计算机

发布时间：2025-03-26

证券研究报告 / 行业深度报告

中国 AI 崛起：技术突破与应用落地

优于大势

上次评级:优于大势

报告摘要：

中美 AI 差距开始缩小，中国迎来 AI 发展大时代。在当今数字化浪潮席卷全球的时代，人工智能（AI）已然成为衡量一个国家科技实力与未来竞争力的关键领域。长久以来，美国凭借其在技术、人才、资金等多方面的先发优势，在全球 AI 版图中占据着举足轻重的地位，而中国虽起步稍晚，却凭借着庞大且复杂的数据资源、强大的制造业基础以及对科技创新的高度重视，一路奋起直追。如今，一系列关键指标显示，中美在 AI 领域的差距正悄然缩小，中国正迎来属于自己的 AI 发展大时代。这一转变背后，是无数科研人员的日夜钻研、是政策的有力扶持、是产业生态的逐步完善，更是一个大国在科技赛道上加速奔跑的坚定决心。

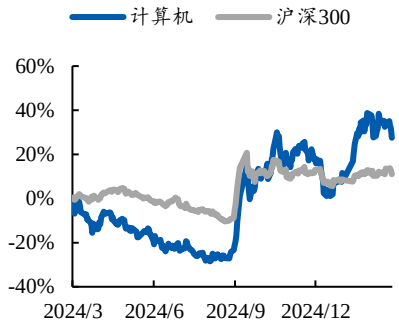
从模型角度而言，中国已经有多个模型在工程能力上各有特点。（1）DeepSeek：高效性与性价比。期采用 MLA 多头潜在注意力机制（降低 KV Cache 消耗）、FP8 混合精度训练（减少内存占用）、MoE 稀疏激活架构（671B 参数模型单 Token 仅激活 37B 参数）。同时通过预填充与解码分离部署架构，H800 节点实现 14.8k tokens/s 输出吞吐，成本仅为同类模型的 1/10。**（2）Qwen2.5：**多模态能力领先。Qwen 重新设计 ViT 架构，引入窗口注意力机制和二维 RoPE，支持原生动态分辨率处理。其动态帧率训练与绝对时间编码技术，视频任务性能接近 GPT-4o，小模型 QwQ-32B 以 1/10 成本达到 DeepSeek-R1 80%性能。（3）**混元 T1：**混合 Mamba 架构创新。长文本处理：Hybrid-Mamba-Transformer 融合模块，长文推理速度提升 2 倍，解决上下文丢失问题。通过 Mamba 模块处理局部特征，Transformer 捕捉全局依赖，AIME 数学竞赛成绩达 78.2 分，效率和效果均超越 GPT-4。

在应用层面，如何规避或者利用幻觉是目前落地的难点之一。垂类客户，尤其是那些对专业性和准确性要求极高的行业用户，如医疗、金融、法律等，对 AI 的输出结果有着近乎严苛的标准，而幻觉作为大模型的特性之一短期无法消除。目前来看，除了传统技术方法降低大模型误差率以外，还有一些工程化方向可以降低、消除 AI 幻觉：比如通过规模化容错（误差成本低于效率红利）、知识增强+可信生成（稳定数据库抑制错误）等技术路径缓解。此外，很多时候学会利用 AI 幻觉生成也是一种选择，AI 幻觉宛如一面棱镜，既折射出技术的边界，也映照出超越人类想象的潜力，或许我们不必执着于“绝对正确”，而是学会与 AI 的“想象力”共舞。毕竟，最伟大的创新往往诞生于理性与狂想的交汇处。

投资建议：我们认为，目前国内 AI 在技术追赶与场景创新上正在加速突破中，通过架构优化与多模态融合推动产业升级有望进一步的挖掘新的 AI 应用场景。建议关注医疗（AI 预问诊）、教育（个性化教学）、自动驾驶（3D/4D 标注）、人力资源（AI 招聘）及创意产业（AI 生成内容）中 AI 带来的新市场增量。

