

科技行业

DeepSeek 爆火后，影响几何？

- ▶ DeepSeek 性能表现几乎追平美国龙头
- ▶ 全球人工智能发展加速，中美差距缩小
- ▶ 开源和闭源竞争加剧
- ▶ 训练算力需求短期内降温

深度求索 (DeepSeek)，一家位于杭州的人工智能初创公司，发布了全新强化学习版本 DeepSeek 模型 DeepSeek R1 及相关应用程序，其性能几乎追平与现有领先 AI 模型，但成本却非常低，这引发市场对大规模算力部署和资金投入必要性的担忧。通过强化学习，R1 模型在数学、编码、自然语言推理等领域拥有卓越的表现。据 DeepSeek 介绍，其 V3 模型 (R1 之前的模型) 的训练成本估计仅为 Llama 3 或 GPT-4 的 8-9% 左右。

重点关注股票：我们建议关注国内算力股，包括中芯国际 (981 HK/688981 CH, 优于大市/中性)、北方华创 (002371 CH, 优于大市)、中微 (688012 CH, 优于大市)、深南电路 (002916 CH, 优于大市)、中际旭创 (300308 CH, 优于大市)、寒武纪 (688256 CH, 未评级)、海光 (688041 CH, 未评级)；服务器和数据中心供应商包括联想 (992 HK, 优于大市)、中兴通讯 (763 HK/000063 CH, 优于大市/优于大市)、浪潮信息 (000977 CH, 未评级)、紫光股份 (000938 CH, 未评级)、工业富联 (601138 CH, 未评级)。

全球 AI 发展或进一步加速，边缘 AI 及应用将是下一轮竞争的焦点。DeepSeek 的成功表明，通过使用强化学习和 MoE (专家混合模型) 技术，训练过程可以更加高效和节省成本。当然，这需要更高水平的编码和工程设计。同时，它的推理能力几乎与 GPT-o1 等现模型相当。在 MoE 下，数千亿个参数被分类为各个领域的数百名专家，因此在推理时只激活相关领域的某些专家，这大大提高了计算和推理效率，并降低了训练成本。考虑到 OpenAI、Meta、谷歌、微软等人工智能巨头拥有巨大的计算能力，这种 MoE 的成功使用可能会大大加速人工智能的发展。边缘 AI 一直备受关注，因为安全和隐私问题对于用户，尤其是企业和政府来说是不可忽视的。将人工智能模型从云端转移到边缘可以解决这一问题，但需要本地设备上更多的计算能力。自 2023 年底以来，AI 电脑和 AI 智能手机已成为一种趋势。在 CES 2025 上占据了焦点的 AI 眼镜也进入了战场。

开源与闭源的竞争愈演愈烈。DeepSeek 是开源 AI 模型之一，其代码和模型结构对用户和社区开放，可供查看和编辑。大部分开源模型可以免费使用，但如果不支持云端使用，则可能需要一定的部署成本。Meta 的 Llama 3.1 和 Gemma-7B 是两种开源模型。而 GPT 是一种闭源模型，其源代码和结构不公开，通常需要许可证才能使用，费用相对较高。其他闭源模型包括百度的文心一言、苹果的 Siri、快手的可灵等。开源与闭源之间的竞争

可以追溯到几十年前，涉及许多领域，例如计算机操作系统中的 Linux (开放) 与 Windows (封闭)、智能手机操作系统中的 Android (开放) 与 iOS (封闭)、指令集中的 RISC-V (开放) 与 x86 (封闭) 等。在 DeepSeek R1 之前，大语言模型的市场份额和性能主要由 GPT 等闭源模型主导。随着 DeepSeek R1 的表现追平同业龙头，开源模型和封闭模型之间的竞争可能会迅速升温。

训练算力需求热潮降温。据 DeepSeek，V3 模型的训练过程仅需 280 万 GPU 小时左右，成本约 560 万美元 (假设使用 H800，每 GPU 小时的租赁成本为 2 美元)，仅为 Llama 3.1 (约 5,800 万美元) 的 9%，GPT-4 (约 6,300 万美元) 的 8%，Claude 3.5 Sonnet (约 3,000-5,000 万美元) 的 11-18%。如此高的训练效率主要得益于在负载均衡、通信、内存的策略优化以及使用 MoE 模型。AI 模型的训练过程是一个极其耗费算力的过程，需要强大的算力基础设施能力。而作为最常用的训练加速器，NVIDIA 的 AI GPU 需求巨大，而台积电和 SK 海力士的供应有限，一直处于供不应求的状态。同时，其他与训练相关的 ASIC 和通信器件 (例如高端光学模块) 的需求也一直很旺盛。随着 DeepSeek 的成功，高效训练和推理可能是大多数玩家会选择道路。对高性能 GPU 或其他器件的需求可能会在短期内降温，但仍将保持增长，因为 AI 的发展不会停止。

