

全球AI周报

阿里云AI势能大会召开，谷歌计划全年750亿美元资本支出

作者：

分析师 孔蓉 SAC执业证书编号：S1110521020002

分析师 李泽宇 SAC执业证书编号：S1110520110002

分析师 樊程安吉 SAC执业证书编号：S1110524080001



■ 全球AI动态汇总

- 全球云计算巨头在AI领域持续加码，海内外云计算巨头阿里云和谷歌云均召开发布会。谷歌在Cloud Next大会发布第七代TPU Ironwood、Gemini 2.5 Flash等多款AI产品及模型，并联合50多家合作伙伴推出Agent2Agent（A2A）协议，推动AI应用生态建设，**CEO Sundar Pichai宣布到2025年将在服务器和数据中心投入约750亿美元**。阿里云在北京AI势能大会同样聚焦AI商业化落地，展示最新AI技术与行业应用，截至2025年1月底，百炼平台通义API调用企业与开发者超29万，并发布AI Agent Store与“繁花计划”，计划未来三年联合生态伙伴服务百万客户，拓展百亿市场。
- **字节发布人像视频模型DreamActor-M1，即梦AI即将上线**。字节推出全新人像视频生成模型 DreamActor-M1，该模型基于 DiT 架构，采用多控制信号混合驱动架构，仅需一张静态照片和一段驱动视频，即可生成高质量、电影级的视频，能精准保留原图身份特征，捕捉并迁移驱动视频中的动作和表情，呈现高度逼真效果，还可驱动从真人到卡通的不同形象，支持多种画幅，显著增强了单图视频驱动能力的表现力与易用性。
- OpenAI正式为ChatGPT推出全局记忆功能，允许模型在用户授权下调用所有历史对话，以提供更加个性化和连贯的交互体验。该功能能够根据用户的偏好、习惯甚至人生阶段提供定制化回答，例如在写作时匹配用户文风或在学习外语时记住薄弱点。
- 著名AI科学家李飞飞领衔的斯坦福大学人工智能研究所近日发布了最新一期《2025年人工智能指数报告》，**报告指出中美顶级AI大模型性能差距已经由2023年的17.5%大幅缩至0.3%，中美AI差距正显著缩小**。同时，AI正变得高效且普惠，推理成本显著降低。随着小模型性能提升，达到GPT-3.5水平的推理成本在两年间下降280倍，硬件成本以每年30%的速度递减，能效年提升率达40%；开源模型性能快速增长，部分基准测试中与闭源模型的差距从8%缩至1.7%。

■ 投资建议：

- 尽管全球宏观环境存在诸多扰动，包括市场对美国通胀及潜在经济衰退可能抑制企业AI资本开支的担忧，但全球云计算龙头企业仍在AI领域重申投入金额，加强了AI产业信心：谷歌Cloud Next大会，**CEO Sundar Pichai宣布2025年将在服务器和数据中心投入约750亿美元**。中国AI方面，阿里云AI势能召开，发布AI Agent Store与“繁花计划”，计划未来三年联合生态伙伴服务百万客户，拓展百亿市场，国产AI应用有望加速发展；同时，AI科学家李飞飞发布的《2025年人工智能指数报告》显示中美AI差距显著缩小。**我们认为，随着中国本土AI技术能力的持续提升，以及云厂商与生态伙伴的深度合作不断拓展，中国AI应用有望在多个行业加速渗透与商业化落地，助力本土AI生态进入加速发展阶段。我们预计后续AI应用公司有望逐步进入AI带来的增量业绩兑现阶段，建议持续关注AI基础设施需求及AI应用商业化进展。**

➢ 建议关注：

- ✓ **中国AI**：我们持续看好中国本土AI技术能力的持续提升，中国AI应用有望在多个行业加速渗透与商业化落地，助力本土AI生态进入加速发展阶段。互联网板块内受外部政策影响相对小、宏观关联度相对低，短期估值或超跌，基本面短期维持相对强势。智能驾驶方面，2025年4月车展前后的新车发布周期将驱动市场预期升温，叠加车展作为产品曝光爆点带来的短期催化。**建议关注：腾讯控股、阿里巴巴、快手、美团（商社组联合覆盖）、小鹏汽车、小米集团（汽车、电子组联合覆盖）、理想汽车（汽车组联合覆盖）。**
- ✓ **海外AI**：短期宏观扰动不改AI发展长期趋势，谷歌再次重申全年750亿美元资本支出，强化全球科技巨头在AI投入上的长期定力。我们认为，AI基础设施建设与商业化进程持续推进，推理算力需求高增驱动算力产业链景气延续，AI应用端在广告、电商、医疗等场景的落地不断提速。**建议关注：微软、英伟达、亚马逊、谷歌、特斯拉、赛富时、多邻国、ROBLOX、ROBINHOOD、ELASTIC、RUBRIK、SNOWFLAKE、APPOVIN、CONFLUENT、VERTIV。**

