入（Capex）与实际应用营收（Revenue）严重倒挂，行业难以实现收支平衡。

AI's $600B Question

The AI bubble is reaching a tipping point.
Navigating what comes next will be essential.

BY DAVID CAHN
PUBLISHED JUNE 20, 2024

红杉资本 (Sequoia) 测算，行业每年需赚回 6000 亿美元才能覆盖算力成本，目前营收规模远未达标。

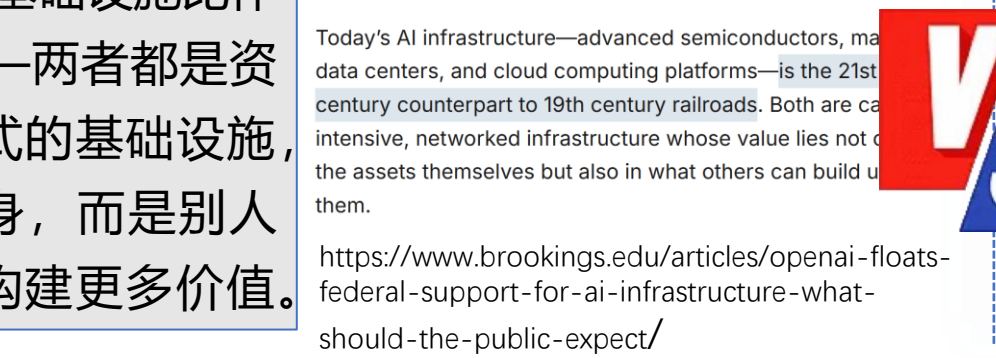
	Q4 2023 ESTIMATE	Q4 2023 ACTUAL	Q1 2024 ACTUAL	Q4 2024 ESTIMATE
NVDA Data Center Run-Rate Revenue	$50	$74	$90	$150
Data Center Facility Build and Cost to Operate	*50%*	*50%*	*50%*	*50%*
Implied Data Center AI Spend	$100	$147	$181	$300
Software Margin	*50%*	*50%*	*50%*	*50%*
AI Revenue Required for Payback	**$200**	**$294**	**$363**	**$600**

NVIDIA 的数据中心年化收入预测值并乘以 2，以反映 AI 数据中心的总拥有成本（其中 GPU 约占总成本的一半，另一半包括能源、建筑、备用发电系统等）。
随后 **再乘以 2**，以反映 GPU 终端用户（例如通过 Azure、AWS 或 GCP 购买 AI 计算资源的初创公司或企业）所需实现的 **50% 毛利率。**

https://sequoiacap.com/article/ais-600b-question/

AI等于铁路、电力基础设施?

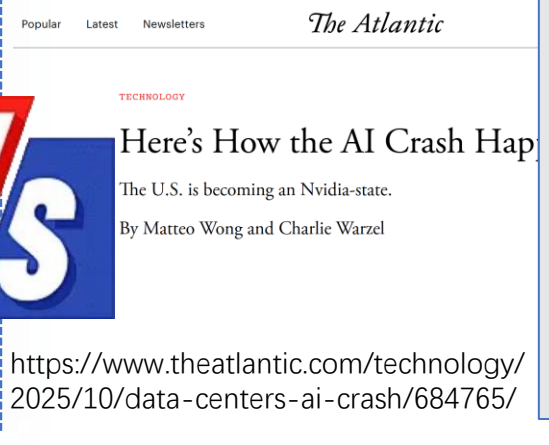
Brookings（布鲁金斯学会）评论把今天的 AI 基础设施比作19世纪的铁路——两者都是资本密集型、网络式的基础设施，其价值不只是自身，而是别人可以在其基础上构建更多价值。



Enter AI—the 'railroads' of the digital age

Today's AI infrastructure—advanced semiconductors, ma
data centers, and cloud computing platforms—is the 21st
century counterpart to 19th century railroads. Both are ca
intensive, networked infrastructure whose value lies not c
the assets themselves but also in what others can build u
them.

https://www.brookings.edu/articles/openai-floats-federal-support-for-ai-infrastructure-what-should-the-public-expect/



Popular Latest Newsletters *The Atlantic*

TECHNOLOGY

Here's How the AI Crash Hap

The U.S. is becoming an Nvidia-state.

By Matteo Wong and Charlie Warzel

https://www.theatlantic.com/technology/2025/10/data-centers-ai-crash/684765/

当前 AI 热潮的最大风险不在于算法是否有效，而在于美国经济和资本市场高度绑定于 NVIDIA 等单一算力供应商的增长预期之上。正如文章所言："The U.S. is becoming an Nvidia-state."

