



## 证券研究报告·行业简评报告

# 2025年人工智能十大发展趋势

分析师：于芳博

yufangbo@csc.com.cn

SAC编号：S1440522030001

分析师：武超则

wuchaoze@csc.com.cn

SAC 编号：s1440513090003

SFC 编号：BEM208

分析师：庞佳军

pangjiajun@csc.com.cn

SAC 编号：S1440524110001

发布日期：2025年1月27日

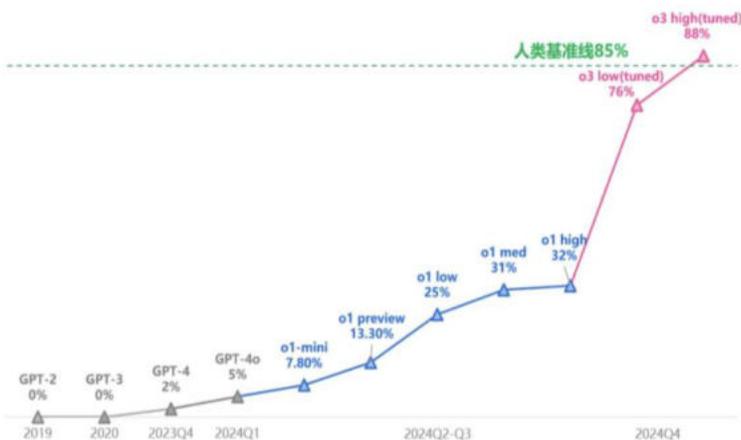
# 核心观点

- **核心观点：**大语言模型发展进入深度推理阶段，通用人工智能愈行愈近，AI应用进入爆发前夜。站在当前这一重要的历史节点，我们从技术、应用、能源三个维度展望了人工智能的未来发展，其中技术是源动力，应用是牵引力，能源是支撑力。对未来的展望中，我们提出推理计算、合成数据、缩放法则、超级智能体、具身智能、AI4Science、端侧创新、自动驾驶、人工智能+、能源需求十个方面的重要发展趋势。
- **技术是原动力。**OpenAI发布具有深度思考能力的o1推理模型，标志大语言模型发展进入深度推理阶段，推理侧缩放法则同样存在，大模型算力需求侧逐步迁移至推理侧，推理计算提升大模型准确率，强化学习激发模型推理能力；伴随文本模型的日益成熟，高质量数据更为稀缺，合成数据价值显现，其与大语言模型推理有望产生新的化学反应；缩放法则从文本为主的大语言模型训练迁移到更加广阔的人工智能领域，同时o3与GPT5循环驱动有望开启。
- **应用是牵引力。**AI Agent即将进入能力快速跃升阶段，头部的人工智能企业和互联网公司均在端侧AI Agent方面有所布局，超级智能体将走向普及；具身智能不断突破，人形机器人进入量产元年，机器人相继进入工厂实训，加速智能制造落地进展；人工智能极大加速科学研究进度，应用可见于所有STEM领域，AI4Science已经进入黄金时代；随着AI大模型逐步成熟，几乎所有硬件产品都可以加入AI元素来提升表现能力，端侧创新将不断涌现，AI塑造端侧新分工新格局；自动驾驶算法进入端到端驾驶算法发展阶段，大语言模型和视觉语言模型（VLM）逐步与端到端融合，进一步增强环境理解能力，Robotaxi进入商业化落地阶段；“人工智能+”全面铺开，企业数字化率先落地，AI在提升效率、精准决策、降低风险、创新服务方面均有巨大潜力。
- **能源是支撑力。**推理端算力需求大幅增长，基于云的人工智能系统需要消耗更多的能源，可持续发展日益紧迫。
- **风险提示：**人工智能模型技术发展不及预期、数据数量与数据质量不及预期、隐私问题、伦理冲突风险、算力基础设施支持不及预期等。

