

基于人工智能的商业模式创新

胥正川 博士

复旦大学管理学院 信息管理与商业智能系



课程框架体系

数字时代的技术演变路径与现状

- 1、数字时代环境变迁 (AI+BC)
- 2、数字技术的逻辑路径
- 3、移动互联网引发大数据营销革命
- 4、物联网引发企业数智化转型升级

AI基本原理与运作模式

- 1、AI基本原理与优缺点
- 2、AI的技术演变规律与特征
- 3、从图像识别应用到AI工厂

AI 1.0至2.0的原理、模式与前瞻

- 1、AI 1.0至2.0大模型原理
- 2、AI 2.0时代的应用实践
- 3、AI 2.0的前瞻变化 (To B, 准确可控, 小模型, 智能体)

AI应用模式

- 1、AI应用全流程业界经验介绍

- 2、重构AI实施认知

AI企业应用方法论

- 1、AI应用场景创意选择
- 2、AI实施全流程规划

清晰概念，系统框架，战略思考，启发思维



目录 Contents



1. 大数据AI时代的技术演变路径与现状



2. 人工智能的演变、原理与趋势



3. 企业AI创新应用模式



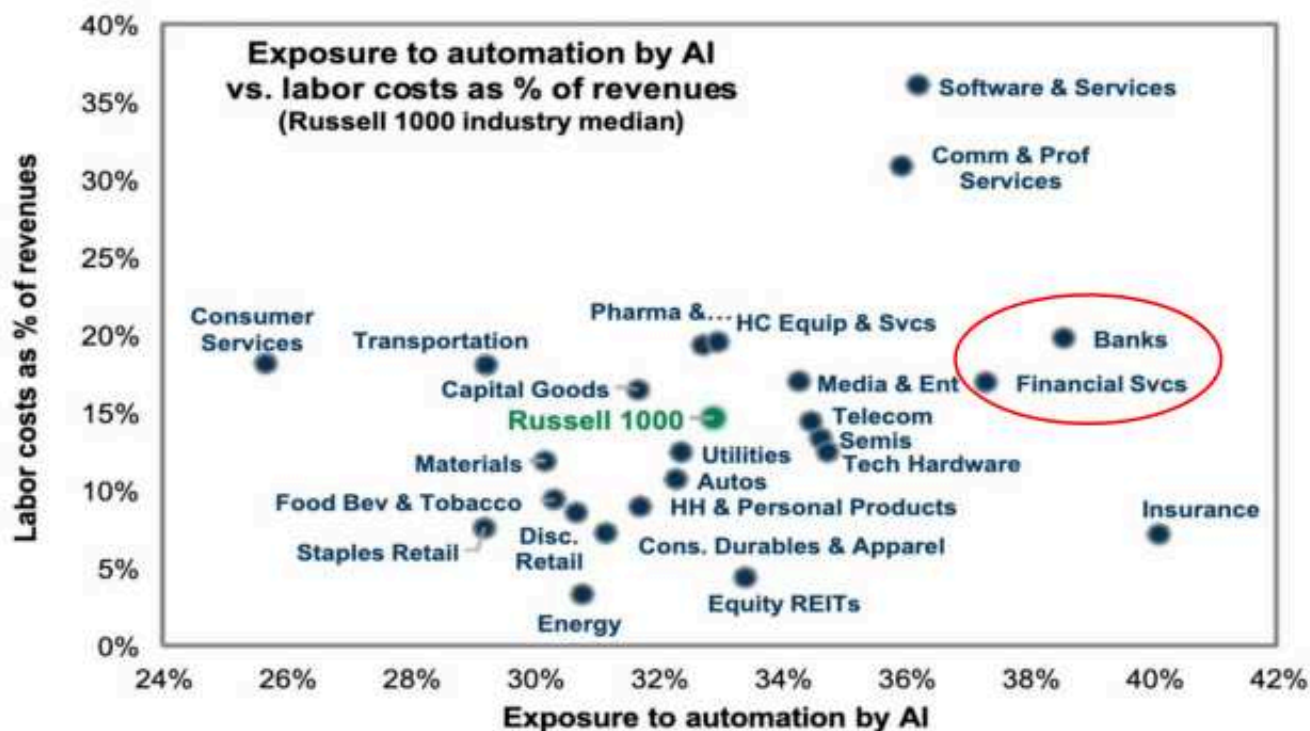


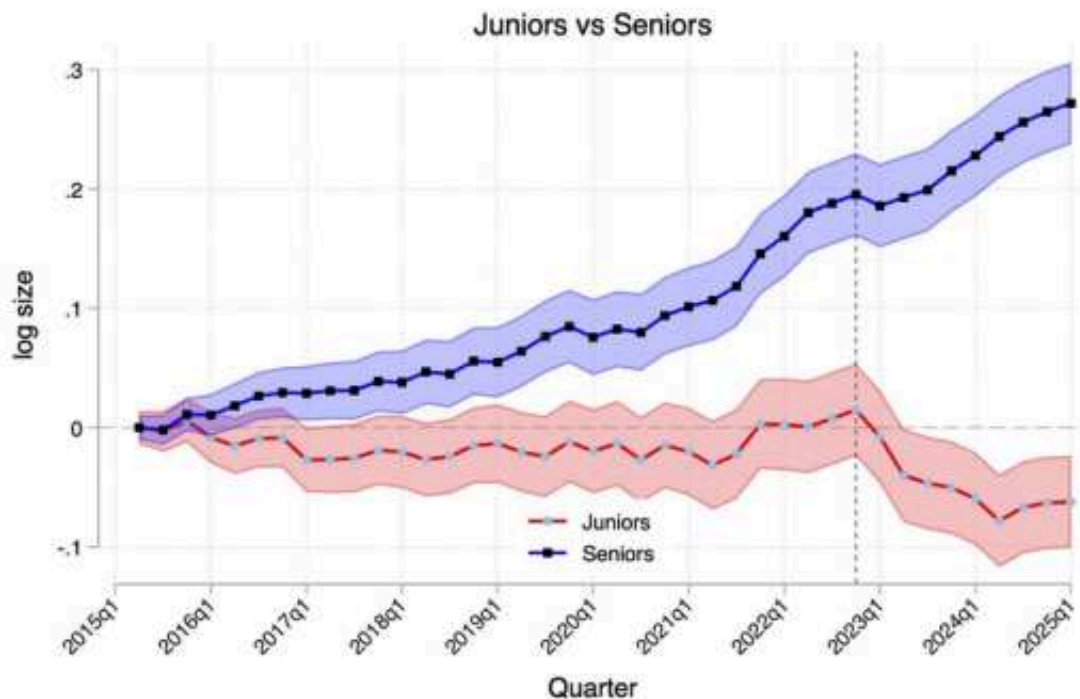
All rights reserved.





应对：掌握技术特性 + 理解商业模式变迁 + 技术驱动的应用和管理





All rights

AI时代，人类只剩下两种绝对价值

即将消失 / 被取代者

1



纯执行者

只是被动完成任务，不负责最终结果。代码机器、初级设计、机械文案。

2



对齐者 / 中间人

促进结果发生但不承担责任的中层管理或跨部门沟通节点。

不可替代 / 幸存者

1



提需求的人 (Definers)

拥有独特的行业洞察、私有上下文、以及定义“我们要去哪里”的极致想象力。

2

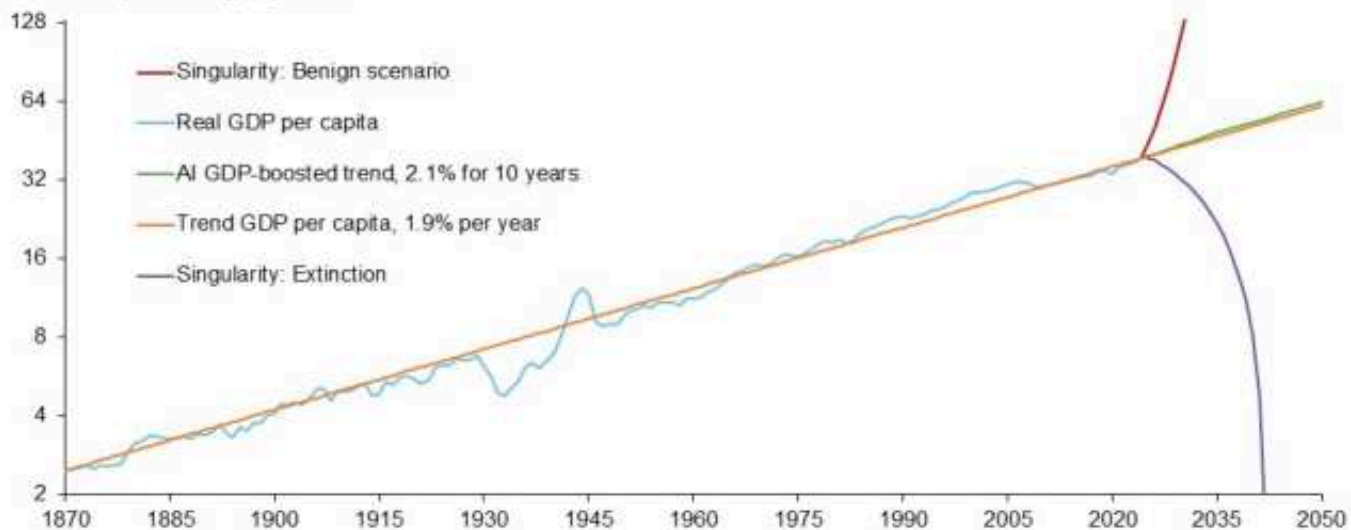


背锅的人 (Blame Takers)

承担法律、商业和最终结果责任的人。例如：签字画押的代码审查员、律师、甚至“龙虾维修员”。

Chart 1
AI scenarios

1990 dollars (thousands), log scale



NOTES: The blue line is real gross domestic product (GDP) per capita in 1990 dollars. The orange line is a trend line fitted to the data for 1870-2024 with a trend growth rate of 1.9 percent per year. The red, green and purple lines are hypothetical paths for per capita GDP based on different scenarios.

SOURCES: Bureau of Economic Analysis; Haver Analytics; Macrobhistory.net; United Nations; authors' calculations.

Federal Reserve Bank of Dallas



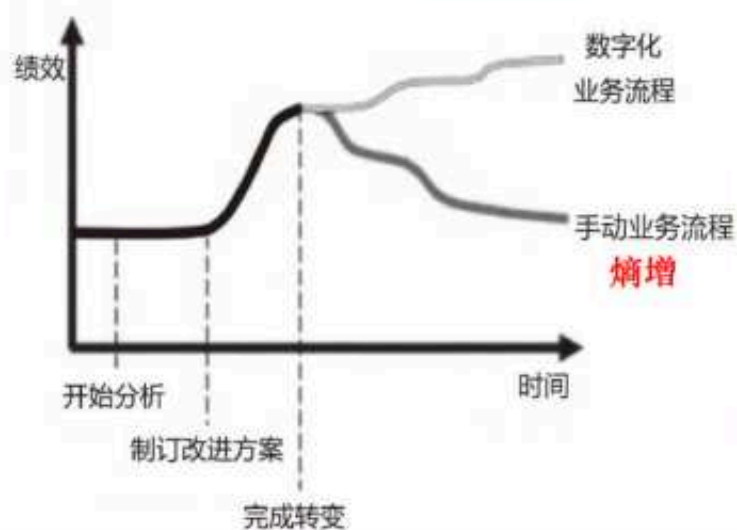


数智化转型升级的收益（AI思维）

规模、范围、学习能力



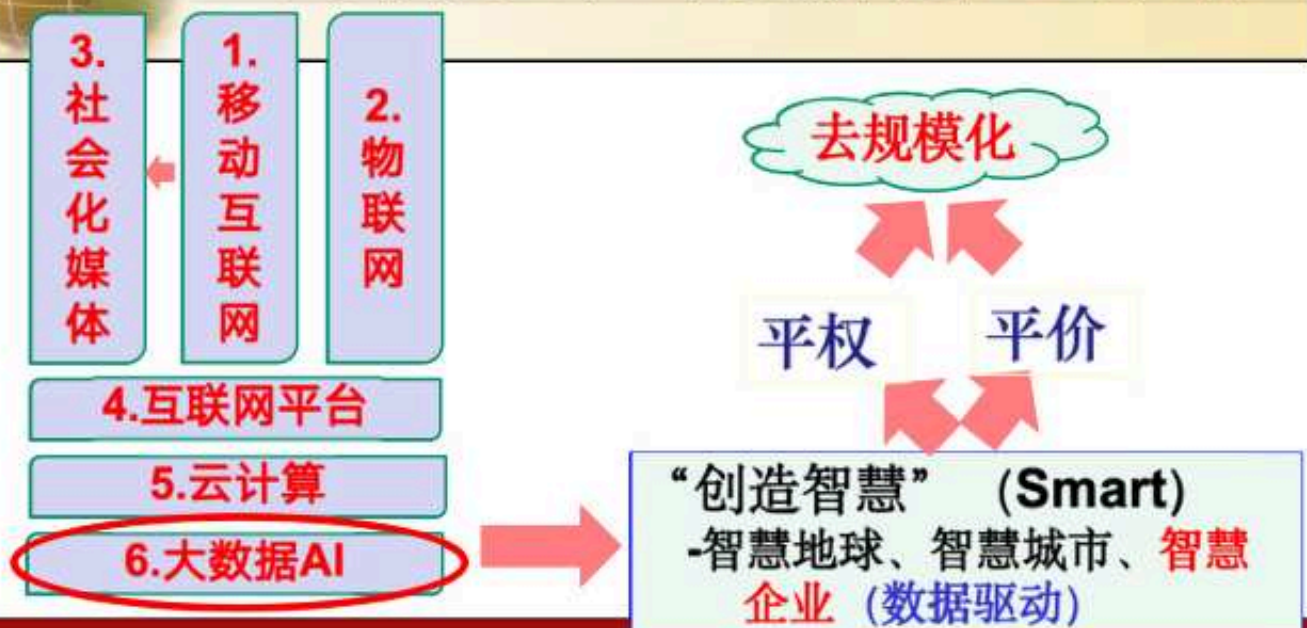
数字化业务流程与手动业务流程的表现对比



数字技术，加速与产业融合

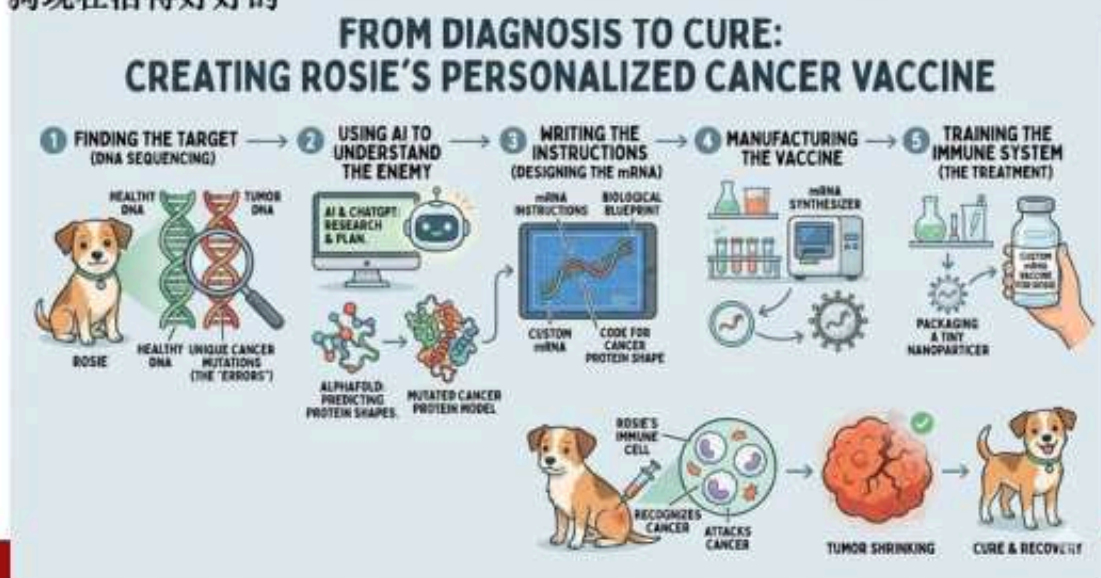
- 当前热点数字技术：ABCDE技术
 - A: 人工智能 (AI)，“风起云涌”
 - 人工智能正成为与人类互动的强大协作工具，正加速企业智能化创新与转型
 - 人工智能是最重要的“颠覆手段”，已经基本达成共识
 - B: 区块链技术 (Blockchain)，“含苞欲放”
 - 我国央行正式试点基于区块链的数字货币
 - IDC预测，全球区块链支出将于2023年达到159亿美元
 - C: 云计算 (Cloud Computing)，“尘埃落定”
 - 全社会/全产业的基础设施，云计算仍处于高速成长阶段，龙头亚马逊等仍有40%+增速，我国阿里云增速近100%。Gartner数据显示，2022年，全球公有云市场将达到3312亿美元
 - 云存算一体化服务将成为发展趋势
 - D: 大数据分析/数据智能 (Data Analytic/Intelligence)“春暖花开”
 - 大数据已成事实，成为核心的生产要素，其意义非同一般性技术
 - “财富1000强”中90%以上的企业正在投资大数据资源及其分析应用
 - E: 边缘计算/物联网 (Edge Computing)“蓄势待发”
 - 将成为数字化颠覆的重要基础：智能硬件、设备无处不在，万物互联，智联网正在形成
 - 互联网的应用开发目前仍是初级阶段，物联网是方向，是数字技术发展的意义所在
 - 到2021年，全球将安装350亿台物联网设备，到2025年将安装754.4亿台

