

2025

DeepSeek开启AI算法变革元年

出品机构：甲子光年智库

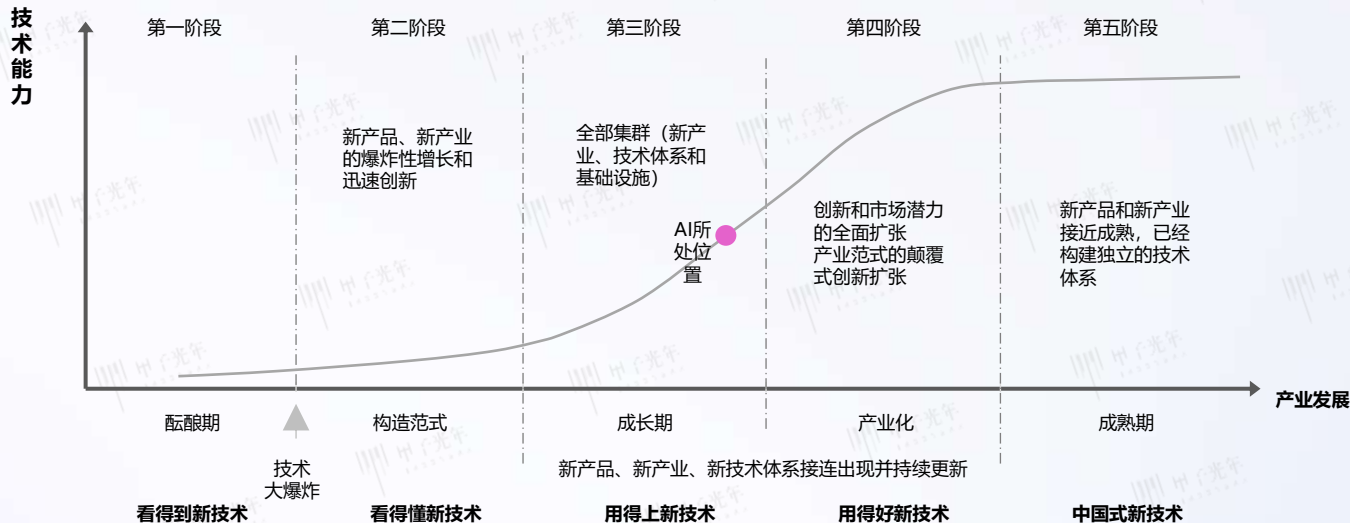
研究团队：宋涛、刘瑶、翟惠宇、何伟康、小麦

发布时间：2025.02

1、人工智能的发展已到战略拐点

- 经过近七十年年的发展，人工智能自出现至今其核心能力一直在提升，到如今已经成为覆盖感知、认知、决策、学习、执行、社会协作能力，未来还会向符合人类情感、伦理与道德观念的智能机器迈进。
- 发展到现在，人工智能的技术能力已经到了一个拐点，即从技术能力驱动向需求应用驱动转型的关键时期。随着技术能力的提升，技术进化引发经济进化，从旧范式抵达新范式。如今人工智能的发展已经处于从第三阶段向第四阶段过渡的时期。人工智能的能力成长历程已经足够长，到了寻找需求的战略拐点。

