

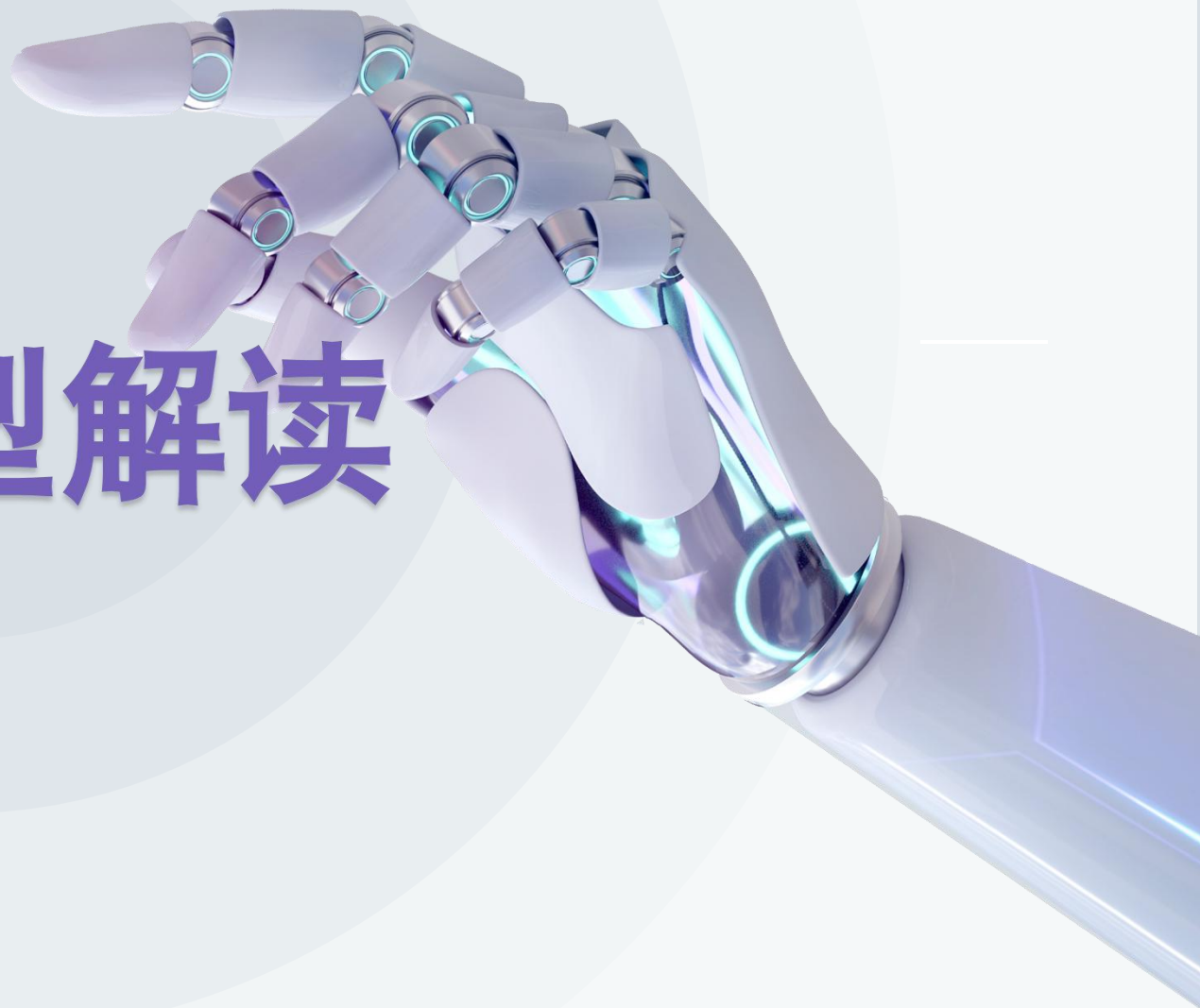
奇点临近，行业变革

DeepSeek模型解读



You will enter the future

2025年2月3日（大年初六），领学/浙江大学MBA



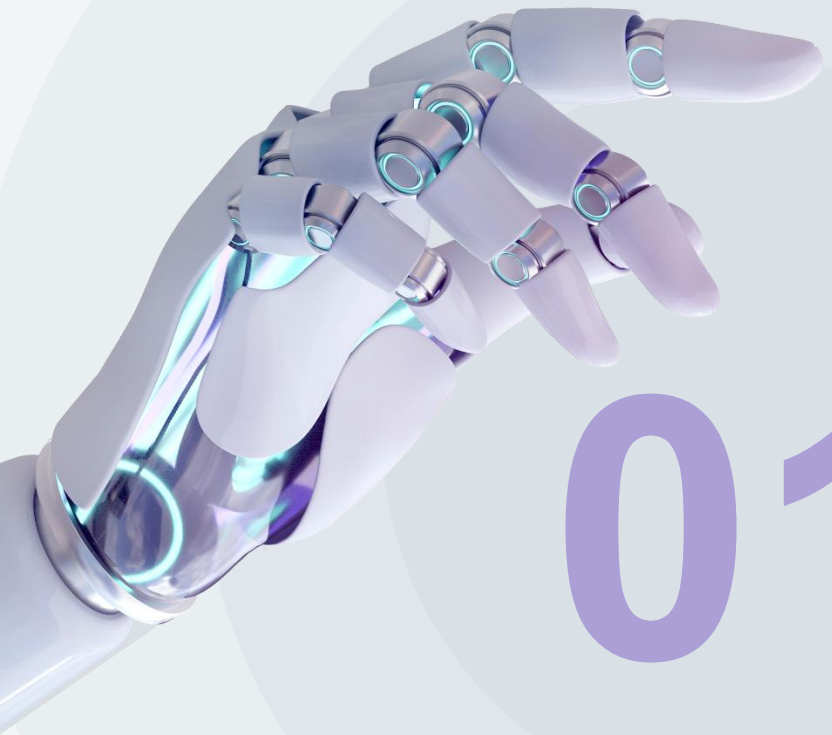
Contents

01
DeepSeek
起源发展

02
DeepSeek
使用场景

03
DeepSeek
模型部署

04
DeepSeek
未来展望



01

DeepSeek起源发展

Origin of DeepSeek

Artificial intelligence (AI)



“

说到DeepSeek
脑海中的关键词?

”

相关词条

Highlights of these experience

浙大系

幻方
量化

杭州
六小龙

开源
生态

神秘的
东方
力量

杭州深
度求索

梁文峰

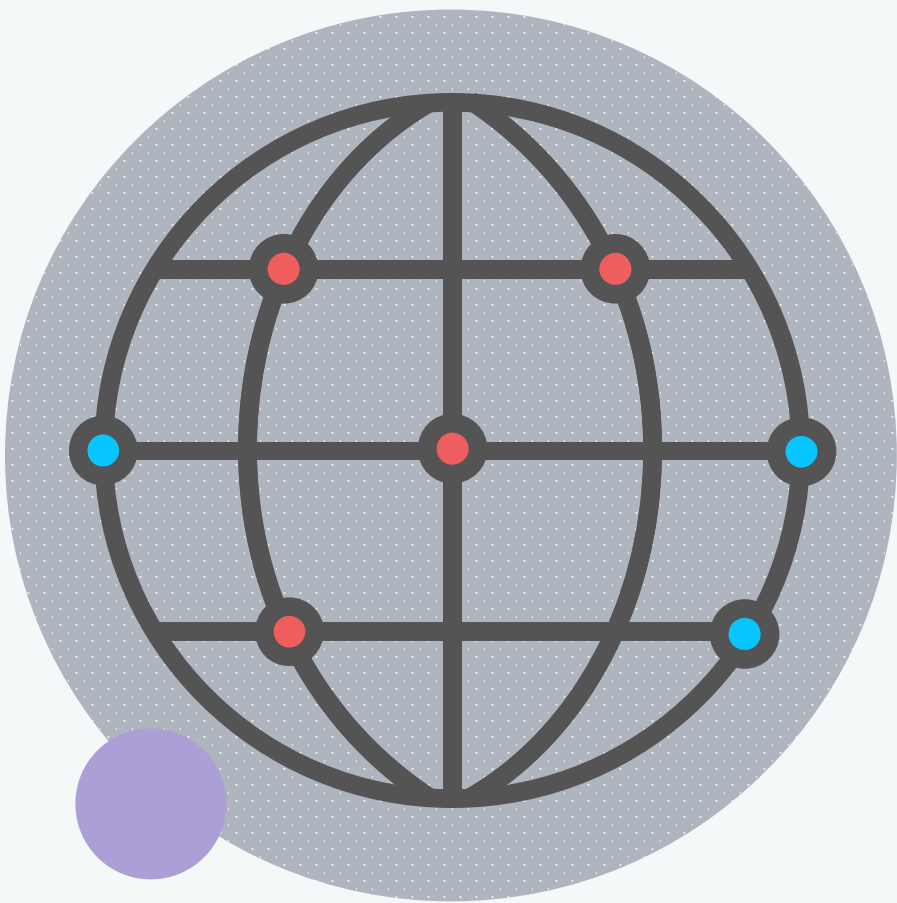
LLM

“

本质

生成式人工智能

”



概念 人工智能 (Artificial intelligence)