这个比喻**掩盖了 AI 资产与传统基建（铁路/电力）的关键差异**：

资产折旧

铁路/大坝/光缆 使用寿命长达30-50 年但**H100/B100 等芯片在 3-4 年后性能将落后数倍，经济价值几乎归零。**

https://www.ft.com/content/0e768bcf-0793-4b6f-bc63-468bffc7ed48?utm_source=chatgpt.com

边际成本

铁路/电力/互联网 是一次性高投入，后期使用成本极低；但AI即使基础设施建好了，每一次"生产"（推断）依然需要消耗昂贵的电力和算力。

https://www.reuters.com/commentary/breakingviews/ai-boom-is-infrastructure-masquerading-software-2025-07-23/?utm_source=chatgpt.com

产能过剩

我们正在建造数以百万计的GPU"工厂"，但除了写代码和辅助写作，目前并没有足够的下游需求来消化这些产能。这与2000年左右的光缆泡沫类似。

https://www.forbes.com/sites/danrunkevicius/2025/03/24/is-the-ai-boom-headed-for-its-dark-fiber-moment/?utm_source=chatgpt.com

OpenAI 走向巅峰或者破产?

2026 是 OpenAI 的生死关键年

1. **财务失控**： 预计 2026 年将烧掉 **170 亿美元**现金；2025 年上半年 OpenAI 的**推断成本已超过营收。**
2. **护城河消失**：Google 的 **Gemini 3** 模型在多项指标上已超越 GPT-5.1，OpenAI 失去了绝对的技术统治力；在欧洲等关键市场，ChatGPT 的消费者订阅增长已出现停滞迹象。
3. **传统科技公司模式靠拢**：为填补资金黑洞，OpenAI 将于 2026 年在 ChatGPT 中**植入广告。**

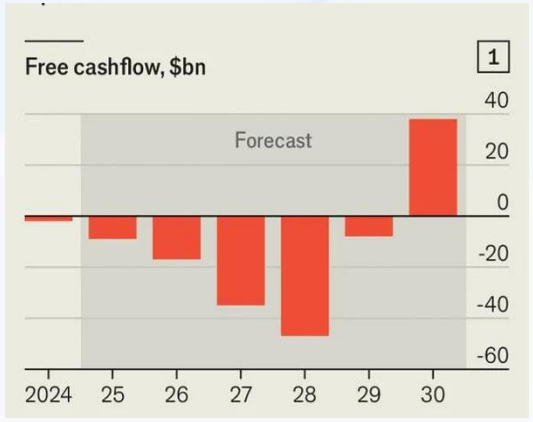
https://eu.36kr.com/en/p/3624813808157699

OpenAI's 2026: Either achieve divinity or go bankrupt

新智元 2026-01-04 17:02

Bubble or singularity?

2026 will be a make-or-break point for OpenAI. Facing an astonishing projected cash hole of $17 billion and a fierce counterattack from Google's Gemini, Altman was forced to activate the "Code Red." On one hand, there is an unprecedented financing plan of hundreds of billions of dollars; on the other hand, there is a financial crisis with the inference cost exceeding the revenue. Is this the inevitable path to AGI, or is it the eve of the bursting of the biggest bubble in Silicon Valley?



AI泡沫论及其崩盘风险?

经济学家Goldfarb和Kirsch评估技术泡沫的四个关键因素，并认为AI在每一点上都高度符合

Goldfarb and Kirsch's framework for evaluating tech bubbles considers four principal factors: the presence of uncertainty, pure plays, novice investors, and narratives around commercial innovations. The authors identify and evaluate the factors involved, and rank their historical examples on a scale of 0 to 8—8 being the most likely to predict a bubble.

Goldfarb和Kirsch评估科技泡沫的框架考虑**四个主要因素**：不确定性的存在、纯粹的业务公司、新手投资者以及围绕商业创新的叙事.....在**0到8分的量表中**，AI是**8分**。

Uncertainty不确定性

AI的长期商业模式、法律风险（如版权诉讼）等都极不明确。一些主要公司（如OpenAI、Anthropic）仍在巨额亏损，且一项MIT研究指出95%采用生成式AI的企业**并未从中获利**。

Pure Plays “纯粹业务”公司

纯粹业务公司是指其命运与某项特定创新的成功与否紧密捆绑的公司。根据评估框架，当一个行业出现大量纯粹业务公司时，它更有可能过热并**产生泡沫**。

四个因素

Novice Investors新手投资者

大批初入市场的散户投资者正通过交易平台涌入AI相关股票，而由于AI是全新领域，所有投资者在某种程度上**都是“新手”**，没有人知道它将如何发展。

Great Narrative 宏大叙事

行业领袖们持续宣扬着这样**一个无可匹敌的叙事**：“通用人工智能很快就能胜任人类的所有工作，并将开启一个我们仅能初步想象的超能技术时代。”

https://www.wired.com/story/ai-bubble-will-burst/


The Atlantic

TECHNOLOGY

Here's How the AI Crash Happens

The U.S. is becoming an Nvidia-state.