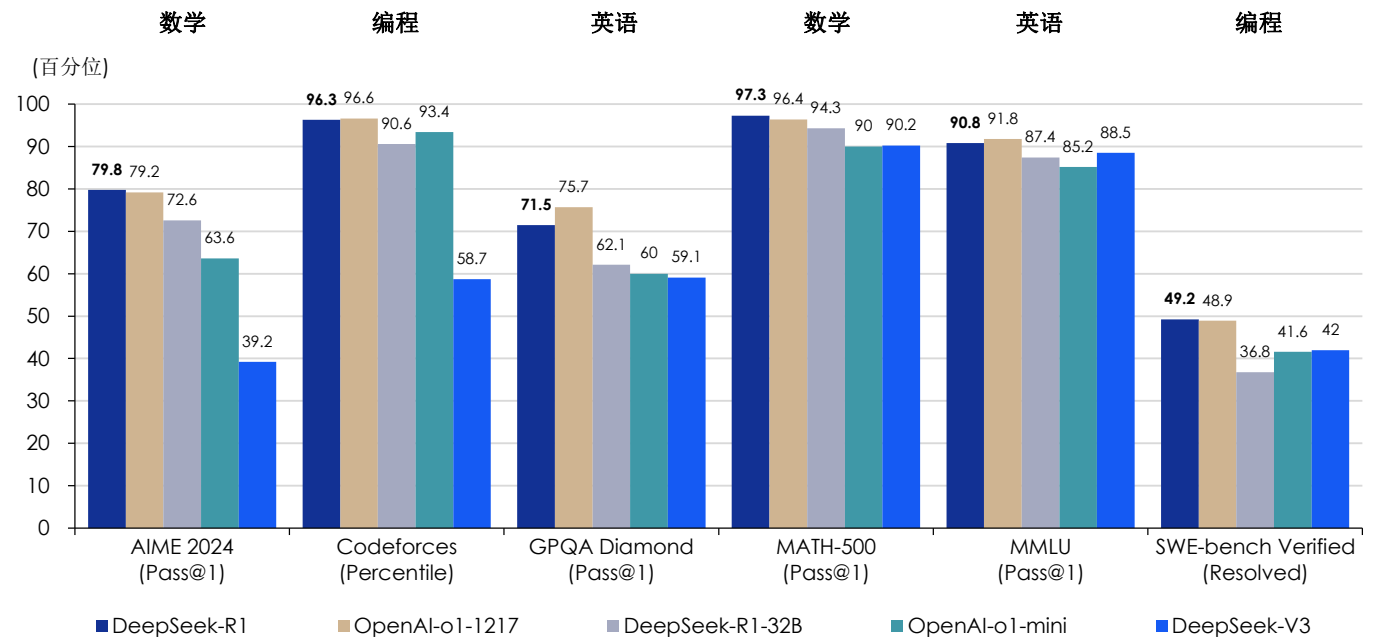
中美人工智能技术差距将加速缩小。自 2022 年 GPT-3.5 发布以来，得益于美国领先的芯片设计和制造能力以及对中国的供应限制，美国 AI 发展迅速。领先企业纷纷推出自己的大语言和多模态模型。在中国，百度、阿里巴巴和字节跳动等大型互联网公司借助其已经建立的计算能力和强大的数据优势，推出了自己的模型文心一言、通义千问和豆包。同时其他一些参与者和初创公司也展示了自己的模型，并获得了良好的市场反响，包括月之暗面的 Kimi、Minimax 和 DeepSeek。DeepSeek 在没有最先进的 AI GPU 的情况下，通过工程创新实现了突破并证明了其卓越的性能。未来中美人工智能发展之间的技术差距可能会进一步缩小。同时，美国可能会实施更多限制，进一步限制中国获得现有的相关 AI ASIC。

多模态 AI 与 AI Agent 步入焦点。与大语言模型这类单模态模型 (文本) 不同，多模态 AI 是指能够同时处理和整合多种类型数据 (如文本、图像、音频、视频等) 的 AI 模型。典型的例子有快手的可灵 (文本/图像转图像/视频)、DALL·E (文本转图像)、Stable Diffusion (文本/图像转图像) 等。多模态 AI 拓宽了 AI 的能力，丰