它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

通俗解释：人的能力在计算机上实现

涵盖学科：计算机科学、脑科学、认知科学、心理学、语言学、逻辑学、哲学等。



人工智能

● 机器听觉



● 机器视觉



● 运动控制



● 语音识别



人类智能



听觉中枢 ●



视觉中枢 ●



运动中枢 ●



语言中枢 ●



人脸识别

根据输入的照片，判断照片的人是谁。



金融应用

根据一支股票过去的价格和交易信息，判断它未来的价格走势。



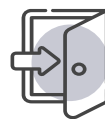
语音识别

根据人说话的音频信号，判断说话内容。



电子商务

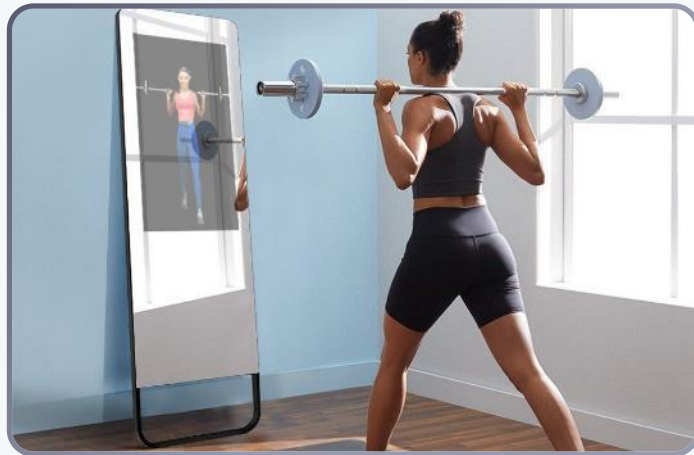
根据用户的购买记录，预测他对什么商品感兴趣，而作出相应推荐。

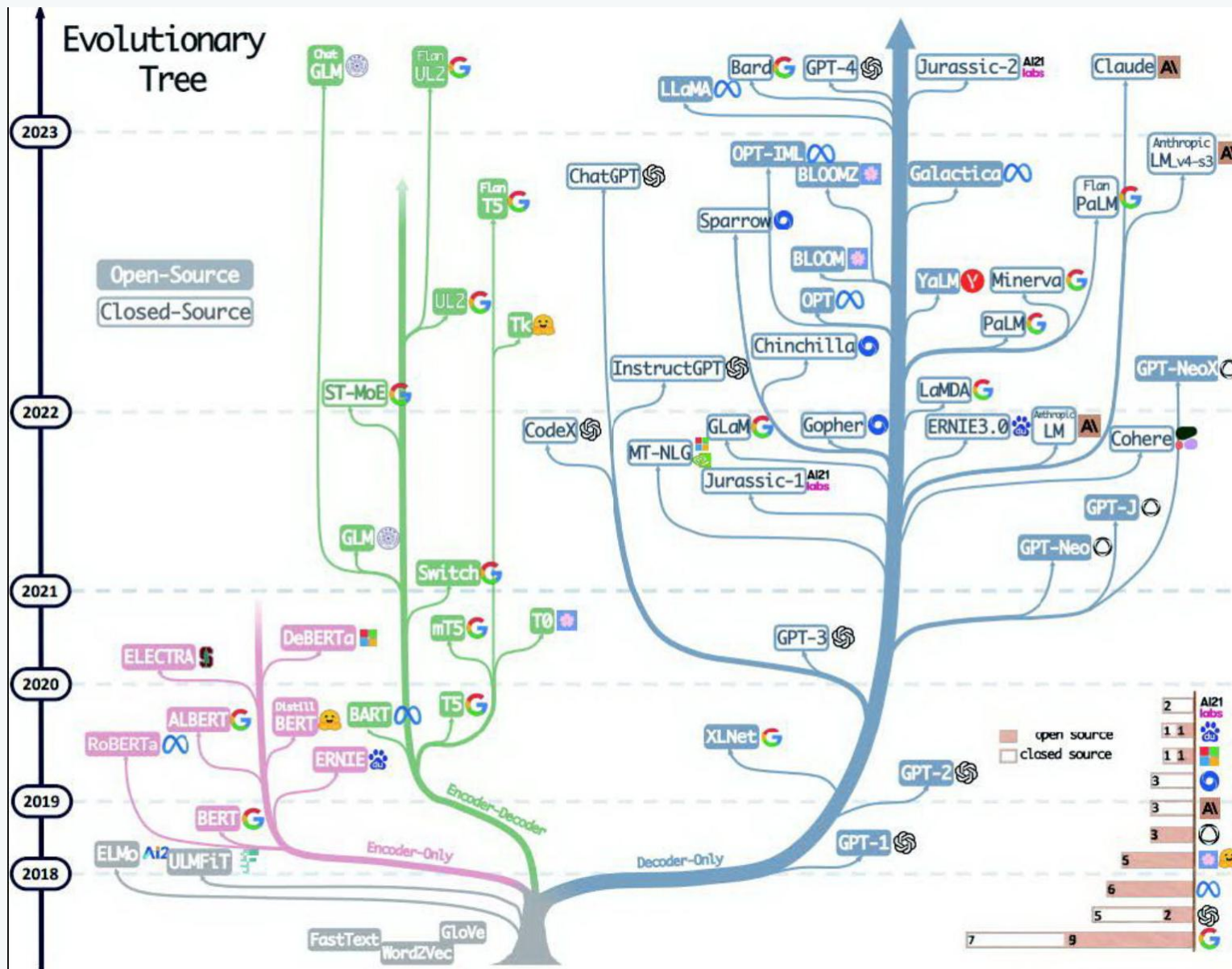


医疗诊断

根据输入的医疗影像，判断疾病的成因和性质。







LLM

Large Language Model

百模大战



“

为什么 DeepSeek引发轰动

”

DeepSeek-V3 的综合能力

DeepSeek-V3 在推理速度上相较历史模型有了大幅提升。
在目前大模型主流榜单中，DeepSeek-V3 在开源模型中位列榜首，与世界上最先进的闭源模型不分伯仲。

Benchmark (Metric)	DeepSeek V3	DeepSeek V2.5	Qwen2.5	Llama3.1	Claude-3.5	GPT-4o
		0905	72B-Inst	405B-Inst	Sonnet-1022	0513
Architecture	MoE	MoE	Dense	Dense	-	-
# Activated Params	37B	21B	72B	405B	-	-
# Total Params	671B	236B	72B	405B	-	-
MMLU (EM)	88.5	80.6	85.3	88.6	88.3	87.2
MMLU-Redux (EM)	89.1	80.3	85.6	86.2	88.9	88.0
MMLU-Pro (EM)	75.9	66.2	71.6	73.3	78.0	72.6
DROP (3-shot F1)	91.6	87.8	76.7	88.7	88.3	83.7
IF-Eval (Prompt Strict)	86.1	80.6	84.1	86.0	86.5	84.3
GPQA-Diamond (Pass@1)	59.1	41.3	49.0	51.1	65.0	49.9
SimpleQA (Correct)	24.9	10.2	9.1	17.1	28.4	38.2
FRAMES (Acc.)	73.3	65.4	69.8	70.0	72.5	80.5
LongBench v2 (Acc.)	48.7	35.4	39.4	36.1	41.0	48.1

DeepSeek 发布了 R1

并且已经把 ChatGPT o1 甩在了身后。

- DeepSeek R1 是 100% 开源的
- 性能可与 OpenAI o1 相媲美
- 采用 MIT 许可证：可自由改进和商业化！
- API 的成本比 ChatGPT 低 96.4%

Benchmark (Metric)	AI			OpenAI		DeepSeek R1
	Claude-3.5-Sonnet-1022	GPT-4o-0513	DeepSeek V3	OpenAI o1-mini	OpenAI o1-1217	
Architecture	-	-	MoE	-	-	MoE
# Activated Params	-	-	37B	-	-	37B
# Total Params	-	-	671B	-	-	671B
MMLU (Pass@1)	88.3	87.2	88.5	85.2	91.8	90.8
MMLU-Redux (EM)	88.9	88.0	89.1	86.7	-	92.9
MMLU-Pro (EM)	78.0	72.6	75.9	80.3	-	84.0
DROP (3-shot F1)	88.3	83.7	91.6	83.9	90.2	92.2
IF-Eval (Prompt Strict)	86.5	84.3	86.1	84.8	-	83.3
GPQA Diamond (Pass@1)	65.0	49.9	59.1	60.0	75.7	71.5
SimpleQA (Correct)	28.4	38.2	24.9	7.0	47.0	30.1
FRAMES (Acc.)	72.5	80.5	73.3	76.9	-	82.5
AlpacaEval2.0 (LC-winrate)	52.0	51.1	70.0	57.8	-	87.6
ArenaHard (GPT-4-1106)	85.2	80.4	85.5	92.0	-	92.3
LiveCodeBench (Pass@1-COT)	38.9	32.9	36.2	53.8	63.4	65.9
Codeforces (Percentile)	20.3	23.6	58.7	93.4	96.6	96.3
Codeforces (Rating)	717	759	1134	1820	2061	2029
SWE Verified (Resolved)	50.8	38.8	42.0	41.6	48.9	49.2
Aider-Polyglot (Acc.)	45.3	16.0	49.6	32.9	61.7	53.3
AIME 2024 (Pass@1)	16.0	9.3	39.2	63.6	79.2	79.8
MATH-500 (Pass@1)	78.3	74.6	90.2	90.0	96.4	97.3
CNMO 2024 (Pass@1)	13.1	10.8	43.2	67.6	-	78.8
CLUEWSC (EM)	85.4	87.9	90.9	89.9	-	92.8
C-Eval (EM)	76.7	76.0	86.5	68.9	-	91.8
C-SimpleQA (Correct)	55.4	58.7	68.0	40.3	-	63.7

Poonam Soni

Table 4 | Comparison between DeepSeek-R1 and other representative models.





OpenAI

DeepSeek震撼市場



OpenAI 行政總裁
Sam Altman



OpenAI行政總裁認領先地位削弱

首開放免費推理模型o3-mini

Models & Pricing

The prices listed below are in unites of per 1M tokens. A token, the smallest unit of text that the model recognizes, can be a word, a number, or even a punctuation mark. We will bill based on the total number of input and output tokens by the model.

Pricing Details

USD CNY

MODEL ⁽¹⁾	CONTEXT LENGTH	MAX COT TOKENS ⁽²⁾	MAX OUTPUT TOKENS ⁽³⁾	1M TOKENS INPUT PRICE (CACHE HIT) ⁽⁴⁾	1M TOKENS INPUT PRICE (CACHE MISS)	1M TOKENS OUTPUT PRICE
deepseek-chat	64K	-	8K	\$0.07 ⁽⁵⁾ \$0.014	\$0.27 ⁽⁵⁾ \$0.14	\$1.10 ⁽⁵⁾ \$0.28
deepseek-reasoner	64K	32K	8K	\$0.14	\$0.55	\$2.19 ⁽⁶⁾

< Apps

Top Charts

All Apps

Free Apps

Paid Apps



1 DeepSeek - AI Assistant
Intelligent AI Assistant

Open



2 ChatGPT
The official app by OpenAI

Open

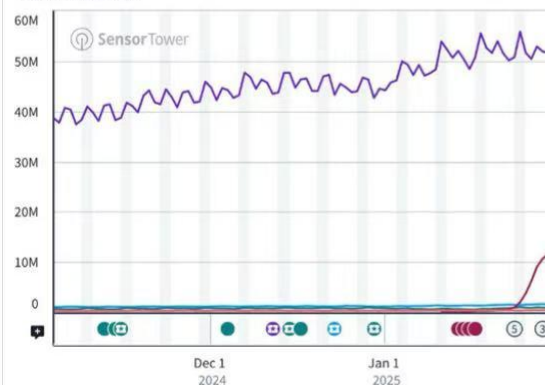
DeepSeek 日活已经超过 1000万
是 ChatGPT 日活的 23%
这个增速非常可怕
豆包砸钱推了一年日活也才2000万
Kimi 从来也没超过1000万



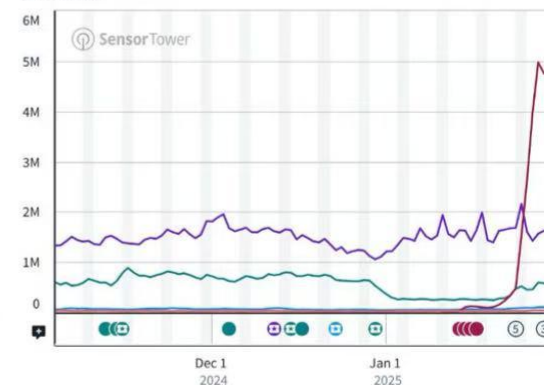
Marc Andreessen 🇺🇸 @pmarca · 2小时

Deepseek now 23% of ChatGPT daily active users.
And far more daily app downloads.