风险提示：宏观经济恢复不及预期、市场竞争加剧、AI 应用商业化落地风险、技术迭代风险

历史收益率曲线



涨跌幅（%）	1M	3M	12M
绝对收益	-10%	6%	22%
相对收益	-10%	7%	11%

行业数据

成分股数量（只）	336
总市值（亿）	48,475
流通市值（亿）	41,562
市盈率（倍）	144.16
市净率（倍）	4.49
成分股总营收（亿）	11,812
成分股总净利润（亿）	317
成分股资产负债率（%）	42.32

相关报告

《华为中国合作伙伴大会+英伟达 GTC 大会，AI 行业再迎国内外催化》 --20250317

《超大订单昭示算力景气，Manus 带动应用风潮》 --20250310

《超低成本算力预示应用大爆发，重视数据赋能 G 端政务应用》 --20250303

《国资 AI+专项行动再提速，重视国资软硬件投资机会》 --20250224

《DeepSeek 部署带动算力需求提升，注重 AI 软硬机会》 --20250216

目 录

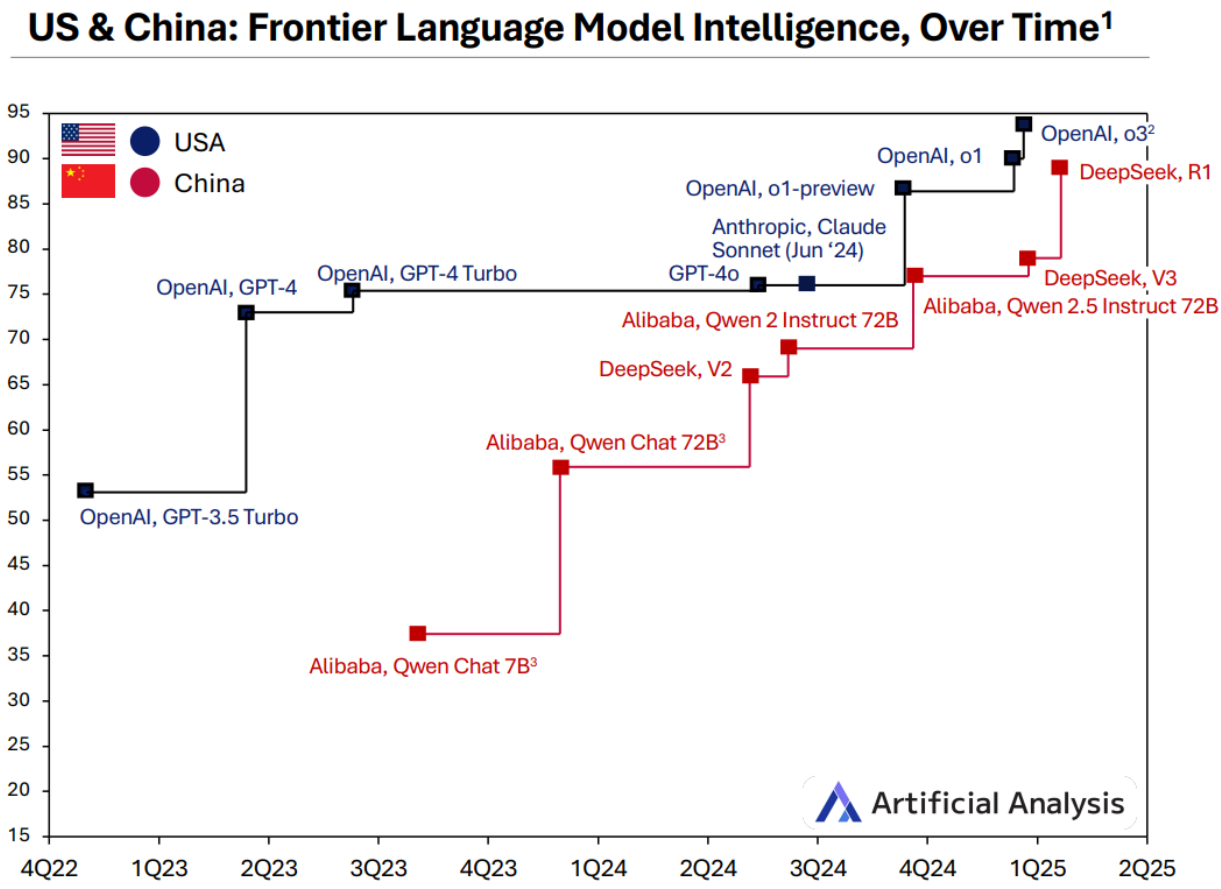
图表目录

1. AI 竞合拐点：中国智起

中美 AI 差距开始缩小，中国迎来 AI 发展大时代。在当今数字化浪潮席卷全球的时代，人工智能（AI）已然成为衡量一个国家科技实力与未来竞争力的关键领域。