By Matteo Wong and Charlie Warzel

无论AI公司成功（实现超级智能，导致大规模失业和经济震荡）还是失败（投资泡沫破裂引发金融海啸），都可能对社会造成巨大痛苦和混乱。

投入与产出脱节

与人工智能相关的支出如今对美国GDP增长的贡献已超过所有消费者支出的总和。

those AI expenditures accounted for 92 percent of GDP growth during the first half of 2025. Since the launch of ChatGPT, in late 2022, the tech

数据中心支出与经济其他领域之间的巨大差距。人工智能领域的历史性投资与其相对微薄的收入之间存在巨大差异。

between the historic investments in AI and the tech's relatively modest revenues. For instance, according to *The Information*, OpenAI likely made $4

Data-center build-outs aren't the same as subprime mortgages. Still, th
plenty of precarity baked into these investments. Data centers deterior
rapidly, unlike the more durable infrastructure of canals, railroads, or e
fiber-optic cables. Many of the chips inside these buildings become ob
within a few years, when Nvidia and its competitors release the next w
bleeding-edge AI hardware. Meanwhile, the returns on scaling up chat
are, at present, diminishing. The improvements made by each new AI
are becoming smaller and smaller, making the idea that Silicon Valley
spend its way to superintelligence more tenuous by the day.

这些投资中隐含了大量的风险。数据中心老化速度极快，与运河、铁路甚至光纤电缆这些更持久的基础设施形成鲜明对比。随着英伟达及其竞争对手推出新一代尖端人工智能硬件，这些建筑内部的许多芯片在短短几年内就会过时。与此同时，目前扩大聊天机器人规模所带来的回报正在递减。每一款新人工智能模型的提升效果正在变得越来越小，这使得硅谷通过烧钱来实现超强人工智能的想法日益显得站不住脚。

两种灾难性结局

如果因为AI公司未能兑现承诺而导致科技股下跌...将导致金融损失蔓延至养老基金、共同基金、保险公司和普通投资者

causing the financial damage to spread to pension funds, mutual funds, insurance companies, and everyday investors. As capital flees the market, non-

如果AI公司兑现了其巨额投资，可能意味着创造出一种...消灭无数工作岗位的技术...；如果它们失败了，也可能引发前所未有的金融动荡。

(Perhaps we will be unable to adapt at all.) If they fail, there will likely be unprecedented financial turmoil as well.

访问时间：2026.1.21

正确认识AI的现状及其角色

关于“自主智能”：

- AI完全自主的说法更多是科幻憧憬，是科幻电影中赋予机器的“人格”。
- 现有技术路径下，并不存在真正意义上的“自主智能”，AI不能自行
其行为仍**来源于人类预**

“拟人化”叙事已成为AI界常态：

- 把描述人类思维和智能的术语用于AI，诸如“理解”“思考”“意识”“心智”甚至“硅基生命”“数字员工”等，既误导公众，也导致错误的科
自身认知能力为目标的科研。
工智能：目标为何?

何为智能化?

- **本质上应该是AI-powered，AI化。**
- **不能因为简称，忘了“人”才是地球上真正的智能体!**
- *一个例子*：大多数设计工作均是人类的智力性创造性活动，是智能的体现，如果将计算机辅助设计或AI设计称为“智能设计”，那么，人类的设计算什么?

- 2024年诺贝尔经济学奖
Johnson）著作《权力与
设计是尽可能地取代人类
）和西蒙• 约翰逊（Simon
展已经误入歧途，许多算法的
不是取代人类”。

- **人类的认知能力是人类**
知能力，加上语言工具的使用，可以相互交流、汇聚群智，使得人类能够成为地球的主宰。
- **我们可以接受机器在感知智能方面超越人类，但在认知领域，人类的主体地位必须坚守!**