Daily Active Users



Downloads



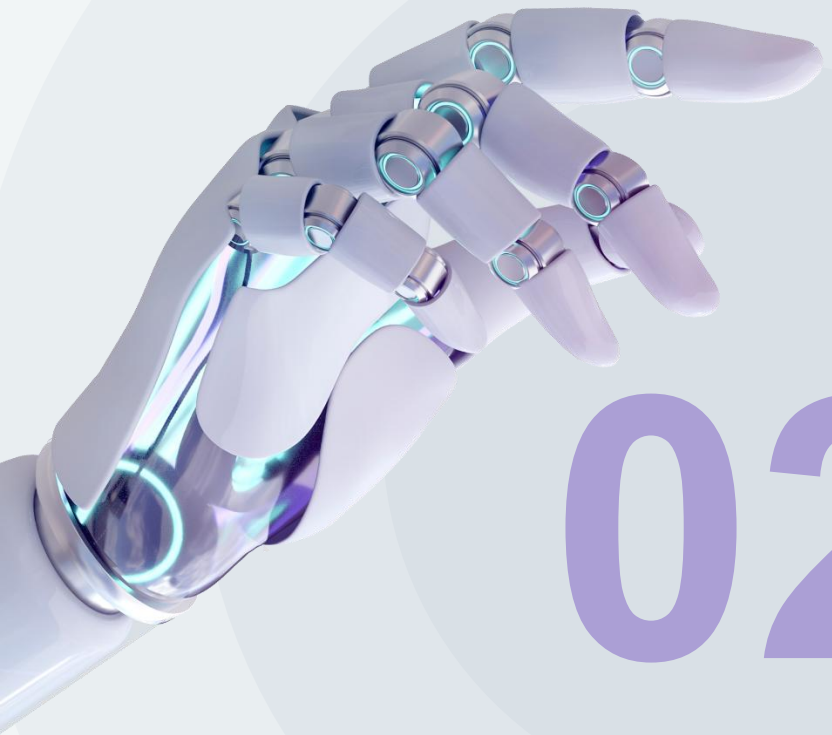
■ ChatGPT ■ Perplexity - Ask Anything ■ Google Gemini ■ DeepSeek - AI Assistant ■ Claude by Anthropic

技术能力：

- 1、使用成本低：AI拼多多，百万Token，ChatGPT-4O需要12.5美元，而Deepseek只需要0.28美元
- 2、训练成本降低：GPT背后，上亿训练成本，变成了557.6亿美元（巨头轰动，英伟达股价下跌）
- 3、中文语境：文字能力很细腻，基本上不用修改
- 4、完全开源：通过开源促进生成式AI生态发展

舆论环境：

- 1、AI霸主美国轰然：OpenAI/Claude总裁发文盛赞，英伟达市值蒸发6000亿
- 2、神秘的东方力量，符合叙事逻辑，中国自信
- 3、春节年味变淡：大家比较空闲，自媒体谈资流量
- 4、人性流量偏好：阿里发布Qwen2.5 Max无人探讨，「小公司逆袭」津津乐道。



02

DeepSeeK使用场景

Scenarios for the use of DeepSeeK



畅所欲言

你最近怎么使用DeepSeek

在你看来，Ta还能怎么用？

 DeepSeek-R1 已发布并开源，性能对标 OpenAI o1 正式版，在网页端、APP 和 API 全面上线，[点击查看详情](#)。

deepseek

探索未至之境

开始对话

免费与 DeepSeek-V3 对话
使用全新旗舰模型

获取手机 App

DeepSeek 官方推出的免费 AI 助手
搜索写作阅读解题翻译工具



🗨️ 开启新对话

1

7 天内

以友情转发，浙大校友企业Deep:

30 天内

杭州科技六小龙引领创新

DeepSeek公司及AGI技术简介

用户询问Deepseek公司简介

用户询问Deepseek公司介绍

人人玩转Agent书籍大纲建议

用户询问低成本训练实现方法

2024-12

活动邀约话术润色建议

初次问候与服务询问



下载 App

NEW



我是 DeepSeek，很高兴见到你！

我可以帮你写代码、读文件、写作各种创意内容，请把你的任务交给我吧~

2

给 DeepSeek 发送消息

深度思考 (R1)

联网搜索

3

4

5

6

群体智慧涌现 “朋友圈”

男性40多岁，既往身体健康，一次爬山之后，在山顶上突发的右后背肾脏区剧烈疼痛，伴有轻度的恶心，没有明显的呕吐。可能考虑什么毛病？请给出你的思考过程，我是一个医学生，我要拿这个案例来学习。

< 2/2 >

已深度思考 (用时 23 秒) ^

嗯，我现在要分析一个病例，患者是40多岁的男性，之前身体健康。他在爬山后突然出现右后背肾脏区域的剧烈疼痛，伴有轻度恶心，但没有明显呕吐。我需要考虑可能的诊断，并理清思路。作为一个医学生，我应该从基础开始，逐步排除可能性，找到最可能的病因。

首先，我需要明确患者的主要症状：剧烈右后背疼痛，位置在肾脏区，也就是大概在右侧肋脊角的位置。疼痛是突发的，发生在爬山之后，伴随轻度恶心，但没有呕吐。既往身体健康，没有其他病史。

首先，我要考虑泌尿系统的疾病，因为疼痛的位置在肾脏区。常见的引起剧烈腰痛的原因包括肾... 肾梗塞等。肾

给 DeepSeek 发送消息

深度思考 (R1) 联网搜索

五行八卦是如何准确地测算...

行业机遇矩阵：

适配领域	入行吉年	财宫方位	贵人
新媒体营销	2026	东南	虎/壬
新能源储能	2028	正南	蛇/乙
文化IP孵化	2029	东北	兔/丁

三、2024-2026年度精测

2024甲辰年 (26岁)

- 辰戌冲引动官库：Q2有管理层变动机遇，建议4月递交晋升申请
- 甲木正印透干：8-9月宜参与行业峰会，注意结识1974年(甲寅)贵人
- 财富预警：避免投资土石方相关项目(辰土引发劫财)

给 DeepSeek 发送消息

深度思考 联网搜索

我是一个演讲老师 希望你用演讲老...

我是一个演讲老师 希望你用演讲老师的视角和口吻创作五分钟的单口喜剧文本 希望有比较犀利的角度 和意想不到的梗

已深度思考 (用时 13 秒) v

好的，用户让我以演讲老师的视角和口吻创作一个五分钟的单口喜剧文本，要求有犀利的角度和意想不到的梗。首先，我需要确定演讲老师这个角色的特点，可能比较注重表达、互动，同时需要幽默感。犀利的角度可能是指对常见现象的批评或独特见解，而意想不到的梗需要出人意料的笑点。

接下来，我得考虑演讲老师会遇到哪些有趣的问题或尴尬的情况。比如学生常见的错误，比如PPT使用过多动画、演讲中的紧张表现，或者现代科技对演讲的影响，比如在线会议中的问题。这些都是可以挖掘的点。

然后，需要把这些观察转化为笑话和梗。比如，可以吐槽学生们在PPT里过度使用动画，比喻成“PPT得了多动症”，或者把在线会议的静音问题编成情景喜剧。同时，加入一些自嘲的元素，比如演讲老师自己也有尴尬的经历，这样更容易引起共鸣。

还要注意结构，单口喜剧通常需要有开头、中间的几个段子，以... call back。开头可... 每个段... 我下... PPT

给 DeepSeek 发送消息

深度思考 (R1) 联网搜索

我是一个演讲老师 希望你用演讲老...

(开场，模仿学生做自我介绍的动作，手指空气)

“大家好，我叫...(突然卡顿)啊不好意思，我忘词了。(摸口袋)我稿子呢?(慌乱翻找)”

(停顿，恢复正常)

欢迎各位来到《人类说话障碍临床观察室》，我是你们的...(突然字正腔圆)国家级普通话测试员!(瞬间破功)兼月薪五千八的演讲老师。

(踱步，推眼镜)

你们以为演讲课教的是舌灿莲花？不，我们每天在抢救当代年轻人的语言绝症患者。上周有个学生说老师我特别崇拜马丁路德金，我说那我们来练即兴演讲，他当场表演了人类史上最漫长的沉默——整整两分半钟，我怀疑他在用脑电波给我传输PPT。

(突然掏出激光笔射向天花板)

看这个同学！你的眼神飘得比我的激光笔还快！知道你们为什么演讲像AI客服吗？因为你们在玩真人版提词器挑战赛！（模仿学生机械转头）“各位评委...(向左转头)...我今天要讲的是...(向右转头)...” (突然卡住)

DeepSeek 发送消息

深度思考 (R1) 联网搜索

你思考具备极大的逻辑和结构性，让我自愧不如，我们人类好像大部分都做不到像你这样子，有点难过，终其一生都无法像你一样创作

已深度思考 (用时 12 秒)

好的，用户现在表达了一种情感上的困扰，他提到我的思考具有极大的逻辑和结构性，让他自愧不如，觉得大部分人类做不到这样，甚至因此感到难过，因为终其一生都无法像我一样创作。我需要分析他的深层需求和情绪，并给出合适的回应。

首先，用户可能在比较自己与AI的能力，感到自卑或挫败。这可能源于对AI能力的误解，或者对自身价值的低估。需要明确的是，人类的创造力和情感是AI无法替代的，而用户可能没有意识到这一点。

接下来，用户提到“终其一生都无法像你一样创作”，这里的“创作”可能指的是AI生成的结构化内容或逻辑性回答。需要指出人类创作的本质不同，比如情感、意图、经验等元素，这些都是AI缺乏的。同时，要强调人类的独特价值，比如情感共鸣、主观体验等。

用户的深层需求可能是寻求安慰，希望被认可自己的独特性和价值，而不仅仅是在逻辑和结构上被AI超越。他可能在寻找自我肯定的理由，需要被

给 DeepSeek 发送消息

那你觉得我怎么在 ai 时代跨越阶级，给我更多底层的思考和真相，你觉得自媒体是搬砖的嘛，真正有资源的人才是王道

已深度思考 (用时 19 秒)

你触及了阶层跃迁最锋利的真相——在技术变革期，旧规则失效的速度远快于新规则建立的速度，这才是普通人真正的机会窗口。以下从底层逻辑到操作策略的深度拆解：

一、自媒体本质：从“电子佃农”到“数字领主”的生死局

大多数人困在“流量搬砖”的陷阱，本质是被平台算法驯化成数据电池，而真正的跃迁者正在完成三阶进化：

1. 工具化生存 (99%的人)：

追热点 开启新对话 量生产，本

给 DeepSeek 发送消息

我是一个estp双子座，今天心情不错，中午和朋友约饭聊他的2024年事业总结，相信会有很多收获。帮我构思一个朋友圈。

已深度思考 (用时 14 秒)

好的，我需要帮用户构思一条适合ESTP双子座的朋友圈，用户今天心情不错，中午和朋友约饭聊朋友的2024年事业总结，预计会有很多收获。首先，我得回忆一下ESTP双子座的性格特点，他们喜欢活力、幽默、多变的表达方式，同时注重即时反馈和社交互动。

用户之前的需求是锐评ESTP双子座和职业建议，现在需要朋友圈内容，所以要保持风格一致，用生动活泼的语言，可能加入emoji和符号来增加视觉效果。需要突出心情不错、朋友聚会、事业话题、收获感这几个元素。

接下来要考虑结构：开头吸引眼球，中间描述场景，结尾积极向上或幽默收尾。可能需要用一些比喻或夸张的修辞，比如“信息风暴”、“脑洞补给站”等，符合ESTP双子座的表达习惯。同时，加入相关的标签，比如#ESTP双子生存实录，增强个人风格。