富了输入、处理和输出的数据类型，可广泛应用于医疗、制造、AR/VR 等多个领域。再进一步，AI Agent 出现了。AI Agent 是一种能够感知环境、收集和处理信息、做出决策并做出反应的 AI 系统。它的主要特征有：(1) 独立，不受人为干扰；(2) 交互，与用户和其他系统通信；(3) 自适应，根据环境和变化调整行为；(4) 目标驱动，优化路径和行为以达到某一目标。当然，AI

Agent 需要多模态 AI，因此需要更多时间来开发。Agent AI 潜力巨大，可用于人形机器人、智能家居、自动驾驶、医疗保健、客户服务等领域。

DeepSeek 模型与 OpenAI-o1 模型的性能表现比较



资料来源: DeepSeek API 技术文件, 建银国际

AI 模型 API 之间价格比较 (均以美元记录)

模型	发布时间	上下文长度	最大 CoT * tokens	最大输出 tokens	每百万 tokens 输入价格 (cache hit)	每百万 tokens 输入价格 (cache miss)	每百万 tokens 输出价格
DeepSeek R1	2025 年 1 月	64k	32k	8k	\$0.14	\$0.55	\$2.19
DeepSeek V3	2024 年 12 月	64k	-	8k	\$0.07	\$0.27	\$1.10
GPT-4	2023 年 3 月	8k	-	2k	\$30.00	\$30.00	\$60.00
GPT-4-32k	2023 年 7 月	32k	-	2k	\$60.00	\$60.00	\$120.00
GPT-4 Turbo	2023 年 11 月	128k	-	2k	\$10.00	\$10.00	\$30.00
GPT-3.5 Turbo	2022 年 11 月	16k	-	4k	\$0.50	\$0.50	\$1.50
Gemini Pro	2023 年 12 月	32k	-	32k	\$0.50	\$0.50	\$1.50
Claude 3.5 Sonnet	2024 年 6 月	200k	-	4k	\$3.00	\$3.00	\$15.00
百度文心	2023 年 3 月	8k	-	8k	c.\$0.11 (RMB0.8)	c.\$0.11 (RMB0.8)	c.\$0.28 (RMB2.0)
阿里通义	2023 年 9 月	8k	-	8k	c.\$0.28 (RMB2.0)	c.\$0.28 (RMB2.0)	c.\$0.83 (RMB6.0)

注: (*) 思维链 (Chain of Thought): R1 模型在给出最终答案输出之前的推理和思考过程和内容

资料来源: DeepSeek, the United Nation University, 格隆汇, 建银国际估计

评级定义:

优于大市 — 于未来 12 个月预期回报为高于 10%
中性 — 于未来 12 个月预期回报在 -10% 至 10% 之间
弱于大市 — 于未来 12 个月预期回报低于 -10%

分析师证明:

本文作者谨此声明: (i) 本文发表的所有观点均正确地反映作者有关任何及所有提及的证券或发行人的个人观点, 并以独立方式撰写; (ii) 其报酬没有任何部分曾经, 目前或将来会直接或间接与本文发表的特定建议或观点有关; 及 (iii) 该等作者没有获得与所提及的证券或发行人相关且可能影响该等建议的内幕信息/非公开的价格敏感数据。本文作者进一步确定 (i) 他们或其各自的关联人士 (定义见证券及期货事务监察委员会持牌人或注册人操守准则) 没有在本文发行日期之前的 30 个日历日内曾买卖或交易过本文所提及的股票, 或在本文发布后 3 个工作日 (定义见《证券及期货条例》(香港法例第 571 章)) 内将买卖或交易本文所提及的股票; (ii) 他们或其各自的关联人士并非本文提及的任何公司的雇员; 及 (iii) 他们或其各自的关联人士没有拥有本报告提及的证券的任何金融利益。

免责声明:

本文由建银国际证券有限公司编写。建银国际证券有限公司为建银国际 (控股) 有限公司 (「建银国际控股」) 和中国建设银行股份有限公司 (「建行」) 全资附属公司。本文内容之信息相信从可靠之来源所得, 但建银国际证券有限公司, 其关联公司及/或附属公司 (统称「建银国际证券」) 不对任何人士或任何用途就本文信息的完整性或准确性或适切性作出任何形式担保、陈述及保证 (不论明示或默示)。当中的意见及预测为我们于本文日的判断, 并可更改, 而无需事前通知。建银国际会酌情更新其研究报告, 但可能会受到不同监管的阻碍。除个别行业报告为定期出版, 大多数报告均根据分析师的判断视情况不定期出版。预测、预期及估值在本质上是推理性的, 且可能以一系列偶发事件为基础。读者不应将本文中的任何预测、预期及估值视为建银国际证券或以其名义作出的陈述或担保, 或认为该等预测、预期或估值, 或基本假设将实现。投资涉及风险, 过去的表现并不反映未来业绩。本文的信息并非旨在对任何有意投资者构成或被视为法律、财务、会计、商业、投资、税务或任何专业意见, 因此不应因而作为依据。本文仅作参考资讯用途, 在任何司法管辖权下的地方均不应被视为购买或销售任何产品、投资、证券、交易策略或任何类别的金融工具的要约或招揽。建银国际不对收件人就本报告中涉及的证券的可用性 (或相关投资) 做任何陈述。本文中提及的证券并非适合在所有司法管辖权下的地方或对某些类别的投资者进行销售。建银国际证券及其它任何人仕对使用本文或其内容或任何与此相关产生的任何其它情况所引发的任何损害或损失 (不论直接的、间接的、偶然的、示范性的、补偿性的、惩罚性的、特殊的或相应发生的) 概不负上任何形式的责任。本文所提及的证券、金融工具或策略并不一定适合所有投资者。本文作出的意见及建议并没有考虑有意投资者的财政情况、投资目标或特殊需要, 亦非拟向有意投资者作出特定证券、金融工具及策略的建议。本文的收件人应仅将本文作为其做出投资决定时的其中一个考虑因素, 并应自行对本文所提及的公司之业务、财务状况及前景作出独立的调查。读者应审慎注意 (i) 本文所提及的证券的价格和价值以及来自该等证券的收益可能有所波动; (ii) 过去表现不反应未来业绩; (iii) 本文中的任何分析、评级及建议为长期性质的 (至少 12 个月), 且与有关证券或公司可能出现的表现的短期评估无关联。在任何情况下, 未来实际业绩可能与本文所作的任何前瞻性声明存在重大分歧; (iv) 未来回报不受保证, 且本金可能受到损失; 以及 (v) 汇率波动可能对本文提及的证券或相关工具的价值、价格或收益产生不利影响。应注意的是, 本文仅覆盖此处特定的证券或公司, 且不会延伸至此外的衍生工具, 该等衍生工具的价值可能受到诸多因素的影响, 且可能与相关证券的价值无关。该等工具的交易存在风险, 并不适合所有投资者。尽管建银国际证券已采取合理的谨慎措施确保本文内容提及的事实属正确、而前瞻性声明、意见及预期均基于公正合理的假设上, 建银国际证券不能对该等事实及假设作独立的复核, 且建银国际证券概不对其准确性、完整性或正确性负责, 亦不作任何陈述或保证 (不论明示或默示)。除非特别声明, 本文提及的证券价格均为当地市场收盘价及仅供参考而已。没有情况表明任何交易可以或可能依照上述价格进行, 且任何价格并不须反映建银国际证券的内部账簿及记录或理论性模型基础的估值, 且可能基于某些假设。不同假设可能导致显著不同的结果。任何此处归于第三方的声明均代表了建银国际对由该第三方公开或通过认购服务提供的数据、信息以及/或意见的阐释, 且此用途及阐释未被第三方审阅或核准。除获得所涉及到的第三方书面同意外, 禁止以任何形式复制及分发该等第三方的资料内容。收件人须对本文所载之信息的相关性、准确性及足够性作其各自的判断, 并在认为有需要或适当时就此作出独立调查。收件人如对本文内容有任何疑问, 应征求独立法律、财务、会计、商业、投资和/或税务意见并在做出投资决定前使其信纳有关投资符合自己的投资目标和投资界限。

使用超链接至本文提及的其它网站或资源 (如有) 的风险由使用者自负。这些链接仅以方便和提供信息为目的, 且该等网站内容或资源不构成本文的一部分。该等网站提供的内容、精确性、意见以及其它链接未经过建银国际证券的调查、核实、监测或核准。建银国际证券明确拒绝为该等网站出现的信息承担任何责任, 且不对其完整性、准确性、适当性、可用性及安全性作任何担保、陈述及保证 (不论明示或默示)。在进行任何线上或线下访问或与这些第三方进行交易前, 使用者须对该等网站全权负责查询、调查和作风险评估。使用者对自身通过或在该等网站上进行的活动自负风险。建银国际证券对使用者可能转发或被要求通过该等网站向第三方提供的任何信息的安全性不做担保。使用者已被认为不可不撤销地放弃就访问或在该等网站上进行互动所造成的损失向建银国际证券索赔。