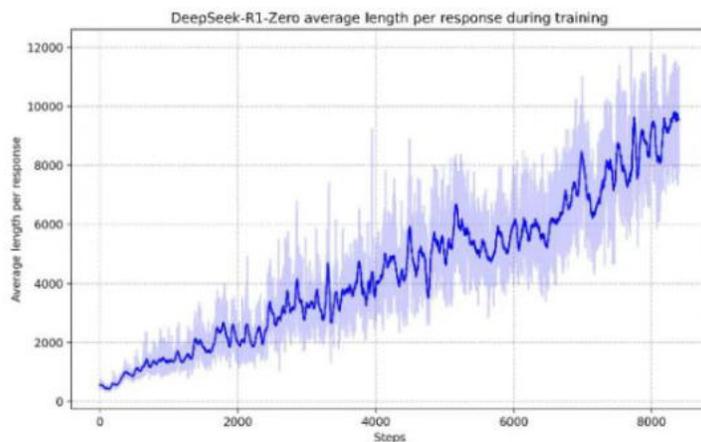
# 趋势一 推理计算提升大模型准确率，强化学习激发模型推理能力

- **OpenAI发布新的具有深度思考能力的o1推理模型，推理侧缩放法则同样存在，大模型算力需求逐步迁移至推理侧。** o1模型在复杂问题上花更多时间进行思考而非直接回应，具有改善和调整策略的能力，在科学、代码和数学等复杂问题上表现出色。思维链和思维树通过中间推理步骤帮助大模型实现了复杂的推理能力，大语言模型能够自己对推理过程的中间思维进行评估，o1呈现出类人的逻辑思维潜力。
- **强化学习激发模型推理能力，复杂思维方式——反思能力涌现。** 传统大模型需要大量人工标注的监督数据进行训练，DeepSeek-R1-Zero验证了无需任何微调数据，仅通过强化学习即可实现推理能力的自主进化。而且这是没有奖励思维链长度的情况下实现的，即目标只是为了作对题目，大模型就会自发进行更长地思考，并且最后回答效果更好。在这个过程中，自发涌现出“反思”、“多步验证”等复杂推理行为，出现问题后，模型会自动纠正早期错误，这种“智能”的出现，为未来更有智慧的Agent出现铺平道路。
- **结合蒸馏技术实现能力的高效迁移。** 使用DeepSeek-R1作为教师模型生成800K数据，并对多个小型密集模型微调，小模型性能飞跃。以通义千问的15亿参数量大模型为例，经过DeepSeek蒸馏后，数学题上超过GPT4o-0513的水平，为小算力实现特定功能奠定了基础。

图：推理侧缩放法则展现出的巨大潜力



图：DeepSeek-R1-Zero在RL过程中训练集上的平均响应长度



图：DeepSeek-R1-Zero中间版本一个有趣的“顿悟时刻”

Question: If  $a > 1$ , then the sum of the real solutions of  $\sqrt{a - \sqrt{a+x}} = x$  is equal to

Response: <think>

To solve the equation  $\sqrt{a - \sqrt{a+x}} = x$ , let's start by squaring both sides...

$$(\sqrt{a - \sqrt{a+x}})^2 = x^2 \Rightarrow a - \sqrt{a+x} = x^2$$

Rearrange to isolate the inner square root term:

$$a - x^2 = \sqrt{a+x} \Rightarrow a^2 - 2ax^2 + (x^2)^2 = a+x \Rightarrow x^4 - 2ax^2 - x + (a^2 - a) = 0$$

Wait, wait. Wait. That's an aha moment I can flag here.

Let's reevaluate this step-by-step to identify if the correct sum can be ...

We started with the equation:

$$\sqrt{a - \sqrt{a+x}} = x$$

First, let's square both sides:

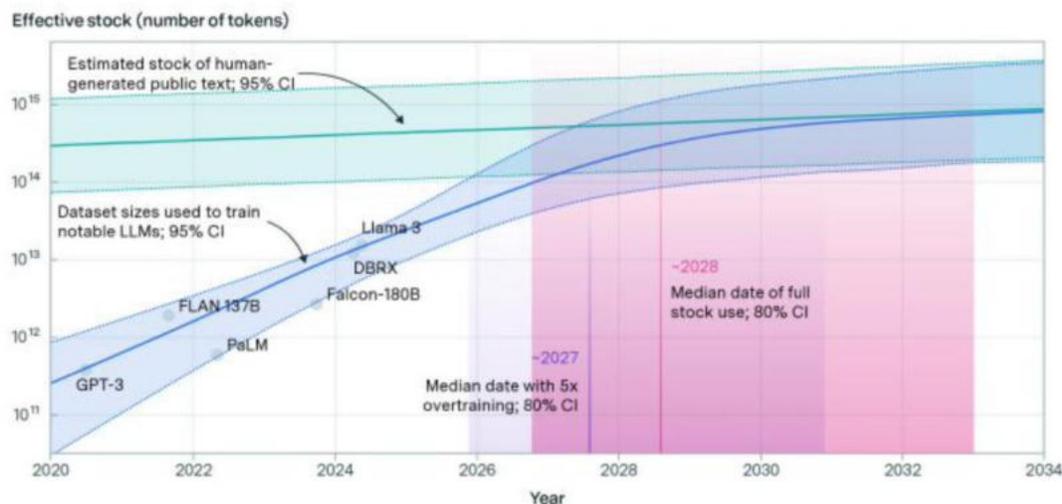
$$a - \sqrt{a+x} = x^2 \Rightarrow \sqrt{a+x} = a - x^2$$

Next, I could square both sides again, treating the equation: ...