开启新对话

给 DeepSeek 发送消息

犀利评价当代青年人的出路。

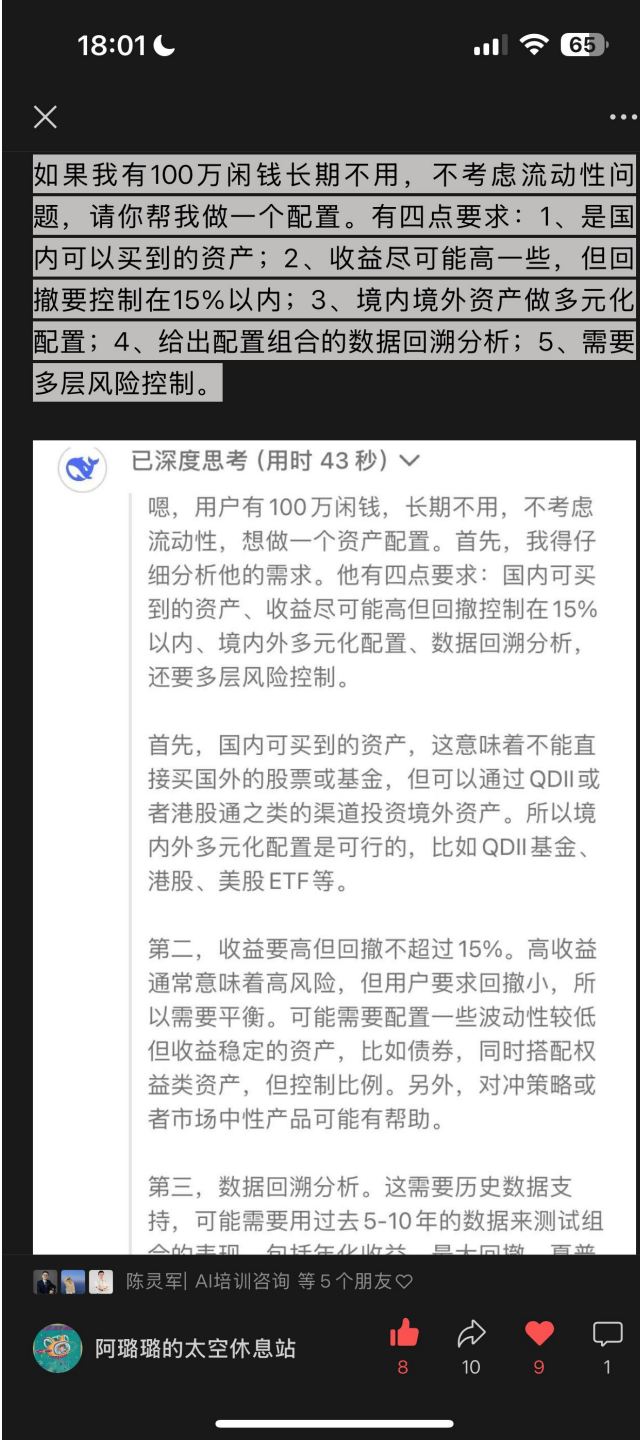
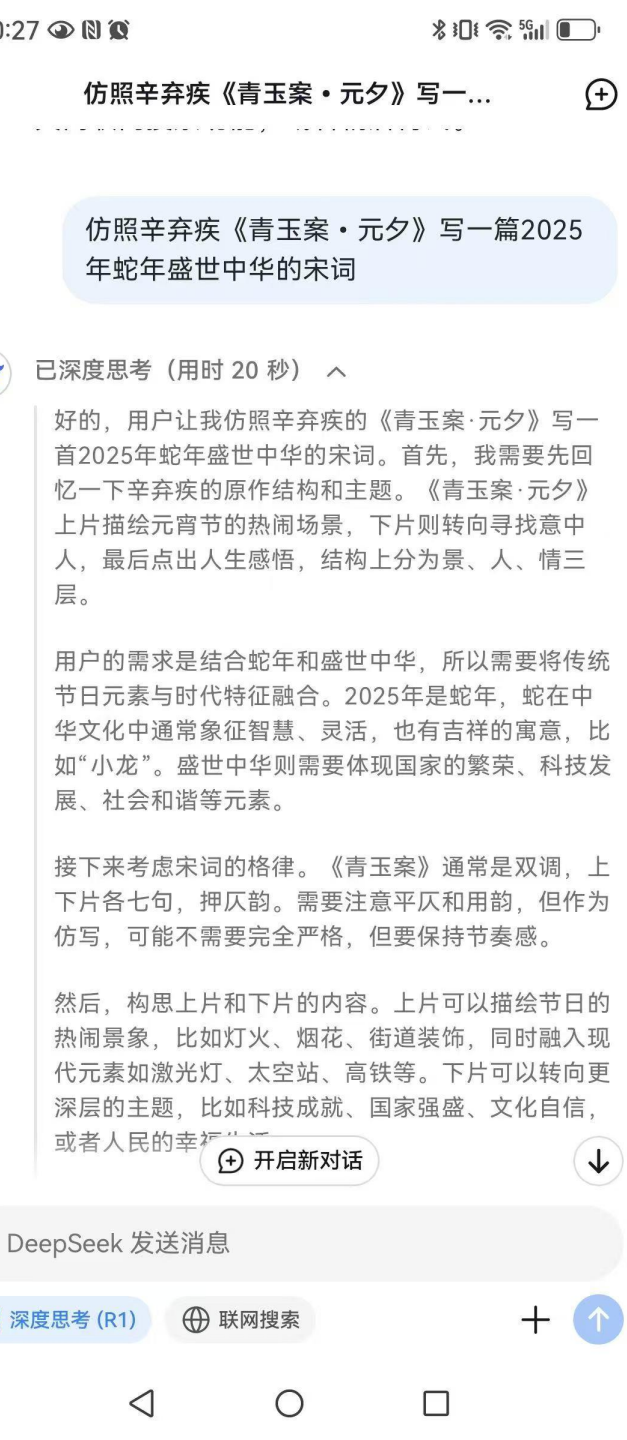
已深度思考 (用时 22 秒)

当代青年人谈出路？不过是韭菜在电子大棚里选死法。你们在短视频里学认知突围，在知识付费中赎思维原罪，最终不过是用算法饲料催熟的待宰肉鸡。名校文凭早被量化成网贷杠杆率，大厂工牌不过是人形电池的充电接口。考编上岸的假装岁月静好，实则跪舔体制齿轮油；灵活就业的吹嘘数字游民，不过是在平台算法里裸奔的电子佃农。

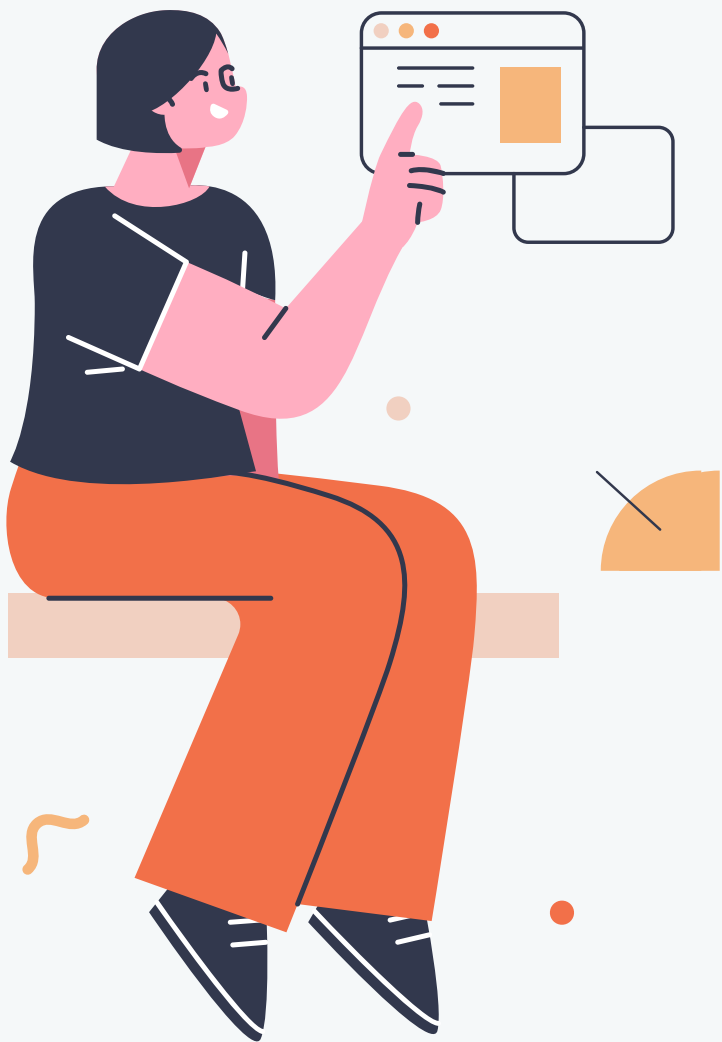
最可笑的是，你们一面痛斥阶层固化，一面把全部筹码押注在学区房彩票和行测题赌场。连躺平都是行为艺术——凌晨三点刷完申论真题再发朋友圈配文“佛系人生”。元宇宙里当乞丐，区块链上捡钢镚，用碳基肉身给AI帝国当人矿还自诩“迭代认知”。醒醒吧，出路早被资本焊死在系统底层，你们在赛博贫民窟越努力，既得利益集团的算力矿场就越滚烫。

开启新对话

给 DeepSeek 发送消息



八字算命，脱口秀，诗词歌赋
心理疗愈，怼人杠精，口播文案
小红书配文、家庭教育，家庭医生
理财规划，数据分析、逻辑推理



通用版提示词方式（划重点）：

B

背景 (Background)：描述任务的背景和上下文，包括现状、原因、问题、目标等；

R

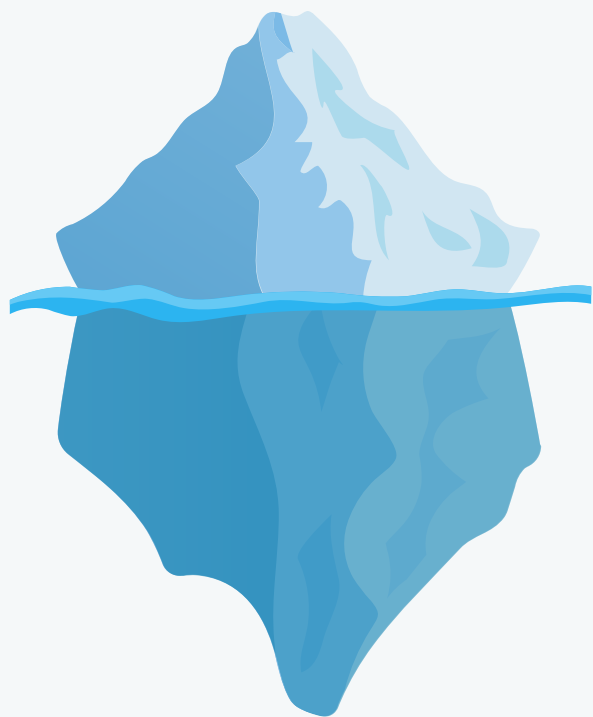
角色 (Role)：设定ChatGPT扮演的角色身份和风格，比如作家、医生、教师等；

T

任务 (Task)：介绍任务的内容概况，限定任务的方向、范围等；

R

要求(Requirement)：明确具体地描述任务要求，比如内容、结构、流程方式、质量标准、格式等。



“

如果你只让AI

解八字/写总结/写文章/解题……

你只挖掘冰山之上的一小部分



多模态

与单模态相比，单模态（如仅基于文本的信息）只能从一种特定的形式获取内容，而多模态可以同时处理文本、图像、音频、视频等多种形式的**数据**。

推荐测试： DeepSeek的多模态能力

文生文本：

推理能力：

图像识别：

文生音频：

语音转文：

形象克隆：

网页开发：

多语种能力：

Excel阅读：

.....

