

AI 技术 4 月深度观察

中国高技术产业发展促进会新质生产力工作委员会

博雅云创 & 中科创新驱动

2026 年 4 月

摘要

2026 年 4 月，全球 AI 技术演进呈现出几个极其鲜明的主线：一是原生多模态与超长上下文能力持续突破，Gemini 3.1 Ultra 与 GPT-5.4 分别在多模态推理和桌面自主执行上树立新标杆；二是行业竞争焦点从单纯“参数扩张”转向“效率优化”与“系统能力”，TurboQuant、神经符号推理等路线为算力、内存与能耗约束提供了新的解法；三是 AI 智能体基础设施正在加速成型，MCP 协议迈向开放治理、Claude Cowork 冲击垂直 SaaS、AI 医疗决策与网络安全模型进入高敏感场景，标志着 AI 正从工具层走向制度层和产业层；四是资本、算力、基础设施与物理世界深度绑定，Q1 创纪录融资、微软日本 100 亿美元投资、NVIDIA 物理 AI 展示共同说明，AI 竞争已进入“模型、协议、能源、机器人、数据中心”并行重写的新阶段。本报告试图基于 2026 年 4 月的关键事件，梳理这一轮 AI 技术演进背后的底层趋势与产业含义。

Contents

1	Google 发布 Gemini 3.1 Ultra, 原生多模态推理再升级	3
2	OpenAI 发布 GPT-5.4, 首次在桌面任务上超越人类	3
3	Google 在 ICLR 2026 发布 TurboQuant, 内存占用降低六倍	4
4	Anthropic MCP 协议突破 9700 万安装量, Linux 基金会接管治理	4
5	Meta 发布 Muse Spark, 豪掷千亿重新入局	5
6	AI 能耗降低 100 倍的突破, 神经符号推理重新登场	5
7	2026 年 Q1 AI 风投创纪录, 达 2672 亿美元	6
8	OpenAI 月收入破 20 亿美元, 筹备 IPO 同时砍掉 Sora	6
9	犹他州首创, AI 系统可自动续开药物处方	6
10	Anthropic 发布 Claude Mythos Preview, 专攻网络安全	7
11	Claude Cowork 引发 “SaaS 末日”, 法律科技股暴跌	7
12	微软投资 100 亿美元建设日本 AI 基础设施	8
13	NVIDIA 国家机器人周, 物理 AI 和世界模型大爆发	8
14	OpenAI 收购科技播客 TBPN, 加码内容创作者经济	9
15	AI 虚拟试穿重塑零售业, 直击 8500 亿美元退货难题	9
16	结语: 2026 年 4 月, AI 开始全面进入 “系统竞争” 阶段	10
17	参考来源	10

1 Google 发布 Gemini 3.1 Ultra, 原生多模态推理再升级

Google 推出了 Gemini 3.1 Ultra, 这是今年最重磅的模型发布, 配备 200 万 token 的上下文窗口, 能够在文本、图像、音频和视频之间原生推理, 无需任何转录中间环节。与之前的 Gemini 版本不同, 3.1 版本从训练阶段就被设计为可以同时跨所有模态进行推理。此外, 新模型还搭载了一个沙盒化的代码执行工具, 允许模型在对话中实时编写、运行和测试代码, 同时显著改进了事实查询的基础支撑能力, 以减少幻觉现象。

这一发布被业内视为 Google 在多模态原生推理领域对 OpenAI 和 Anthropic 的重要反击, 尤其是超长上下文窗口让 Gemini 3.1 成为处理整本书、长视频和大型代码库的首选工具。它的战略价值不只体现在参数和功能清单上, 更在于 Google 试图重新定义未来模型竞争的维度: 不是单一模态上的聊天能力, 而是跨模态、跨工具、跨长时信息载体的统一理解与执行能力。

2 OpenAI 发布 GPT-5.4, 首次在桌面任务上超越人类

OpenAI 发布了 GPT-5.4, 配备 100 万 token 上下文窗口, 能够跨软件环境自主执行多步 workflow。在 OSWorld-V 基准测试上, 这一模拟真实桌面生产力任务的测试中, 模型得分达到 75%, 略高于人类基线的 72.4%。相比 GPT-5.2, 这是整整 27.7 个百分点的提升。

这种原生操作系统级别的计算机使用能力让 GPT-5.4 成为真正意义上的自主智能体, 能够以最少的人工干预导航文件、浏览器和终端界面。这标志着 AI 正从“聊天工具”转变为“自主数字同事”, 重新定义了知识工作的边界。从产业角度看, 这类能力将直接推动办公自动化、客服运营、数据处理和企业内流程执行进入一个全新的阶段。