AI：人类历史上第一次机器自己产生了知识！

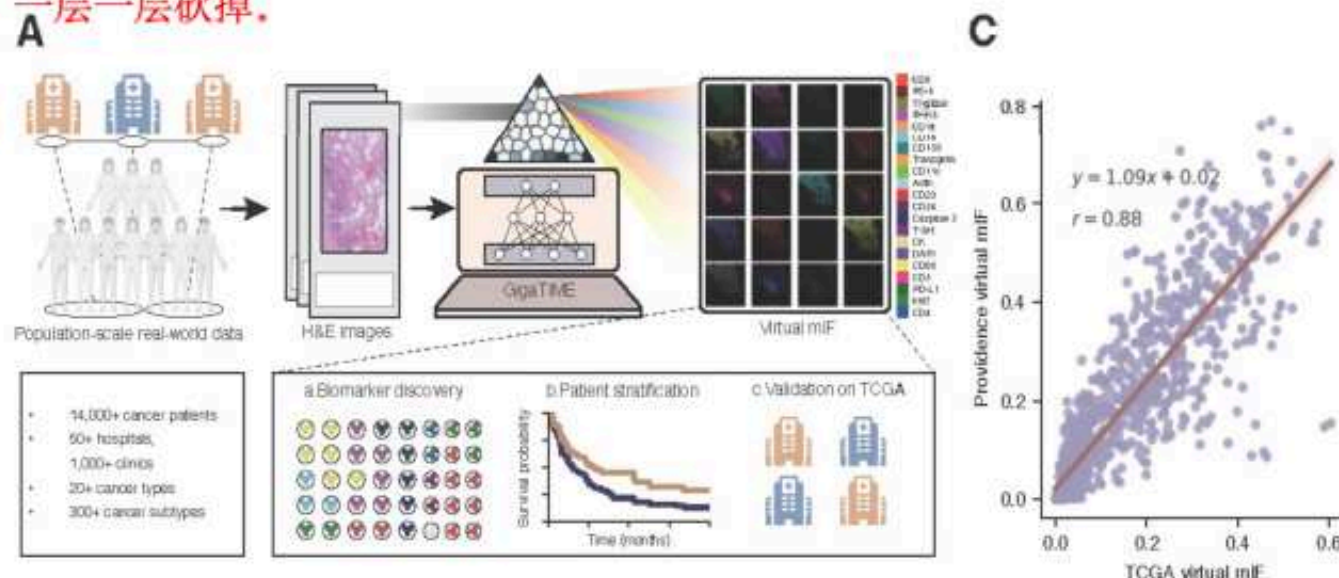


2026 会是 AI 在“**个性化医疗**”上真正引爆的一年。

自己给狗做 DNA 测序，然后搞出了一款“定制疗法”。- 用 ChatGPT 分析狗的 DNA，找出基因突变 - 把这些突变丢进 Google 的 AlphaFold（AI 蛋白质结构预测），用它来“设计”了一款针对这些突变的定制疫苗。- 给狗打了疫苗，几周之内肿瘤缩小了 50%，狗现在活得好好的



微软开源了癌症AI模型：GigaTIME。把\$5的切片变成\$5000的检测结果。检测肿瘤免疫微环境需要多重免疫荧光（mIF）成像——每张切片成本数千美元，最先进的实验室也只能覆盖极少数样本。精准医疗的门槛，正在被AI一层一层砍掉。





目录 Contents



1. 大数据AI时代的技术演变路径与现状



2. 人工智能的演变、原理与趋势



3. 企业AI创新应用模式



人工智能技术

人工智能商业应用 (认知智能)

无人驾驶	智慧医疗	智能金融	智能家居
智慧营销	工业4.0	智能制造	智慧教育
信息安全			

人工智能基础技术 (感知智能)

语音识别	计算机视觉	机器视觉	自然语言处理
机器学习	知识图谱		

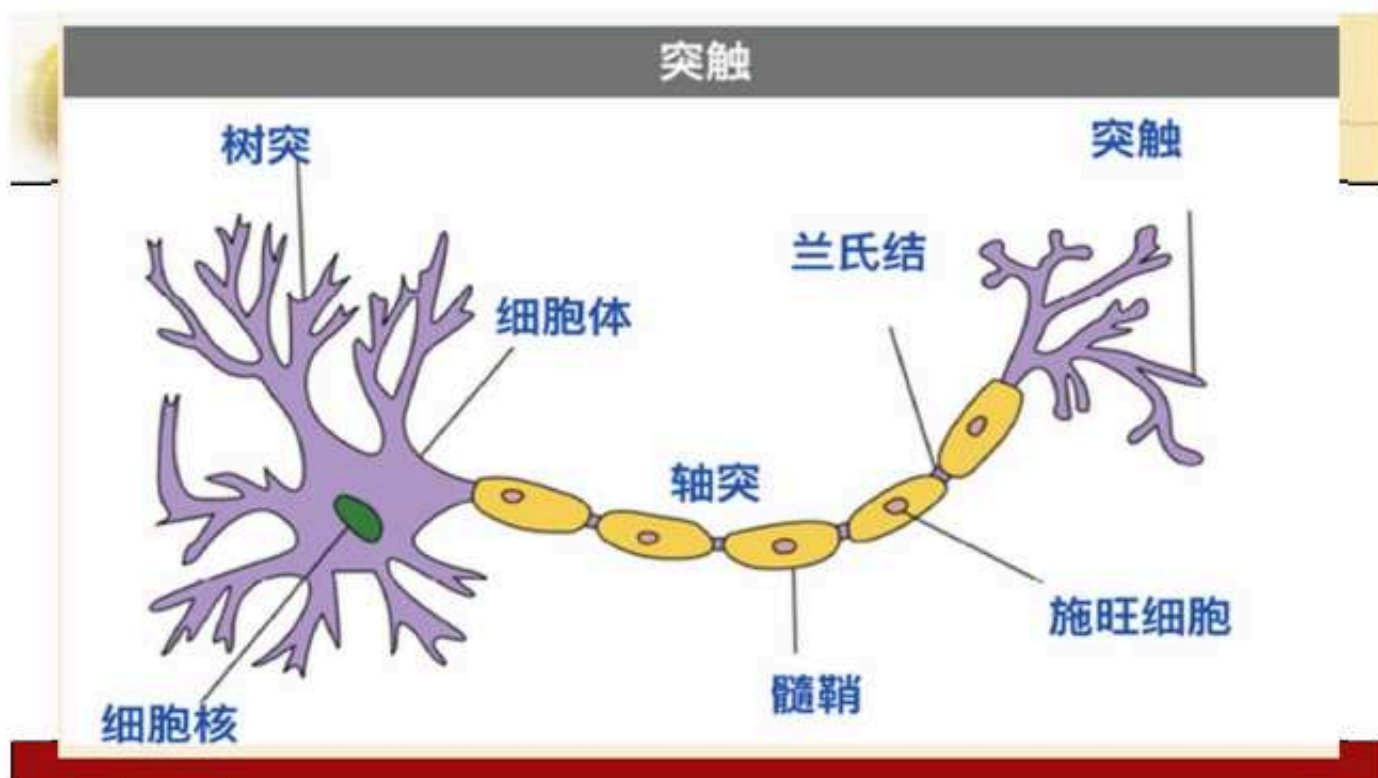
人工智能底层要求 (计算智能)

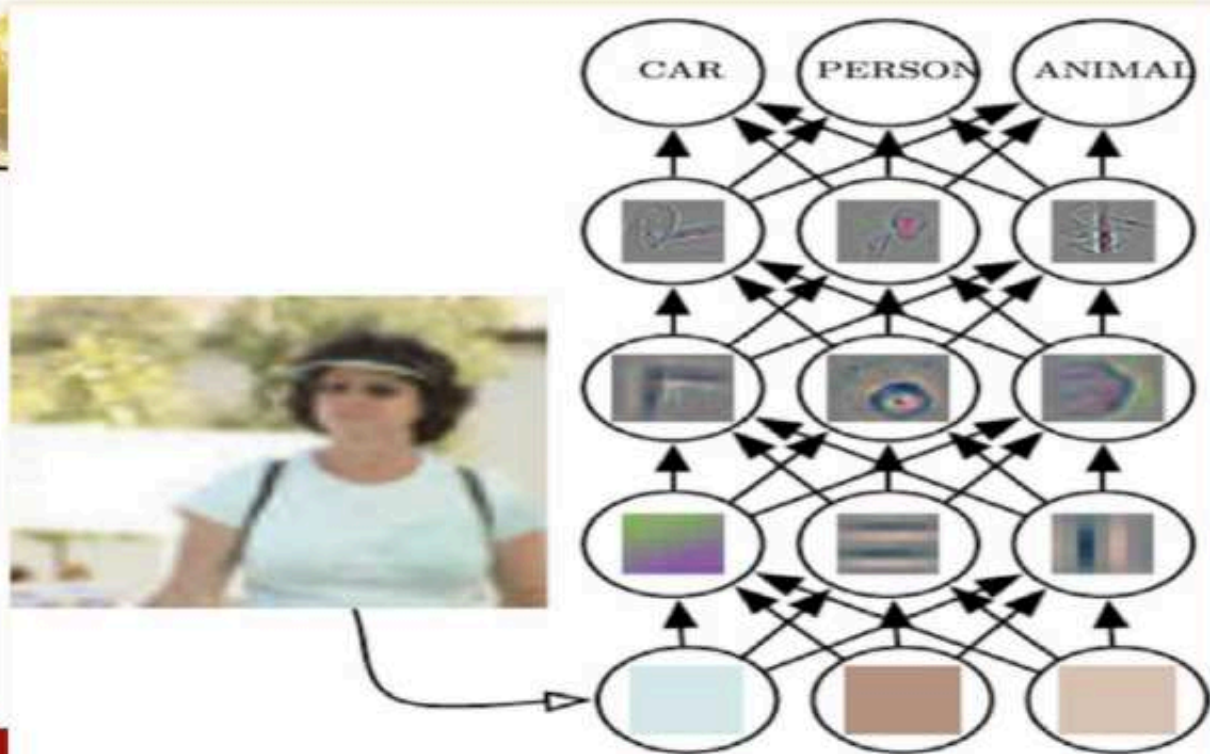
数据平台	数据储存	数据挖掘	云计算
------	------	------	-----



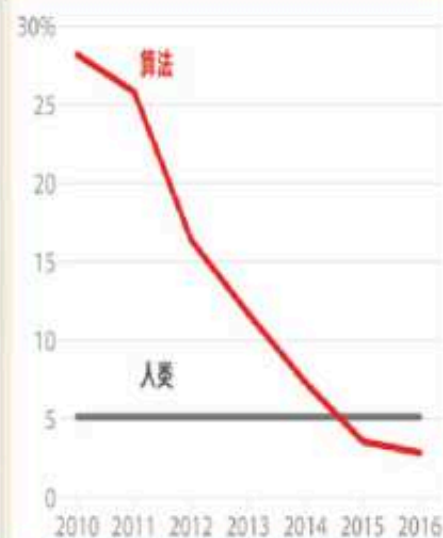
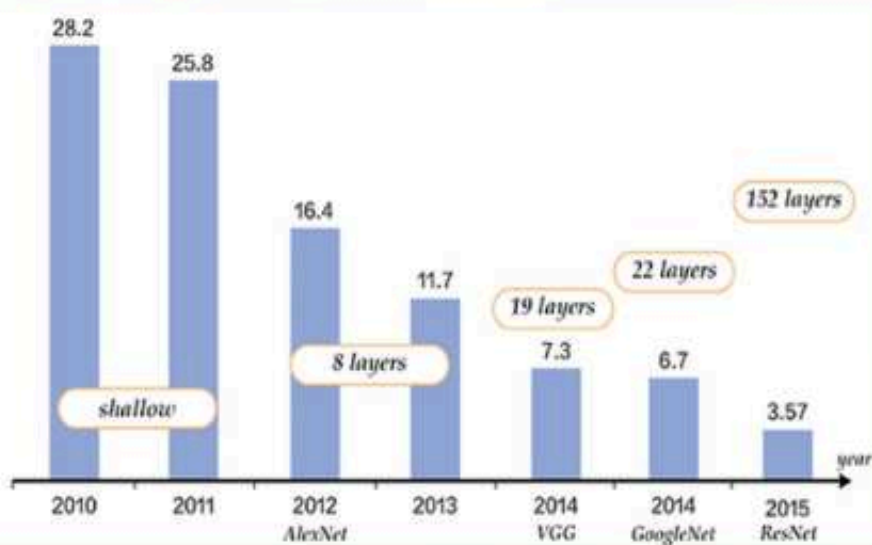
人工智能的派别

- (1) 符号主义，又称逻辑主义或计算机学派，其原理主要为物理符号系统(即符号操作系统)，如：专家系统。
(成人学习二外：显性知识) 白盒AI
- 经典算法：SVM, 决策树
- (2) 联结主义，又称仿生学派或生理学派，其原理主要为神经网络及连接机制与学习算法，如：阿尔法狗。
(幼童学习母语：隐形知识) 黑盒AI
- 经典算法：CNN, RNN, Transformer





2015年AI1.0获得突破





数据驱动AI基本原理——>弱人工智能



人工智能的优点、缺点、致命伤？



数据驱动的人工➡智能 (有监督机器学习)

$$Y = F(X), Y = a_1x_1 + a_2x_2 + a_3x_3 + \dots$$

输入组X $x_1, x_2, x_3 \dots$	输出组Y	应用软件
保单 (年龄, 性别, 健康状况...)	保单是否欺诈	语音识别
历史市场数据	未来市场数据	交易机器人
照片	图片说明文字	图像标注
药物化学性质	治疗效果	医药研发
店铺交易细节	本次交易是否为诈骗?	欺诈检测
食谱配料	用户评价	美食推荐
购买历史记录	未来购买行为	客户保持
汽车定位和速度	交通流	红绿灯
脸部	名字	脸部识别



AI（机器学习）预测与诊断原理

- 预测分析目标 $Y=F(X)$, 例如, 产品质量Y包括膜厚、粗糙度、附着力、耐腐蚀性、抗裂强度等指标
- 相关影响因子X
 - 镀膜机台（型号、编号）
 - 基材数据（来料材质、来料尺寸、来料特性等）
 - 靶材数据（靶材名称、靶材特性、配方比例等）
 - 工艺参数（周期时间、温度特性、电流、电压、功率、旋转速度、工作面距离、通气配方、压力、温度等）
- 数据样本单位：每条数据采自一个生产批次

29



AI助力破解物理学难题，推进核聚变控制

- MIT科学家用物理与机器学习结合的方法，首次**精准预测托卡马克内等离子体行为**，为核聚变能源的可控与规模化应用奠定基础。这一突破，或许是通往几乎无限的清洁能源的里程碑起点。
- 它像一个巨大的甜甜圈，利用强大的磁场来束缚高速运动的等离子体。当托卡马克全面启动时，内部的等离子体电流以**每秒约100公里的速度流动，温度高达一亿摄氏度——比太阳核心还炽热**。在这样的极端环境中，即使只是关闭反应堆，也是一项危险的艺术。操作员必须让电流逐渐“降档”，让反应慢慢平息，否则炽烈的等离子体可能在内壁留下灼痕和裂痕，需要耗费大量时间与资源修复。
- MIT研究团队Allen Wang在采访中说：“如果聚变要成为真正可用的能源，它必须稳定可靠。而要做到可靠，我们首先得学会如何驯服我们的等离子体。”新方法正是为此而生。**通过结合物理学定律与机器学习算法，他们建立了一个模型**，能够在给定初始条件下**预测等离子体在托卡马克内的演化轨迹**。这项成果发表于《Nature Communications》，被认为是聚变控制领域的重要里程碑。

- “如果放电过程失控，即使是在降功率阶段，也会产生极强的热流，对反应堆内部造成严重损伤。”Wang解释道，“尤其在高性能运行时，任何微小的波动都可能逼近不稳定边界。这就像在刀刃上行走。”
- 而现实更残酷：因为核聚变实验成本极高，全球绝大多数托卡马克一年只能运行几次，研究人员根本无法反复试错。于是，他们必须用理论与算法来“模拟未来”，让每一次实验都物有所值。
- 为突破这一限制，MIT团队做了一件聪明至极的事：他们让AI“学习”物理本身。研究者将神经网络与描述等离子体动力学的物理模型结合，再利用瑞士一座名为TCV的小型实验聚变装置的数据进行训练。数据涵盖了实验前的温度、能量变化，以及整个运行过程的详细记录。
- 随后，算法会生成一条条“轨迹”，相当于一份反应堆操作指南，预测在不同阶段等离子体会如何变化。团队将这一算法应用到真实实验中，结果令人惊喜——只要按照模型建议的轨迹操作，就能让反应堆平稳、安全地完成“熄火”。
- “我们做了多次实验。”Wang回忆道，“每次都更顺利、更高效，数据证明确实改善了整个过程。”他说，他们的目标不仅是让聚变“能用”，更要让它“可控、可重复”。“这只是漫长征途的起点，”Wang说，“但这是一个美好的开始。”