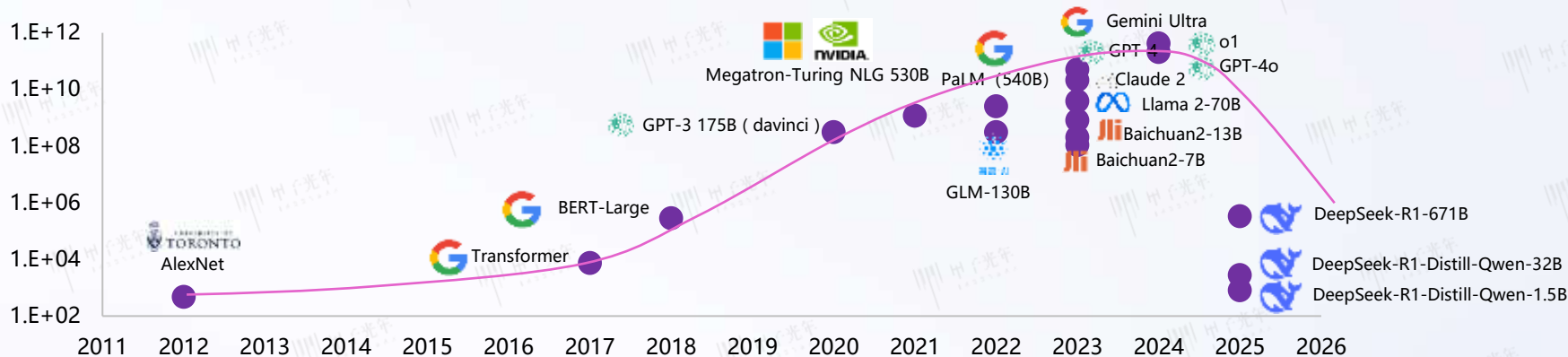
图1：技术革命的周期性示意图



2、算力拐点：DeepSeek的出现，意味着算力效率拐点显现

DeepSeek通过优化算法架构，显著提升了算力利用效率，打破了算力至上的传统认知

训练算力 (petaFLOPs)



2、算力拐点： DeepSeek的出现，意味着算力效率拐点显现

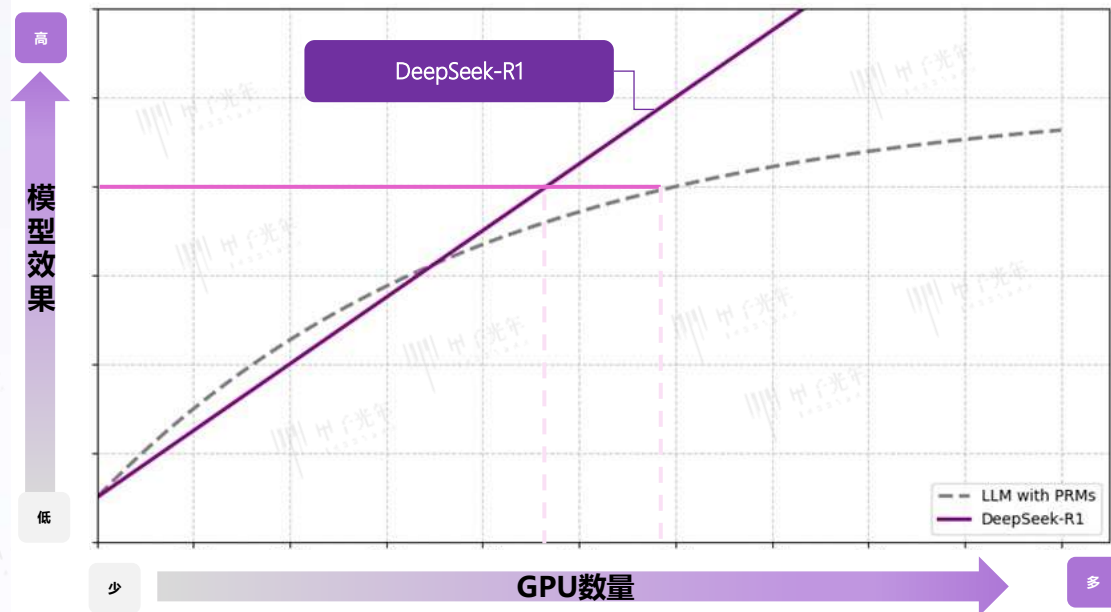
DeepSeek通过重新设计AI Infra，验证“性价比”训练路径

DeepSeek-R1的惊艳之处是通过重新设计训练流程、以少量SFT数据+多轮强化学习的办法，在提高了模型准确性的同时，也显著降低了内存占用和计算开销。

DeepSeek-R1提供了一种低成本训练的方法，而不是说只能通过低成本来进行训练。

反而通过优化，DeepSeek-R1 可能实现了算力与性能的近似线性关系。每增加一张 GPU，模型推理能力可稳定提升，无需依赖复杂的外部监督机制。这一特性直接验证了“算力即性能”的 Scaling Law，为AI芯片市场及AI Infra市场提供了明确的增量需求逻辑。