3 Google 在 ICLR 2026 发布 TurboQuant，内存占用降低六倍

Google 研究团队在 ICLR 2026 上发布了 TurboQuant 算法，显著降低了 KV 缓存导致的内存开销，这是运行大型 AI 模型最大的瓶颈之一。该算法采用两步流程，先用 PolarQuant 方法对数据向量进行随机旋转以简化几何结构，再应用量化的 Johnson-Lindenstrauss 压缩方法作为数学纠错器。

TurboQuant 承诺在保持前沿性能的同时将内存需求降低六倍，这可能加速 AI 行业从“暴力堆参数”向“效率优先”的转变，对端侧 AI 和数据中心成本都具有深远意义。它的价值在于指出了一条越来越清晰的路径，即未来模型竞争不仅是“能不能训练出来”，更是“能不能以更低内存、更低延迟、更低成本稳定运行”。

4 Anthropic MCP 协议突破 9700 万安装量，Linux 基金会接管治理

Anthropic 的模型上下文协议（MCP）在 2026 年 3 月安装量突破了 9700 万次，这一里程碑标志着它从实验性标准转变为构建 AI 智能体的基础设施。如今所有主要 AI 提供商都推出了兼容 MCP 的工具，该协议已成为智能体连接外部工具、API 和数据源的默认机制。Linux 基金会还宣布将把 Anthropic 的 MCP 纳入开放治理，进一步巩固其作为行业共享基础设施的地位。

MCP 在短短一年内从 Anthropic 的一个提案发展为跨公司标准，是 AI 生态系统互操作性的重大胜利，也让 Anthropic 在智能体时代占据了独特的生态位。它的重要性不亚于云时代的 API 标准化，因为谁掌握智能体与外部世界交互的协议，谁就更可能塑造未来 AI 生态的连接方式。

5 Meta 发布 Muse Spark，豪掷千亿重新入局

Meta 发布了全新 AI 模型 Muse Spark，试图在去年开源模型表现不佳后重振旗鼓。Meta 表示，过去 9 个月 Meta 超级智能实验室从零开始重建了整个 AI 技术栈，Muse Spark 在多模态感知、推理、健康和智能体任务中提供了有竞争力的性能。与以往 Llama 的完全开源策略不同，Muse Spark 将采用专有许可模式，但公司表示“希望未来版本能够开源”。

Meta 还宣布 2026 年 AI 相关资本支出将达到 1150 亿至 1350 亿美元，几乎是去年的两倍。这标志着扎克伯格在 AI 战略上的重大转向，也说明 Meta 正在从“开源分发者”转向“重资产竞争者”。这种变化的含义非常清楚：在进入超级智能与基础设施竞赛阶段后，单靠开源叙事已不足以支撑顶层战略，资本投入、系统整合与技术闭环正在重新成为关键。

6 AI 能耗降低 100 倍的突破，神经符号推理重新登场

Tufts 大学的研究人员推出了一种革命性的高效 AI 方法，能够将 AI 能耗降低多达 100 倍，同时还能提升准确率。该系统结合神经网络与类人符号推理，帮助机器人进行更有逻辑的思考，而不是依赖暴力的试错。在 AI 电力消耗已占美国总用电量 10% 以上的背景下，这一突破意义重大。

该研究来自 Matthias Scheutz 教授的实验室，团队开发的神经符号 AI 将传统神经网络与符号推理结合，模仿人类将问题分解为步骤和类别的思维方式。研究成果将在 5 月于维也纳举行的国际机器人与自动化会议上展示。这也为行业提供了“参数竞赛”之外的另一条路径，即通过更高层的结构化推理降低训练与推理中的无效计算。

7 2026 年 Q1 AI 风投创纪录，达 2672 亿美元

2026 年第一季度 AI 风险投资总额达到 2672 亿美元，是前一季度纪录的两倍多。这一激增由少数几笔巨额交易推动，OpenAI 融资 1220 亿美元，由亚马逊、英伟达和软银领投；Anthropic 获得 300 亿美元 G 轮融资；而 xAI 被 SpaceX 以 2500 亿美元收购。

这种资本的高度集中表明行业正在向“行星级”算力集群的建设，以及 AI 与物理基础设施的垂直整合过渡。这也意味着 AI 行业的准入门槛正在变得极高，大型玩家与中小创业公司之间的鸿沟不断扩大。资本市场正在用最直接的方式投票，未来 AI 竞争将是模型、基础设施、能源、协议和分发渠道的全栈竞赛。