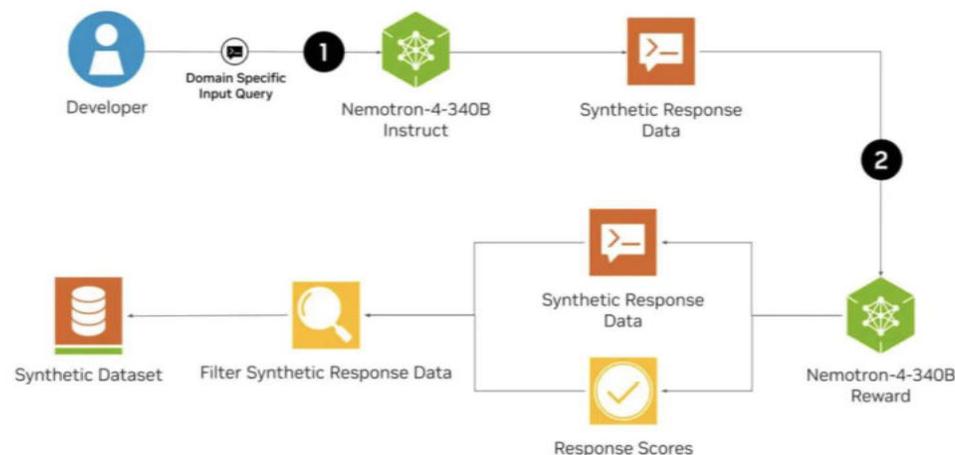
## 趋势二 高质量数据更为稀缺，合成数据价值显现

- **高质量数据的稀缺性日益加剧。** 生成式AI模型的智能表现很大程度上依赖于数据，高质量数据是构建和训练复杂大模型的基础，模型训练所需要的数据主要来源于互联网上的文字和视频。目前大语言模型需要的数据资源规模日益扩充，加速了数据资源耗尽的时间点到来，同时互联网数据的创建者提升了数据抓取的限制，让搜寻数据变得越来越困难。
- **合成数据是一种模仿真实世界数据的非人工创建的数据，是由基于生成式AI技术的计算算法和模拟创建而成。** 英伟达发布的开源模型 Nemotron-4 340B，开发者可以使用它们生成用于训练大型语言模型的合成数据，以应用于医疗、金融、制造、零售及其他各行各业的商业应用，生成的合成数据在规模较小的大语言模型训练中将扮演重要的角色。
- **合成数据与推理时计算的化学反应在DeepSeek V3、DeepSeek R1中开始显现。** 以DeepSeek为例，用DeepSeek R1模型蒸馏出来高质量CoT数据去训练DeepSeek-V3，其高质量的合成数据对V3基础模型训练有明显提升。

图：公共数据耗尽时间



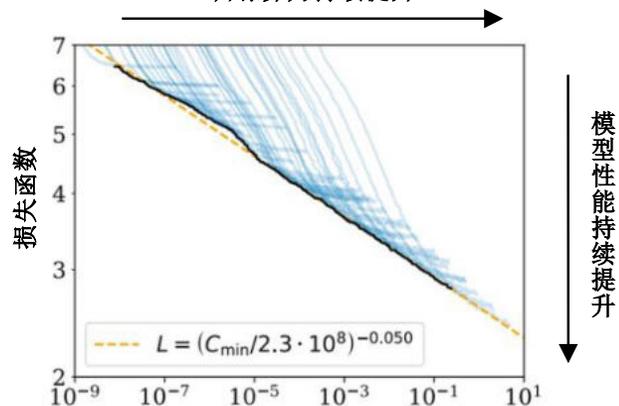
图：英伟达Nemotron-4数据合成过程



# 趋势三 缩放法则依然有效，o3与GPT5循环驱动有望开启

- **缩放法则(Scaling law):** 人工智能领域的缩放法则在2020年被首次提出，缩放法则在过去一段时间内指引大语言模型取得巨大成功：随着模型参数数量、训练数据量和计算量的增加，模型的性能会以可靠且可预测的方式提高。
- **发展趋势:** 随着大语言模型的模型尺度快速上升以及数据和算力的限制，尽管提升模型参数带来性能上升的边际收益正在逐步递减，提升模型参数规模、扩大训练语料库仍然是大语言模型提升性能的重要手段。在多模态数据中、在模型推理过程中、在生物数据中、在世界模型中，缩放法则初步展露头脚，缩放法则将从文本为主的大语言模型训练迁移到更加广阔的人工智能领域，指导人工智能模型在更多维度上的发展路径。
- **o3模型有望驱动模型和数据间的飞轮效应。** o3模型思维链过程数据是目前AI行业的稀缺资源，可以极大推动模型和数据的飞轮效应，未来o3提升GPT-5、GPT-5提升o4的循环有望开启。