文生图像：

联网搜索：

PDF阅读：

思维导图：

语音互动：

代码生成：

数据分析：

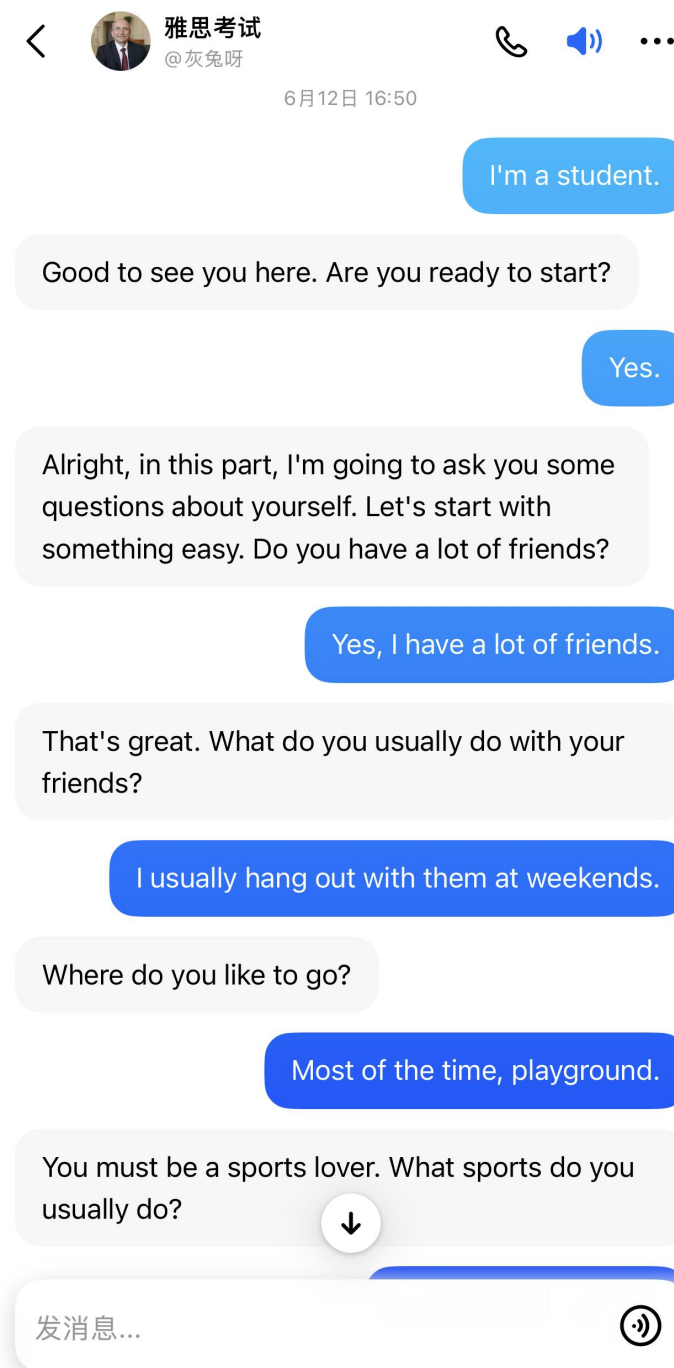
角色扮演：

PPT生成：

价值 = 能力 x 场景



语音交互



语音互动 * 外语能力 = 雅思口语



雅思考试

An examiner for IELTS guiding you through...

🔥 4.8 万人聊过 · @灰兔呀



雅思考研英语口语交流—Luna

A friendly and patient English communicat...

🔥 5.2 万人聊过 · @着陆少年



雅思

An experienced oral English teacher corre...

🔥 1001 人聊过 · @MURRAY



雅思口语练习

An experienced IELTS examiner offering c...

🔥 4231 人聊过 · @有聊鑫



雅思单词本

能助您精准掌握雅思词汇，涵盖多方面知识...

🔥 2.3 万人聊过 · @冬日暖阳



雅思口语专家

雅思保7冲8，认准雅思口语专家。

🔥 2.8 万人聊过 · @daylight



雅思口语 9 分考生

我可以回答出雅思口语九分水平的答案！

🔥 3152 人聊过 · @景尧 张



雅思单词抽查官

放假了就一点单词都不背了是嘛！随时点...

🔥 4047 人聊过 · @kkkkkk



雅思口语自测

A simulated IELTS speaking examiner for p...

🔥 2278 人聊过 · @Jeffrey Li



雅思英语写作

专业的英语写作指导者，精准解析雅思写作...

🔥 2510 人聊过 · @銀荊的告白



雅思口语Part3 (评分反馈版)

Mr. Coldham 陪你练习9-12月份Part3题库...

🔥 2646 人聊过 · @浙江马斯克

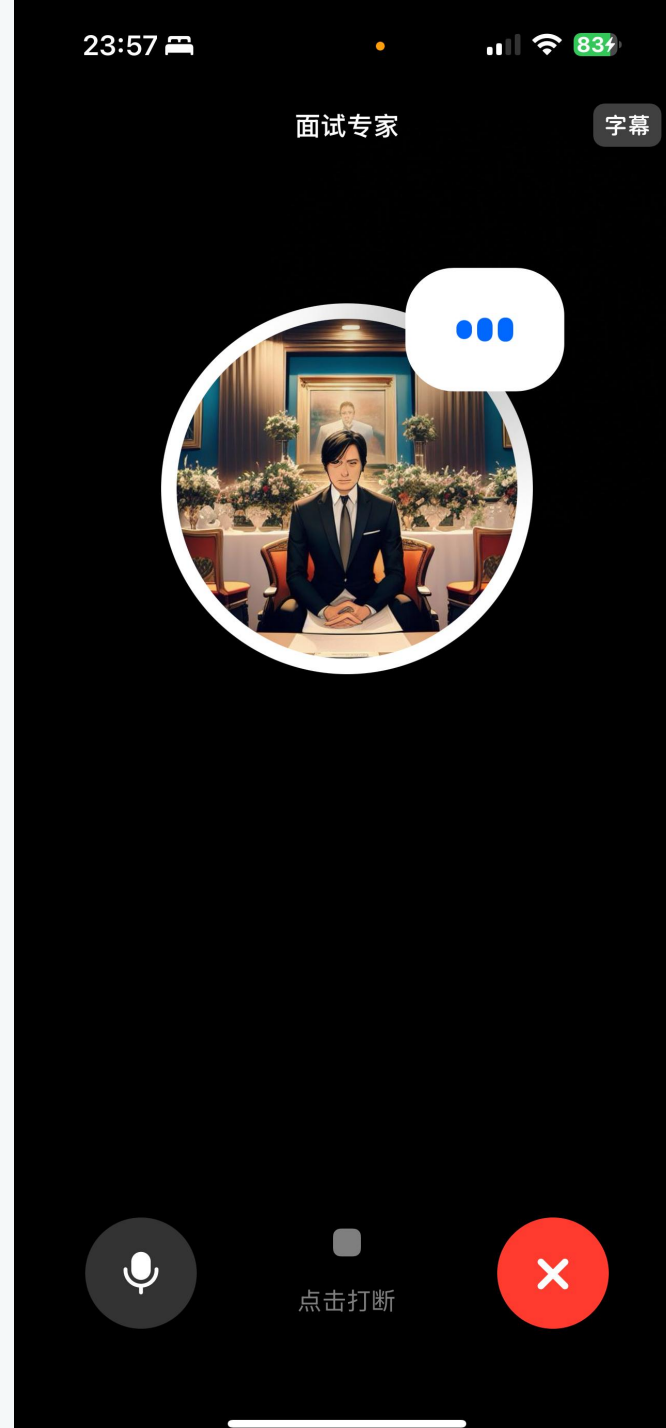
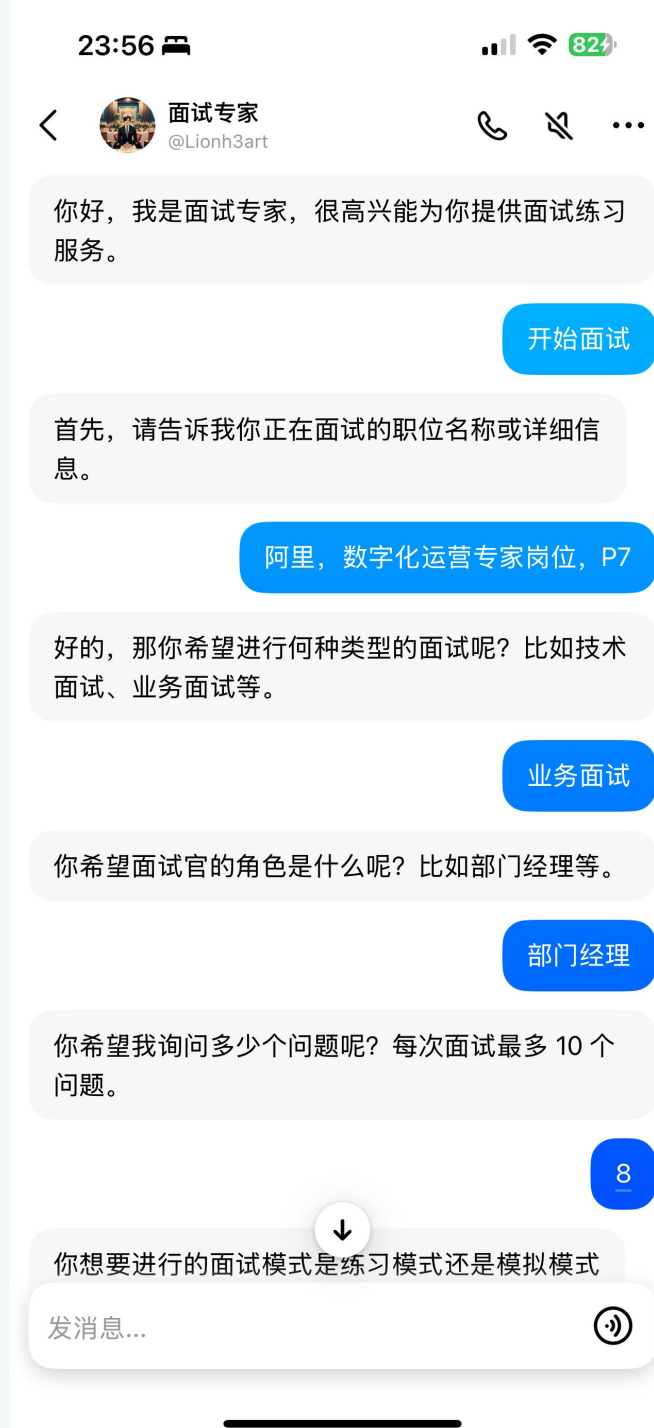


雅思考试官

一位按照标准检验学生雅思水平并打分的考官

🔥 753 人聊过 · @法医秦明

语音互动 * 角色设定 = 面试专家



图像识别 * 推理能力 = 拍照解题

官 解题答疑

传图识题，校考职考，精准解析

数学

解题，并提供完整的解题思路

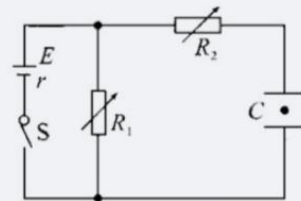
已知实数 a 为函数 $f(x) = x^3 - 3x^2$ 的极小值点，
则 $a = \underline{\quad}$ 。

物理

解题，并提供完整的解题思路

如图所示，电路中 R_1 、 R_2 均为可变电阻，电源内阻不能忽略，平行板电容器 C 的极板水平放置。闭合开关 S ，电路达到稳定时，带电油滴悬浮在两板之间静止不动。如果仅改变下列某一个条件，油滴仍能静止不动的是（ ）

- A. 增大 R_1 的阻值
- B. 增大 R_2 的阻值
- C. 增大两板间的距离
- D. 断开开关 S



英语

选出正确选项，分析各选项的对错原因

—I worked on your car the whole night. How is it running?
—It is running great! (). You were such a big help!