DeepSeek-R1的效果示意：更少的GPU（或其他AI芯片）带来同样的效果



备注：该图仅为示意，不代表实际情况具体数量与效果的对比

*该图为无编码基础分析师通过DeepSeek R1协作利用python编码生成

3、数据拐点：AI基础大模型的参数量迎来拐点

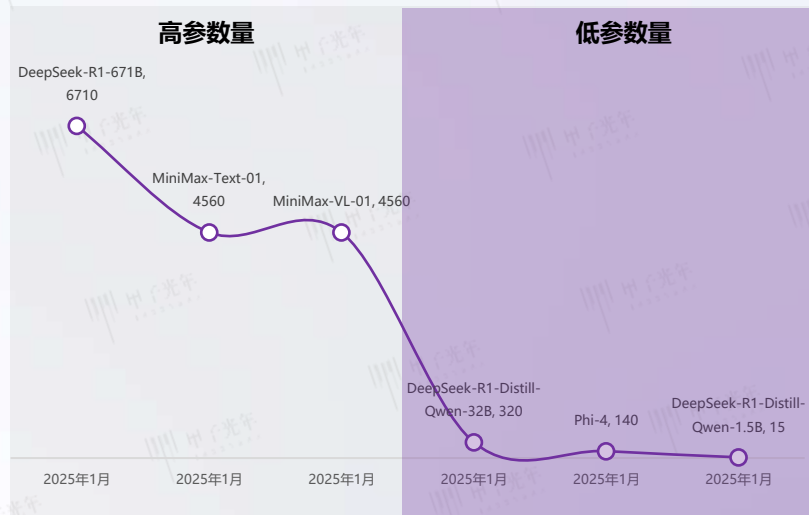
2025年发布的大模型，都具有低参数量的特征，为本地化部署到AI终端运行提供了可能