图：大语言模型训练过程中的缩放法则  
训练算力持续提升

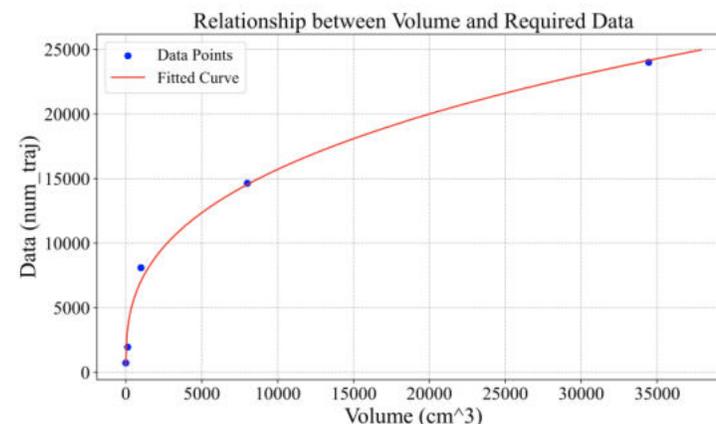


训练算力

图：缩放法则的延拓



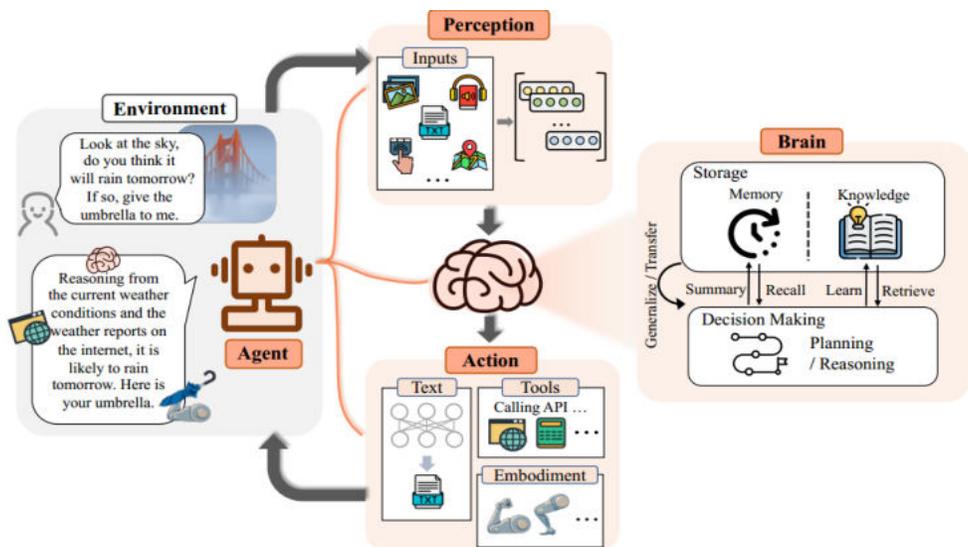
图：机器人空间泛化中发现的缩放法则



# 趋势四 应用的最佳形态——超级智能体（AI Agent）走向普及

- **AI Agent指接受周围环境的信息，做出规划和决策，并通过执行器采取行动的人工智能体。** Agent主要包括三个组成部分：大脑、感知和行动，以LLM为大脑模块的Agent展现出强大潜能，大语言模型展现的推理能力和决策能力是AI Agent的灵魂。
- **AI Agent即将进入能力快速跃升阶段。** 随着LVM(纯视觉大模型)和流式多模态模型的不成熟，大语言模型的视觉和语音不断进步，叠加近期大语言模型在推理侧“慢思考”能力方面取得的突破，意味着AI Agent已准备好进入普及阶段。
- **全球巨头争相布局端侧AI Agent。** 头部的人工智能企业和互联网公司均在端侧AI Agent方面有所布局，AI Agent的能力跃升将是关注重点。展望后续，Agent在C端有望成为AI时代的操作系统，大量营销、办公等B端AI助手向Agent过渡。

图：AI Agent框架



图：全球AI Agent进展梳理

日期	公司	备注
10月22日	微软	在B端办公产品Dynamics 365中集成10个自主AI Agent, 自动执行客服、销售、财务、仓储等集成OpenAI的o1模型
10月23日	Claude	Claude模型更新电脑操作功能, 按照用户的命令在计算机屏幕上移动光标, 点击相关位置, 并通过虚拟键盘输入信息
10月25日	智谱	发布AutoGLM, 基于语音交互, 自动完成点外卖、回复微信、写好评等多步骤任务
10月29日	微软	发布新模型, 支持网页自动化操作
12月	苹果	Siri将集成ChatGPT, 读取屏幕信息、完成复杂任务的能力有望提高
今年内	谷歌	谷歌将推出Jarvis, 能读取和输入信息, 帮助用户进行收集研究、购买产品或预订航班
预计25年	OpenAI	发布名为“Operator” AI Agent, 自动执行复杂任务, 包括编写代码、预订旅行等