风险提示：宏观经济不及预期；企业盈利不及预期；通胀上行压力。

全球AI动态汇总

1.1 科技大厂动态更新-阿里云AI势能大会召开，国产AI应用全面加速

- 2025年4月9日，阿里云在北京召开AI势能大会，旨在加速AI应用在企业市场的落地。大会展示了阿里云在人工智能领域的最新技术进展及行业应用成果，同时发布了多项技术升级和生态合作计划。截至2025年1月底，阿里云百炼平台中调用通义API的企业和开发者已超过29万，覆盖代码开发、硬件制造、智能座舱、金融服务、药物研发等多个领域。
- 阿里云在会上展示了多个行业应用案例，包括地平线基于阿里云智算集群训练智驾模型，中华财险通过通义灵码实现研发人员AI编程，叮咚买菜利用通义千问模型优化售前咨询和供应链管理等。此外，阿里云还支撑2025年总台春晚实现史上最广覆盖。
- 技术层面，阿里云基于自研“飞天+CIPU”架构，全面升级AI云计算基础设施。第九代ECS实例性能提升20%，并支持TDX机密虚拟机；智能计算平台“灵骏”实现故障自愈，显著提升GPU集群训练效率。人工智能平台PAI面向MoE架构及推理模型全面升级，推理速度和资源利用率显著提高。
- 在生态合作方面，阿里云预告了AI Agent Store愿景，计划通过开放生态模式推动AI普及。同时，阿里云发布“繁花计划”，推出AI应用与服务市场，计划未来三年与生态伙伴共同服务百万云上客户，拓展百亿商机，助力中国AI生态繁荣发展。

图：阿里云 AI Agent Store 计划



图：阿里云宣布“繁花计划”



1.2 科技大厂动态更新-字节发布人像视频模型DreamActor-M1，即梦AI即将上线

- 字节跳动智能创作团队继数字人技术 Omnihuman-1 引发行业关注后，再度推出全新人像视频生成模型 DreamActor-M1，该模型基于 DiT 架构，采用多控制信号混合驱动架构，仅需一张静态照片和一段驱动视频，即可生成高质量、电影级的视频，能精准保留原图身份特征，捕捉并迁移驱动视频中的动作和表情，呈现高度逼真效果，还可驱动从真人到卡通的不同形象，支持多种画幅，显著增强了单图视频驱动能力的表现力与易用性。
- 在技术层面，表情迁移运用隐式 Face Motion Tokenizer 进行自监督训练以精准建模表情细节，动作迁移采用 3D 头部球体和 3D 身体骨架实现肢体动作精准迁移，基于 DiT 架构通过大量数据多阶段训练实现各种画幅镜头的精准复刻。相较于其他相关工作，DreamActor-M1 在人物 ID 保持、动作表情还原、视频生成质量等方面表现更优。模型将于近期上线即梦 AI，用户可在“数字人”-“动作模仿”功能下体验，新功能在生成内容逼真度、画风支持、画面比例支持等方面将大幅优化，且平台会对输出视频进行严格安全审核并添加“AI 生成”水印。