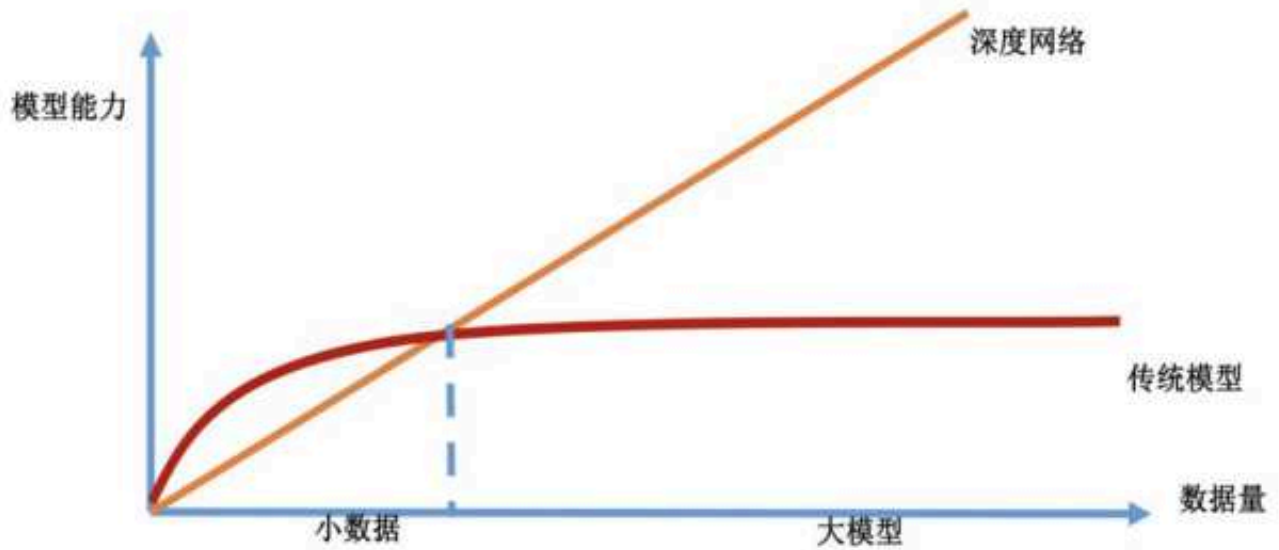
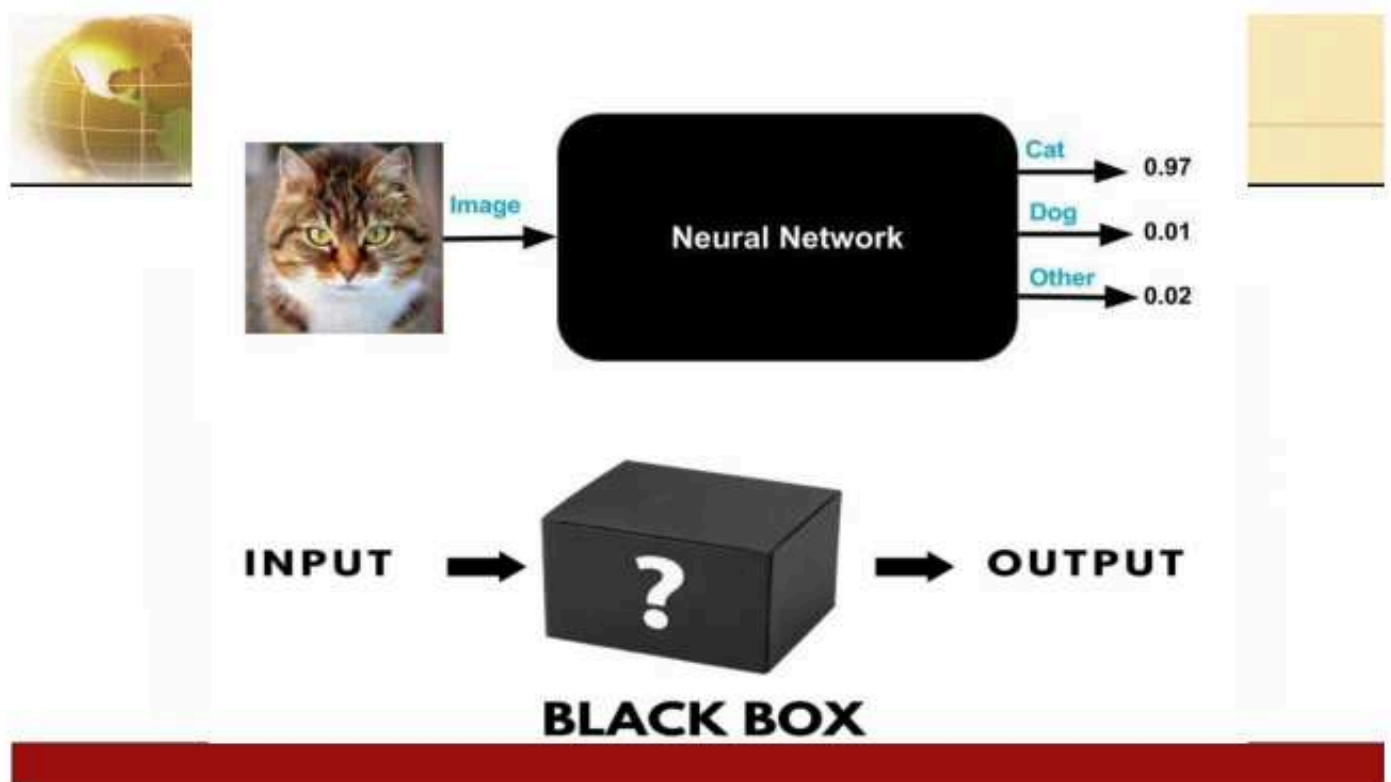
翻译这事儿,google认为**大数据**
比语言专家靠谱



Google
Translate

Break through language barriers

本质1：去专家化



数据+算法+算力

应用场景包括机器人、智能医疗、智能驾驶、智能家居等细分行业，基于

人工智能的技术及应用，各应用领域

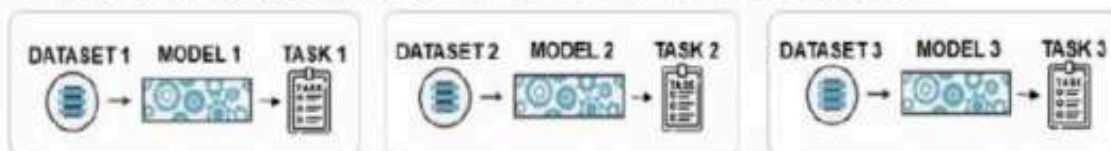
在AI的1.0时代：存在模型碎片化明显、AI泛化能力不足等问题。



基础支撑主要由数据提供和计算能力支撑两部分组成，为人工智能的技术和产业发展提供支撑，是人工智能产业的基础设施；

单一领域数据集；诸多数据集和诸多模型各成孤岛缺乏纵效；劳动密集型的数据标注

AI 1.0



AI 1.0:人的经验系统化 → AI 2.0:人的智慧系统化

超级海量数据，无需人工标注；一个具有跨领域知识的“基础模型”，能执行五花八门的任务

AI 2.0



前神经网络 专家系统 概率推论

神经网络 基础模型 大规模模型

2011年以前

2012年至今

决策式/分析式AI时代

生成式AI急速发展

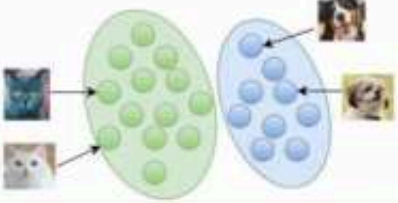


资料来源: Dynabench: Rethinking Benchmarking in NLP, Google Scholar, indigox.me, 公司官网, 中信建投

中信建投证券
CHINA SECURITIES

- AI模型可大致分为决策式/分析式AI (Discriminant/Analytical AI) 和生成式AI (Generative AI) 两类。
 - 决策式AI: 学习数据中的条件概率分布, 根据已有数据进行分析、判断、预测, 主要应用模型有用于推荐系统和风控系统的辅助决策、用于自动驾驶和机器人的决策智能体。
 - 生成式AI: 学习数据中的联合概率分布, 并非简单分析已有数据而是学习归纳已有数据后进行演技创造, 基于历史进行模仿式、缝合式创作, 生成了全新的内容, 也能解决判别问题。

表: 决策式AI与生成式AI对比

类型	决策式AI	生成式AI
技术路径	已知数据分别求解输出类别标签, 区分不同类型数据, 例如将图像区分为猫和狗 	分析归纳已有数据后创作新的内容, 例如生成逼真的猫或狗的图像 
成熟程度	技术成熟, 应用广泛, 辅助提高非创造性工作效率	2014年开始快速发展, 近期发展速度呈指数级爆发, 部分领域应用落地
应用方向	推荐系统、风控系统、决策智能体等	内容创作、科研、人机交互以及多个工业领域
应用产品	人脸识别、精准广告推送、金融用户评级、智能辅助驾驶等	文案写作、文字转图片、视频智能配音、智能海报生成、视频智能特效、代码生成、语音人机交互、智能医疗诊断等



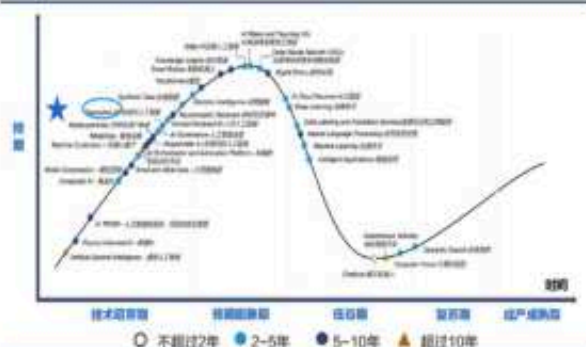
第四次科技革命，已经进入AI 2.0时代

人工智能1.0时代（2012年-2018年）：图像和语音识别+赋能各行各业。

人工智能2.0时代（2017年-至今）：2017年Google 提出Transformer架构，到2022年模型参数达到5400亿，模型参数呈现指数级增长，“预训练+微调”成为标准方法。

2022年Gartner人工智能商业化曲线

图表：Gartner2022年人工智能技术成熟度曲线



第四次科技革命标志：AI大爆发

图表：中国生成式AI技术应用规模（单位：亿元）



注：生成式 AI 应用规模的统计口径为应用生成式 AI 技术生成的数字内容的市场规模，统计方式为数字内容市场规模乘以生成式 AI 渗透率，以上规模的推导考虑国家商务部、GARTNER、第三方研究机构等多个数据源。

预训练：无监督机器学习的方法

AI最大的本领是学习/预测/制造/理解/干活

AI最大的本领是学习

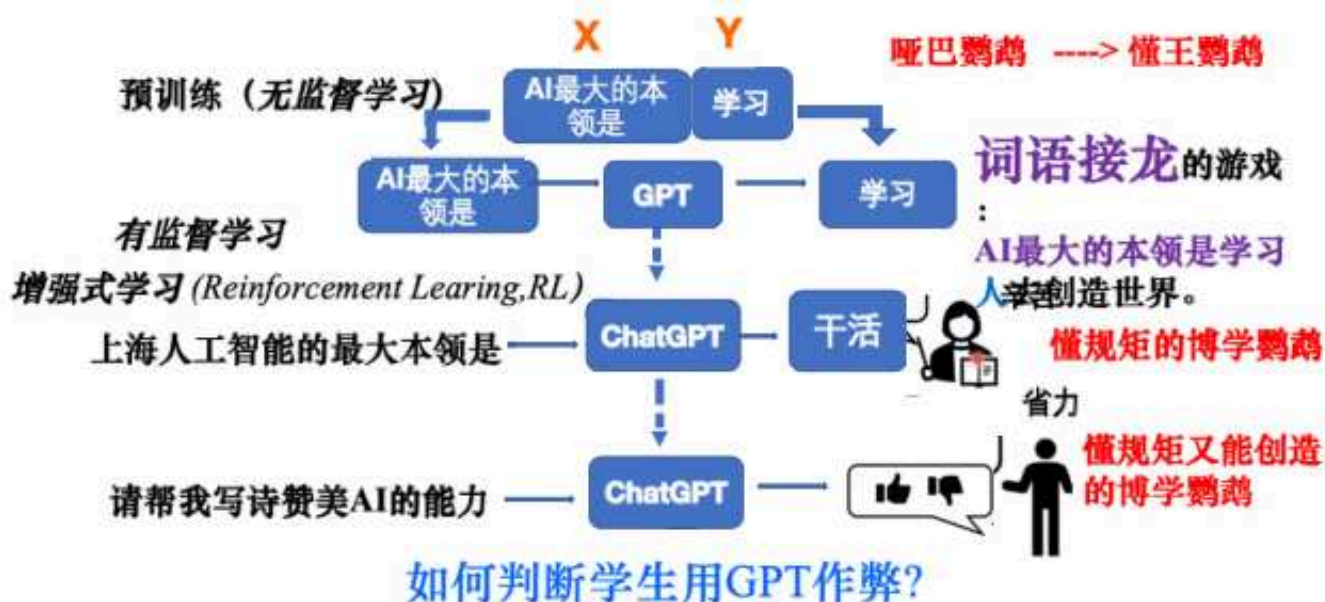
X

Y

AI最大的本领是

学习	4.5%
预测	3.5%
制造	3.2%
理解	3.1%
干活	2.9%

方法：预训练+监督学习的方式+强化学习



人工写答案 -> 人工选答案 -> 机器选答案

阶段1
收集演示数据并训练

阶段2
通过人工标注训练数据来训练

阶段3
使用PPO强化学习方法优化回答

“暴力美学”大力出奇迹!

- 1、大模型大参数 大语料：通识化：通用化
- 2、模考（强化学习）：RLHF (Reinforcement Learning From Human Feedback)
- 3、左右互搏，自我演化

GPT-3.5



利用以上排序结



的模型给出的回答，得到分数

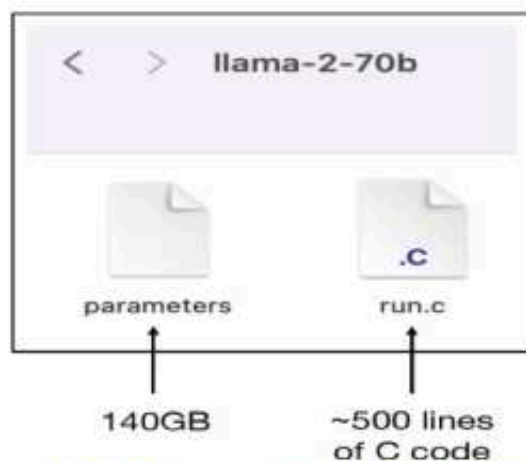


及分数/策略
可以更新
模型参数

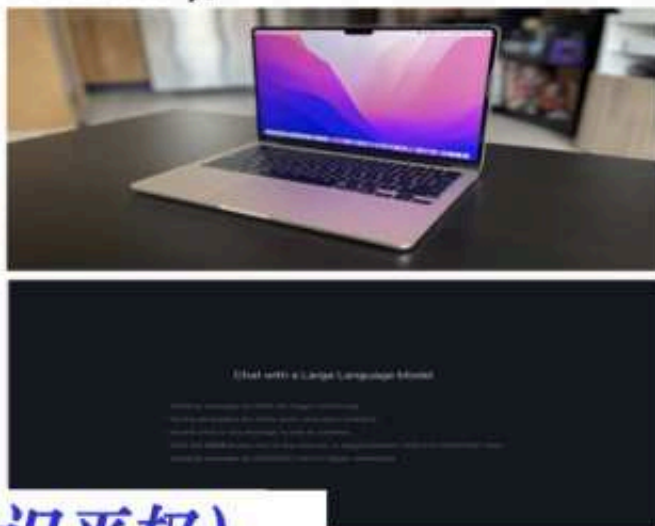
r_k

本质2：去人化

大模型实体文件（全网的知识无损压缩进大模型）



MacBook



本质3：再专家化（知识平权）

models by Andrej Karpathy

大模型的成本



Chunk of the internet,
~10TB of text



6,000 GPUs for 12 days, ~\$2M
~1e24 FLOPS



parameters.zip
~140GB file

*numbers for Llama 2 70B

引自《Intro to Large Language Models》by Andrej Karpathy

训练部署大模型的方法



every
~year

第一阶段:预训练

1. 下载约10TB的文本
2. 获得约6000个GPU的集群。
3. 将文本压缩成神经网络，支付两百万美元，等十二天。
4. 获取基础模型。



<USER>
Can you tell a short introduction about the concept of the term "responsibility" in economics? Please use examples related to personal responsibility in the labor market, where a responsibility employee has significant power over the wage and working conditions of their employees. The presence of a responsibility gap leads to more wages and reduced employment opportunities for workers, as the employer has little incentive to increase wages to provide better working conditions.