2024年之前，最大参数量一直在增长

图：AI预训练模型的参数规模呈现走势



2025年发布的大模型开始分化为两极

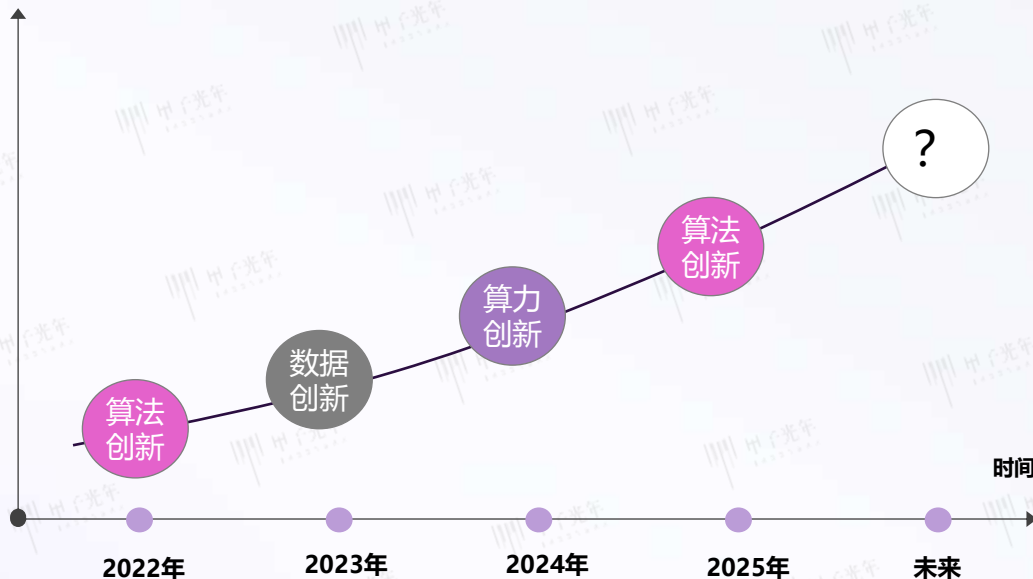


4、技术路径的循环：算法创新再次成为新的突破点

AI技术创新一直在围绕核心三要素在动态循环，2025年再次进入算法创新阶段

图：每年AI技术创新发展的方向围绕三要素迭代展开

AI技术成熟度



核心三要素的创新循环

算力、数据、算法的三角创新体系，在动态循环之中再次进入算法创新阶段。

- 2022年：算法创新为主，ChatGPT发布，引发Transformer架构的风潮迭起
- 2023年：数据创新为主，数据合成、数据标注等成为高质量数据集建设的热点方向
- 2024年：算力创新为主，算力迈向超万卡时代，算力运营商等产业新物种诞生
- 2025年：再次进入算法创新阶段