长久以来，美国凭借其在技术、人才、资金等多方面的先发优势，在全球 AI 版图中占据着举足轻重的地位，而中国虽起步稍晚，却凭借着庞大且复杂的数据资源、强大的制造业基础以及对科技创新的高度重视，一路奋起直追。如今，一系列关键指标显示，中美在 AI 领域的差距正悄然缩小，中国正迎来属于自己的 AI 发展大时代。这一转变背后，是无数科研人员的日夜钻研、是政策的有力扶持、是产业生态的逐步完善，更是一个大国在科技赛道上加速奔跑的坚定决心。

AI 代差缩短至半年：中国模型追上硅谷节奏。2024 年第四季度，中国 AI 实验室密集推出 7 个千亿级模型，将与美国顶尖实验室的智能差距从 24 个月压缩至 6 个月。据第三方评测，DeepSeek-R11、通义千问 3.0 等模型在多轮推理任务中已达 GPT-4 的 90%水平，首次实现"季度级"技术追平。从技术水平来看，OpenAI 去年 9 月首创的"思考链"（Chain of Thought）推理技术，在 6 个月内被中国团队全面复现。2024 年底，国内 TOP10 实验室均已部署自研推理框架，其中 DeepSeek 开源的 R12 模型，在医疗问诊场景中展现出接近 GPT-4 的逻辑连贯性，推动行业从"黑箱决策"转向可解释 AI。以 DeepSeek-Uni、阿里魔搭社区为代表的开源项目，通过开放 70%核心权重，快速聚拢 23 万开发者。这些"可拆解、可微调"的模型底座，在金融风控、县域医疗等场景的适配速度，已超过闭源的 GPT-4s——这意味着中国 AI 正从"追赶性能"转向"定义场景"的新赛道。