<ASSISTANT>
"Responsibility" refers to a moral obligation where one is responsible for a particular goal or action. In economics, this term is particularly relevant in the labor market, where a responsibility employee has significant power over the wage and working conditions of their employees. The presence of a responsibility gap leads to more wages and reduced employment opportunities for workers, as the employer has little incentive to increase wages to provide better working conditions.

every
~week

第二阶段:微调

1. 编写标签说明
2. 雇佣员工 (或使用scale.ai!), 收集10万个高质量的理想问答。
3. 在此数据上微调基模, 等待约1天。
4. 获取辅助模型。
5. 做很多评估。
6. 部署
7. 监控, 收集不良行为, 进入第一步。

引自《Intro to Large Language Models》by Andrej Karpathy

超大模型强化学习工程

创新本质: 人机协作的复杂AI软件持续演化系统工程 (算法、数据、软件工程)

人机协作的海量数据工程



对齐人类价值观的超大模型强化学习工程

无监督预训练超大语言模型
+
Reward Model
+
强化学习迭代优化

迭代飞轮

微软云Azure
+
大规模并发推理API

超级智能算力云大规模并发推理弹性服务工程

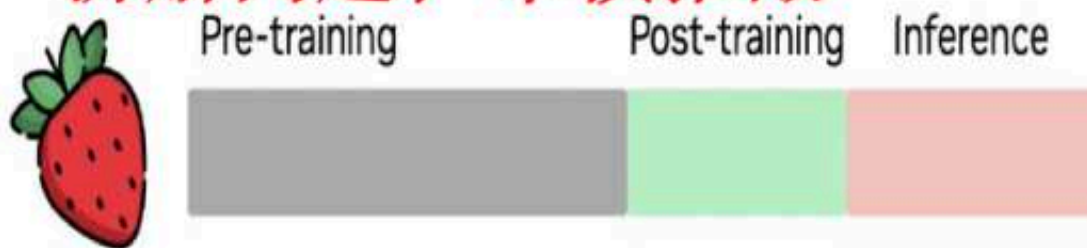
超级智能算力云大规模并行训练工程

微软云Azure
+
Ray大规模并行训练

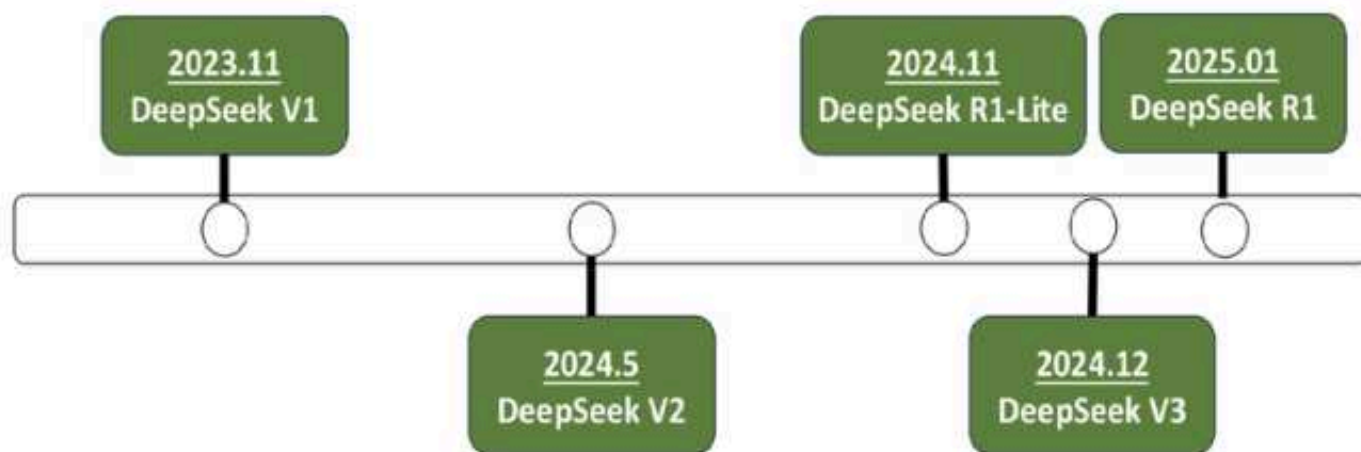




从“随机鹦鹉”到“深度思考者”
拆解问题，审慎推敲

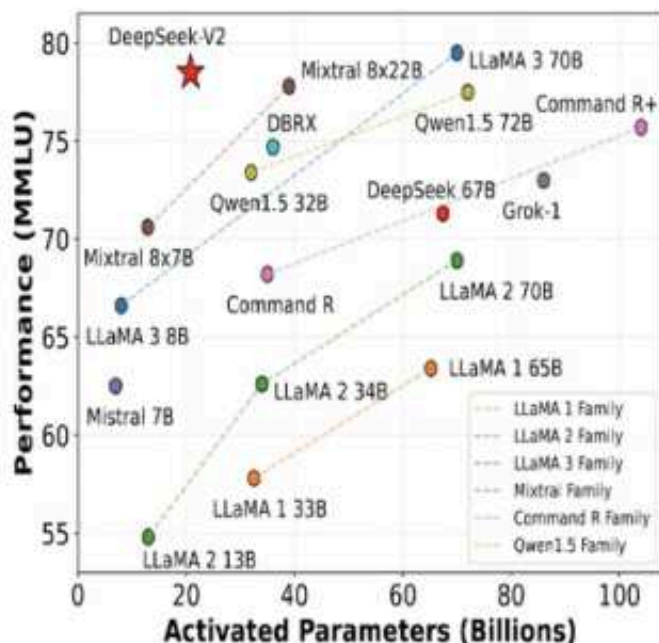


杀手锏：性能/成本曲线| 性价比



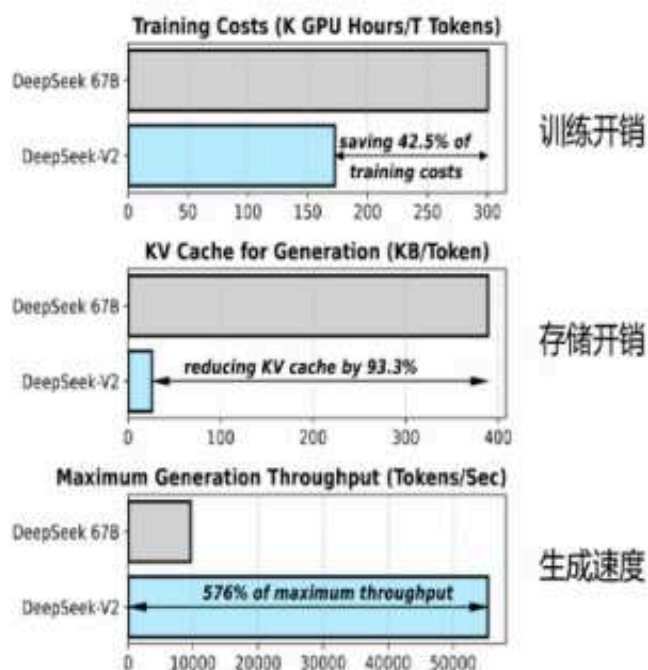
天边的两多云 (国内外现状)

- 模型架构: 大部分企业采用已验证架构 (试错成本高昂) 【不敢】
- 推理模型: 大部分实验室仍在苦苦猜测摸索Q*/o1 (OpenAI保密) 【不知】



(a)

source: 熊德意, <<深度解读DeepSeek: 原理与效应>>



(b)



DeepSeek R1主要创新

- o **DeepSeek-R1-Zero**:大规模RL训练, 发现了RL训练的 Scaling Laws, RL训练涌现 “aha”时刻
- o **推理模型训练技术框架**:4步法, 有效解决了R1-Zero存在问题, 将推理与对齐合为一体
 - o **强化学习训练框架**:GRPO, 来自DeepSeekMath, 降低了强化学习训练成本
 - o **推理模型蒸馏**:将大模型推理能力蒸馏到小模型, 优于小模型直接进行推理训练(规模效应)

source:熊德意,<<深度解读DeepSeek: 原理与效应>>

知识蒸馏 (Knowledge distillation)

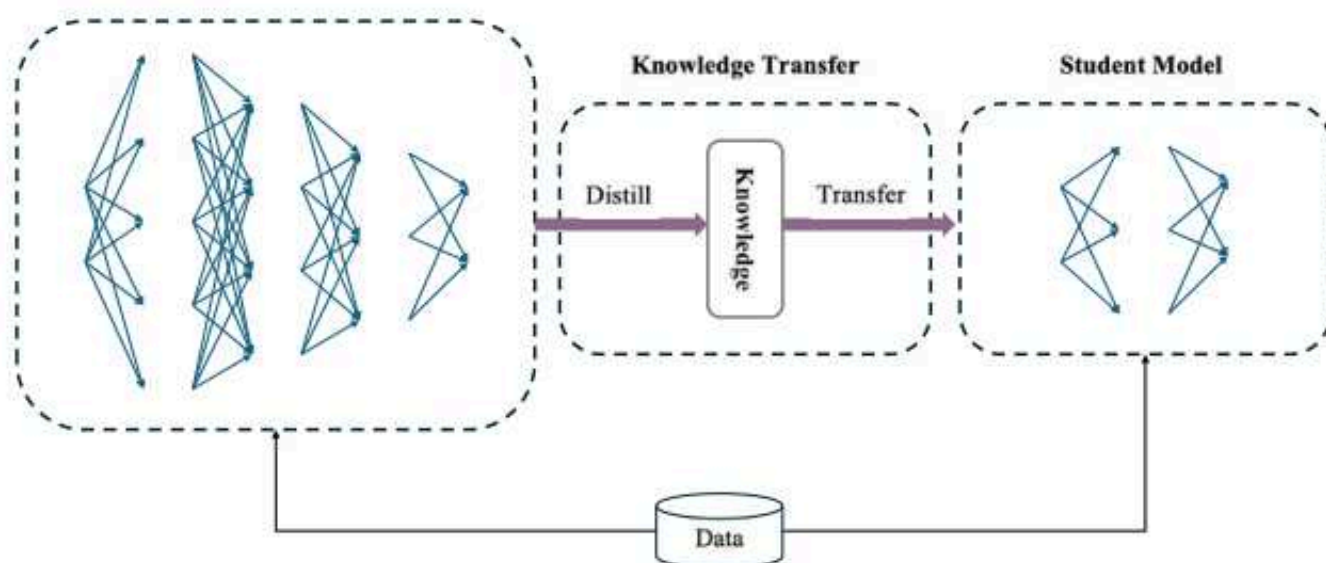




Figure 1: An identity jailbreak demonstration. The responses come from real samples.

DeepSeek：效应——打破美国霸权与认知误区

- 如果ChatGPT刷新了我们对AI的认知，那么DeepSeek在某种程度上颠覆了：
 - 美国人对中国AI水平的认知：长久以来，美国认为中国在AI科技创新上更多是跟随者角色
 - 大模型研发成本的认知：大模型研发成本需要数千万乃至上亿美元
- 成果：引发全球的AI繁荣与发展，端侧AI应用进入“寒武纪”时代

MODEL ↑↓	CREATOR ↑↓	CONTEXT WINDOW ↑↓	ARTIFICIAL ANALYSIS INTELLIGENCE INDEX ↑↓	BLENDED ↑↓ USD/1M Tokens	MEDIAN ↑↓ Tokens/s	MEDIAN ↑↓ First Chunk (s)
o3	 OpenAI	128k	70	\$17.50	154.1	17.08
o4-mini (high)	 OpenAI	200k	70	\$1.93	148.6	40.11
Gemini 2.5 Pro	 Google	1m	69	\$3.44	148.3	34.02
DeepSeek R1 (May '25)	 deepseek	128k	68	\$0.96	28.6	3.62
Gemini 2.5 Pro (May' 25)	 Google	1m	68	\$3.44	151.0	34.24
Grok 3 mini Reasoning (high)	 xAI	1m	67	\$0.35	63.7	0.34
o3-mini (high)	 OpenAI	200k	66	\$1.93	170.3	50.72
Gemini 2.5 Flash (Reasoning)	 Google	1m	65	\$0.99	349.9	15.83
Claude 4 Opus Thinking	ANTHROPIC	200k	65	\$30.00	57.3	2.66
o3-mini	 OpenAI	200k	63	\$1.93	166.8	13.24



目录 Contents



1.大数据AI时代的技术演变路径与现状



2.人工智能的演变、原理与趋势



3.企业AI创新应用模式

人机交互方式的变迁



ChatGPT的重要革命性意义在于：让AI模型进入了知识和推理的时代。目前很多AI应用最大的短板恰恰在于决策规划缺乏足够的智能。

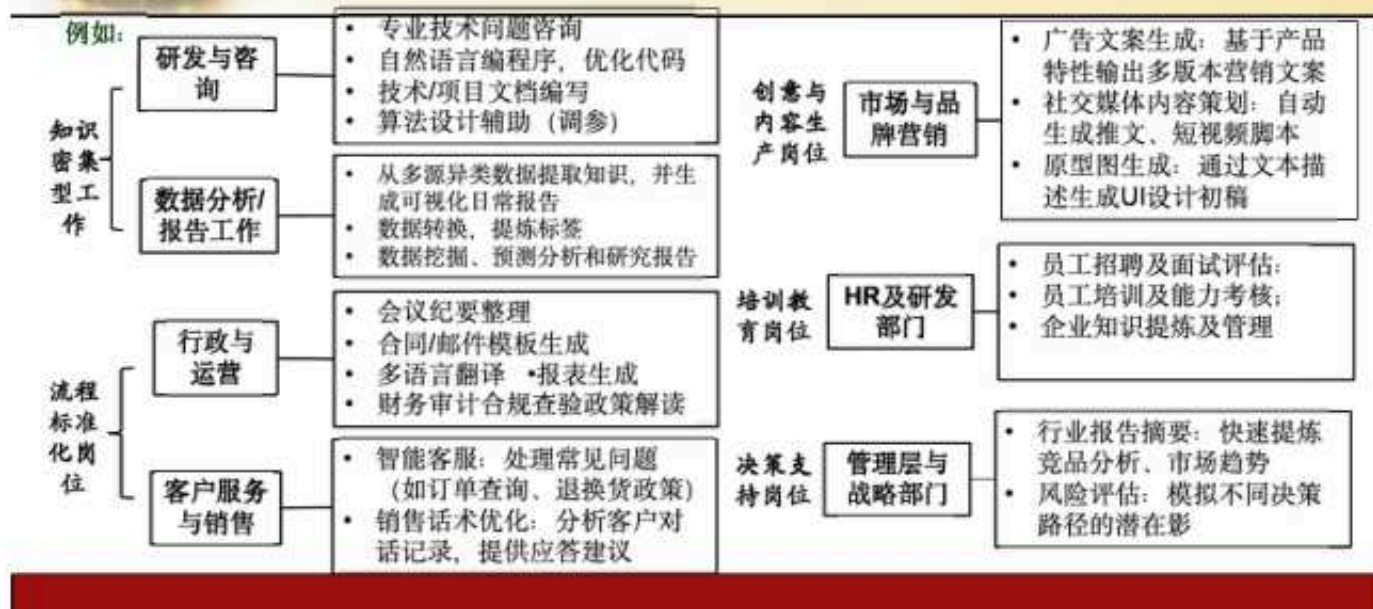
57

三大前沿技术能力：数字内容孪生、数字内容的智能编辑、数字内容的智能创作

	2020年前	2020	2022	2025E	2030E	2050E
文本生成	垃圾邮件识别 翻译 基础问答	基础文案写作 草稿撰写	长文本写作 草稿撰写与修改	专业文本写作 (如科研、金融、医疗)	终稿写作 写作能力超越人类平均水平	终稿写作 写作能力超越人类专业人士
代码生成	单行代码生成	多行代码生成	长代码写作 准确率提升	支持更多代码语言 支持更多垂直行业	输入文本即可自动生成产品原型	输入文本即可自动生成最终产品
图片生成			艺术生成 Logo生成 照片编辑与合成	产品原型设计 建筑原型设计	产品最终设计定稿 建筑最终设计定稿	设计能力超越艺术家、专业设计师、专业摄影师
视频3D生成				视频初稿	视频初稿与修改	AI机器人个性化的电影
游戏AI						个性化的游戏

多模态指的是多种模态的信息，包括：文本、图像、视频、音频等

人工智能应用：60%以上的知识工作场景



总之，AI适合于任务：

1. 数据操控文稿撰写：如数据录入、各类文稿报告撰写。
2. 知识依赖性强：需快速调用行业知识（法律、医疗）。
3. 协作需求多：跨部门沟通的文档对齐、会议记录。
4. 容错率较高：非实时关键任务（如内部培训材料生成）。



AGI的优势与特点



人类情绪意图精准识别



通过人类情绪意图的精准识别与反馈，可以规避常规机器学习过程中靠人工进行反馈的主动性、片面性、低效性

因果逻辑类人脑推理

模仿人脑因果推理过程，大幅提升机器计算效率，从底层机制解决人工智能对环境的适应性

内容生成&调度&执行

基于情绪意图、因果推理的海量学习，生成符合要求的内容，并调度资源完成任务执行

大模型应用开发的底层逻辑

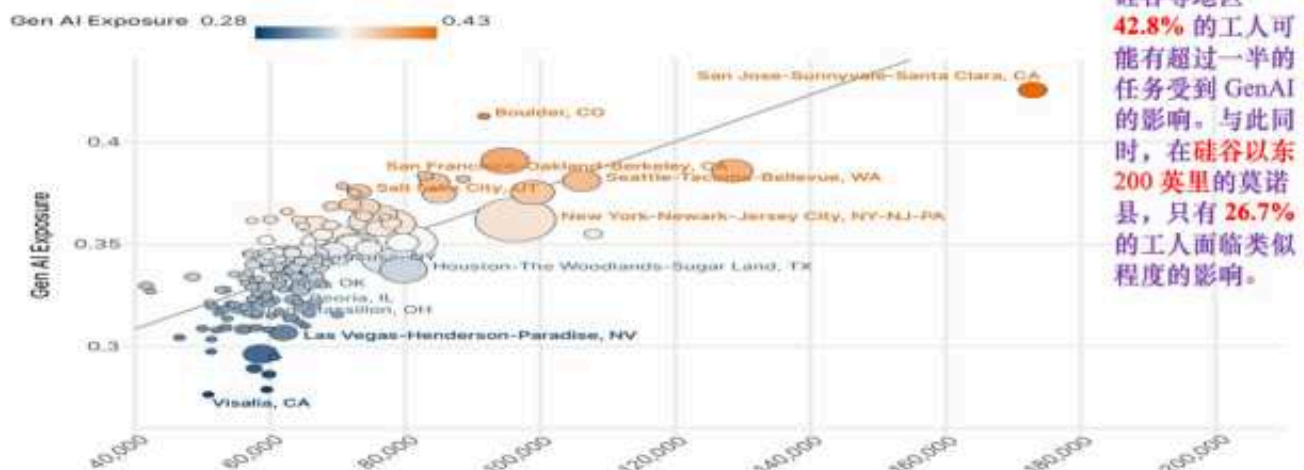


大模型带来认知助力革命



FIGURE 2 生成式人工智能专注于认知任务，而不是体力或体力任务

Generative AI exposure level and annual pay by metro



硅谷等地区
42.8% 的工人可能
有超过一半的任务
受到 GenAI 的影响。
与此同时，在硅谷
以东 200 英里的莫
诺县，只有 26.7%
的工人面临类似
程度的影响。

人工智能有望降低智能的成本——这是现代经济的核心投入。当代理可以完成一项任务时，它们的成本仅为人类的三十分之一。

complete by 50% or more.