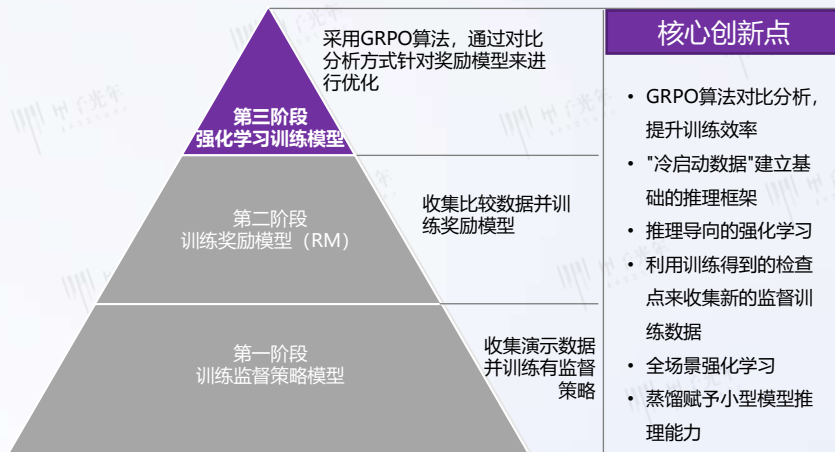
5、算法变革元年：DeepSeek的推理模型开启算法变革的元年

R1-zero完全基于RL（强化学习）进行训练，以第三阶段为主，通过自我学习来提高性能

图1：传统AI基础大模型的训练过程三个阶段都注重



图2：DeepSeek聚焦的训练过程核心在第三阶段

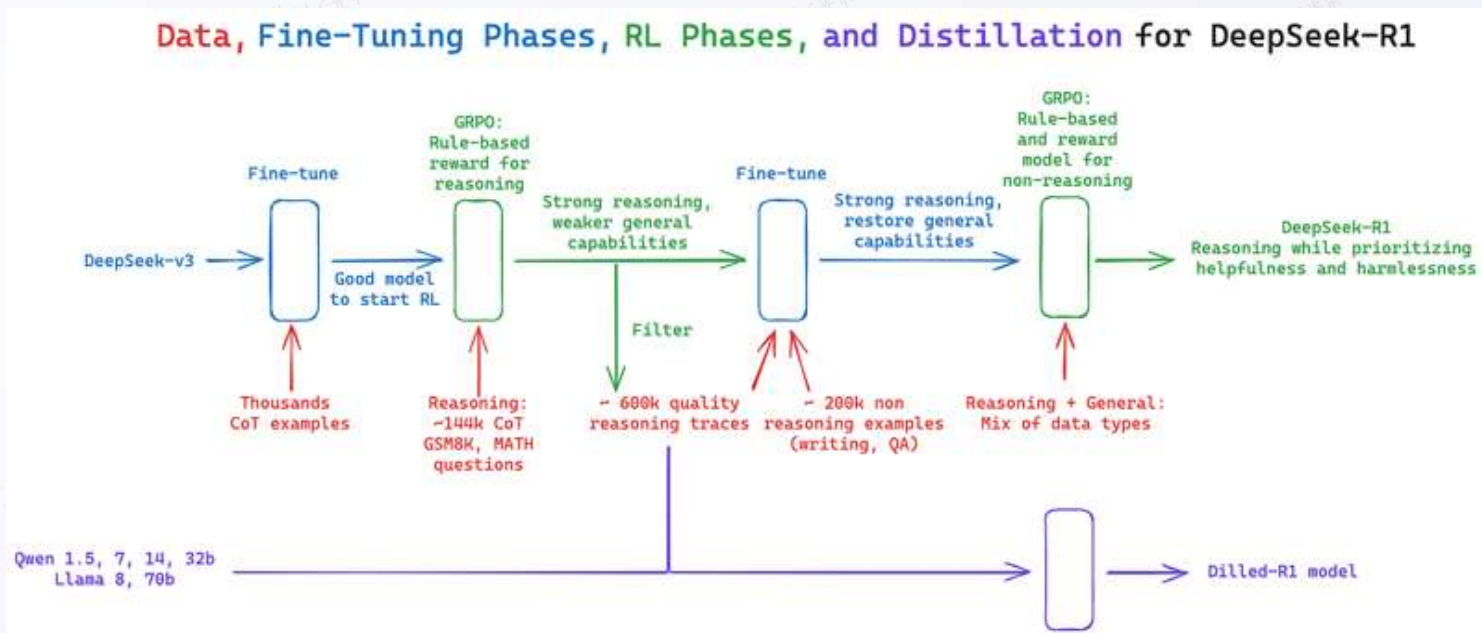


5、算法变革元年：DeepSeek的推理模型开启算法变革的元年

DeepSeek-R1的训练流程：通过RL强化LLM的推理能力

- ❑ R1-zero完全基于RL（强化学习）进行训练，未使用任何监督训练或人类反馈，能够通过自我学习来提高性能。
- ❑ R1是在 R1-zero 的基础上，通过少量冷启动数据进行微调，提高了输出质量和可读性。