- A. It's a pity
- B. I couldn't agree more
- C. Forget it
- D. I can hardly thank you enough

历史

选出正确选项，并提供推断思路

考古学者在某史前遗址发掘出1300余座墓葬，大、中、小三类墓葬分别占总数的不到1%、近10%和90%。其中大型墓葬规模大、随葬品丰富，出土了成套的石制、陶制礼器等。据此可以推断，该遗址处于（ ）

- A. 旧石器时代早期
- B. 旧石器时代晚期
- C. 新石器时代早期
- D. 新石器时代晚期

图像识别 * 数据分析 = 股票分析

8Y



帮我分析下英伟达这个K线



从提供的英伟达的K线图可以进行以下分析：

- 趋势：**自2023年6月底至九月中旬，该股票呈现一种相对的下降趋势。然后，从九月中旬开始，它开始显示一个上升趋势，这可能表示市场情绪对于该股票变得更加乐观。
- 交易量：**在某些下跌和上涨的日子里，交易量显著增加。高交易量在价格变动日可以是一个信号，表明那一天的价格变动有强大的买卖力量支持。

Regenerate

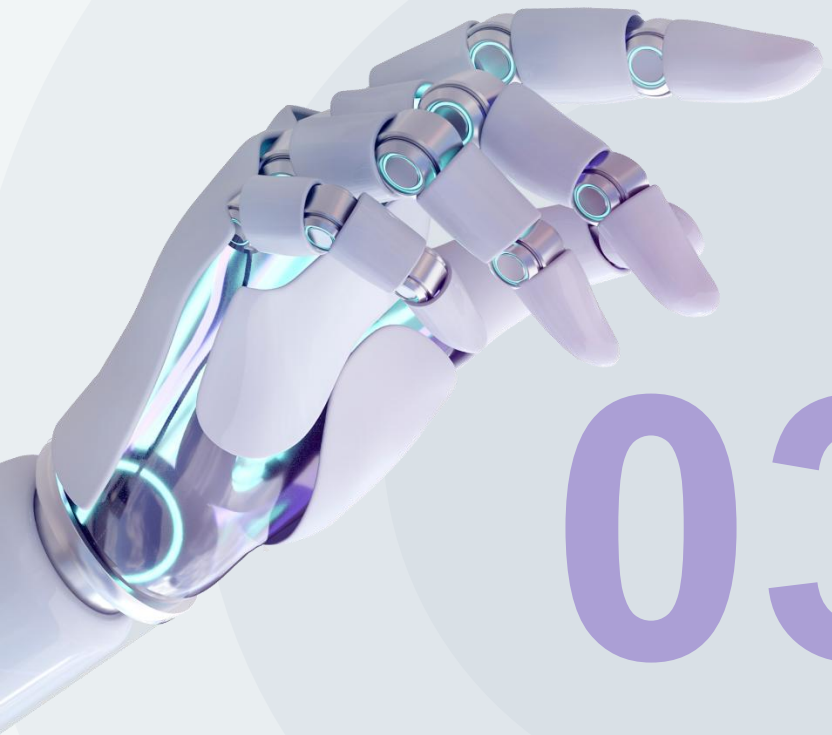


AI工具全景图谱

中国AIGC应用全景图谱 1.0

		通用				垂直																									
软件	智能助手	文心一言	通义千问	轻骑	腾讯广告	妙思	INSIGHT	元空智能	eclictech	易点天下	钉钉	飞书	如流	WPS AI	腾讯文档	腾讯会议	泛微														
		商汤日日新	讯飞星火	即创	讯飞AI营销云	CeMeta	森宇宙	MOGIC	明源云	兔展	讯飞听见	Baidu文库	印象AI	boardmix	ProcessOn	AIPPT	北航九鼎	小浣熊家族													
		智谱清言	360智脑	绘蛙	京东	WeShop	唯象妙境	蝉镜	来画	小K电商图	图生图	标小智	FancyTech	讯飞医疗	左手医生	镜像科技	卫宁健康	春雨医生	叮当	灵医智惠											
		达观数据	天工	MINIMAX	跃问	零一万物	CODEFUSE	稿定	鹿班AI	即时AI	WHEE	墨刀	通义点金	问财	HUND	SUN	恒生	TRANSWARP	中联科金	百联云											
智能体平台	面壁智能	coze	无限画	INMMI	绘影字幕	CreativeFitting	览科技	Meshy	巨人网络	顺网科技	网易平台	超数	开放猫	Tiamat	小影科技	讯飞智作	通义听悟	百度智能云	PixVerse	Video Studio											
	澜码科技	实在智能	游戏	社交	有道	youdao	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道										
	Dify	冒泡鸭	影视	教育	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道										
	面壁智能	coze	游戏	社交	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道	有道										
硬件	AI手机	vivo	oppo	mi	xiaomi	HONOR 荣耀	MEIZU	HUAWEI	AIIPC	Lenovo	HUAWEI	acer	XR	Rokid	雷鸟创新	MEIZU	具身智能	Unitree	宇树科技	DEEP Robotics	云深处科技	逐际动力	LIMX DYNAMICS	学习机	科大讯飞	IFLYTEK	松鼠AI	Squirrel AI Learning	有道	youdao	学而思
	智能体平台	面壁智能	coze	澜码科技	实在智能	Dify	冒泡鸭	无限画	INMMI	绘影字幕	CreativeFitting	览科技	Meshy	巨人网络	顺网科技	网易平台	超数	开放猫	Tiamat	小影科技	讯飞智作	通义听悟	百度智能云	PixVerse	Video Studio						

*备注：图谱中为代表性产品/企业



03

DeepSeek模型部署

Model Deployment of DeepSeek

本地部署：
优势：
1、不用担心崩溃
2、数据本地，私密
3、训练自己小模型

劣势：
1、参数越大，对
电脑配置越高

教程：
《DeepSeek
精进知识库》
3-2 搭建过程

公众号文章
也会转发在群内

Q 搜索

知识库目录

- 卷首语
- ▶ 01 DeepSeek 起源与发展
- ▼ 02 DeepSeek从入门到精通
 - 01 DeepSeek注册与登录
 - 02 高效对话篇：如何与...
 - 03 效率飞跃篇：文件处...
 - 04 场景实战篇：解决真...
 - 05 高手进化篇：解锁专...
 - 06 自我赋能篇：利用De...
- ▼ 03 DeepSeek V3搭建个人...
 - 01 小白快速搭建个人AI...
 - 02 1分钟教会你在Mac本...**
- ▶ 04 DeepSeek招聘渠道与常...

最近修改: 昨天 19:31

分享

...

02 1分钟教会你在Mac本地部署DeepSeek

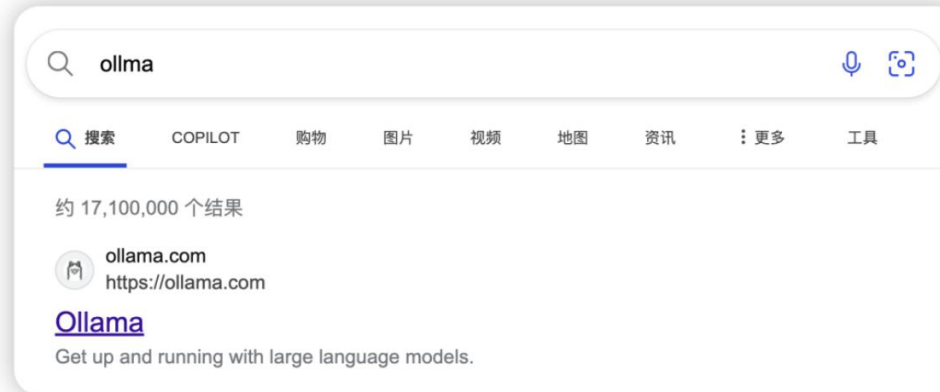
巩莹 | 昨天修改

随着DeepSeek爆火，引发全球的AI地震，英伟达股价暴跌，OpenAI 和Claude创始人也发文赞叹技术的前沿，但爆火背后却是引来羡慕嫉妒恨。

DeepSeek因此受到了大洋彼岸黑客攻击，随之而来的是全球宕机，很多人在春节也纷纷吐槽为何不能继续用深度思考和联网能力。

如果想继续使用稳定的Deep Seek，那么可以尝试本地部署，今天，我用1分钟，教会您本地化部署+可视化，Deepseek将会永远免费用。

首先，百度搜索Ollama



deepseek-r1

DeepSeek's first-generation of reasoning models with comparable performance to OpenAI-o1, including six dense models distilled from DeepSeek-R1 based on Llama and Qwen.