与在不同时间或在不同市场环境下覆盖的其它证券相比, 建银国际根据要求可能向个别客户提供专注于特定证券前景的专门的研究产品或服务。虽然在此情况下表达的观点可能不总是与分析师出版的研究报告中的长期观点一致, 但建银国际证券有防止选择性披露的程序, 并在观点改变时向相关读者更新。建银国际证券亦有辨别和管理与研究业务及服务有关联的潜在利益冲突的程序, 亦有中国墙的程序确保任何机密及/或价格敏感信息可被恰当处理。建银国际证券将尽力遵守这方面的相关法律和法规。但是, 收件者亦应注意中国建设银行、建银国际证券及其附属机构以及/或其高级职员、董事及雇员可能会与本文所提及的证券发行人进行业务往来, 包括投行业务或直接投资业务, 或不时自行及/或代表其客户持有该等证券 (或在任何相关的投资中) 的权益 (和/或迟后增加或处置)。因此, 投资者应注意建银国际证券可能存在影响本文客观性的利益冲突, 而建银国际证券将不会因此而负上任何责任。此外, 本文所载信息可能与建银国际证券关联人士, 或中国建设银行的其它成员, 或建银国际控股集团子公司发表的意见不同或相反。本文中所提及的产品、投资、证券、金融工具或行业板块, 仅为建银国际的机构和专业客户所提供, 本文不应提供给零售客户。

如任何司法管辖权区的法例或条例禁止或限制建银国际证券向您提供本文，您并非本文的目标发送对象。在阅读本文前，您应确认建银国际证券根据有关法例和条例可向您提供有关投资的研究资料，且您被允许并有资格接收和阅读本文。尤其本文只可分发给依据美国证券法许可建银国际证券分发的指定美国人士，但不能以其它方式直接或间接分发或传送至美国或任何其他美国人士。任何本文接收者中的美国人士（根据美国《交易法》或 1986 年《美国国内税收法（修订）》的定义），如希望根据此处提供的信息进行任何证券买卖或相关金融工具的交易，则此类交易仅限转交与，且仅应通过建银国际海外（美国）公司进行。除非根据适用法律，本文同样不能直接或间接分发或传送至日本和加拿大，及中国境内的公众人士（就本文而言，不包括香港、澳门及台湾）。严禁以任何方式擅自将本研究报告的全部或任何部分再复制或分发予任何人士，建银国际证券对第三方再发本研究报告的行为概不承担任何责任。若本文已经通过电子途径，如电子邮件分发，则该途径并非保证安全或无误，因为信息可能被拦截、损坏、丢失、破坏、迟达或不完整，或包含病毒。因此，建银国际证券对电子传送对本文可能产生的任何错误，或内容遗漏不负责任。本文中所载的披露由建银国际证券编制，应按香港法律管辖及解释。

在新加坡的分发:本文中的信息/研究由建银国际（新加坡）有限公司（建银国际新加坡）（公司注册号 201531408W）的外国分支机构建银国际证券有限公司（“建银国际证券”）提供。建银国际新加坡持有新加坡金融管理局颁发的资本市场服务牌照，可用于基金管理、资本市场产品交易（特别是证券和集体投资计划）并提供托管服务，并且是新加坡 Financial Advisers Act 规定的通过发布或颁发研究分析或研究报告的豁免财务顾问。建银国际证券可以根据新加坡 Financial Advisers Regulations 第 32C 条之与建银国际新加坡的安排分发信息/研究报告。建银国际证券在香港获得了提供此类信息/研究报告的许可。新加坡接收者应通过+65-68071880 与建银国际新加坡联系，解决分发此信息引起的或与之相关的事宜。

©建银国际证券有限公司 2025。本文所使用的标志、标识和徽章及公司名称「建银国际证券有限公司」均属于中国建设银行、建银国际控股和/或建银国际证券的注册及非注册商标。版权所有，违者必究。除非另有说明，本文中的所有材料的版权为建银国际证券所有。未经建银国际证券的书面同意，本文或本文中的任何部分不得复制、出售或再次派发。

以上声明以英文版本为准。

建银国际证券有限公司
香港中环干诺道中 3 号中国建设银行大厦 12 楼
电话:(852)39118000/传真:(852)25370097