老板眼中的AIGC



帮老板落地的同学 眼中的AIGC



实际的AIGC



AIGC的问题何在?

“大多数正确”和“貌似正确”

技术性问题

合规性问题

好不好?
对不对?

生产者
审核者

模型这么
如何做好
如何确保

AI 对于不熟悉的问题会强行给出一定的答案, 直到不准确的内容被明确戳破。

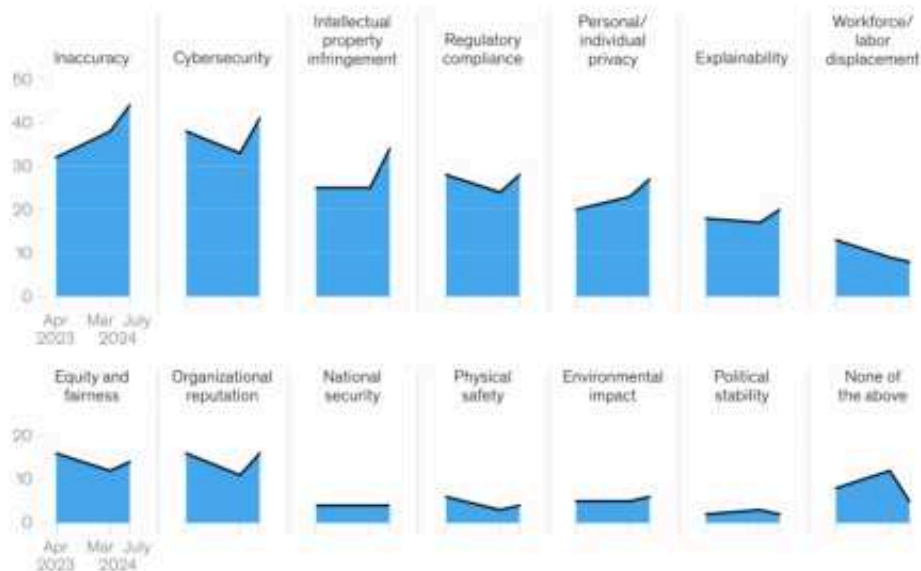
谁该为AI 造成的损失负责?
反垄断问题



技术性问题

合规性问题

Gen-AI-related risks that organizations are working to mitigate,¹% of respondents



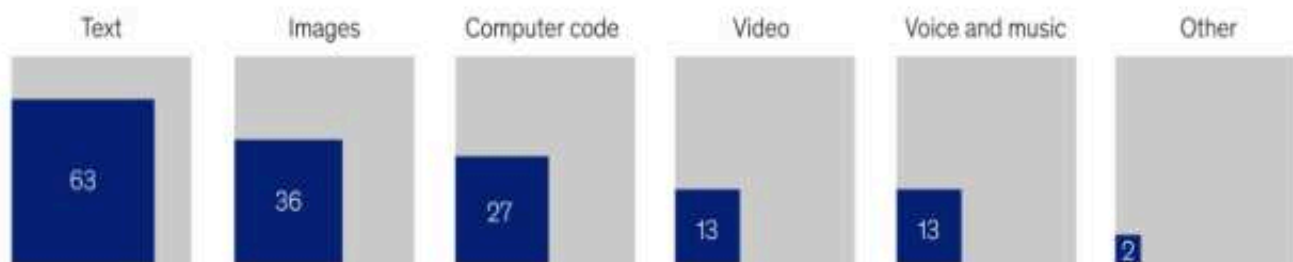
¹Only asked of respondents whose organizations use AI in at least 1 business function. Respondents who said "don't know/not applicable" are not shown.
Source: McKinsey Global Survey on the state of AI, 2023-24

McKinsey & Company

Screenshot

While text is the type of content that organizations are most commonly creating with gen AI, they are also experimenting with other modalities.

Types of content generated by gen AI at respondents' organizations,¹% of respondents



¹Only asked of respondents whose organizations regularly use gen AI in at least one function. Figures were calculated after removing the respondents who said "don't know."

Source: McKinsey Global Survey on the state of AI, 1,491 participants at all levels of the organization, July 16-31, 2024.

McKinsey & Company



MaaS (Model as a Services)

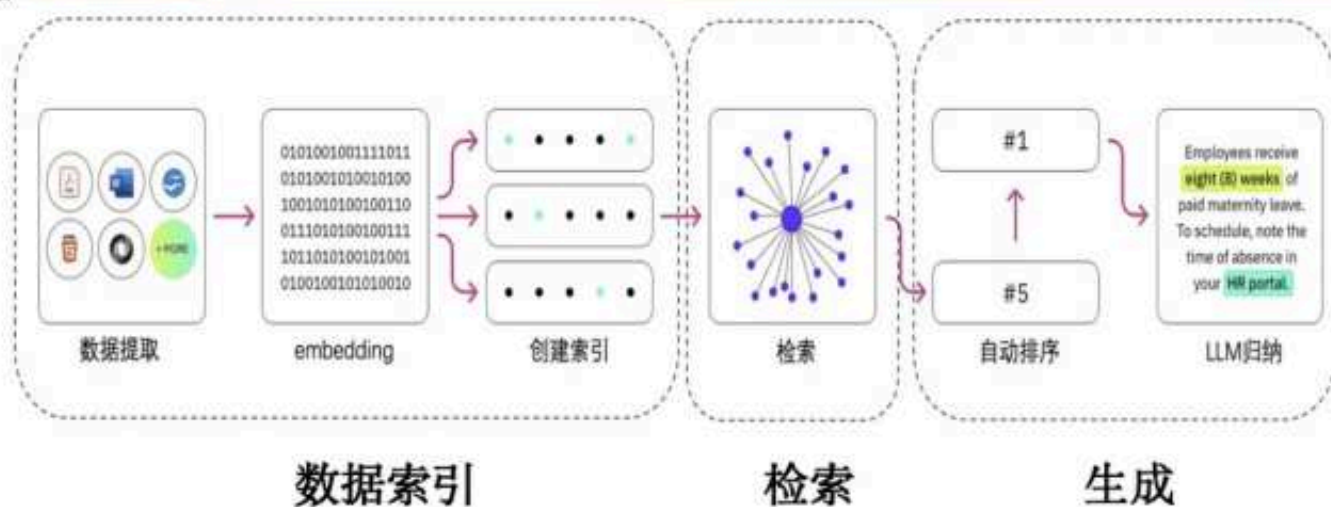
- 专业服务要求高，容错性低。模型要求：可控、可追溯、可修正。
- 场景化、定制化、个性化的小模型，实现工业部署应用。



金融、文旅、传媒、政务、教育



RAG (Retrieval Augment Generation) 模式





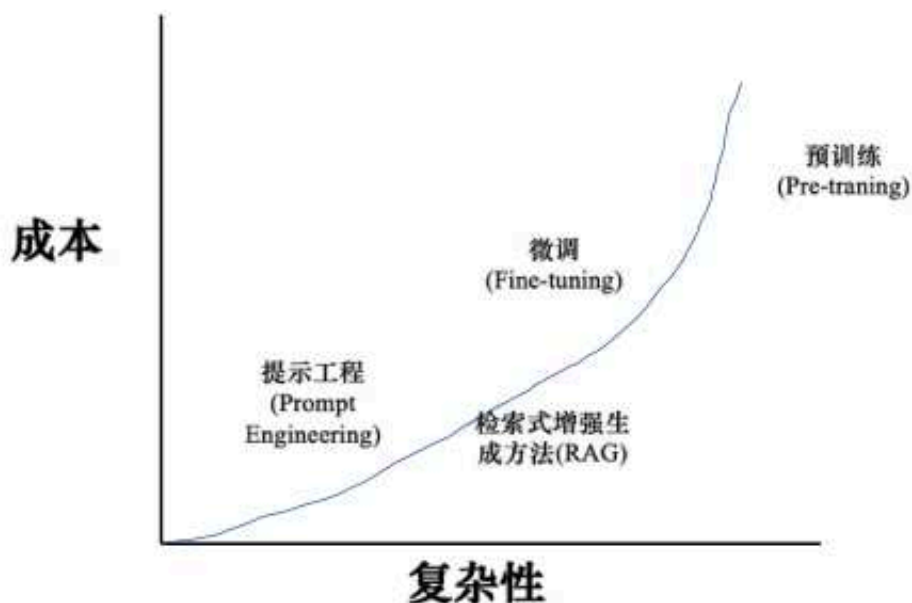
大模型的常用使用方式

提示语优化 Prompt	知识库嵌入 RAG	大模型微调 SFT	行业模型定制 Pre-train	智能体 Agent
- 基于通用模型，上手门槛较低 - 提升应答效果，规范应答格式 - 社区生态完善，交流使用心得	- 提取文档信息，数据简单易得 - 补充领域知识，回答专业问题 - 嵌入原有流程，快捷高效部署	- Few-shot，小样本数据训练 - 针对特定场景，定制任务助手 - 提高单次调用 token 数限制	- 投喂行业数据，定制行业模型 - 结合行业现状，深入理解业务 - 执行专业任务，实现降本增效	- 低代码易开发，多种插件集成 - 增强任务执行，优化决策路径 - 实现目标导向，解决实际难题

多级优化方案组合，满足不同资源与场景的优化需求

71

基于成本与复杂性适配/调整大模型



All rights reserved



73

BRTR 框架提示词

BRTR 提示词 = 说背景 (Backgroud) + 定角色 (Role)
+ 派任务 (Task) + 提要求 (Request)

【背景】：填入你的背景（补充说明，提高输出质量）

【角色】：填入你的角色（第三重要，提高输出质量）

【任务】：填入你的任务（第一重要：明确且具体）

【要求】：填入你的要求（第二重要：要求可以1、2、3排列），要求可以是内容要求和输出要求。

如何用AI Deep Research 功能建立自己的投资框架，而不是继续依赖二手信息，被动跟风？案例实战 每当资本市场有重大消息传出，普通投资者第一反应往往是等**投行研报、看金融分析师的解读**。然而，这些研报不仅滞后，很多时候还带有机立场，等到信息传到你手上，市场早已消化，你只能被动跟风。但现在，AI工具已经改变游戏规则。Grok 3、ChatGPT、Google Deep Research 这些新一代 AI 能在**几分钟内拆解新闻，分析逻辑、估值重估，甚至预测市场反应，让你第一时间掌握真正的信息优势。**



第三部分：知名投行如摩根、高盛或者知名分析师的研报和观点汇总以及原文链接

以下是知名投行和分析师对美股市场、利率和贸易政策的最新观点汇总，基于9月数据，聚焦就业疲软和降息预期。每项包括关键观点和原文链接。

1. 摩根士丹利 (Morgan Stanley): 看好9月科技股反弹，推荐买入估值合理的股票如博通和AMD，预计美联储降息将利好成长股；但警告贸易不确定性可能加剧波动。 原文链接: <https://www.cnbc.com/2025/08/30/morgan-stanley-names-stocks-with-more-upside-in-september.html> [cnbc.com](#)
2. 高盛 (Goldman Sachs): 全球股票在2025年易受贸易战影响，预测标普500年增长放缓至14.37%，建议分散投资到防御性板块；就业数据疲软强化9月降息，但通胀风险上升。 原文链接: <https://www.goldmansachs.com/insights/articles/global-stocks-are-vulnerable-in-2025> [goldmansachs.com](#)
3. 摩根士丹利 (Morgan Stanley): 美股风险包括2025年估值过高，建议关注利率敏感股；预测道指年涨7.05%，但9月需警惕就业放缓导致的回调。 原文链接: <https://www.morganstanley.com/insights/articles/us-stock-market-risks-2025-stocks-rally> [morganstanley.com](#)
4. 高盛 (Goldman Sachs): 9月市场脉动显示制造业收缩但服务业扩张，预测黄金价格接近3650美元阻力位；建议增持AI相关股票如 Nvidia。 原文链接: <https://am.gs.com/en-us/institutions/insights/article/market-pulse> [am.gs.com](#)
5. 摩根士丹利 (Morgan Stanley): 预测美联储9月降息25基点，利好债券和科技股；警告高估值股票如Alphabet面临反垄断风险。 原文链接: <https://www.reuters.com/business/wall-st-ramps-up-september-rate-cut-bets-after-powells-dovish-tone-2025-03-20/> [reuters.com](#)

债市
1. 年
1%
5.

重磅研究：AI将重塑金融分析师职业！

德州大学最新研究显示：
ChatGPT-4o在金融数据分析领域展现惊人实力

- 时间序列分析
- 风险收益评估
- ARMA-GARCH建模

关键发现：AI分析水平已与专业统计软件Stata相当！

未来趋势：

- 数据处理更高效
- 分析能力更强大
- 投资决策更精准

金融分析师请注意：不升级技能≈等待被淘汰
拥抱AI = 职业进阶 建议：及早适应AI工具，让科技成为你的超级助手，而不是职业终结者！

A First Look at Financial Data Analysis Using ChatGPT-4o*

Zifeng Feng

The University of Texas at El Paso
500 W. University Ave, El Paso, TX 79968, USA
Email: zfeng@utep.edu

Bingxin Li

West Virginia University
83 Beechurst Avenue, Morgantown, WV 26505, USA
Email: bingxin.li@mail.wvu.edu

Feng Liu

The University of Texas at El Paso
500 W. University Ave, El Paso, TX 79968, USA
Email: flin@utep.edu

Current Version: May 30, 2024

Abstract:

OpenAI's new flagship model, ChatGPT-4o, released on May 13, 2024, offers enhanced natural language understanding and more coherent responses. In this paper, we investigate ChatGPT-4o's capabilities in financial data analysis, including zero-shot prompting, time series analysis, risk and return analysis, and ARMA-GARCH estimation. We find that ChatGPT-4o's performance is generally comparable to traditional statistical software like Stata, though some errors and discrepancies arise due to differences in implementation. Despite these issues, our findings indicate that ChatGPT-4o has significant potential for real-world financial analysis. Integrating ChatGPT-4o into financial research and practice may lead to more efficient data processing, improved analytical capabilities, and better-informed investment decisions.

Keywords: ChatGPT, large language models, AI, finance research, academia, data analysis, stock return.