图：DreamActor-M1效果演示



图：DreamActor-M1效果演示



1.3 科技大厂动态更新-谷歌云Next 25大会召开，多款重磅产品发布

- 4月10日，谷歌举行谷歌云 Next 大会。谷歌发布了一系列涵盖基础设施、模型、平台工具及 AI Agent 的新产品与服务。谷歌 CEO Sundar Pichai 宣布了对未来的重大投资，计划 2025 年投入约 750 亿美元资本支出用于服务器和数据中心，以支持 AI 与云业务。
- **Ironwood：专为推理模型设计的全新 TPU 芯片。**谷歌推出了第七代张量处理单元（TPU）Ironwood，这是其首款专为推理工作负载设计的芯片，标志着AI芯片设计理念的重大转变。Ironwood在计算性能、内存容量和能源效率上实现了显著提升。在完整集群配置下，其FP8峰值算力可达42.5 Exaflops，单芯片峰值算力为4614 TFLOPS。内存子系统升级至192GB高带宽内存（HBM），带宽提升至7.2 Tbps，大幅优化了推理延迟和复杂模型处理效率。此外，Ironwood的每瓦性能相较于上一代提升了两倍，能效比2018年推出的TPU高出近30倍，并配备先进液冷技术以支持高负载运行。
- **Agent2Agent：AI Agent互操作性的新标准。**谷歌联合了超过 50 家技术合作伙伴与服务提供商共同发起并推出了Agent2Agent（A2A）协议，旨在解决当前AI Agent在不同平台和框架中“碎片化”导致的协作难题。A2A是一个开放的通信规范和互操作框架，支持不同供应商开发的AI代理通过标准化机制进行发现、通信、协作和任务管理。其核心功能包括能力发现（通过“Agent Card”描述功能）、任务管理（定义任务生命周期和状态更新）、协作通信（传递上下文和中间结果）、用户体验协商以及基于现有网络协议的安全性支持。A2A与Anthropic的模型上下文协议（MCP）互补，前者专注于代理间的直接协作，后者关注代理与环境的交互。
- **多款模型更新，Gemini 2.5 Flash 即将推出：**Gemini 2.5 Flash 在保持低延迟和高性价比的同时，融入了强大的思考能力。谷歌宣布其优化后的 Gemini 2.5 Flash 模型即将部署至 Vertex AI 平台，该模型以低延迟与成本效益为核心优势，支持实时摘要及文档探索场景，并通过动态推理层级调整与响应时长自适应机制提升复杂查询处理能力。同时，谷歌在多模态生成领域取得突破性进展：Imagen 3 文本转图像模型通过强化对象移除算法在 LMArena 基准测试中排名第一，Chirp 3 音频生成系统新增基于10秒样本的定制语音合成及多人对话转录功能，Lyria 成为全球首个企业级文本转音乐模型，可生成30秒高保真音频片段；Veo 2 视频生成工具则推出智能擦除、画面扩展、运镜优化及自动补帧功能，覆盖从内容修复到跨平台适配的全流程需求。

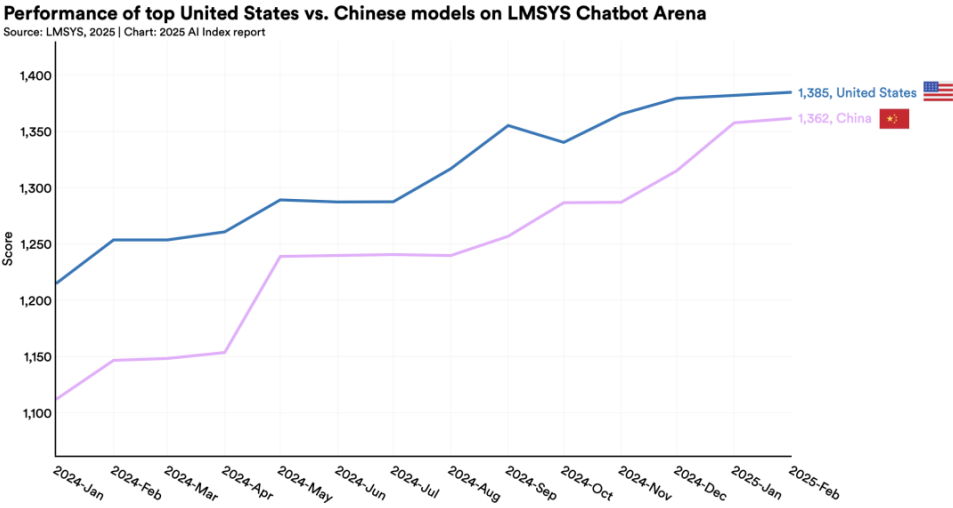
1.4 科技大厂动态更新-OpenAI推出全局记忆功能，并官宣GPT-4将被4o完全替代

- **OpenAI正式为ChatGPT推出全局记忆功能**，允许模型在用户授权下调用所有历史对话，以提供更加个性化和连贯的交互体验。该功能能够根据用户的偏好、习惯甚至人生阶段提供定制化回答，例如在写作时匹配用户文风或在学习外语时记住薄弱点。用户可在设置中随时关闭记忆功能或禁用历史对话调用，首批仅限Plus和Pro用户使用（欧洲经济区、英国等地区暂不可用），Team、企业及教育版用户需等待数周。OpenAI CEO Sam Altman表示全局记忆功能的推出有望使得AI系统能更加了解用户新闻，并变得极其有用和高度个性化。
- **OpenAI官宣GPT-4本月底退役，由4o完全替代**。OpenAI在其网站上表示，从4月30日起，GPT-4将被GPT-4o完全替代，但GPT-4将继续通过API提供使用。OpenAI表示在面对面的评估中，4o在写作、编码、STEM等方面持续超过GPT-4。据The Verge报道，OpenAI将推出新的AI模型，其中将包括GPT-4.1，将是4o多模态模型的改进版。OpenAI同时还将推出更小的GPT-4.1 mini和nano版本，以及o3“推理”模型和新的名为o4-mini的推理模型。