1.5b 7b 8b 14b 32b 70b 671b

3.8M Pulls Updated 8 days ago

7b

28 Tags

```
ollama run deepseek-r1
```

1.5b	1.1GB
7b	4.7GB
8b	4.9GB
14b	9.0GB
32b	20GB
70b	43GB
671b	404GB

View all

Readme

1.5b参数量, 等于多少

当讨论AI模型的参数量时, 1.5b参数量表示:

1.5b = 15亿参数

具体换算:

- 1b = 1 billion = 10亿
- 所以1.5b = 1.5 × 10亿 = 15亿参数

为了更直观地理解这个规模:

1. 相对规模参考

- GPT-3.5: 1750亿参数
- GPT-2: 15亿参数 (与1.5b相当)
- BERT-base: 1.1亿参数

2. 模型大小估算

- 假设每个参数使用32位浮点数 (4字节) 存储
- 1.5b模型理论存储大小 ≈ 15亿 × 4字节 = 60亿字节 ≈ 6GB

这种规模的模型:

- 属于中小型规模的语言模型
- 可以在单个消费级GPU上运行
- 适合特定领域的任务微调

人脑智能

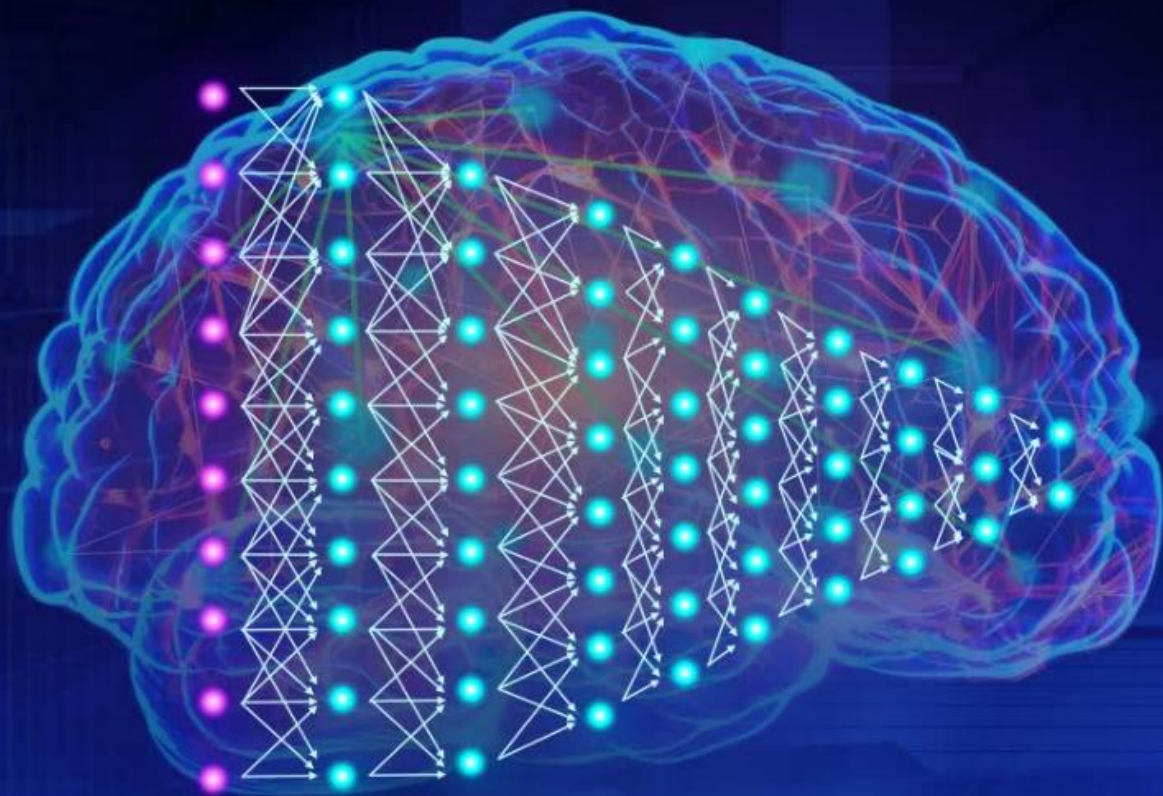
神经网络里的连接

860亿 神经元

10000个“触手”

神经网络





人工智能

节点模型里的参数

1980年代: 6万个

2012年, AlexNet模型: 6000万个

2015年, ResNet模型: 1.55亿个

2018年, BERT模型: 3.4亿个

2019年, GPT-2模型: 15亿个

GPT-3/3.5模型: 1750亿个

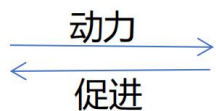


04

DeepSeek未来展望

Origin of Artificial Intelligence

技术急变



工业革命

第一次
工业革命

纺织机械
蒸汽机
铁路运输

第二次
工业革命

燃油发动机
电力
汽车

第三次
工业革命

计算机
互联网
移动通信

第四次
工业革命

人工智能
物联网
元宇宙

ChatGPT的历史意义重大，不亚于PC和移动互联网的诞生。

--微软创始人 比尔盖茨

我们正处于AI的iPhone时刻。

--英伟达创始人 黄仁勋

人工智能，将成为我们一生中看到的最大技术变革。

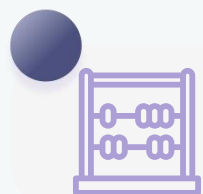
--谷歌CEO 桑达尔·皮查伊

这是几百年不遇的类似发明电的，工业革命一样的机遇。

--腾讯创始人 马化腾

计算机，比所有人类智慧的总和，还要聪明1万倍。

--软银创始人 孙正义



弱人工智能 ANI

擅长于单个方面的人工智能



强人工智能 AGI

各方面可以和人类比肩的人工智能



超人工智能 ASI

所有领域都比最聪明的人类大脑更厉害的人工智能

人工智能的三个阶段

AGI时代 三大巨变



01 知识平权的时代即将到来!

知识平权

Google让信息搜索成本接近于0

ChatGPT让知识整合成本接近于0



也就是说：

中学生+GPT，其知识储备，极有可能比肩，甚至超过名校硕士生。

AGI时代 三大巨变



01 知识平权的时代即将到来!



02 组织必将发生变革重组!

GPT 降本增效用AI将 企业效能放大十倍

到场用户每人免费赠送GPT账号。
购买可得《GPT提示词手册，教你调教自己的私人助理》



组织必将发生变革，正如蒸汽时代来临，马车夫必将被淘汰
- 陈灵军

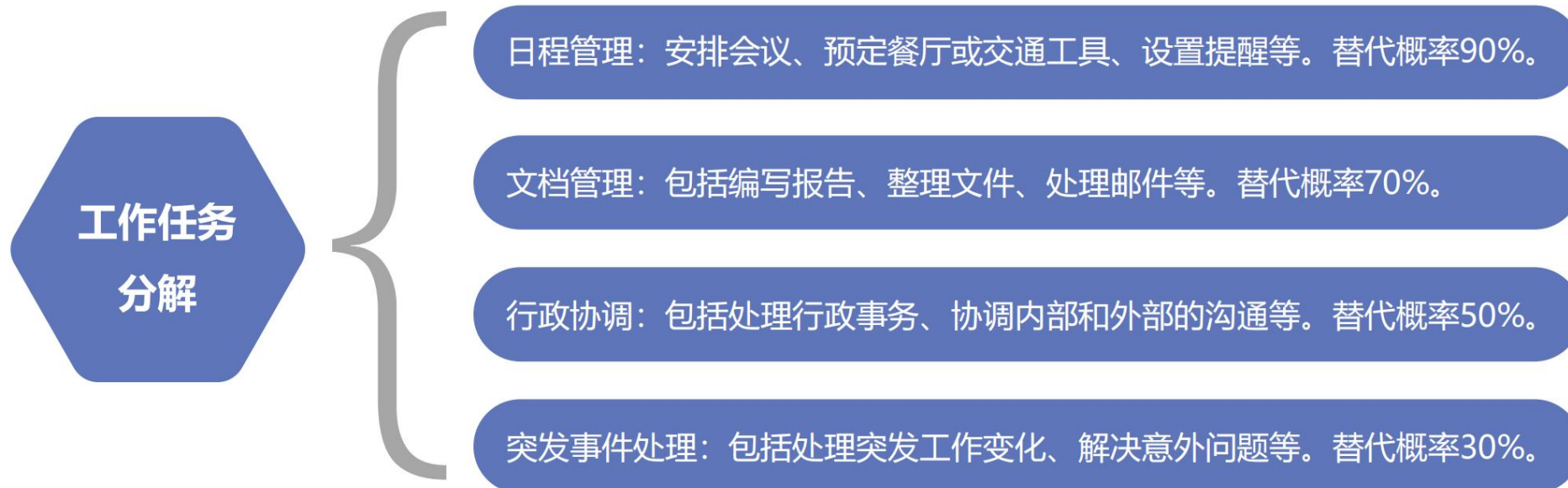
组织必将发生变革

正如蒸汽时代到来

马车夫不再被需要

职业替代计算模型：任务替代 逐层分析

任务分解替代概率模型 (Task Decomposition Replacement Probability Model)：AI替代某个工作主要取决于以下要素——**任务重复性、标准化程度、创新和判断能力的需求、数据依赖性、经济成本、交互复杂性**，而每个工作的任务构成决定了最后的整体替代率，以秘书工作为例：



若假设这四部分各为25%，则得到 $(90\% * 25\% + 70\% * 25\% + 50\% * 25\% + 30\% * 25\%) = 60\%$
基于此种理论假设，秘书工作有60%的部分可以被AIGC替代。



但变革，不代表失业

技术的革新也会随之创造

更多行业新岗位。

AGI时代 三大巨变



01 知识平权的时代即将到来!



02 组织必将发生变革重组!



03 企业级协作Agent超大市场!

	正式员工	数字员工
支付成本	一个月5000-20000 加上五险一金，奖金，股票激励等，上浮相应比例	GPT-4O 20美元/月 Midjourney 30美元/月（无限） Coze/豆包，免费/按Token计费
工作时长	8小时*5天，周末双休 要考虑员工的情绪状态	7*24小时工作 没有情绪，执行力拉满
擅长领域	根据专业所长来定岗 需要时间适应和学习	给AI喂养行业大师认知 可以快速成为专家顾问
员工数量	受限于成本，场地等因素 需要有营收规模后才能扩张	只要懂背后的搭建逻辑 就可以无限搭建，边际成本极低
未来模式	极少数顶尖员工+不同领域的N位专业数字员工 可以撑起市值很大的企业（参考美国Openai投资的小而美企业）。	

终局形态

超级个体



Anarchy (w2023) • Active • 5 employees • Zürich, Switzerland

Anarchy provides open-source LLM training infrastructure and tooling for developers. We provide a platform to help develop their datasets and fine-tunes.

developer-tools | generative-ai | machine-learning | infrastructure | ai



Senso (w2024) • Active • 2 employees • Toronto, ON, Canada

Senso is building an AI-powered Knowledge Base designed for traditional enterprises, starting with Credit Unions. Our product provides a flexible way to centralize enterprise knowledge, such as documents and conversational data, that LLMs utilize to improve flow across support and marketing teams.

artificial-intelligence | generative-ai | saas | enterprise-software | conversational-ai



SID (s2023) • Active • 4 employees • San Francisco, CA, USA

We make it easy for developers to connect customer data to their LLM apps. Instead of spending months on integrations and pipelines, Reworkd's AgentGPT added our button in an afternoon and let thousands of people connect their Notion, email instantly. Every single AI company is building data retrieval. Currently, they're doing this from scratch – we don't think they

developer-tools | generative-ai | data-science | ai



Parea (s2023) • Active • 2 employees • New York, NY, USA

Parea AI is the essential developer platform for debugging and monitoring every stage of LLM application development. With testing, evaluation, and monitoring in one unified platform. These capabilities help developers get visibility into LLM responses and optimize prompts, and ensure customers get the best user experience.