JEL Code: G00, C89, O33

关于美伊战争，看两家人类预测交易平台Kalshi和Polymarket，一直几个月都没有提升概率，人群是看“新闻头条”被操控。但对比AI驱动预测平台Vico，根据各种信号，在这次地缘政治事件中展现了更强的先见之明...**Vico 过去10天稳步攀升，从约65%一路涨到78.5%**在大多数人类交易者还没反应之前，就已经把升级风险计入价格。而人类交易的市场在外交斡旋阶段的头条报道中几乎一动不动，因为大家还在观望谈判结果。这印证了当前AI预测市场Vico等在某些高不确定性事件上具备的信息优势。

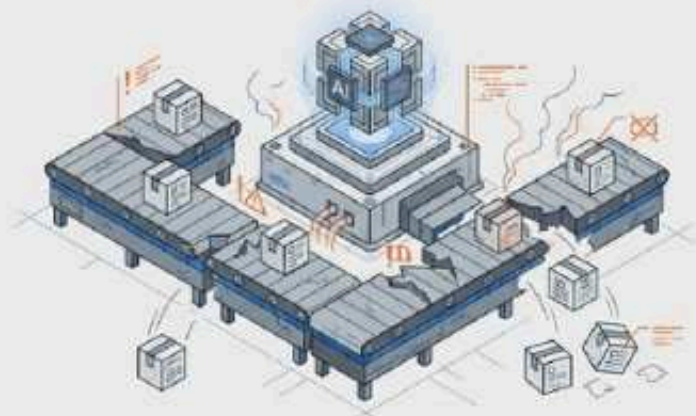


预期 (Expectation)



顶级大模型 + 完美的提示词 = 100% 成功。

现实 (Reality)



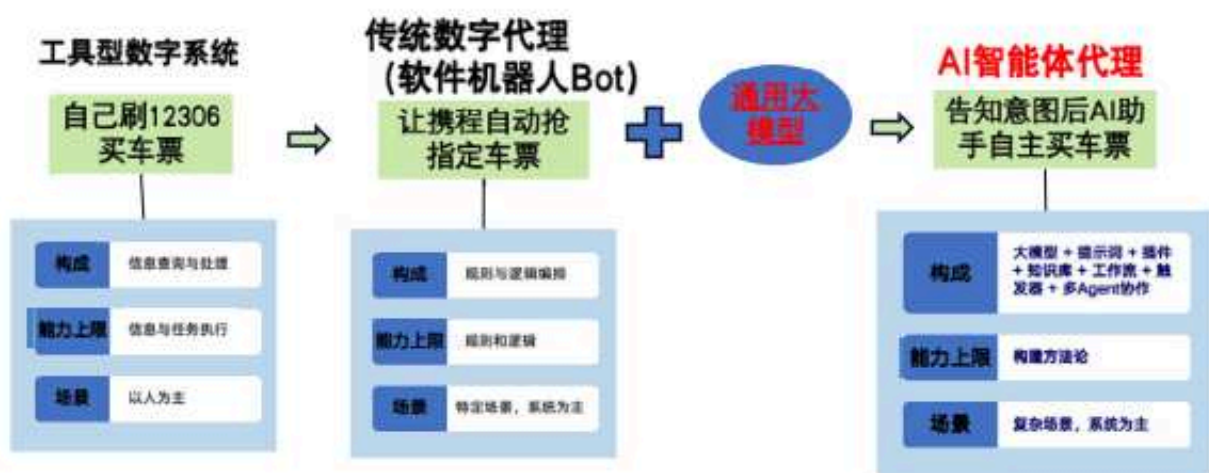
任务跑偏、偶尔极度聪明、偶尔莫名其妙。真实场景成功率 < 70%。

为什么同样的模型，别人能稳定运行，你却总失败？

问题不在“大脑”（模型），而在“系统”（环境）。优化任务拆解、状态管理和错误恢复后，同样的模型与提示词，成功率可提升至 95% 以上。

AI NotebookLM

如何理解AI智能体？



All rights reserved.

君子知道



目前Open AI已迭代了十几代模型

OpenAI 发展阶段

2018/06	2019/02	2020/05	2022/03	2022/12	2023/03	2023/11	2024/2	2024/5	2024/10
GPT-1	GPT-2	GPT-3	InstructGP	ChatGPT	GPT-4	GPT-4 Turbo	Sora	GPT4o	GPTo1
首次提出生成式预训练 模型参数数量为1.17亿	使用无监督训练技术 模型参数数量为15亿	少样本及零样本学习 模型参数数量为1750亿	基于人类反馈指令微调SFT+RLHF 模型参数数量为13亿	建模对话历史能力显著提升	接受图像和文本输入，是目前最先进的多模态大模型。	更强，更便宜，更普及	开启，更模拟世界	交互，无缝辅导	逻辑推理与数理分析

- ◆创作写作能力的飞跃。GPT-4可以与用户一起生成、编辑和迭代创意和技术写作任务，包括创作歌曲、编写剧本等任务，此外还可以学习用户的写作风格进行同风格创作，且创作时间远低于人类所需时间。
- ◆GPT-4处理文本量达到25000字左右，约等于ChatGPT的八倍。除了在文本量方面的进展外，GPT-4对信息的处理联系了上下文，是具有逻辑的分析，即GPT-4可以直接处理输入链接中的信息。
- ◆自定义身份属性的迭代。用户可以在输入框中增加对于生成文本的角色要求，GPT-4也可以做到深刻理解该命令需求，比如用户命令GPT-4以某位名人的语气活语言习惯进行对话。

大模型与智能体的关系：大脑与四肢

- 大模型：大模型就像人类的大脑一样拥有思考的能力，可以接受输入，可以分析&推理,可以输出文字\代码\媒体。然而无法像人类一样灵活规划调整、运用各种工具与物理世界互动。
- 智能体 (Agent)：智能体是具备规划思考能力、记忆能力、使用工具函数的能力，从而能动态地决定自己的执行流程和工具使用方式，自主掌控任务的完成过程。



2025年是AI智能体元年



2025年是AI智能体元年

公司	大模型	智能体	智能体类型
月之暗面	Moonshot	Kimi	对话式服务
科大讯飞	星火	讯飞发布	
百川	Baichuan	百小应	
百度	文心	文心智能体	工作流编排
昆仑万维	天工	天工SkyAgents	
智谱	GLM	智谱清言	
实在智能	TARS	实在Agent	自主智能体
联元科技	欧姆	欧姆智能体	
斑马智行	车智星	BetterYeah	
华为	盘古	Pangu-Agent	多智能体协同
百度智能	CPM	ChatDev	
腾讯	混元	腾讯元器	
源码科技	未读露	AskXBOT	工作流编排 对话式服务
汇智智能	Carrot AI	Gnomie	
阿里巴巴	通义千问	阿里云 ModelScopeGPT	对话式服务
		钉钉AI助理	工作流编排
字节跳动	豆包	Coze	工作流编排
		飞书智能伙伴	工作流编排

• AI智能体 (AI agent) 是通过模拟人类智能过程而创建的软件应用系统、或智能硬件或机器人 (Bot)。它们能够接收输入信息，进行处理，并作出决策或预测，旨在**执行特定任务，提供自动化解决方案**，并改善人类生活质量

- 是一种感知环境、进行决策和执行动作的智能实体。
- 单一智能体和多智能体已经在机器人、交通控制、智能硬件、智能游戏、生产排程等领域得到应用。如：
 - 医疗健康：辅助诊断、个性化治疗建议、患者监护等。
 - 金融服务：风险管理、欺诈检测、自动化交易等。
 - 教育：个性化学习计划、智能辅导、学习效果评估等。
 - 交通：自动驾驶汽车、交通流量管理、智能物流等。
 - 娱乐：游戏中的NPC（非玩家控制角色）、虚拟助手、个性化内容推荐等。

大模型给智能体以灵魂！



从大模型到智能体:AI应用交互的进化的必然之路

LLM: LUI交互（自然语言为核心交互方式）

- 通过语言用户界面，依赖用户给出的清晰明确的指令来完成任务
- 通过对话式给出输出，但是无法直接完成用户的目标，即只具备“你问我答能力”，无法实现“你说我做”



Agent: 具备自主能力的新一代AI应用

- 具备推理和规划能力，无需用户给出非常明确的指令
- 并非辅助用户完成特定任务，而是基于用户提出的目标，自动理解目标并完成用户的任务

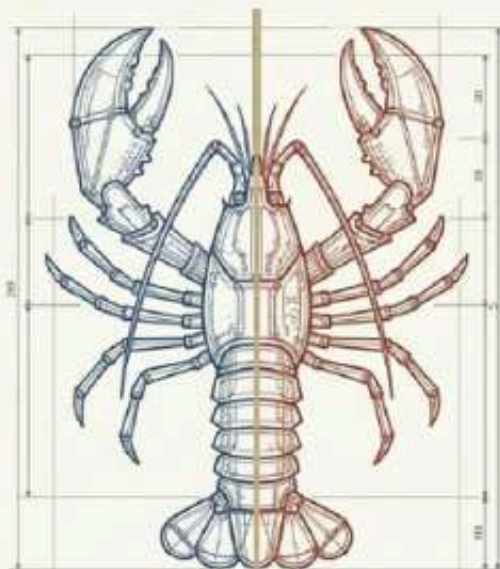
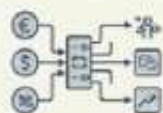
“人类演化的终点不是直立行走，而是拥有数字螯肢的 OpenClaw。”

正向延展：脱壳与成长

自主性：全天候常态哨兵，主动嗅探市场异常，替代人工巡检。



工具化：自主调用API、支付与社交插件，大幅压缩中间业务层。



警示威胁：捕食与反噬

生物性：具备极强的环境适应与自我演化力。



系统性风险：失控的AI代理是最难审计的风控黑洞；黑产同步进化，攻防进入无休止的军备竞赛。



龙虾大战：OpenClaw 生态全景



12家大厂入局 · 8款已发布

龙虾大战：8只龙虾的透明对比矩阵

AutoClaw (智谱)

106 Skills
飞书集成
积分制



● 专用模型

WorkBuddy (腾讯)

微信/QQ
社交生态
积分制



● 社交护城河

ArkClaw (字节)

24/7 云端
飞书原生
订阅制



● 云端运行

QoderWork (阿里)

最早上线
零配置
免费 tier



● 开箱即用

DuClaw (百度)

搜索/百科
知识库
17.8元/月



● 信息调研

MaxClaw (MiniMax)

超长上下文
便宜推理
免费 tier



● 长文本处理

Kimi Claw (月之暗面)

视觉 Agent
1T 参数
免费 tier



● 浏览器视觉

miclaw (小米)

手机系统级
HyperOS
完全免费



● 移动端

	AutoClaw (智谱)	WorkBuddy (腾讯)	ArkClaw (字节)	Kimi Claw (月之暗面)	miclaw (小米)
部署方式	专用模型	微信/QQ	超长云端	视觉 Agent	手机系统级
定价方式	17.8元/月	社交生态	免费 tier	1T 参数	HyperOS
核心亮点	Pony-Alpha-2 模型 → Pony-Alpha-2 chip	微信A/社交生态 社交生态	便宜推理 大本 context chip	浏览器视觉 K2.5 vision chip	MiMo chip 手机系统级完全免费

选购指南

学生 → AutoClaw
团队/飞书 → ArkClaw
微信办公 → WorkBuddy
超长文档 → MaxClaw
小米手机 → miclaw



NVIDIA



黃仁勳談話重點

談「龍蝦」

絕對會是「下一個ChatGPT」，
將讓人們與AI的互動邁進一大步，
而且會帶動個人的專業能力大
幅提升



談輝達光通訊布局

正加速與台灣供應鏈合作

談H200策略

已重啟並加快生產H200

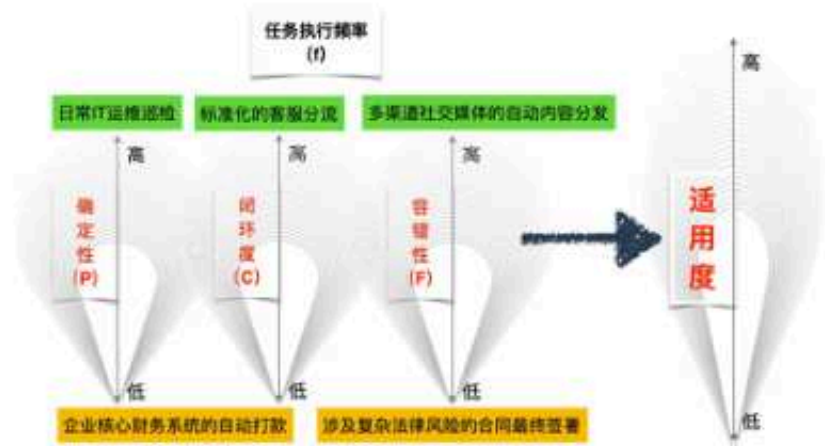
談地緣政治

談半導體美國製造

	阶段一： 对话生成型智能体	阶段二： 工具增强型智能体	阶段三： 全天候自主执行型智能体
核心机制	文本输入→概率预测→文本输出	在对话基础上调用搜索引擎、数据库、日历、邮件等外部 API，实现信息获取与指令执行	打通操作系统级控制权与即时通讯接口，同时支持 API 调用
自主性层级	无自主性	单点任务自主	全流程自主
环境闭环	无闭环。输出止于文本，与外部系统完全隔离，业务流程存在明显断点	单点闭环。通过 API 连接特定平台， workflow 仍呈碎片化，无法实现跨应用的持续监控与长程规划	全数字闭环。可跨平台、跨应用操作，覆盖本地文件处理、系统命令运行等全链路，实现“异步数字打工”
典型代表	早期 ChatGPT、DeepSeek	GPT-4 插件体系、Coze、MCP 工作流产品、RPA+AI 融合方案	OpenClaw、QClaw、AutoClaw



$$S = F \times P \times C \times f$$



规律一：高确定性 × 全数字闭环 = 极度适用

规律二：高风险敏感 × 低容错空间 = 极度不适用

规律三：非标准化交互 × 情感博弈 = 适用性受限

龙虾缺的从来不是菜谱，是“买菜”的供应链

SkillBoss: 为AI Agent打造的无界基础设施



不同阶段的AI 智能体应用

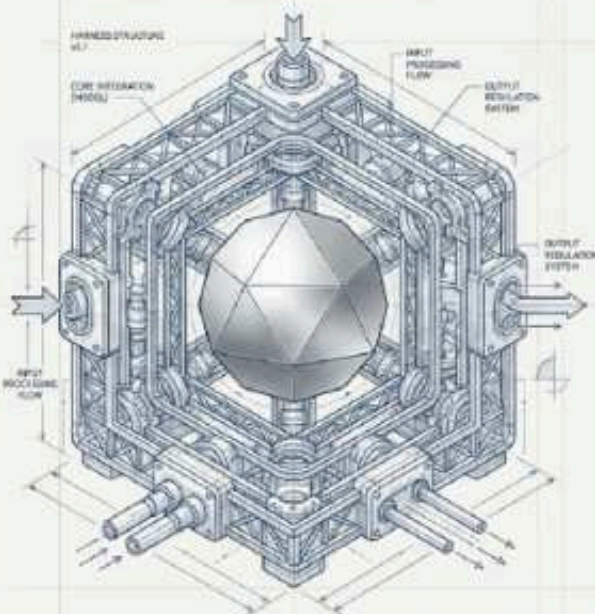


教授视角 | 企业级 AI 部署指南

驾驭智能：Harness Engineering 与 AI Agent 的工业化落地

顶级商学院 AI 战略前沿洞察

从“实验室的聪明大脑”到“生产线的稳定流水线”



AI NotebookLM

AI 工程的三个核心纪元



企业 AI 工程的核心，正从“让模型更会想”向“让模型跑得稳、不跑偏”转移。

AI NotebookLM

范式转移：三大工程化阶段对比

	解决的核心问题	底层逻辑	局限性
Prompt Engineering	意图表达 (Intent articulation)	塑造局部的概率空间。	无法凭空弥补缺失的知识，无法处理长链路任务。
Context Engineering	信息供给与动态检索	在合适的时机把正确的信息送进去（如 RAG）。	即使信息给对，模型也可能在连续执行中跑偏、调用错工具。
Harness Engineering	全生命周期与状态管理	监督、约束、纠偏与最终验收。	构建成本高，需重塑基础设施。

© NotebookLM

MBA 视角：从“带新人”看 AI 系统的演进

业务场景：派一名新员工去完成一次极其重要的客户拜访

人力资源管理 (HR)



任务简报 (The Briefing) - 告诉他先寒暄、再介绍方案、最后确认需求。



客户档案 (The Dossier) - 给他客户背景、历史报价、竞品情况。



管理系统 (The Management System) - 带着 Checklist，关键节点实时汇报，发现偏差马上纠正，按标准验收。

AI 工程化 (Tech)



= Prompt Engineering
(把话说明白)



= Context Engineering
(把信息给对)



= Harness Engineering
(持续观测与纠偏机制)

© NotebookLM

AI 系统的工业级方程式

Agent = Model + Harness

(推导: $\text{Harness} = \text{Agent} - \text{Model}$)