目录

积极拥抱并用好当前的AI

基于深度学习的数据智能正在发挥广泛的赋值赋能作用，对各个学科、领域带来新的发展视角，甚至是颠覆性的影响。以软件学科为例：

基于基金委重大研究计划建议

国家自然科学基金重大研究计划立项建议

面向人机物融合的

智能化软件基础研究

主管学部：交叉科学部

相关学部：信息、数理、工材、管理

汇 报 人：梅宏、吕建、王怀民

2024年9月27日·北京-国家自然科学基金委员会

汇报内容

① 立项背景与依据

② 总体目标与核心科学问题

③ 研究基础、条件与队伍

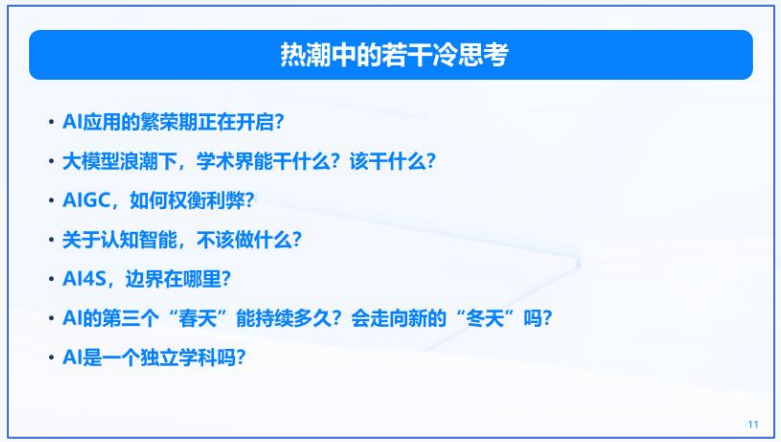
④ 计划框架与组织方式

指导专家组名单

序号	指导专家组	姓名	年龄	职称	研究领域	单位
1	组长	梅宏	61	教授 院士	软件体系结构	北京大学
2	副组长	吕建	64	教授 院士	计算机系统结构	南京大学
3	成员	王怀民	62	教授 院士	智能软件的工程方法	国防科技大学
4	成员	蒋昌俊	62	教授 院士	网络计算	同济大学
5	成员	李克强	61	教授 院士	智能网联汽车	清华大学
6	成员	顾斌	56	研究员	航天控制软件	航天五院502所
7	成员	包刚	60	教授 院士	应用数学	浙江大学
8	成员	韩旭	56	教授	复杂装备先进设计理论与方法	河北工业大学
9	成员	吴俊杰	45	教授	数据科学、智能管理	北京航空航天大学

撰写工作组成员：
马晓星、毛晓光
彭　鑫、刘譞哲

积极拥抱并用好当前的AI



AI应用的繁荣期正在开启?

- AI+ 或 AI for X，期望很大!
- 雷声隆隆，但雨点还不大。还需要一段时期的**探索、磨合和积累**。
  - **探索**：行业真需求。数字化、网络化后，才可能智能化。
  - **磨合**：提升可信性。现有技术路径下，至少安全攸关领域难应用。
  - **积累**：获取全数据。关键是跨越足够时间尺度!
- 当前的问题：
  - 泡沫太大：仍处于Hype Cycle的高峰阶段，喧嚣埋没理性，需要一个冷静期。
  - 以偏概全：对成功个案不顾前提地放大、泛化，过度承诺。
  - 期望过高：用户神化AI的预期效果，提出难以实现的需求。

当前如果有彷徨，能做些什么？—— **积累数据**
- 可采尽采
- 能存尽存

12

如何用好大语言模型?

预训练+精调（fine tuning）获得的大语言模型，其能力本质上是基于数据归纳的推断（inference），当前其最大的问题是“幻觉”。一方面，需要探讨如何不断提升模型自身的准确性和可信性，另一方面，需要通过软件工具构建大模型系统，提升可靠性。

路径一：

- 以大语言模型作为核心模块，结合检索、校验、计算、推理等辅助工具模块，形成一个**更可靠的AI系统**。
  - 将LLM与检索模块结合的RAG（Retrieval Augmented Generation，检索增强生成）；
  - 通过CoT（Chain of Thought，思维链）让模型分步推理，以提高准确性、可信性；
  - 调用其他验证过的“计算”“推理”模块。
- 面向具体任务，围绕大模型构建代理（Agent）。

路径二：

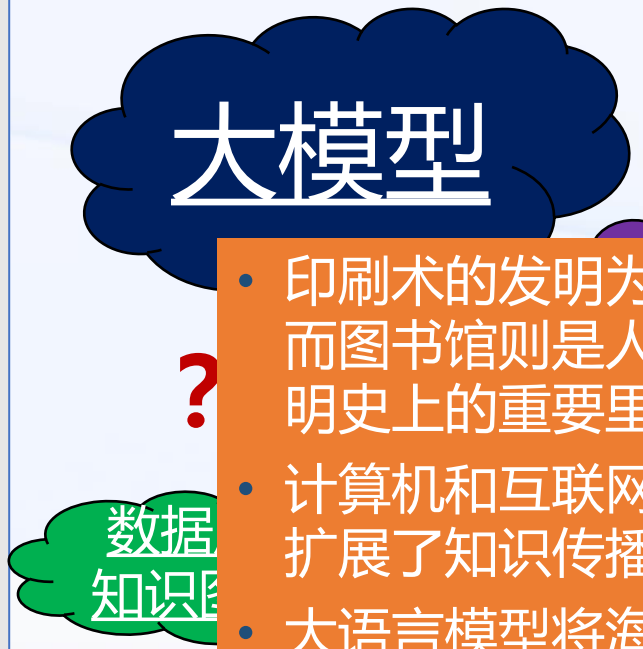
- 构建专门的知识库或知识图谱来**沉淀**经过验证的推断。形成知识的不断累积。

如何用好大语言模型?

预训练+精调（fine tuning）获得的大语言模型，其能力本质上是基于数据归纳的推断（inference），当前其最大的问题是“幻觉”。一方面，需要探讨如何不断提升模型自身的准确性和可信性，另一方面，需要通过软件工具构建大模型系统，提升可靠性。

路径一：

- 以大语言模型作为核心模块，结合检索、校验、计算、推理等辅助工具模块，形成一个**更可靠的AI系统**。
  - 将LLM与检索模块结合的RAG（Retrieval Augmented Generation，检索增强生成）；
  - 通过CoT（Chain of Thought，思维链）让模型分步推理，以提高准确性、可信性；
  - 调用其他验证过的“计算”“推理”模块。
- 面向具体任务，围绕大模型构建代理（Agent）。