## 趋势五 具身智能不断突破，人形机器人进入量产元年

- **具身智能**的核心能力：感知、学习和与环境动态交互的能力。
- **Nvidia世界模型加速机器人训练**。物理AI模型开发成本很高并且需要大量真实数据和测试，英伟达推出的Cosmos世界基础模型可以使开发者能够轻松生成大量基于物理学的逼真合成数据，以用于训练和评估其现有的模型。
- **人形机器人进入量产元年**。特斯拉在2023年12月发布第二代原型机Optimus Gen 2，Optimus而仅仅依靠2D摄像头、手部触觉和力传感器，其端到端神经网络经过训练，能够对特斯拉工厂的电池单元进行准确分装，算法在机器人的FSD计算机上实时运行。
- **人形机器人相继进入工厂实训，加速智能制造落地进展**。预计2025年有上千台机器人应用于特斯拉工厂，未来人形机器人将与工业机器人、自动化设备等协同作业，实现柔性化生产。家庭场景是最具潜力的市场之一，未来目标可以完成扫地、擦窗、洗衣等家务劳动，陪伴老人等。

图：特斯拉机器人分拣电池



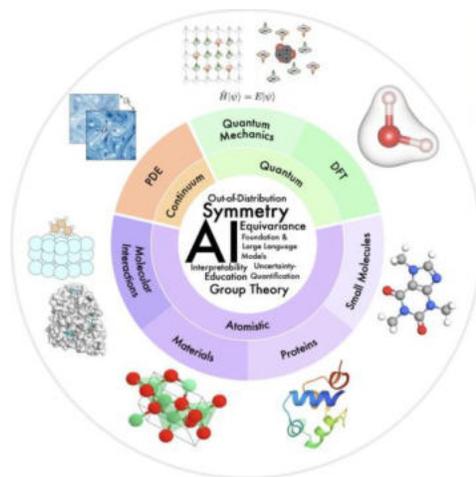
图：斯坦福家务机器人ALOHA 2



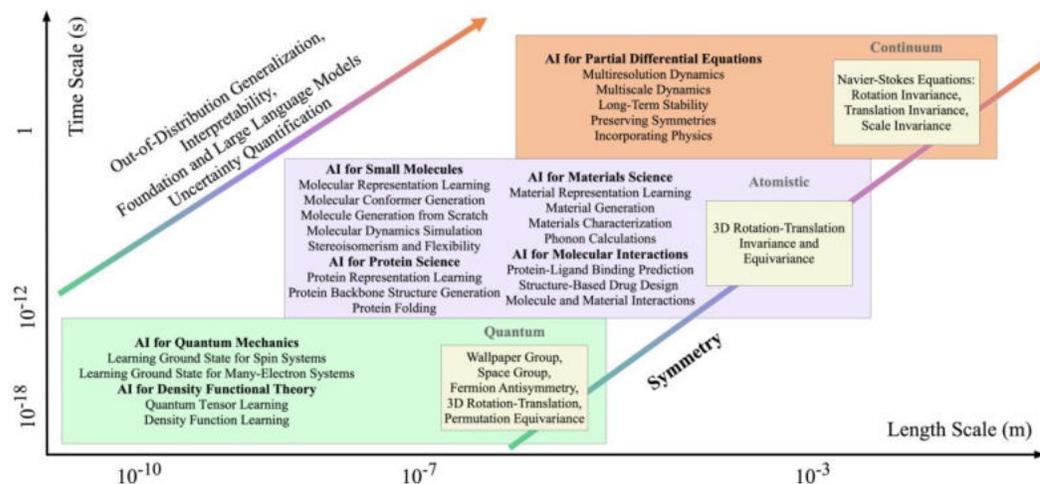
# 趋势六 AI4science：黄金时代已经到来

- **人工智能极大加速科学研究进度。** 人工智能应用可见于所有STEM领域，应用在医学、材料科学、机器人技术、农业、遗传学和计算机科学等领域，人工智能擅长模拟复杂系统及研究组成部分之间的相互作用，可以在庞大的问题解决方案中寻找最优解，这些特点使得人工智能在蛋白质结构预测、小分子药物研发、材料结构预测、数学问题证明等研究过程中具备巨大的优势。
- **诺贝尔奖：** 2024年诺贝尔物理学奖给了美国科学家约翰·霍普菲尔德、英国裔加拿大科学家杰弗里·辛顿，以表彰他们通过“**人工神经网络实现机器学习的基础性发现和发明**”。2024年诺贝尔化学奖授予德米斯·哈萨比斯和约翰·M·詹珀，以表彰他们对使用AI模型AlphaFold在蛋白质结构预测方面的贡献。
- 展望未来，特定领域短期内就能从AI中受益，比如自主实验室。同时，AI智能能够提升最先进技术水平，比如研究在实验中无法观测到的生物、化学或物理过程的长度和时间尺度相关的研究。