DeepSeek-R1的训练过程

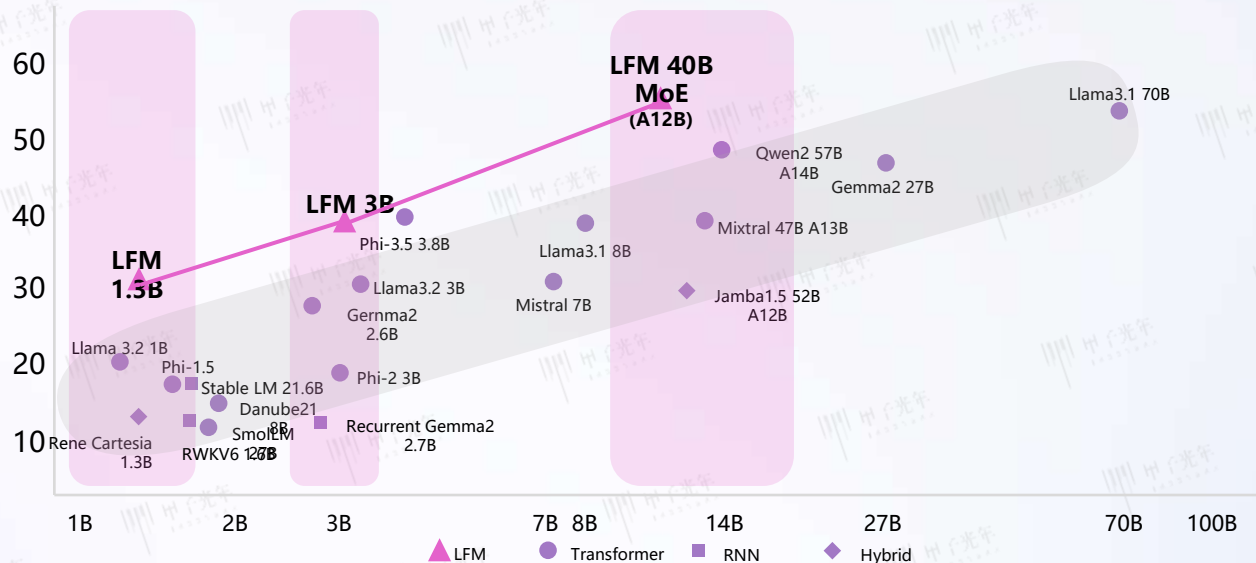


6、算法变革的第二条路径：非Transformer架构的算法模型

- 除了基于Transformer架构的算法创新之外，非Transformer架构的新算法也成为近期新出现的重点发展方向。
- 一种新架构模型LFM（Liquid Foundation Model），超越了同等规模的Transformer模型如Llama 3.2。它用的是一种液态神经网络（LNN），其计算单元植根于动态系统理论、信号处理和数值线性代数。其小巧便携的特性使得它能够直接部署在手机上进行文档和书籍等分析。
- 因此，甲子光年智库判断，2025年将是算法变革的元年。未来最适合人类使用的大模型，可能会是非Transformer架构的算法模型。

图：LFM架构模型性能超越同等规模的Transformer模型

MMLU-Pro测试分数



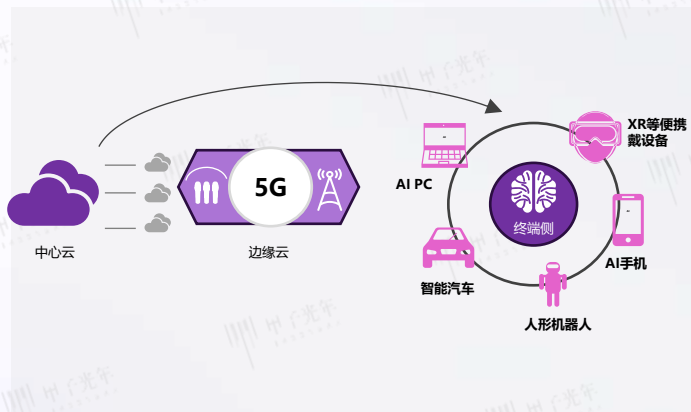
活跃参数数量

6、算法变革的第二条路径：非Transformer架构的算法模型

AI终端推动端侧模型和端侧算力需求增加，小参数量模型需求爆发推动算法变革到来

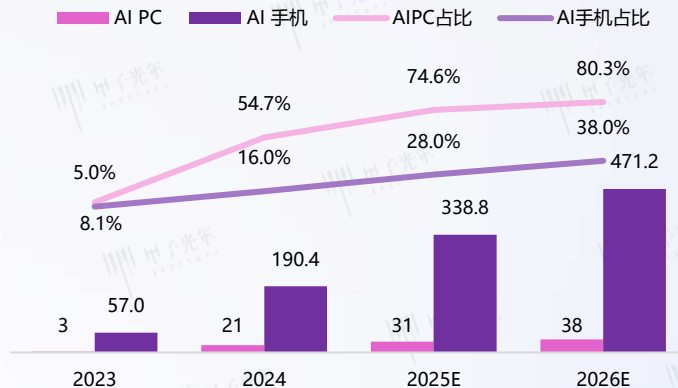
AI处理重心正向端侧转移，催生小语言模型的需求爆发

图1：AI处理重心向端侧转移



2024年AI终端已经开始爆发

图2：2023-2026年中国AIPC出货量及占比（百万台）



7、算法变革推动AI超级产品蕴生：

2024城头变幻大王旗，2025年AI超级应用开始出现，DAU达2161万

- 为2024年的AI超级应用的头把交椅处于不断变动之中。甲子光年智库统计汇总了近13个月中访问量最高的产品访问量数据，发现第一名的AI产品是处于不断变动状态的，可以用一句话来形容。那就是城头变幻大王旗。基本上每个解读的头把交椅都会发生变化。
- 2025年的开年，DeepSeek用短短二十天的时间就实现了DAU2161万的数据，超过了ChatGPT发布之处的数据表现，具有成为AI超级应用的潜力。