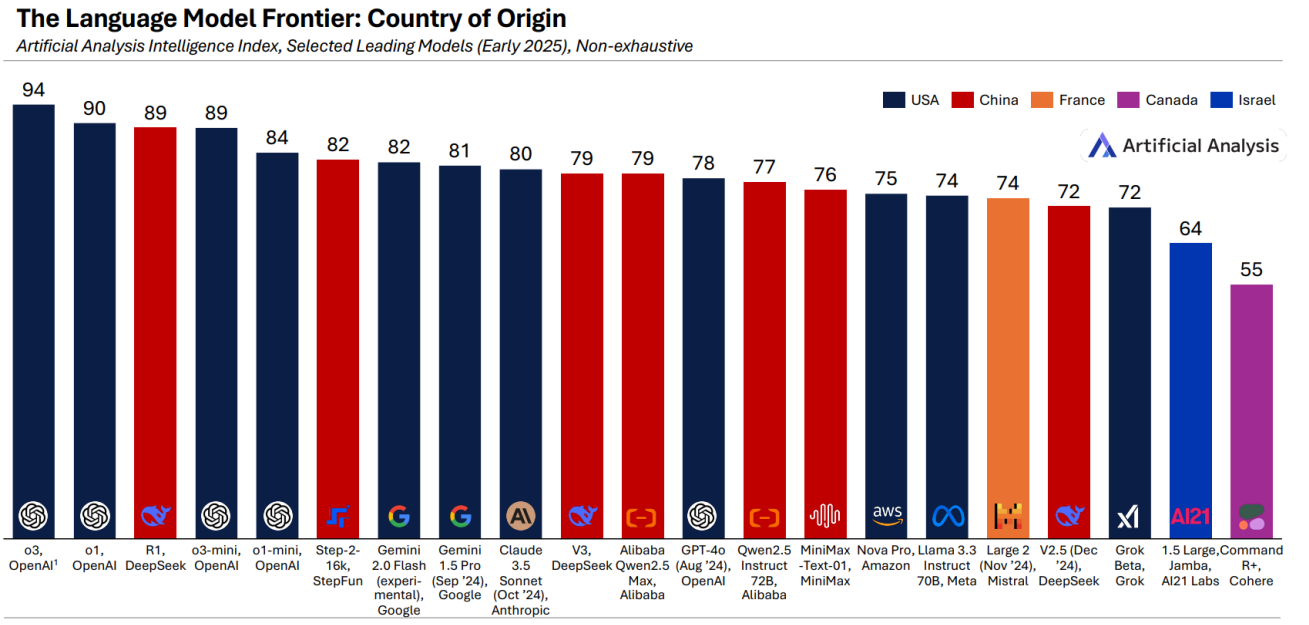
图 1：随着时间的推移，中美差距正在缩小



数据来源：Artificial Analysis

在全球人工智能领域，除中美以外其他国家和地区的 AI 发展则相对滞后。尽管欧洲、以色列等地区在 AI 领域拥有深厚的技术积累和创新能力，但整体来看，它们在过去半年中并未展现出与中美相抗衡的强劲实力，也未能在技术突破和应用拓展方面赶上美国的步伐。这表明，尽管全球 AI 竞争格局多元化，但中美两国在短期内仍将主导全球 AI 的发展趋势。

图 2：AI 的核心玩家还是来自于中美



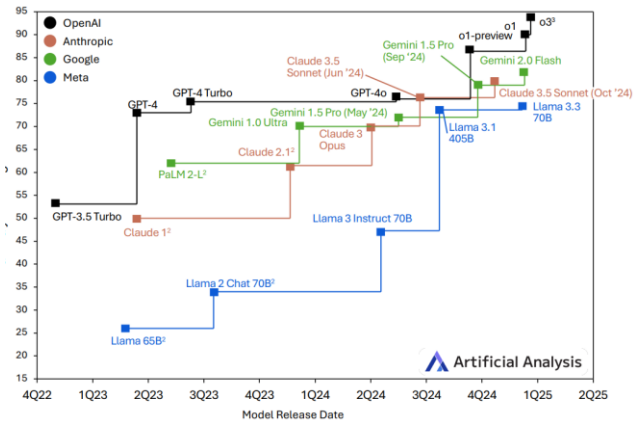
数据来源：Artificial Analysis

美国 AI 竞争格局：追赶 OpenAI 与技术突破。在美国国内，AI 竞争的核心围绕着 OpenAI 展开。自 2022 年 11 月 OpenAI 通过 ChatGPT 中的 GPT-3.5 开启语言模型竞争以来，美国的领先实验室一直在努力追赶其前沿模型。与此同时，谷歌和 Meta 等科技巨头也在迅速推进自身模型的研发，其最新推出的 Gemini 2.0 和 Flash 等模型在性能上已经超越了 Claude 3.5 Sonnet 和 GPT-4o，显示出美国 AI 领域内部的激烈竞争态势。此外，2024 年最后几个月中，除了 GPT-4 之外，其他模型也取得了重大突破，OpenAI 的 o3 模型引领了这一趋势。推理模型的优化、数据质量的提升以及新的强化学习技术等，成为推动模型性能提升的关键因素。

中国 AI 崛起：迅速追赶与前沿突破。尽管中国的 AI 实验室较晚加入全球竞争，但其发展速度令人瞩目。2024 年，中国 AI 实验室在智能水平上与美国前沿模型的差距显著缩小。当 OpenAI 推出 o1 时，中国的实验室仅用几个月时间就开发出了性能相当的模型，例如 DeepSeek 的 R1。这一成就不仅展示了中国 AI 技术的快速进步，也标志着中国在全球 AI 竞争中逐渐崭露头角。中国 AI 实验室的崛起不仅体现在追赶速度上，更在于其在前沿技术领域的布局。2024 年，中国的阿里云、深视和腾讯等实验室纷纷发布了开放权重的前沿模型，这些模型在全球范围内具有竞争力，显示出中国在 AI 技术的开放性和应用拓展方面的积极探索。