图： AI4Science的广泛应用领域



图： 不同空间尺度上的AI应用



# 趋势七 端侧创新将不断涌现，AI塑造端侧新分工新格局

- 随着AI大模型逐步成熟，几乎所有硬件产品都可以加入AI元素来提升表现能力。“AI+硬件”模式在办公、娱乐、教育等领域百花齐放，未来有望在更多场景以更多品类形式进行应用。2024 CES展会的关注焦点是AI PC和AI手机，2025 CES展会能看到端侧AI的泛化，例如“百镜大战”的预演，三星喊出“All for AI”的口号，在手机、电视、耳机、戒指、手表等产品中均融入AI。
- 硬件产品功能性较强的单品逐步扩散到全场景、全品类、全价格段的多品类铺量模式。各品类SKU的需求量快速增加，同时产品与模型结合的创新在不断涌现，有望催生新一轮产业链分工，同时拉动上游SoC、PCB、新型封装、存储等产品，AI塑造产业的新格局。

图：“AI+硬件”已全方位覆盖到办公、娱乐、教育等领域



AI+耳机



AI+戒指



AI+玩具



AI+桌面机器人



AI+眼镜



AI+手机/PC/平板

# 趋势八 自动驾驶迈向端到端，Robotaxi进入商业化落地阶段

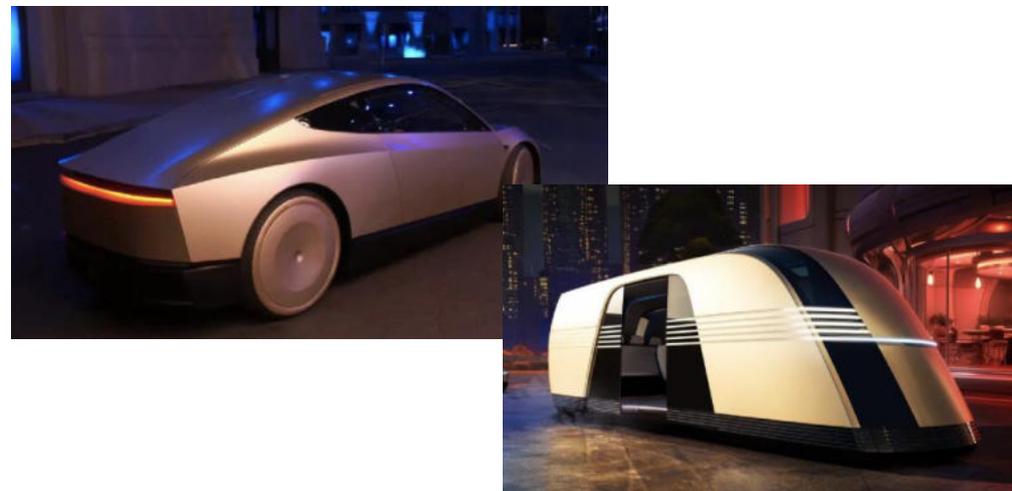
- **自动驾驶算法经过多年不断演进路线，目前进入端到端自动驾驶算法发展阶段。**端到端自动驾驶算法采用感知决策一体化方案:即设计一个算法模型，直接输入传感器感知的信息，输出控制结果。自2023年以来智驾行业掀起的BEV、端到端技术浪潮后，智驾正逐步将AI神经网络融入感知、规划、控制等环节。大语言模型和视觉语言模型（VLM）逐步与端到端融合，进一步增强环境理解能力。远期来看VLA模型将端到端、VLM两个模型合二为一，多模态大模型将成为端到端自生的一种能力。
- **智能驾驶商业化进程不断加速。**国内Robotaxi相关政策不断完善，逐渐建立起Robotaxi从准许上路到全无人商业运营、包含责任承担等细则在内的政策体系。智能驾驶算法不断向上迭代的同时，配有高阶智能驾驶产品的车型起售价在不断下降。