图：近一年中各月AI产品访问量最高的产品数据走势图



8、算法变革推动AI普惠应用

DeepSeek的模型实现AI时代的“小米模式”，实现技术和价格的双重普惠

- 如果说Transformer的应用标志着基础模型时代的开始（基础模型的庞大规模和应用范围突飞猛进），可以称为AI技术的“iPhone时刻”。那么DeepSeek从建立之初的“价格屠夫”模式就赢得了开发者市场的关注，可以看做是AI时代的“小米模式”。

2010-2016年 中国智能手机出货量（亿台）



短期影响

小米1代手机	相同点	DeepSeek R1
直接挑战了国际品牌的高价策略国际品牌（如三星、HTC）高价策略（3000-4000元）	挑战高价策略	直接挑战OpenAI的闭源技术模式及OpenAI O1的价格策略（O1正式版需要200美元）
将高性能智能手机推向大众市场	受众群体扩大，普惠开发者	让更多中国普通用户使用推理型模型，中文语言能力较海外更强，同时具有极强的代码及数学能力

长期影响

小米1代手机	相同点	DeepSeek R1
带动带动京东方（屏幕）、舜宇光学（镜头）、比亚迪电子（代工）等本土供应链企业技术升级，降低全行业制造成本	成本重构，打破资源垄断	训练效率提升与推理成本压缩，带动一系列本土AI infra建立产业生态（如硅基流动与华为云联合提供解决方案）
小米通过红米系列覆盖下沉市场，移动互联网应用服务更多样人群	推动垂直场景、长尾应用爆发	从“科技巨头试验”走向“千行百业刚需”，构建更多元的开发生态

8、算法变革推动AI普惠应用

DeepSeek的“高性价比”有望解锁具身智能的理解与推理

- DS不仅在推理模型上实现了突破，在多模态方面也保持了进步，并于近日开源发布了Janus-Pro多模态模型。
- 基于算法工程方面的优化，DS能够实现性价比更高的模型推理能力（价格低、效率高、性能强、可部署在端侧平台），
- DS有望以强大的推理能力和多模态感知能力，重塑机器人交互与决策，同时通过低成本、高效率的解决方案，加速具身智能的普及与应用。