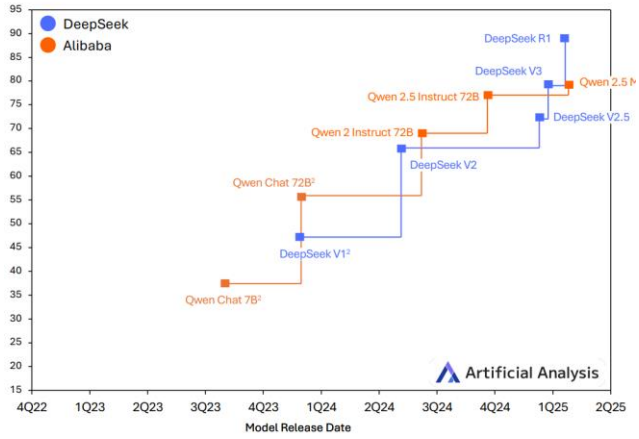
进入 2025 年初，包括阿里云、深视、明略、腾讯、智谱和通义在内的多家中国 AI 实验室陆续发布了前沿推理模型。这些模型的发布速度和频率表明，中国 AI 实验室在 2025 年已经不再是全球竞争中的追赶者，而是成为全球 AI 发展的重要参与者和潜在领导者。

图 3：美国 AI 进化图（Open AI、Google、Anthropic、Meta）



数据来源：Artificial Analysis

图 4：中国 AI 进化图（Deepseek ，阿里巴巴）



数据来源：Artificial Analysis

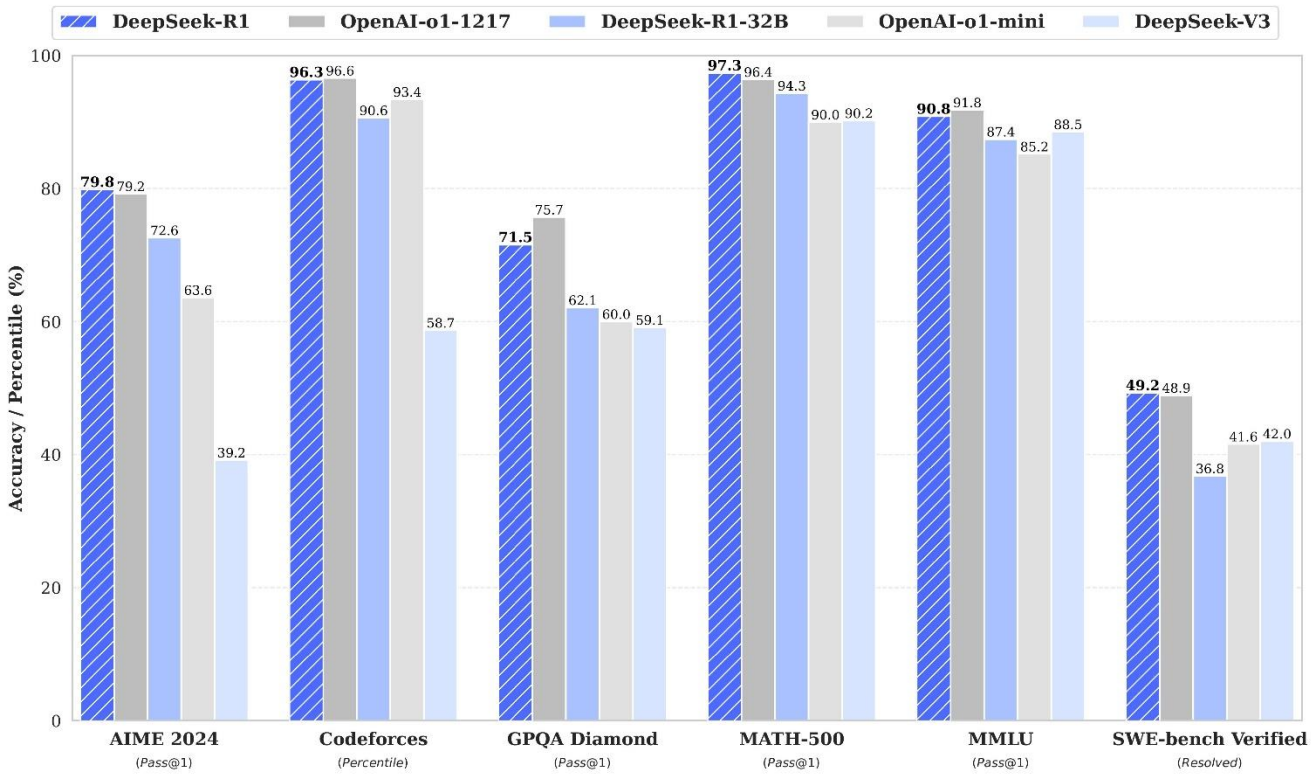
2. 中国 AI 的工程创新，从 Deepseek、Qwen 还有混元说起

2.1. DeepSeek：技术创新+高性价比

DeepSeek 在多个关键方面展现出了显著的优势和创新性。DeepSeek-R1 模型在训练过程中采用了独特的技术路径，尤其是在“点火”环节，所需的启动数据量远低于传统模型。这种低数据需求的特点使得模型的训练门槛大幅降低，能够在有限的资源和数据条件下快速启动并进入高效训练阶段，这对于资源有限的研究机构和企业来说具有巨大的吸引力。并且 DeepSeek-R1 采用了复杂且高效的强化学习（RL）技术。强化学习作为一种先进的训练方法，能够在无需大量标注数据的情况下，通过与环境的交互自主学习最优策略。这种技术不仅提高了模型的训练效率，还显著降低了训练成本。与传统的监督学习相比，强化学习能够更自然地引导模型生成强大的推理能力和复杂的思维链。DeepSeek-R1 通过强化学习训练后，能够自主产生连贯且逻辑性强的推理路径，这种能力在处理复杂的推理任务时表现得尤为突出。