路径二：

- 构建专门的知识库或知识图谱来**沉淀**经过验证的推断，形成知识的不

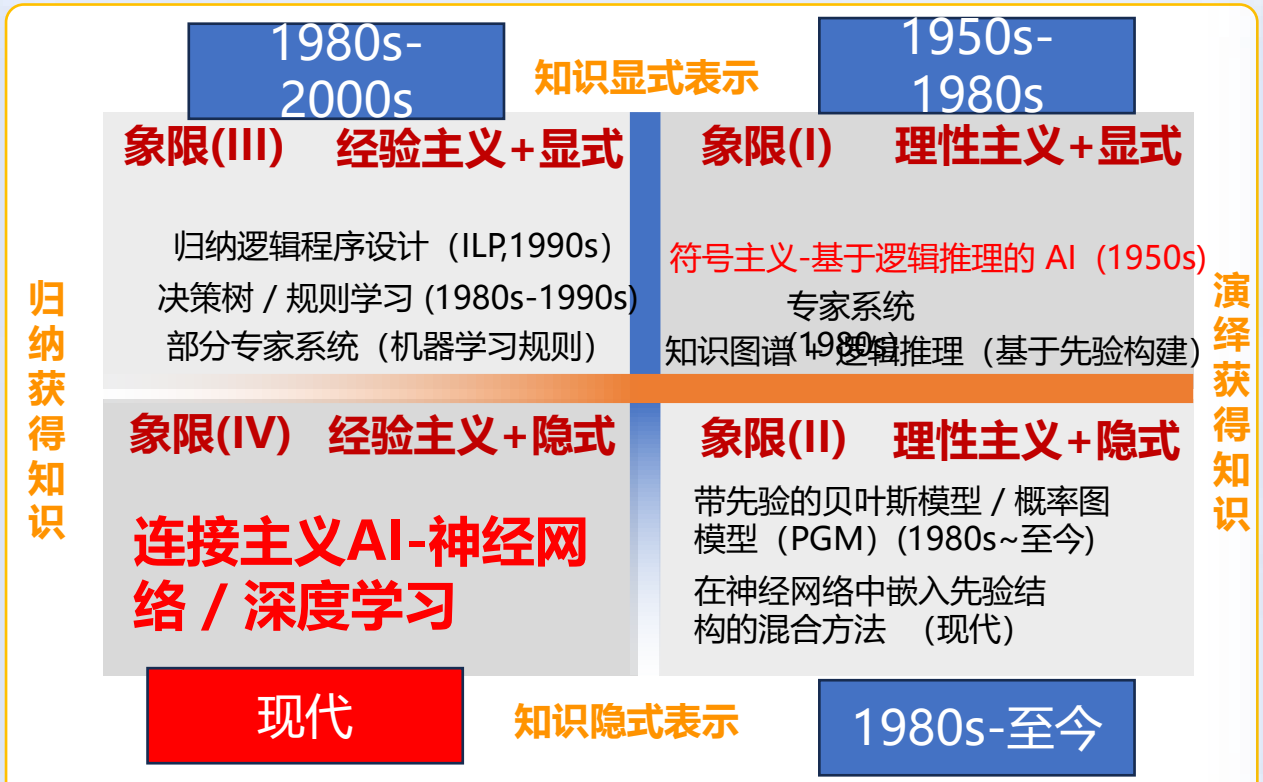
- 印刷术的发明为知识的记载和传播提供了有效工具，而图书馆则是人类知识积累和分享的基础，是人类文明史上的重要里程碑！
- 计算机和互联网的出现，实现了知识的数字化，极大扩展了知识传播分享的广度和效率。
- 大语言模型将海量知识“压缩”为一个“**动态知识库**”，并以灵活的交互方式供用户“按需”使用。它不仅存储信息知识，还能根据需求动态生成答案、分析和综述。是**新一代的知识基础设施。**

AI研究的多样性正在丧失

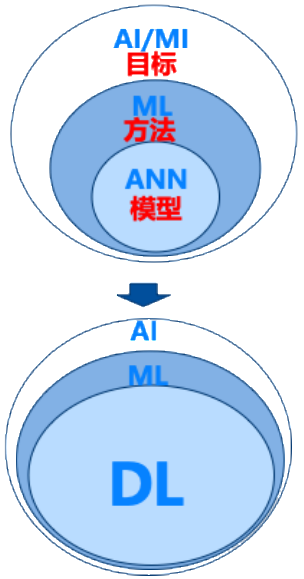
1956年达特茅斯会议七大核心问题——包括自动计算机、语言、神经网络、计算规模理论、自我改进、抽象化、随机性与创造性。

符号主义 连接主义 行为主义

达特茅斯会议AI的主要目标：
模拟人类的高级智能。



AI≈机器学习≈深度学习



当前语境下的AI已经失去了研究的多样性！

数据智能是否是人工智能发展的唯一路径?