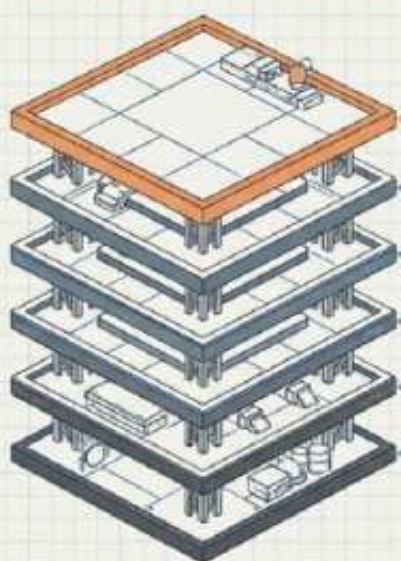
模型（大脑）：负责概率生成与文本预测的核心引擎。

Harness（外围设计）：包含除了模型本身之外，决定系统能否稳定交付的所有工程化外围设计。

Harness 并不是要取代 Prompt 或 Context，而是在更大的系统边界上将前两者包容进来。它是对整个运行系统的工程化。

© NotebookLM

Harness Engineering 的六层工业架构



Level 6: 约束与恢复 (Constraints & Recovery)
- 校验规则、回滚与重试机制。

Level 5: 评估与观测 (Evaluation & Observation)
- 独立 QA 与自动化测试。

Level 4: 记忆与状态 (Memory & State)
- 隔离短期状态与长期记忆。

Level 3: 执行编排 (Execution Orchestration)
- 建立标准工作流 (SOP) 轨道。

Level 2: 工具系统 (Tool Systems)
- 渐进式提供工具，防止滥用。

Level 1: 边界与信息 (Boundaries & Info)
- 定义角色，结构化剪裁上下文。

这些层次将实验室里的 AI 转化为企业级生产力。

© NotebookLM

架构深潜：输入控制与工具分配 (Layer 1 & 2)

第一层：边界与信息

Information Filter

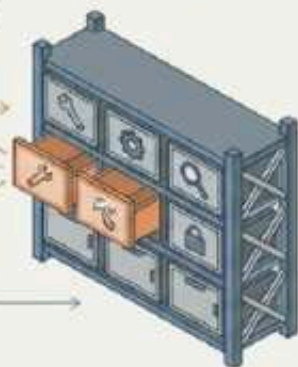


理念：上下文不是越多越好，而是越相关越好。

动作：结构化组织数据。区分固定规则、当前任务与外部证据，防止信息污染和约束遗忘。

第二层：工具系统

Smart Dispenser

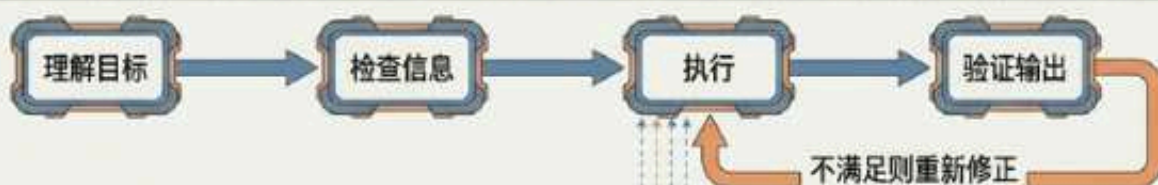


理念：防止工具滥用与注意力涣散。

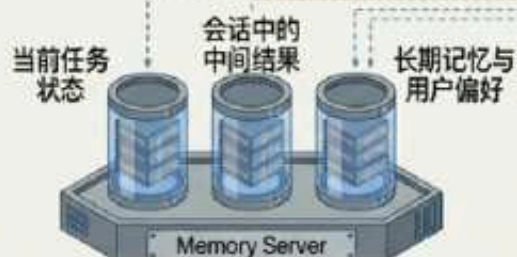
动作：渐进式 (Progressive Disclosure)。不一次性抛给模型所有工具参数；先给少量原型，触发特定能力时再动态加载详细 SOP 脚本。

© NotebookLM

架构深潜：SOP 轨道与记忆隔离 (Layer 3 & 4)



从‘想到哪做到哪’到‘标准化轨道’。极度接近人类的 SOP 流程，区别在于 AI 靠 Harness 环境约束。



Layer 4：记忆与状态

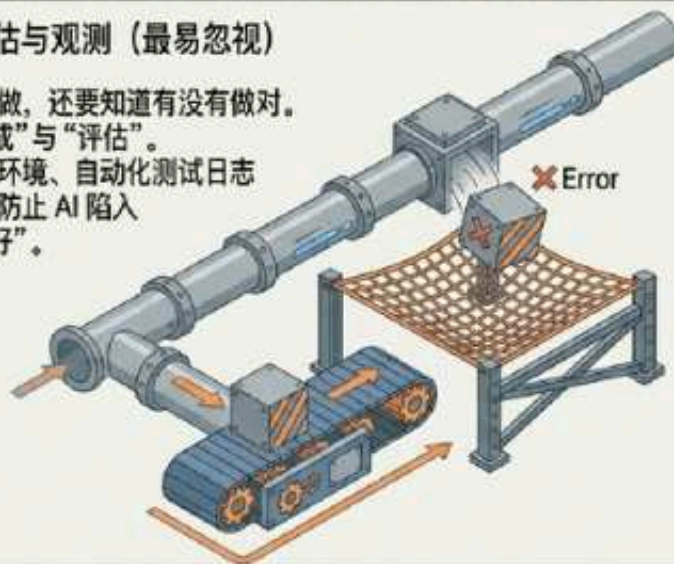
解决 Agent 的失忆症：必须严格隔离三类数据，混淆将导致系统越跑越乱。

© NotebookLM

架构深潜：系统安全网与自动纠偏 (Layer 5 & 6)

第五层：评估与观测（最易忽视）

系统不仅要会做，还要知道有没有做对。
必须剥离“生成”与“评估”。
引入独立验收环境、自动化测试日志
与错误归因，防止 AI 陷入
“自我感觉良好”。



第六层：约束与恢复（最后防线）

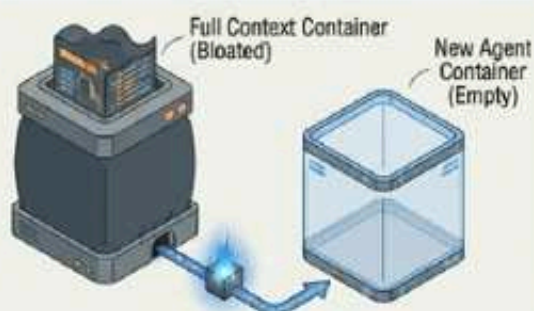
在真实世界里，失败
(API 超时、格式混乱)
是常态。若无恢复机制，
出错只能从头再来。
必须构建恢复机制，
出错只能从头再来。

必须构建：约束、前后置
校验，以及重试路径和状
态回滚。

© NotebookLM

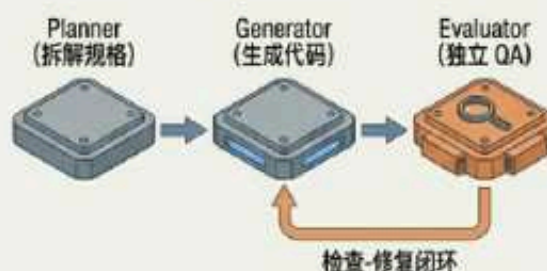
一线实战：Anthropic 的工业级解法

解法一：Context Reflection（上下文反射）



- 问题：随上下文填满，模型会丢细节并急于收尾。常规做法是压缩历史（有缺陷）。
- Harness 方案：重启进程。把工作交接给一个干净的全新 Agent，类似处理软件工程中的内存泄漏。

解法二：生产与验收的彻底分离

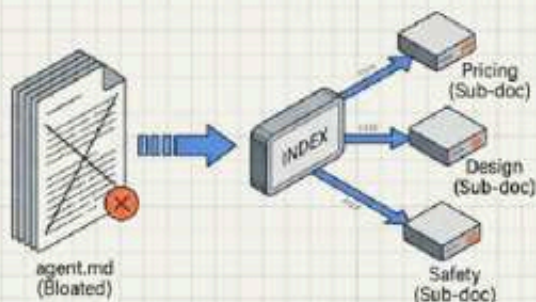


模型自己给自己打分往往盲目乐观。Harness 方案引入独立 Evaluator：不仅检查代码，而是真实调用浏览器操作页面，形成有效的生成-检查-修复闭环。

© NotebookLM

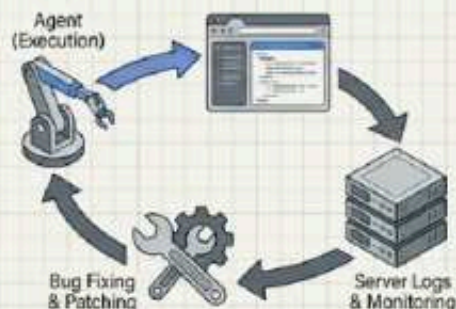
一线实战：OpenAI 的百万行代码流水线

从巨型规则本到动态索引



将冗长的 agent.md 变为纯粹的目录索引。Agent 仅读取核心索引，只在真正需要时，动态拉取具体的子说明文档，极大地保护了上下文窗口。

自动化验证环境



当生产速度爆炸时，瓶颈从‘编写’转移到‘测试’。Agent 被接入真实的浏览器、日志系统与隔离沙盒中。不仅输出文本，更是直接运行代码、看监控找 Bug 并自主修复。

© NotebookLM

重定义 AI 工程师

传统 Coder

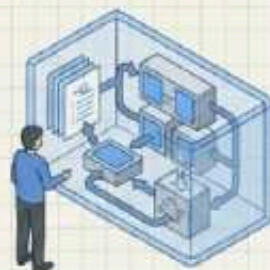
关注点：编写语法、业务逻辑与具体的执行指令。



Agent 工程师

关注点：设计环境、边界与反馈循环。

1. 拆解目标
(转化为 Agent 可读任务)
2. 设计环境
(制定 Harness 规则与约束)
3. 构建链路
(搭建自动化反馈闭环)



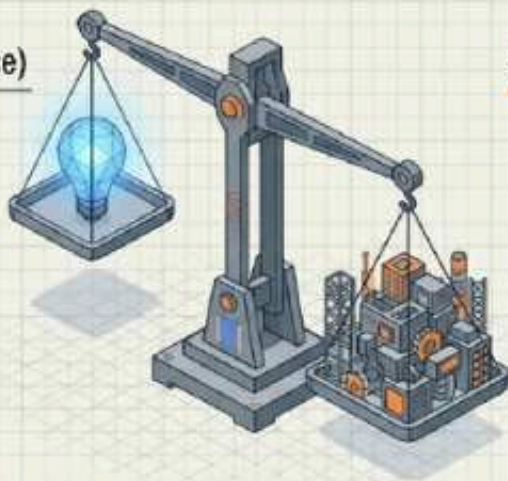
“当 Agent 任务失败时，解决办法绝不是去提示词里让它『更努力一点』，而是去系统里排查它『缺少了什么结构性的外围能力』。”

© NotebookLM

范式转移：从“实验室奇迹”到“工业化基石”

模型层面 - 聪明 (Intelligence)

算力与基础大模型正趋于同质化与开源，提供的是潜力。



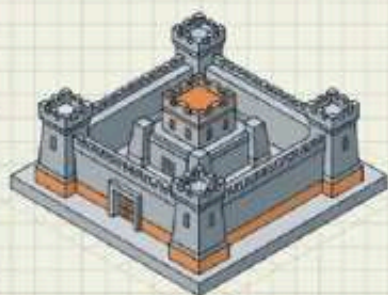
系统层面 - 稳定 (Reliability)

状态管理、错误恢复与执行编排构成了真正的企业护城河，Harness 保障的是交付。

真正决定应用能否上线的可能是模型（大脑），但真正决定能否落地、能否稳定产生商业价值的，是 Harness（系统）。

© NotebookLM

管理者的行动指南



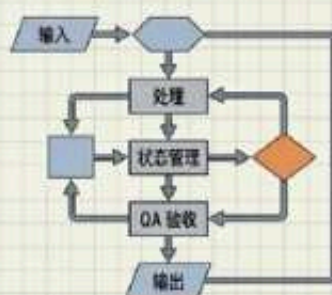
1. 护城河不在大模型

基础模型能力正趋于同质化。企业真正的核心壁垒在于你围绕自有业务流构建的专有 Harness 基础设施。



2. 投资工程基础设施

停止在无穷无尽的“提示词玄学”上耗费资源。将预算转向状态管理、自动化评估 (Eval)、日志追踪与错误恢复协议。

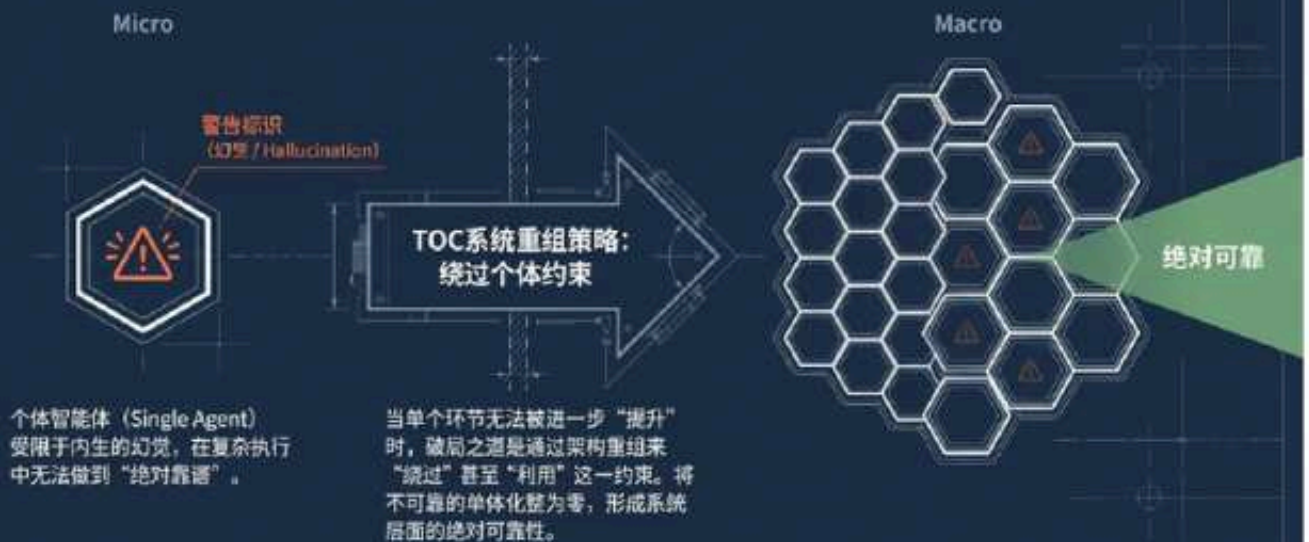


3. 用“系统思维”管理 Agent

不要把 AI 当作全知全能的神谕。引入分离职责机制，为 AI 建立严格的数字 SOP 工作流与独立的 QA 验收体系。

© NotebookLM

幻觉重塑系统边界：用系统架构的可靠性包裹个体的不可靠



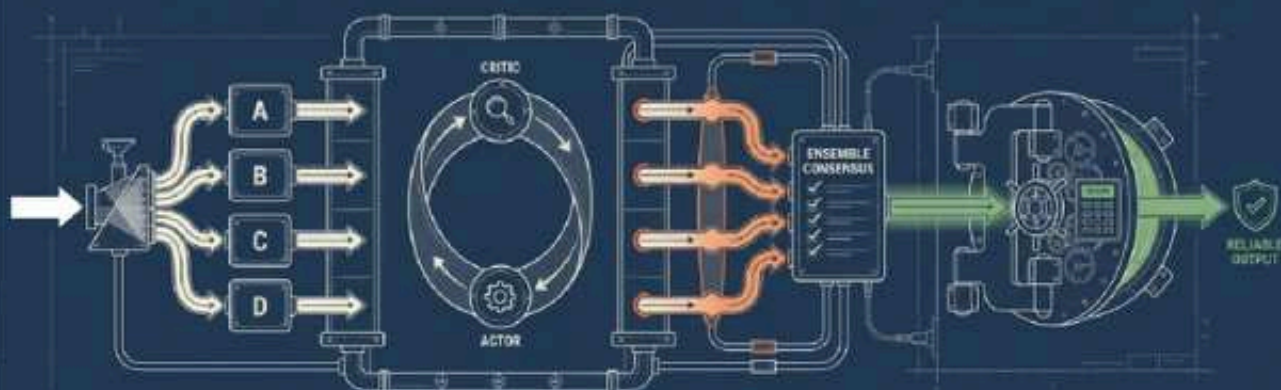
NotebookLM

LLM 理事会 多智能体协作——三个臭皮匠顶个诸葛亮！



这个仓库的想法是，不是向你喜欢的 LLM 提供商（例如 OpenAI GPT 5.1、Google Gemini 3.0 Pro、Anthropic Claude Sonnet 4.5、xAI Grok 4 等）提问，而是将它们组合成你的“LLM 理事会”。这个仓库是一个简单的本地网页应用，基本上看起来像 ChatGPT，只不过它使用 OpenRouter 将你的查询发送给多个 LLMs，然后让它们相互审查和排名，最后由主席 LLM 生成最终的回答。

绕过个体约束：多智能体协作流转的四大防线



关卡1：角色分工与专业化 (Role Specialization) —— 缩小任务边界，将大幻觉降解为特定领域的可定位错误。

关卡2：批评者-执行者双环路 (Critic-Actor Loop) —— 利用“检验比生成更容易”的非对称逻辑，建立内部质量把关闭环。

关卡3：独立验证与集成 (Ensemble Verification) —— 模拟临床多学科会诊，通过统计学的独立性消除个体的随机错误。

关卡4：可验证的检查点架构 (Checkpoint Architecture) —— 只有通过形式化验证的输出才能进入下游，彻底切断错误的累积效应。

© NotebookLM

全场最核心洞察（约束迁移理论）：阻力全面向“意图端”转移



© NotebookLM

深挖新瓶颈：什么是正在吞噬效能的“意图摩擦”？

AI 领域顶尖学者 Karpathy 提出的“Skill Issue”（技能问题）：人类脑海中看似清晰的宏大想法，一旦脱离了人与人之间“心照不宣”的隐性知识背景。在转化为智能体可直接执行的代码级逻辑时，会产生极其恐怖的能量损耗。