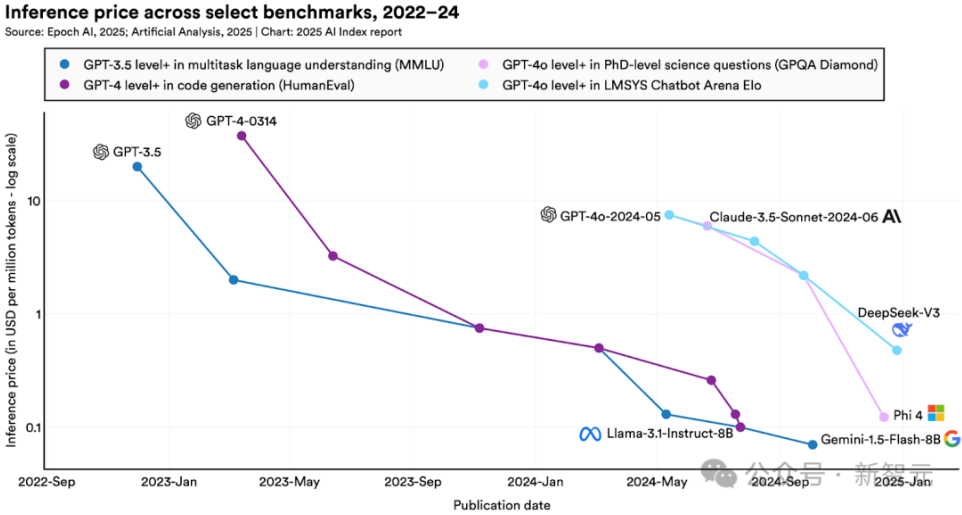
2.1 AI相关研究-《2025年人工智能指数报告》发布，中美AI差距正显著缩小

- 著名AI科学家李飞飞领衔的斯坦福大学人工智能研究所近日发布了最新一期《2025年人工智能指数报告》，报告指出中美顶级AI大模型性能差距已经由2023年的17.5%大幅缩至0.3%，接近抹平。报告显示，2024年重要大模型（Notable Models）中，美国入选40个，中国15个。其中，谷歌、OpenAI入选7个并列第一，阿里入选6个，按照模型贡献度位列全球第三。在业内顶级专家评选出的32项“2024年AI领域重要发布”中，中国的阿里Qwen2、Qwen2.5及DeepSeek-V3三大发布上榜。
- AI正变得高效且普惠，推理成本降低280倍。报告显示随着小模型性能提升，达到GPT-3.5水平的推理成本在两年间下降280倍，硬件成本以每年30%的速度递减，能效年提升率达40%。同时，开源模型性能快速增长，部分基准测试中与闭源模型的差距从8%缩至1.7%。小模型性能显著提升，参数下降142倍。2022年，在大规模多任务语言理解（MMLU）基准测试中，得分超60%的最小模型是 PaLM，参数量为5400亿。到了2024年，微软Phi-3-mini仅用38亿参数，就取得了同样的实力，两年多的时间里模型参数减少了142倍。

图：美国与中国顶尖模型在LMSYS聊天机器人竞技场的表现



图：推理价格持续下降



分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

一般声明

除非另有规定，本报告中的所有材料版权均属天风证券股份有限公司（已获中国证监会许可的证券投资咨询业务资格）及其附属机构（以下统称“天风证券”）。未经天风证券事先书面授权，不得以任何方式修改、发送或者复制本报告及其所包含的材料、内容。所有本报告中使用的商标、服务标识及标记均为天风证券的商标、服务标识及标记。

本报告是机密的，仅供我们的客户使用，天风证券不因收件人收到本报告而视其为天风证券的客户。本报告中的信息均来源于我们认为可靠的已公开资料，但天风证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，天风证券及/或其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，天风证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。

天风证券的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。天风证券没有将此意见及建议向报告所有接收者进行更新的义务。天风证券的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

特别声明

在法律许可的情况下，天风证券可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。因此，投资者应当考虑到天风证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突，投资者请勿将本报告视为投资或其他决定的唯一参考依据。

投资评级声明

类别	说明	评级	体系
股票投资评级	自报告日后的6个月内，相对同期沪深300指数的涨跌幅	买入	预期股价相对收益20%以上
		增持	预期股价相对收益10%-20%
		持有	预期股价相对收益-10%-10%
		卖出	预期股价相对收益-10%以下
行业投资评级	自报告日后的6个月内，相对同期沪深300指数的涨跌幅	强于大市	预期行业指数涨幅5%以上
		中性	预期行业指数涨幅-5%-5%
		弱于大市	预期行业指数涨幅-5%以下