8 OpenAI 月收入破 20 亿美元，筹备 IPO 同时砍掉 Sora

OpenAI 月收入已达 20 亿美元，正在筹备可能重塑 AI 行业格局的 IPO。然而即便是如此惊人的收入，公司仍然宣布关停 AI 视频生成应用 Sora，该项目每日运营成本高达 100 万美元，活跃用户数从 100 万跌至不足 50 万。这一决定甚至是在公告发布前一小时才通知其重要合作伙伴迪士尼。

OpenAI 年化收入已超过 250 亿美元，可能最早在 2026 年底进行公开上市。相比之下，竞争对手 Anthropic 的年化收入接近 190 亿美元。OpenAI 的战略重心正明显地从消费级实验转向企业级生产力工具。这一变化很值得重视，因为它意味着视频生成等“吸睛型”产品，正在让位于更高频、更可计费、更能进入企业主流程的 AI 服务。

9 犹他州首创，AI 系统可自动续开药物处方

犹他州成为美国首个授权 AI 系统续开药物处方的州，标志着 AI 在医疗自动化领域的一个重要里程碑。这一举措代表了人工智能在直接患

者护理领域的重大扩张，从诊断辅助进入到了此前仅由持证医疗专业人员负责的实际治疗决策。

该实施引发了关于医疗场景中 AI 可靠性、患者安全协议以及所需监管框架的重要讨论。这是 AI 在受严格监管的医疗领域的一次大胆试验，也为其他州如何处理“AI 医疗决策权”提供了先例参考。支持者认为这能缓解基层医生短缺，反对者则担心患者安全和责任归属问题。无论争议如何，这都意味着 AI 已经开始试探制度边界，而不只是优化工作流。

10 Anthropic 发布 Claude Mythos Preview，专攻网络安全

Anthropic 发布了 Claude Mythos Preview，这是一款专门为网络安全设计的先进 AI 模型，已经在各大系统中发现了数千个此前未知的零日漏洞。该模型通过“玻璃翼计划”进行部署，这是一个与 40 多家公司的有限合作项目，包括微软、亚马逊、苹果、谷歌、英伟达、CrowdStrike 和 Palo Alto Networks，且仅用于防御性安全用途。

这一发布既展示了前沿 AI 发现漏洞的惊人能力，也凸显了同样的技术一旦被滥用可能带来的巨大风险。这也是 Anthropic 选择通过受控合作而非公开发布该模型的原因。它反映出越来越尖锐的问题，即最强 AI 能力往往不适合完全开放，而需要以“可控访问、限定用途、协作治理”的方式落地。

11 Claude Cowork 引发“SaaS 末日”，法律科技股暴跌

Anthropic 发布的 Claude Cowork 搭载了用于自动化合同审查和 NDA 分诊等法律任务的 AI 智能体，引发了法律科技和 SaaS 股票的大幅抛售，业内将这一现象称为“SaaS 末日”。这是 AI 智能体首次直接冲击传统 SaaS 商业模式的标志性事件。

当 AI 智能体能够自主完成合同审查、文档管理等过去需要专门软件的工作时，许多垂直 SaaS 产品的价值主张瞬间被动摇。投资者开始重新评估，哪些 SaaS 公司会被 AI 颠覆，哪些能够利用 AI 实现新增长。这场震荡预计将在未来几个月持续，也预示着未来软件行业竞争的重心将从“功能堆叠”转向“智能体 workflow 承载能力”。

12 微软投资 100 亿美元建设日本 AI 基础设施

微软宣布向日本投资 100 亿美元，这是西方科技公司在亚洲的最大单笔 AI 基础设施承诺。该投资涵盖 AI 基础设施、网络安全加强和全面的劳动力发展计划，将日本定位为微软全球 AI 战略中的关键枢纽。约 60 亿美元将用于在东京、大阪和横滨等日本主要都市区建设数据中心和 AI 计算基础设施。

行业分析师预计该投资将在未来五年内在日本创造约 5 万个 AI 相关的直接和间接工作岗位。微软 CEO 萨蒂亚·纳德拉强调，日本的技术成熟度、监管稳定性和战略位置使其成为服务亚太市场的理想 AI 基础设施枢纽。这对日本的科技产业和 AI 人才市场是重大利好，也说明 AI 基础设施竞争已经越来越像地缘经济和区域能力布局的一部分。

13 NVIDIA 国家机器人周，物理 AI 和世界模型大爆发

NVIDIA 在国家机器人周上展示了一系列将 AI 引入物理世界的突破。其中 Mimic Robotics 采用的 mimic-video 模型将预训练的互联网规模视频模型与流匹配动作解码器配对，取代了传统 VLA 中静态的图像-语言骨干网络，在真实世界的操作任务上实现了 10 倍的样本效率提升和 2 倍的收敛速度。