DeepSeek的性价比&效率优势，有望加速具身智能技术发展普及

低成本和高效率对于实现规模化应用和商业化至关重要

降低训练成本

- 以较低的成本训练出高质量的模型

加速研发速度、加速迭代速度、降低开发门槛

降低推理成本

- 性能更强的中小模型蒸馏
- 更低的计算成本在真实硬件上运行

端侧平台部署、提升反应速率、降低计算能耗

DS Janus-Pro - 多模态感知理解的引擎

多模态融合感知

- 能够有效融合视觉和语言信息，使得机器人可以理解复杂的场景描述和人类指令

图像生成

- 可生成高质量的图像，对虚拟环境建模、仿真测试等具有重要意义。

编码器解耦

- 视觉编码方法解耦，对不同的任务进行优化，例如同时提升感知能力和生成能力
- 缓解多模态理解和生成任务之间的冲突

DS R1 - 推理能力的基石

自主决策

- 通过强化学习自主探索，发展出复杂的推理能力
- 实现机器人自主规划任务和解决复杂问题的能力

链式思考

- 链式思考能力：帮助机器人在执行任务时分解步骤，进行逻辑推理，从而更可靠地完成目标

通用性

- 不针对特定任务，追求通用推理能力
- 适应不同具身场景

低成本训练

- 冷启动训练减少了对大量标注数据的依赖，降低数据成本，并解决了R1-zero的可读性问题

8、算法变革推动AI普惠应用

推理成本的颠覆性降低，将会推动C端产品的大多数应用场景进入实际落地阶段

- 大模型的价格正在快速下降，而开放权重也在加速，并为开发者提供更多选择。OpenAI 的 o1 每输出 100 万令牌收费 60 美元，而 DeepSeek-R1 只需 2.19 美元。这将近 30 倍的价差，让算力成本不再是模型门槛，正在让基础模型层“平民化”。
- 推理成本的颠覆性降低，将会推动C端产品在大多数应用场景得到落地。尤其是在星云象限和星团象限之中的应用场景，将会得到全面的落地。

图：甲子星空坐标系—生成式AI在C端应用场景的商业潜力评估



备注说明：

1. 满意度：一个用来衡量个人或组织对特定产品、服务、经验或事件的满意程度的指标。当使用体验大于预期时会获得满意感。
2. NPS值：净推荐值，是评估用户口碑的一种方式。评估用户/客户向朋友、其他人推荐公司的产品或服务的可能性的指标。

9、全球AI科技竞争进入G2时代

DeepSeek将成为特朗普制定人工智能政策的重要参考，预示着AI进入G2竞争时代

- 2025年1月23日，刚就任的美国总统特朗普签署了一项行政命令，废除了拜登政府“有害的”人工智能政策，要求一个跨部门小组在6个月内为“人工智能行动计划”（Artificial Intelligence Action Plan）。DeepSeek的出现和美国政商各届的反应将成为特朗普政府制定计划的重要参考。

拜登政府

《关于安全、可靠和可信地开发和人工智能的行政命令》

- 2023年10月，拜登总统签署了《关于安全、可靠和可信地开发和人工智能的行政命令》，该行政命令的目的是确保AI技术的发展既能促进创新，又能有效地应对技术带来的潜在风险。该行政令也是美国迄今为止最全面的人工智能监管原则，标志着美国政府在定义AI监管和问责制方面的关键一步。
- 主要内容和目标：加强AI安全性和可靠性；保护隐私和数据安全；消除AI中的偏见和歧视；促进AI的可解释性和透明度；确保AI技术的创新和竞争力；跨部门合作与国际合作；AI的国家安全考量；监管和监督机制；社会影响评估。
- 主要特点：该行政令要求AI开发者在公开发布任何可能引发风险的AI系统之前，必须根据《国防生产法》向政府提交安全测试报告。意在通过政府主导的方式，全面规范人工智能的开发与应用。此外，该行政令明确提出了推动AI安全性和透明性的重要措施。规定了各机构制定测试人工智能系统的标准和指标、人工智能模型的安全测试、责任追究机制、算法透明度要求，并纳入了限制偏见和歧视的条款，体现对隐私和公平的保护的强监管立场。

特朗普政府在人工智能领域的计划和对华反制措施（待定）

- 调动企业力量发展人工智能：相比拜登政府坚持通过政府干预进行对人工智能监管，特朗普政府更倾向于依靠企业和资本的力量发展人工智能，特朗普会通过发动和建立“企业家朋友圈”，搭建起从基础层、中间层到应用层的“四梁八柱”。
- 限制对美国人工智能大模型使用：美国政府等相关部门高度怀疑中国大模型“窃取”了美国大模型企业的技术和训练成果，
- 扩大对芯片的出口管制：一方面，扩大对英伟达芯片对华出口限制，英伟达对中国的部分“特供型号”有望加入限制名单；另一方面，加强对新加坡等转运国和部分芯片贸易公司的调查和限制。
- 限制中国大模型应用：来自中国的大模型应用将被加入限制名单，有望被从美国及其盟国的app store下架，目前，DeepSeek已经被爱尔兰、意大利等美国的追随国下架，未来可能进一步扩大。