AI研究的多样性正在丧失

AI的第三个“春天”能持续多久？会走向新的“冬天”吗？

- 我回答不了！但是，**至少我不希望**在当前的技术路径上持续这个“春天”。
  - “不可解释性”不符合人类发现知识、发明技术的基本逻辑，希望“知其然并知其所以然”是人的天性，更应该是科学家遵循的基本原则。
  - 以Scaling Law为“信仰”的大模型训练，以过度的资源消
    - Universal Approximation Theorem只是证明了能够构建出如何构造该网络的具体方法，也未说明需要多少神经元或
    - Scaling Law提供了一个“经验参考”，即一个通过观察实“可能”用包含多少参数的模型、用多少的训练数据、用多一个解释“通过扩大模型参数或数据和计算资源来提升模型
  - 不应该将Scaling Law奉为圭臬，走“拼”数据和计算资源
    - 虽然基于当前的技术路径，大模型尚不能“无中生有”，做“蛮力”、追求规模，也极易发展出在覆盖面和复杂度上人

AI的第三个“春天”能持续多久？会走向新的“冬天”吗？

- 科学家在探索自然的过程中，一直在追求为世界建模，遵循的基本准则是**“简而美”**。一些“异曲同工”的说法，阐释了相同的道理
  - 妙言至径，大道至简。—《还金述》，陶埴（五代）
  - 结果应该至简，而不仅仅是相对简化（Everything should be made as simple as possible, but not simpler）。— 爱因斯坦
  - 第一性原理（First principles）。
- 然而，按照Scaling Law产出的结果并不符合这个原则，而且，仅利用大模型通过“黑盒”的方式直接获得结果，而不去探索其背后的原理和规律，不是也不应该是科研锚定的目标！
- 都不希望AI的又一轮“冬天”到来，但是，沿袭当前技术路径，AI的**能力“天花板”**似已隐隐可见！

重述“对当前AI热潮的几点冷思考”，2024中国图灵大会

人工智能未来发展需要跨学科

从神经科学、心理学到其他自然科学，不同学科也许可为理解和推动人工智能发展提供丰富多元的视角。

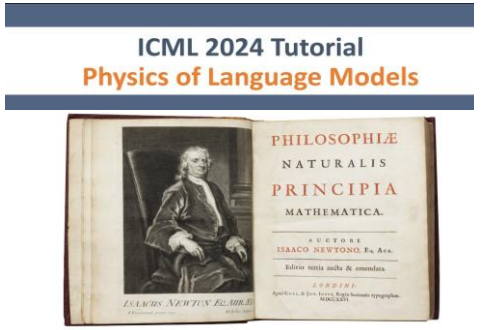
神经元　突触　脑区

脑科学

脑科学在神经元活动、脑区功能及其协同机制等方面取得重要进展，揭示了学习和记忆的生物学基础，特别是突触可塑性在信息处理中的关键作用，以及前额叶皮层、海马体等脑区在认知功能中的核心地位。

心理学

心理学对情绪、注意力与社会交互的研究也表明，情感因素对决策和记忆具有深远影响，而镜像神经元的发现则进一步深化了我们对共情与社会行为的理解。



物理实验设计原则

变化条件　变量控制　重复性

发现通用规则

自然科学

借助自然科学（如物理学）的实验方法对大模型能力的涌现机制进行解释。

探寻
人工智能的
智能产生机理
的方法启示

符号化的重要性

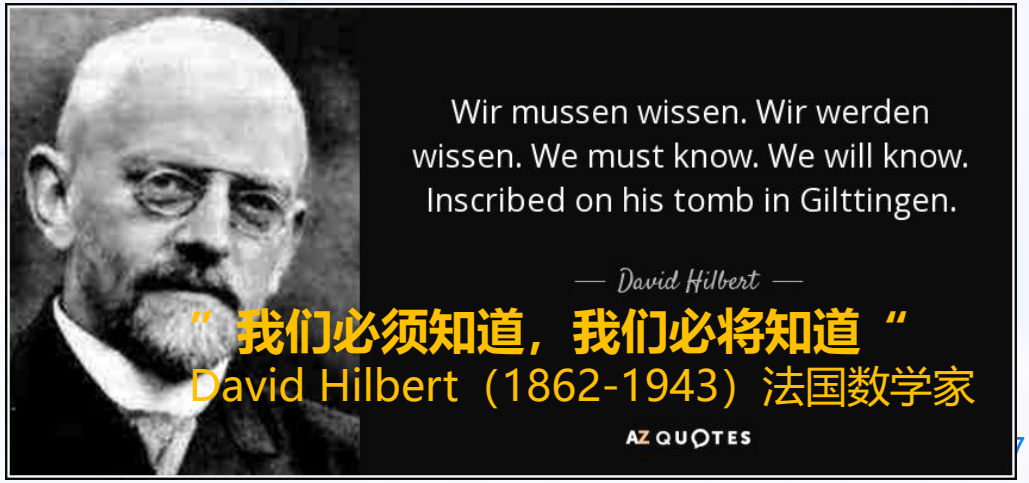
符号化是知识沉淀与传播的关键

- 从古希腊的先哲，到科学革命以后的科学家，在努力探索了解自然的过程中，一直在追求为世界建模，而且是采用“**符号化**”的方式，包括使用自然语言、图形、数学公式、编程语言等。
  - 唯如此，才能实现知识的**交流和传承**！