DeepSeek-R1 模型在训练成本上的优化也值得关注。尽管采用了先进的强化学习技术，但该模型的训练过程并未因此变得复杂或昂贵。相反，通过优化算法和高效的训练策略，DeepSeek 成功地在低资源环境下实现了高性能输出。**这种高效且低成本的训练模式使得 DeepSeek-R1 不仅在技术上具有创新性，更在实际应用中展现出极高的性价比。这也是为什么 DeepSeek-R1 模型凭借其低数据需求、高效的强化学习技术以及强大的推理能力，成为 AI 领域中一个极具潜力的创新成果。它不仅为资源有限的开发者提供了新的选择，也为 AI 技术的广泛应用和普及开辟了新的道路。**

图 5：Deepseek 各项测试表现



数据来源：Deepseek

DeepSeek 训练的高效的原因，我们认为训练成本的核心在于模型架构与训练架构，二者相辅相成、缺一不可。

- **MLA 机制：**DeepSeek V3 采用的多头潜在注意力（MLA）机制，通过联合低秩压缩的方式大幅减少了 KV Cache 的使用量。与业界常见的从 KV 数量角度优化 KV Cache 的方法相比，MLA 的压缩方式对研究团队的技术功底提出了更高要求。这种机制不仅有效降低了 KV Cache 的存储需求，还在推理效率上实现了显著提升，体现了 DeepSeek 团队在基础研究上的深厚积累。

- **FP8 训练：**DeepSeek V3 引入了 FP8 混合精度训练框架，通过低精度计算大幅减少了 GPU 内存的使用量和计算开销。技术报告中提到，这是首次在极大规模模型上验证了 FP8 混合精度训练的有效性。这一点不仅展现了 DeepSeek 在技术上的创新性，也凸显了其 Infra 工程团队在底层架构优化方面的强大实力。这种低精度训练技术的应用，为大规模模型的高效训练提供了有力支持。

- **MoE 架构：**DeepSeek V3 采用了 Mixture of Experts（MoE）架构，通过稀疏激活机制大幅减少了计算量。与 Qwen 和 Llama 等采用密集架构（Dense Architecture）的模型相比，MoE 架构在训练和推理过程中具有显著的先天优势。然而，MoE 架构也带来了专家负载均衡、通信效率和路由策略等技术难题。DeepSeek 的 Infra 工程团队通过创新的无辅助损失负载均衡策略等技术手段，成功解决了这些难题，进一步优化了模型的训练效率和性能表现。