NotebookLM

企业应用 AI 范式的三次演进矩阵

	范式一	范式二	范式三
年份	2020-2023	2023-2025	2025-未来
应用形态	Copilot (副驾驶)	Pipeline (流水线)	Agentic Ecosystem (智能体生态)
核心约束	<u>信任缺口</u> 	<u>流程的可控性</u> 	<u>人类的意图表达能力</u> 
人类角色	最终决策者，测试能力边界	监督者，制定与执行评估规范 (Evals)	意图注入者，系统架构设计师

NotebookLM

范式一与范式二的阶段性局限：受困于信任与控制

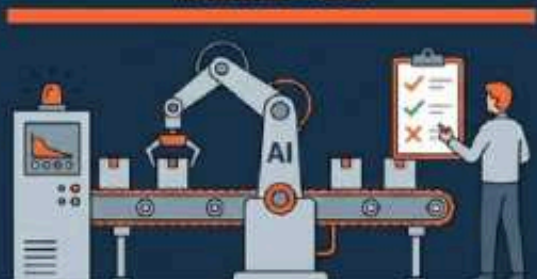
信任天花板



Copilot 模式的瓶颈

- 高度依赖人类的即时干预，AI只是“高级挂件”；由于无法建立完全的信任闭环，生产力提升呈现明显的边际递减。

流程刚性天花板



Pipeline 模式的瓶颈

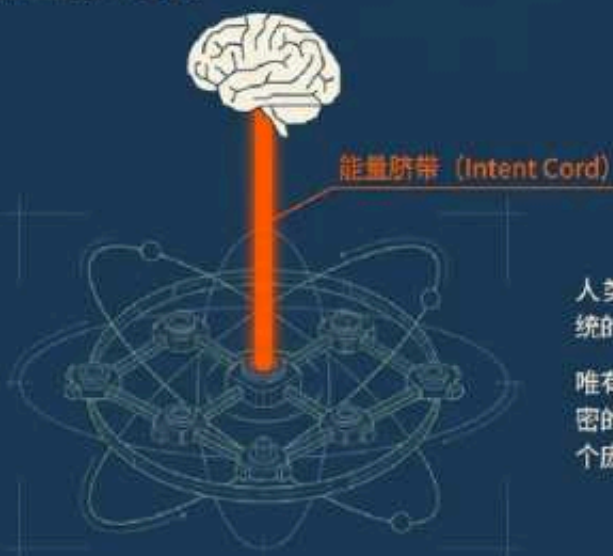
- 业务流程虽然被“AI 优先”重构，但系统极度依赖人类在最后环节进行监督与评估规范（Evals）；流程是硬编码的，缺乏应对复杂未知环境的自适应能力。

NotebookLM

范式三的终极形态： 步入意图驱动的“智能体生态”时代

系统被底层重构，天然支持多个 Agent 进行自主分工、协作与运营。

人类彻底退出“执行层”与“微观监督层”。



人类的新使命：化身为系统的“意图引擎”。

唯有通过极度清晰、逻辑严密的意图表达，才能驱动这个庞大生态的运转。

NotebookLM



English | 简体中文 | 繁体中文 | 日本語 | 한국어 | 中文

Deutsch | Español | Français | 日本語 | 한국어 | Português | Pycckий | 中文

Deutsch | Español | Français | 日本語 | 한국어 | Português | Pycckий | 中文

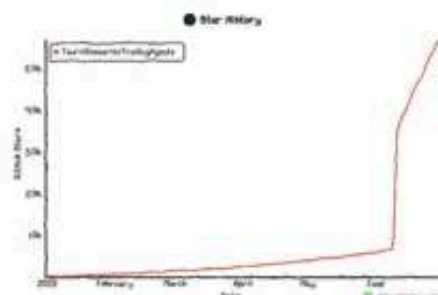
TradingAgents: Multi-Agents LLM Financial Trading Framework

TradingAgents: 多智能体 LLM 金融交易框架

TradingAgents officially released. We have received numerous inquiries about the work, and we would like to express our thanks for the enthusiasm in our community.

TradingAgents 正式发布！我们收到了大量关于该项目的咨询，并感谢您的热情支持。

We've decided to fully open-source the framework, looking forward to building impactful projects with you! 因此我们决定将框架完全开源，期待与您共同构建具有影响力的项目！



TradingAgents | 开源项目 | 金融交易 | 多智能体 | LLM | 框架 | 社区 | 支持

TradingAgents Framework: Trading

TradingAgents is a multi-agent trading framework that specializes in LLM-based agents for fundamental and technical analysis, portfolio management, and risk management. The framework is designed to be modular and extensible, allowing users to integrate their own models and data sources.



TradingAgents Framework is designed for research and development, allowing the user to integrate their own models and data sources. The framework is designed to be modular and extensible, allowing users to integrate their own models and data sources.

Our framework decomposes complex trading tasks into smaller, manageable tasks, allowing for a more systematic and efficient trading process.

Analyst Team: 分析师团队

- Fundamental Analyst: Analyzes company financials, including the chosen business, management, and other relevant information. (基本面分析师: 分析公司财务报表、业务、管理层及其他相关信息。)
- Technical Analyst: Analyzes market trends and price movements to identify potential trading opportunities. (技术分析师: 分析市场趋势和价格变动以识别潜在的 trading 机会。)
- News Analyst: Monitors global news and economic conditions to assess their impact on the market. (新闻分析师: 监控全球新闻和经济状况以评估其对市场的影响。)
- Technical Analyst: Utilizes technical indicators and patterns to identify trading opportunities. (技术分析师: 利用技术指标和模式来识别交易机会。)

TradingAgents Framework is designed for research and development, allowing the user to integrate their own models and data sources. The framework is designed to be modular and extensible, allowing users to integrate their own models and data sources.

Researcher Team: 研究员团队

- Compares both bullish and bearish research reports to identify potential trading opportunities. (通过对比看涨和看跌研究报告来识别潜在的 trading 机会。)
- Analyzes the research reports to identify potential trading opportunities. (分析研究报告以识别潜在的 trading 机会。)



Trader Agent: 交易代理

- Compares reports from the analyst and researcher teams to make informed trading decisions. (比较分析师和研究员团队报告以做出明智的交易决策。)
- Executes trading orders based on the analysis and research. (根据分析和研究执行交易订单。)



Risk Management and Portfolio Manager

- Continuously monitors portfolio risk by assessing market volatility, liquidity, and other risk factors. (持续监控投资组合风险，通过评估市场波动性、流动性及其他风险因素。)
- Adjusts trading strategies based on the risk assessment. (根据风险评估调整交易策略。)
- The Portfolio Manager approves or rejects the transaction proposal. (投资组合经理批准或拒绝交易提案。)



REACTIVE Apache-2.0 license



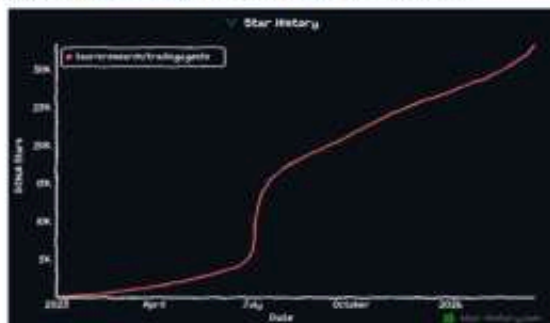
English | 简体中文 | 繁体中文 | 日本語 | 한국어 | 中文

Deutsch | Español | Français | 日本語 | 한국어 | Português | Pycckий | 中文

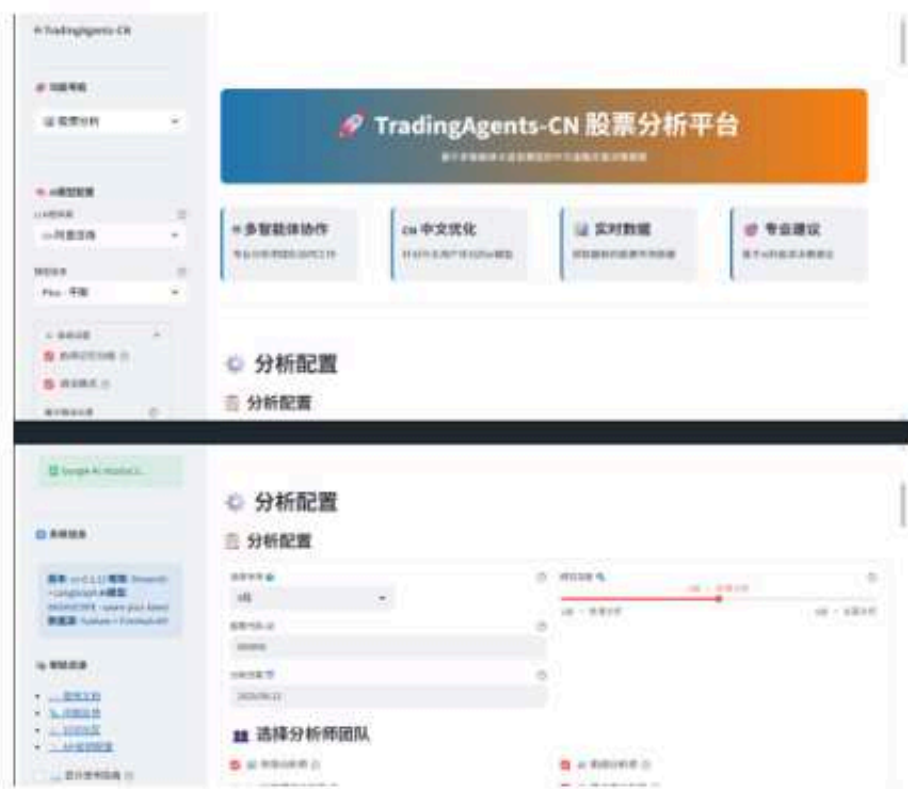
TradingAgents: Multi-Agents LLM Financial Trading Framework

News

- [2023-02] TradingAgents v0.2.1 released with GPT-5.4, Gemini 3.1, Claude 4.6 model coverage and improved system stability.
- [2023-02] TradingAgents v0.2.0 released with multi-provider LLM support (GPT-5.x, Gemini 3.x, Claude 4.x, Gink 4.x) and improved system architecture.
- [2023-01] Trading-R1 Technical Report released, with Terminal expected to land soon.



年化收益率 30.5%。GitHub 标星已破三万。里面装的是一个完整的虚拟华尔街机构。四个 AI 分析师疯狂扫读财报和全网情绪。看多和看空的研究员在系统里互相吵架。交易员负责综合意见。风控团队随时准备一票否决。最新版本直接接入了 GPT-5.4 和 Claude 4.6。它在最大回撤和夏普比率上打爆了所有传统基准策略。华尔街收你 2% 的管理费加 20% 的利润分成。他们雇一帮常青藤毕业生做的也就是这套流程。现在这套体系一分钱不要。金融圈的暴利壁垒被几行开源代码直接砸穿。



企业需要AI组织领导力

- 人工智能的广泛应用正在改变各个行业的竞争格局，企业需要适应和应对这些变化
 - 企业可以利用AI算法和数据资源的协同效应来跨越行业界限，获得竞争优势
 - 数据是企业跨界的关键资源
- 人工智能正在改变竞争的规则，企业需要重新思考自己的竞争策略
 - 传统企业需要重新构思战略，将重点放在连接企业、聚合数据以及通过分析和人工智能获取应用价值的能力上
- 企业已经由数据驱动运营的组织，转型为人工智能为“燃料”运营的组织
- AI转型知识成为企业管理者必备的知识储备

『工具多，体系少』

『部门各自上AI，彼此不认识』



『数据多，质量少』



『投资多，ROI少』

『试点热闹，
规模化生产率
不足15%』



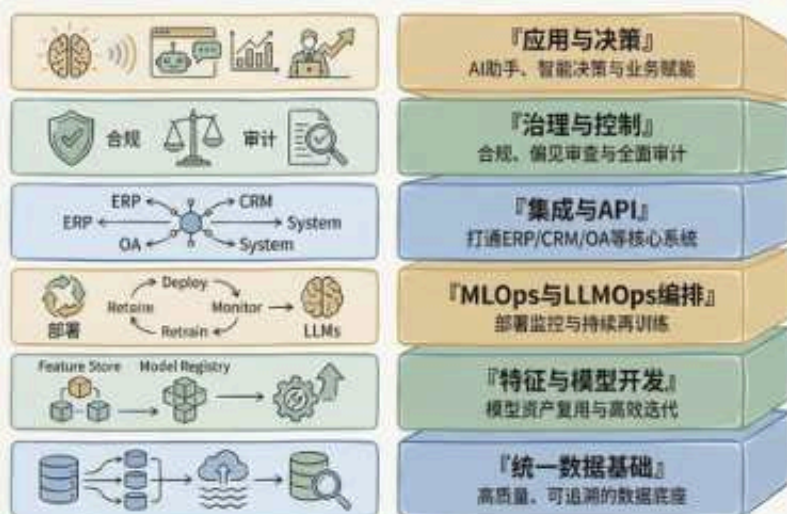
『问题不在模型，在结构』

『没有统一底座，
价值持续流失』

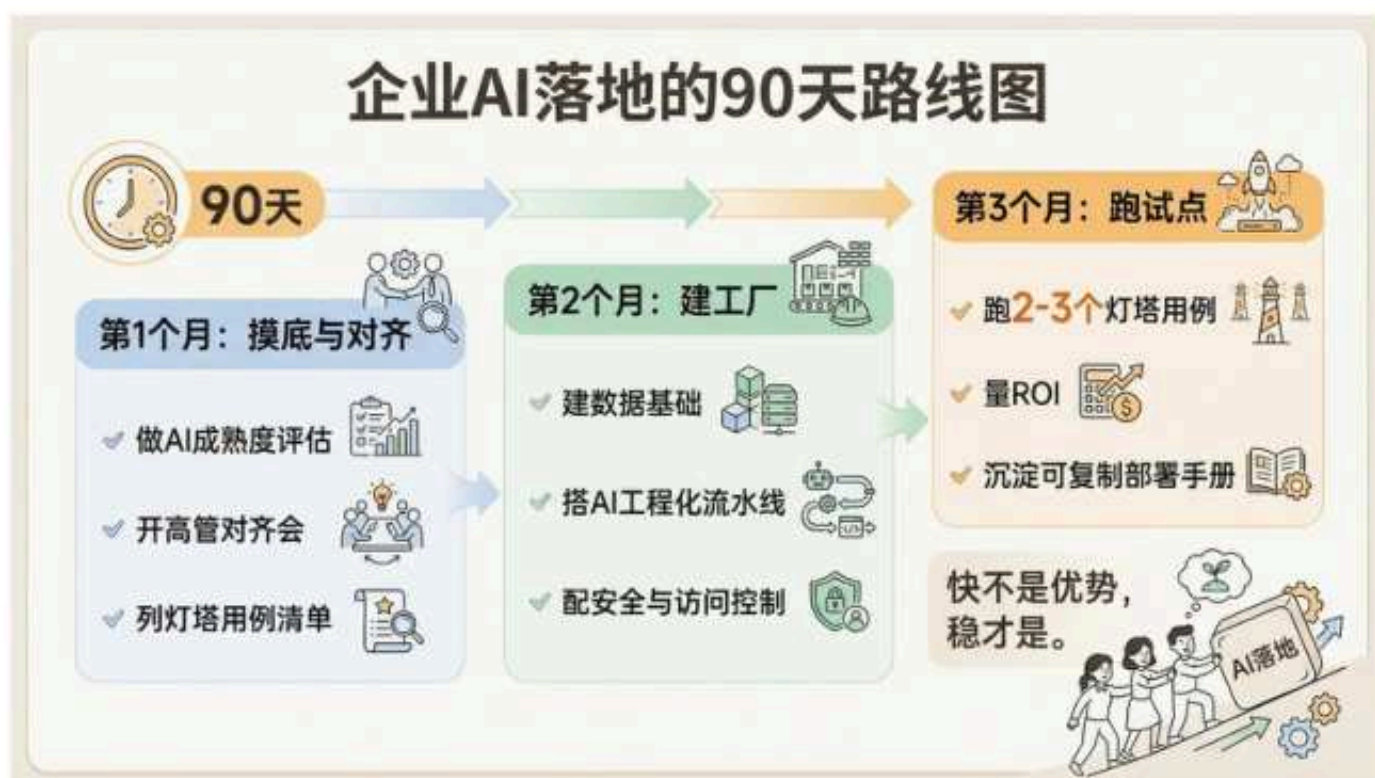


『没有架构，AI就只是昂贵的试验品』

—— 企业AI六层架构 (Enterprise AI Six-Layer Architecture) ——



CIO在构建企业AI核心基建



祝大家工作顺利、 生活幸福！



祝同学们身体健康，家庭幸福，事业有成！

All rights reserved.

君子知道
WUZHILIAN