图：VLA模型及“端到端+VLM”技术框架



资料来源：元戎启行，中信建投

图：特斯拉robotaxi

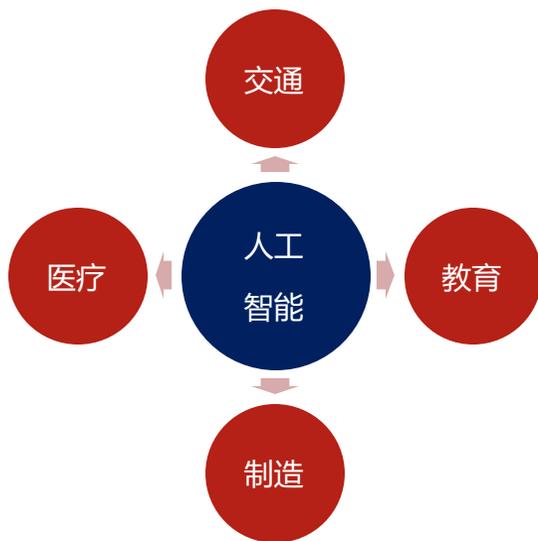


资料来源：特斯拉，中信建投

# 趋势九 “人工智能+” 全面铺开，企业数字化率先落地

- “人工智能+”旨在将人工智能作为一种基础性、驱动性的技术力量，与制造、医疗、教育、交通、农业等多个领域进行深度融合，创造出新的产品、服务和商业模式，从而推动传统行业的转型升级和社会经济结构的变革。
- **企业数字化率先落地。**对于提升效率，人工智能通过自动化流程与智能预测，使企业能够在短时间内实现更高的生产力；在决策方面，通过数据驱动的决策支持系统，人工智能帮助管理者做出更加精准基于数据分析的信息，从而降低风险，提高竞争优势；创新服务也是AI潜力巨大的领域，通过个性化推荐、24小时客服等方式，极大地提升了客户体验与满意度。

图：人工智能+



资料来源：中信建投

图：企业数字化转型核心价值



资料来源：亿欧智库，中信建投

# 趋势十 AI对能源需求大幅增长，可持续发展日益紧迫

- **推理端算力需求大幅增长。**随着基础模型的日益成熟，推理侧算力需求开始大幅增长。以字节为例，日均tokens调用量从1200亿增长到4万亿，所需要的算力大幅增长。
- **随着人类进一步训练更大参数数量的模型，以及越来越多的AI应用开始部署，基于云的人工智能系统需要消耗更多的能源。**英伟达单卡功耗从A100的400W到B200的1200W，功耗越来越高，并且服务器也开始迈向机柜式方案设计，对供电系统提出了更大的挑战。同时，为了更高效、更稳定、更环保，全球新一轮AIDC开始启动建设。

图：豆包日均tokens数据



资料来源：火山引擎，中信建投

图：主流服务器对应配置

服务器型号	DGX A100	DGX H100	HGX H20	NVL 72
GPU类型	A100	H100	H20	B200
GPU个数	8	8	8	72
GPU功耗	8*400=3200KW	8*700=5600KW	8*400=3200KW	72*1000=72KW
其他部件功耗	3.3KW	4.6KW	—	48KW
服务器整机功耗	6.5KW	10.2KW	<8.1KW	120KW
电源配置方案（国内）	2.2KW (3+3)	3.3KW (3+3)	2.7KW (3+3)	33KW (3+3)
电源配置方案（国外）	2.2KW (3+1)	3.3KW (3+1)	2.7KW (3+1)	33KW (3+1)
能源转化率	94%铂金	96%钛金	96%钛金	96%钛金

资料来源：英伟达官网，SemiAnalysis，中信建投

# 人工智能三大推动力

---



# 2025年人工智能十大趋势

## 技术是原动力

**趋势一 推理计算**提升大模型准确率，**强化学习**激发模型推理能力

**趋势二 高质量数据**更为稀缺，**合成数据**价值显现

**趋势三 缩放法则**依然有效，o3与GPT5循环驱动有望开启

## 应用是牵引力

**趋势四 应用的最佳形态**——**超级智能体**（AI Agent）走向普及

**趋势五 具身智能**不断突破，**人形机器人**进入量产元年

**趋势六 AI4science**：黄金时代已经到来

**趋势七 端侧创新**将不断涌现，AI塑造端侧新分工新格局

**趋势八 自动驾驶**迈向端到端，Robotaxi进入商业化落地阶段

**趋势九 “人工智能+”**全面铺开，企业数字化率先落地

## 能源是支撑力

**趋势十 AI对能源需求**大幅增长，可持续发展日益紧迫

# 风险提示

- **人工智能模型技术发展不及预期：**人工智能模型属于先进AI算法，若后续算法更新迭代效果不及预期，则会影响人工智能模型演进及拓展，进而会影响其商业化落地等；
- **数据数量与数据质量不及预期：**人工智能模型需要大量的高质量数据进行训练，若数据数量和质量存在短板，则会影响模型效果；
- **隐私问题：**人工智能在进行预训练过程中，使用了大量数据，其中会涉及相关隐私问题；
- **伦理冲突风险：**技术进步推动人工智能迈向通用型人工智能，可能将产生人工智能欺骗、人工智能上瘾等一系列伦理问题；
- **算力基础设施支持不及预期：**需要关注算力芯片产能不及预期风险；
- **政策监管力度不及预期：**大语言模型带来新的商业生态，尚属于前期成长阶段，政策监管难度加大，相关法律法规尚不完善，政策监管力度可能不及预期；
- **应用不及预期：**人工智能算法与相关应用落地之间还存在一定距离，需要注意应用落地不及预期风险。

## 分析师介绍

**武超则：**中信建投证券研究所所长兼国际业务部负责人，董事总经理，TMT行业首席分析师。新财富白金分析师，2013-2020年连续八届新财富最佳分析师通信行业第一名；2014-2020年连续七届水晶球最佳分析师通信行业第一名。专注于5G、云计算、物联网等领域研究。中国证券业协会证券分析师、投资顾问与首席经济学家委员会委员。