- 人类社会历经漫长发展历史，构建起知识“发现—验证—传播”的一套规范的知识体系和制度。
  - 人类知识主要依靠各领域科学家、专家学者的创新发现获得，并以语言文字的形式记录并呈现，通过各类有效的制度，包括学术界、出版界的共同把关，保障知识的准确性和可信度，并经过长期的实践检验，逐步形成人类知识体系。

- 基于感知数据（音视频等）的学习去建模世界固然重要，当前的“世界模型”“空间智能”“端到端”等概念突显了人们对此的关注，但是，模型必须和符号语言建立**锚定（grounding）**关系，否则，只能用于机器去感知、操纵环境，无益于人类对世界的了解。
  - 动物也具有感知和学习特定规律的能力，但由于缺乏语言或符号系统，它们无法将经验上升为抽象知识，也无法将所学“交流”给同伴。

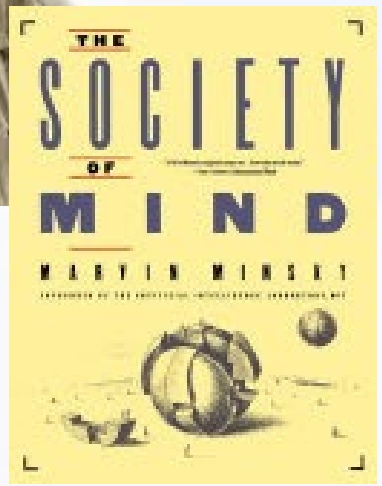
- **不论基于符号体系建模世界在能力上有多么不足（理论上决定于“哥德尔不完备性定理，1931年”），也是我们无法避开的、甚至是唯一可以采用的路径！**