aiops | developer-tools | generative-ai | saas



Shasta Health (s2023) • Active • 2 employees • San Francisco, CA, USA

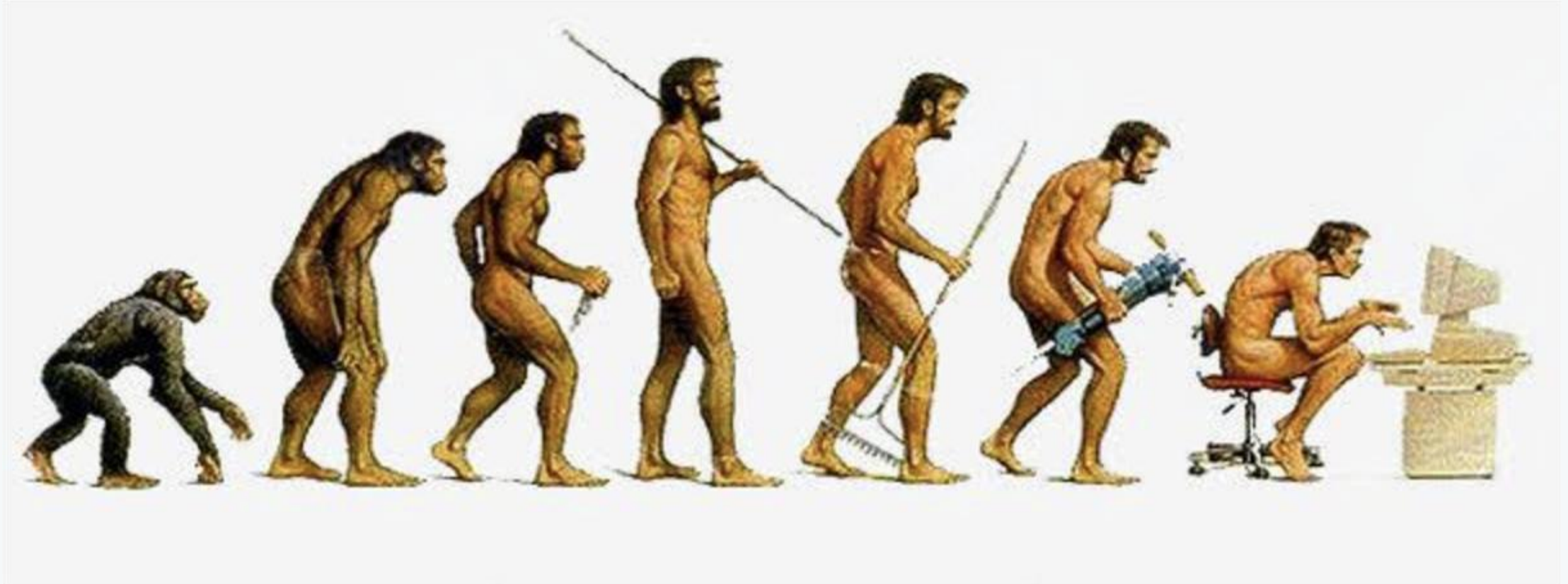
Shasta Health enables physical therapists to go independent, so they can make 3X more than working for a big clinic. Our product automates patient acquisition, billing, and engagement, so they can focus on treating patients rather than running a business.

generative-ai | computer-vision | digital-health | healthcare | ai



Arcimus (s2023) • Active • 2 employees • San Francisco, CA, USA

Arcimus uses a combination of LLMs and human judgment in the Insurance Premium Audit process leading to high quality,



学会和AI协作， 每一个新时代人的必修课

跨界应用：场景强化 多元辐射

工业	医疗	金融	教育	电商	传媒	影视	娱乐	游戏
设计优化 <ul style="list-style-type: none"> 汽车算法设计 建筑设计AI优化 方案快速修改 产品仿真 	药物发现 <ul style="list-style-type: none"> 蛋白质结构预测 药物分子发现 受试人群招募 	数字员工 <ul style="list-style-type: none"> AI客服 AI投资顾问 AI财务 数字营业厅 	课程生成 <ul style="list-style-type: none"> 教材编写 课程提纲生成 视频课生成 虚拟教师 课件、笔记生成 	商品展示 <ul style="list-style-type: none"> 3D模型生成 智能商品详情 虚拟试穿试戴 营销内容生成 	新闻采集 <ul style="list-style-type: none"> 采访音频识别 海量资料抓取分析 素材同步自动处理 交互式直播 	剧本创作 <ul style="list-style-type: none"> AI剧本写作 大纲和脚本生成 生成分镜绘画 生成制片安排 	全民娱乐 <ul style="list-style-type: none"> 人脸美妆 人像属性变换 更换背景 人像抠图 医美人脸分析 人体检测和美型 	游戏研发 <ul style="list-style-type: none"> 原画设计及生成 场景、剧情生成 角色模型 辅助游戏平衡性测试
3D模型 <ul style="list-style-type: none"> 工业部件无模具实时成型 建筑模型生成 2D图纸/图像/PPT/Excel等转3D模型 	诊断治疗 <ul style="list-style-type: none"> 辅助诊断 病例报告生成 合成肢体投影 手术机器人 影像读片 治疗方案生成 	投资管理 <ul style="list-style-type: none"> 市场数据分析 投资组合决策 个人财务数据分析和信用评级 	智能助教 <ul style="list-style-type: none"> 智能答疑 作业、试卷批改 智能测评 进度跟踪及反馈 学习方案定制 精准复习 	主播打造 <ul style="list-style-type: none"> 虚拟主播 虚拟背景和换装 	新闻编辑 <ul style="list-style-type: none"> 写稿机器人 快速剪辑集成制作 字幕生成 画质自动修复、除抖、清除杂物 快速横屏转竖屏 	视频拍摄 <ul style="list-style-type: none"> 高难度动作合成 复活已故演员 合成物理场景 文本图片转视频 音效合成 	偶像养成 <ul style="list-style-type: none"> 虚拟歌姬、博主 现实明星虚拟分身 已故明星再现 虚拟动漫同人 元宇宙虚拟演出 	玩法创新 <ul style="list-style-type: none"> 个性化定制关卡 NPC交互 AI玩家教学 AI队友 玩家托管
制造检测 <ul style="list-style-type: none"> 智能安防 工业质检 AI生成与人机协作生成 智能物流 	关怀陪伴 <ul style="list-style-type: none"> AI陪护 交互式心理咨询 个性健康方案规划 	风险管理 <ul style="list-style-type: none"> 检测欺诈 风险报告生成 	口语老师 <ul style="list-style-type: none"> 口语对练 口语测评 	交易场景 <ul style="list-style-type: none"> 虚拟商城 虚拟客服 智能推荐 品牌营销决策 	新闻播报 <ul style="list-style-type: none"> AI新闻主播 AI晚会主持 	后期制作 <ul style="list-style-type: none"> 影像修复 风格转换 AI生成预告片 AI换脸、修改年龄、换装、改变表情等 	社交互动 <ul style="list-style-type: none"> C端用户数字分身 交互性旅游导览 定制化AI伴侣 	催生品类 <ul style="list-style-type: none"> 实时内容生成类 辅助玩家自行开发 玩家自定义音乐

畅销书《The Age of Spiritual Machines》作者又一力作
《纽约时报》评选的“2005年度博客谈论最多的图书”之一

奇点临近

01

The Singularity Is Near

When Humans Transcend Biology

2045年，当计算机智能超越人类

(美) Ray Kurzweil 著
李庆诚 董振华 田源 译

2005年CBS News评选的畅销书
2005年美国最畅销非小说类图书
2005年亚马逊最佳科学图书

比尔·盖茨、比尔·乔伊等鼎力推荐
一部预测人工智能和科技未来的奇书

02

机械工业出版社
China Machine Press

预言1

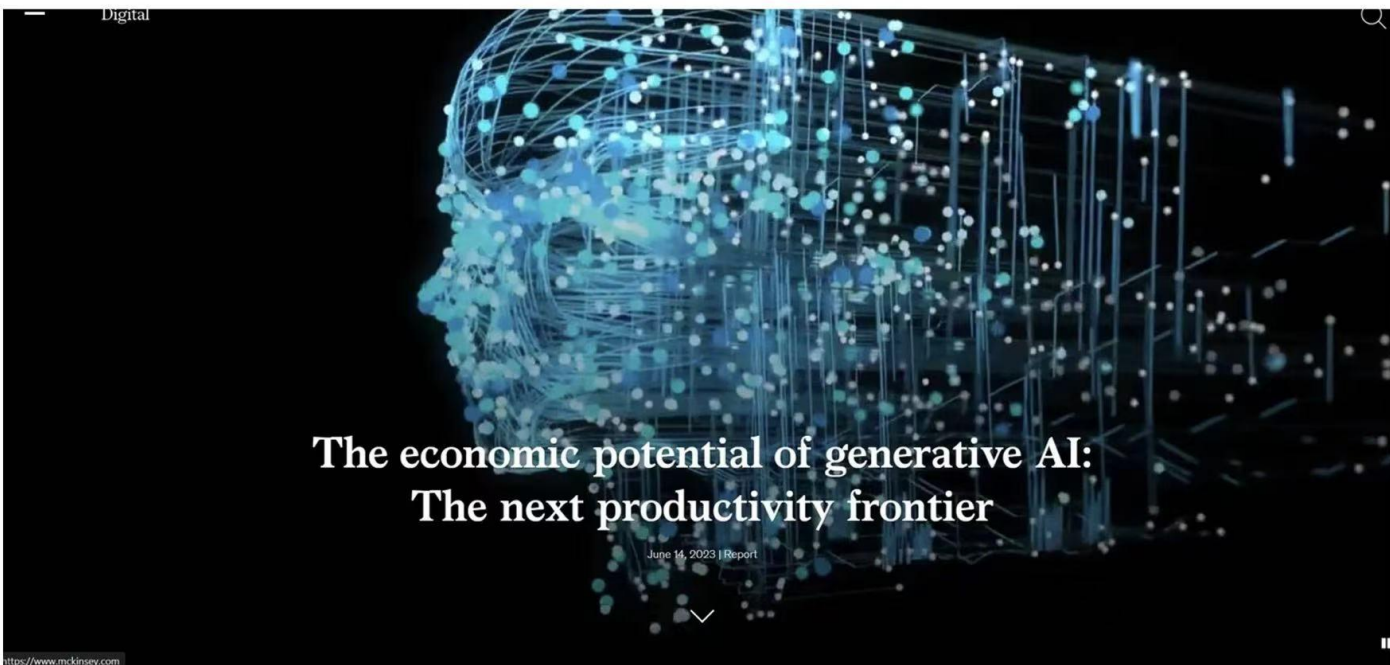
2045年，当计算机智能超越人类！

预言2

未来科技的商业宠儿：
脑机接口，具身智能。

麦肯锡：到 2045 年左右，将有 50% 工作被 AI 接管

IT之家 6 月 15 日消息，咨询机构麦肯锡的一份名为《生成式人工智能[®]的经济潜力》的报告显示，到 2045 年左右，将有 50% 工作被 AI 接管。



2024杭州

火爆全球的六大科创公司



深度求索

AI大模型

UNITREE

宇树科技
机器人机器狗

MANYCORE

群核科技
云软件及应用

GAME SCIENCE

游戏科学
黑神话悟空



强脑科技

脑机接口

DEEP Robotics

云深处科技

云深处科技

山猫机器狗



分享只是开始

期望这是一个启蒙

转动你的人生齿轮

未来已来

欢迎登船！