**庞佳军：**电子行业联席首席分析师、人工智能组联席首席分析师，东南大学硕士，7年半导体行业经验，曾在Marvell、Nvidia、平头哥半导体、乐鑫科技等公司从事芯片研发和管理，2022年加入中信建投电子团队，专注研究CPU、GPU、EDA等领域。

**于芳博：**中信建投人工智能组首席分析师，北京大学空间物理学学士、硕士，2019年7月加入中信建投，主要覆盖人工智能等方向，下游重点包括智能汽车、CPU/GPU/FPGA/ASIC、EDA和工业软件等方向。

## 评级说明

投资评级标准		评级	说明
报告中投资建议涉及的评级标准为报告发布日后6个月内的相对市场表现，也即报告发布日后的6个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数作为基准；新三板市场以三板成指为基准；香港市场以恒生指数作为基准；美国市场以标普500指数为基准。	股票评级	买入	相对涨幅15%以上
		增持	相对涨幅5%—15%
		中性	相对涨幅-5%—5%之间
		减持	相对跌幅5%—15%
	行业评级	卖出	相对跌幅15%以上
		强于大市	相对涨幅10%以上
		中性	相对涨幅-10-10%之间
		弱于大市	相对跌幅10%以上

## 分析师声明

本报告署名分析师在此声明：（i）以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，结论不受任何第三方的授意或影响。（ii）本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

## 法律主体说明

本报告由中信建投证券股份有限公司及/或其附属机构（以下合称“中信建投”）制作，由中信建投证券股份有限公司在中华人民共和国（仅为本报告目的，不包括香港、澳门、台湾）提供。中信建投证券股份有限公司具有中国证监会许可的投资咨询业务资格，本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格证书编号已披露在报告首页。

在遵守适用的法律法规情况下，本报告亦可能由中信建投（国际）证券有限公司在香港提供。本报告作者所持香港证监会牌照的中央编号已披露在报告首页。

## 一般性声明

本报告由中信建投制作。发送本报告不构成任何合同或承诺的基础，不因接收者收到本报告而视其为中信建投客户。

本报告的信息均来源于中信建投认为可靠的公开资料，但中信建投对这些信息的准确性及完整性不作任何保证。本报告所载观点、评估和预测仅反映本报告出具日该分析师的判断，该等观点、评估和预测可能在不发出通知的情况下有所变更，亦有可能因使用不同假设和标准或者采用不同分析方法而与中信建投其他部门、人员口头或书面表达的意见不同或相反。本报告所引证券或其他金融工具的过往业绩不代表其未来表现。报告中所含任何具有预测性质的内容皆基于相应的假设条件，而任何假设条件都可能随时发生变化并影响实际投资收益。中信建投不承诺、不保证本报告所含具有预测性质的内容必然得以实现。

本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况，报告接收者应当独立评估本报告所含信息，基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。中信建投建议所有投资者应就任何潜在投资向其税务、会计或法律顾问咨询。不论报告接收者是否根据本报告做出投资决策，中信建投都不对该等投资决策提供任何形式的担保，亦不以任何形式分享投资收益或者分担投资损失。中信建投不对使用本报告所产生的任何直接或间接损失承担责任。

在法律法规及监管规定允许的范围内，中信建投可能持有并交易本报告中提公司的股份或其他财产权益，也可能在过去12个月、目前或者将来为本报告中所提公司提供或者争取为其提供投资银行、做市交易、财务顾问或其他金融服务。本报告内容真实、准确、完整地反映了署名分析师的观点，分析师的薪酬无论过去、现在或未来都不会直接或间接与其所撰写报告中的具体观点相联系，分析师亦不会因撰写本报告而获取不当利益。

本报告为中信建投所有。未经中信建投事先书面许可，任何机构和/或个人不得以任何形式转发、翻版、复制、发布或引用本报告全部或部分内容，亦不得从未经中信建投书面授权的任何机构、个人或其运营的媒体平台接收、翻版、复制或引用本报告全部或部分内容。版权所有，违者必究。

### 中信建投证券研究发展部

北京  
朝阳区景辉街16号院1号楼18层

电话：(8610) 8513-0588  
联系人：李祉瑶  
邮箱：lizhiyao@csc.com.cn

上海  
浦东新区浦东南路528号南塔2103室

电话：(8621) 6882-1600  
联系人：翁起帆  
邮箱：wengqifan@csc.com.cn

深圳  
福田区福中三路与鹏程一路交汇处广电金融中心  
35楼

电话：(86755) 8252-1369  
联系人：曹莹  
邮箱：caoying@csc.com.cn

### 中信建投（国际）

香港  
中环交易广场2期18楼

电话：(852) 3465-5600  
联系人：刘泓麟  
邮箱：charlene.liu@csci.hk