下一代的AI，连接主义和符号主义的融合？

至少是我个人的期望！ 期待符号主义的回归，回归达特茅斯会议的“初心”：让机器模拟人类的“高级智能”。

Marvin Minsky
（1927-2016）
首个AI领域图灵奖得主，1969
AI奠基人之一

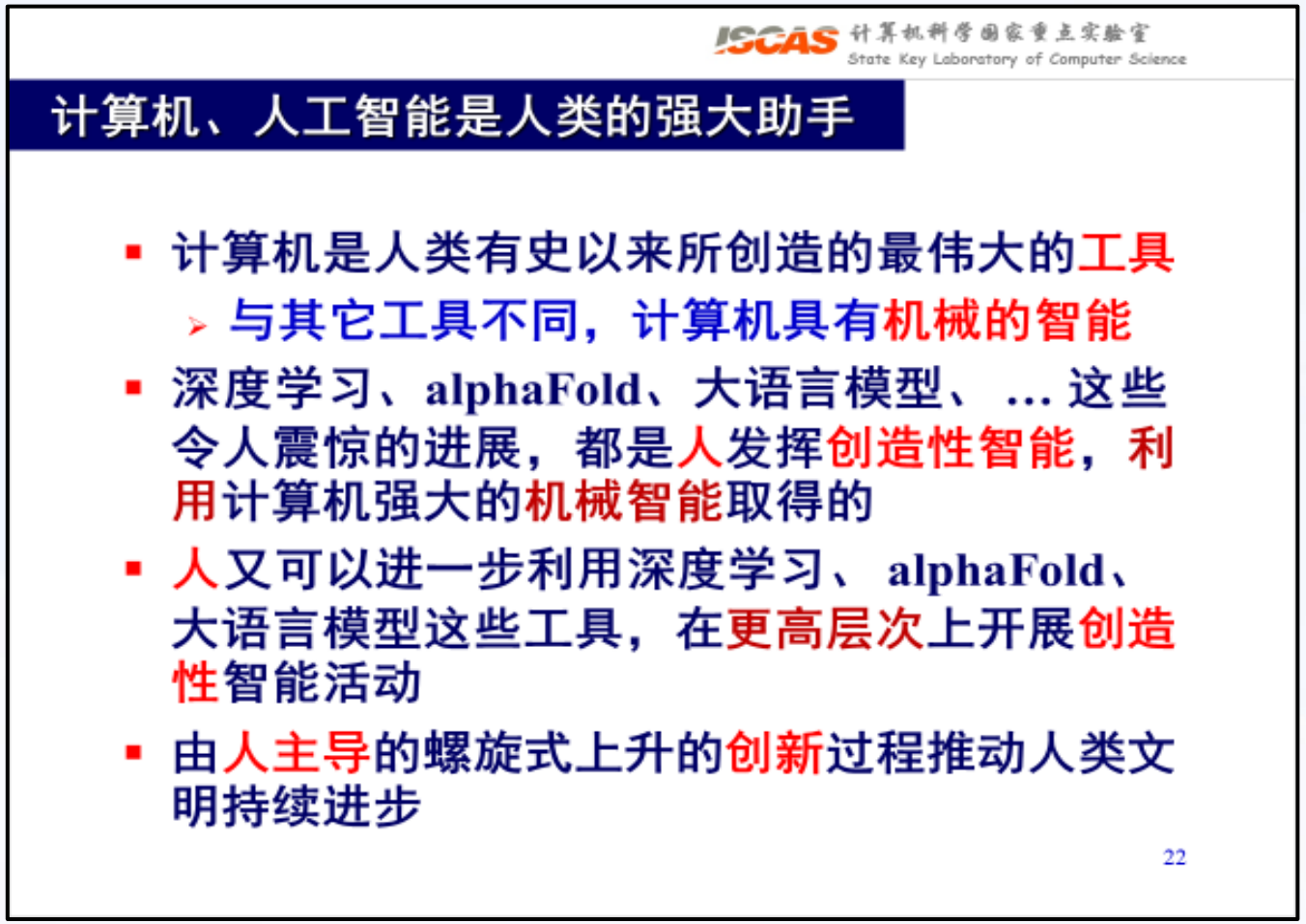


《The Society of Mind》
《心智社会》，1986

“究竟是什么神奇的诀窍使我们变得智能？诀窍就是根本没有什么诀窍。智能的力量源于我们广泛的多样性，而非任何单一的、完美的原则。”（“What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.”）
——《心智社会》，第308页

再次强调AI的角色

- AI的角色：过去是工具，无论是现在还是未来，无论它发展到多么强大，仍然应该定位于“人类的工具”！帮助我们提升工作效率和质量。不能脱离掌控，更不能主动“让位”！



摘自林惠民院士“计算 智能 安全”，2024中国计算机大会

谢谢！

专栏 中国计算机学会通讯 第20卷 第9期 2024年9月

对当前人工智能热潮的几点冷思考

梅 宏
北京大学

如果要票选当前社会经济发展中的热词，“人工智能”（Artificial Intelligence, AI）一词一定可以入选；如果要盘点当前科技界的重大进展，AI 技术的进展也一定会入围，并会名列前茅，虽然其进展主要集中在深度学习这个子领域。用“如日中天”“众星捧月”描述当下 AI 的地位毫不为过！造成这种少见现象的原因不仅仅是 AI 技术本身 10 余年来令人惊艳的突破性进展，还包括 AI 技术对社会经济各行各业，对科学技术各个学科的发展带来的深远甚至颠覆性的影响。在我国，AI 的热潮较国际更盛。

我不是 AI 具体技术的研究者，但作为 AI 领域的大同行，作为 AI 技术发挥作用的基础——计算系统中软件技术的研究者，从当前的热潮中，我看到了太多“炒作”和“非理性”导致的 AI “过热”现象，也对当前 AI 发展技术路径多样性的欠缺萌生了一些担忧。我认为，有必要认真审视和反思当前 AI 的本质及其可能带来的影响，以及人类发展 AI 的目标和路径等问题。为此，在今年 7 月于长沙举行的 2024ACM 中国图灵大会上，我作了一个演讲，题为“对当前人工智能热潮的几点冷思考”。本文根据该演讲整理而成。

AI发展的简短回顾

人类对“用机器模拟人类智能”的追求远早于现代电子计算机的发明。如果将算术计算的能力也视为智能（实际上，在教育还不那么普及的时代，算数能力确是智者的能力），那么历史上的各类计算装置，例如帕斯卡的加法器、莱布尼茨的机械计算器和巴贝奇的差分机，无疑都是对用机器模拟算数能力这一目标的追求。就其第一目标而言，图灵机模型以及第一台电子计算机 ENIAC 均是针对“计算”这类智能的，不过，随着电子计算机在计算能力上超越人类，人们开始不把计算等同于智能了。1950 年，艾伦·图灵（Alan Turing）在其论文[1]的开篇首句即提出：“机器能思考吗？”该问题后来演变为“机器能表现得像人一样吗？”。此即是图灵测试提出的问题基础，也成为后续几十年 AI 研究追求的重大目标，甚至是终极目标。

对机器智能的担忧几乎伴随了追求模拟人类智能的全过程。1942 年，科幻作家艾萨克·阿西莫夫（Isaac Asimov）就在其小说《环舞》（*Runaround*）中就提出了机器人三定律[1]，其目的就是为了警示机器智能可能给人类带来的威胁。

1956 年，“人工智能”一词被提出，在经历两个“春天”和两个“冬天”后，AI 迎来了第三个“春天”。而这个“春天”显得尤为生机勃勃，呈现一片“繁荣”景象。从 2012 年底，在 ImageNet[2]图像分类任务上，AlexNet[3]展现出了深度学习的强大感知能力，2015 年初，ResNet[4]超过人类的平均水平；到

[1] 第一定律是机器人不得伤害人类，或因不作为而让人类受到伤害；第二定律是机器人必须服从人类的命令，除非这些命令违背了第一定律；第三定律是在不违背第一与第二定律的前提下，机器人必须保护自己。