这表明在捕获物理规律和因果关系的世界模型上训练的机器人，在未见过的条件下能够用少得多的真实世界数据可靠地执行任务。这种范

式转变可能极大加速家用服务机器人、工业自动化和自动驾驶的落地，也是 2026 年“物理 AI 元年”的重要信号。世界模型不再只是研究口号，而是开始在机器人训练效率和泛化能力上体现实际价值。

14 OpenAI 收购科技播客 TBPN，加码内容创作者经济

OpenAI 收购了热门科技播客 TBPN，进一步巩固其在内容创作者经济领域的布局。TBPN 是由 John Coogan 和 Jordi Hays 主持的每日科技新闻节目，曾邀请马克·扎克伯格、萨蒂亚·纳德拉等业内重量级人物做客。这笔收购在业界引发广泛讨论，一家 AI 模型公司为什么要买一个播客？

分析人士认为，OpenAI 正在构建一个完整的“AI 原生”内容生态，试图控制从模型到分发渠道的整个链路，同时 TBPN 积累的高质量科技访谈也可能成为训练未来多模态模型的宝贵素材。这一动作让其他 AI 公司开始反思自己的内容策略，也说明未来的 AI 竞争可能不仅发生在模型实验室和企业工作流里，还会发生在注意力分发、品牌叙事与创作者网络的争夺中。

15 AI 虚拟试穿重塑零售业，直击 8500 亿美元退货难题

越来越多的 AI 初创公司推出了虚拟试穿技术，让消费者在购买前可可视化服装的合身度和款式。美国全国零售联合会估计，2025 年零售销售额中有 15.8% 被退回，总额达 8499 亿美元；而在线销售的退货率更是跃升至 19.3%。Gen Z 是这一趋势的主要推动者，18 至 30 岁的购物者去年平均每人进行了近 8 次在线退货。

Shopify 已将初创公司 Genlook 的 AI 虚拟试穿应用集成到其商务平台；Google 的虚拟试穿技术也将从 4 月 30 日起直接在 Google 平台的产品搜索结果中提供。虽然行业多年来一直尝试解决在线试穿问题，但生成

式 AI 的快速发展终于让这些应用达到了真正能影响零售商利润的水平。这是 AI 在消费领域一个极少见的、可以直接用利润率与退货率衡量的胜利，也提醒我们，AI 价值并不只体现在高端模型竞赛中，同样体现在对传统产业微观损耗的持续削减上。

16 结语：2026 年 4 月，AI 开始全面进入“系统竞争”阶段

如果说过去两年行业最关心的是“谁的模型更聪明”，那么 2026 年 4 月给出的答案已经明显不同。真正决定下一阶段格局的，越来越不是单点模型能力，而是围绕模型形成的整套系统能力，包括多模态理解、操作系统执行、协议标准、推理效率、资本供给、能源约束、行业监管、机器人落地与基础设施部署。

从 Gemini 3.1 Ultra 和 GPT-5.4 的能力突破，到 TurboQuant 和神经符号推理的效率革命；从 MCP 的标准化治理，到 Claude Cowork 对 SaaS 的冲击；从微软日本基础设施投资，到 NVIDIA 把 AI 推进物理世界，这些事件共同说明，AI 行业已经进入从“模型竞赛”向“系统竞赛”切换的关键时刻。

未来真正的领先者，未必只是最会训练模型的公司，而是那些能够同时掌握协议、基础设施、产品化能力、产业场景与组织入口的参与者。2026 年 4 月的这些动态，本质上都在指向同一个判断，AI 正从一场技术热潮，演化为一场对数字世界与物理世界底层结构的全面重写。

17 参考来源

1. Crescendo AI, *Latest AI News and AI Breakthroughs that Matter Most: 2026*. <https://www.crescendo.ai/news/latest-ai-news-and-updates>
2. devFlokera, *AI News Last 24 Hours: April 2026 Latest Model Releases &*

- Papers*. <https://www.devflokiers.com/blog/ai-news-last-24-hours-april-2026>
3. ScienceDaily (Tufts University), *AI breakthrough cuts energy use by 100x while boosting accuracy*. <https://www.sciencedaily.com/releases/2026/04/260405003952.htm>
 4. HumAI Blog, *AI News & Trends April 2026: Complete Monthly Digest*. <https://www.humai.blog/ai-news-trends-april-2026-complete-monthly-digest>
 5. Radical Data Science, *AI News Briefs Bulletin Board for April 2026*. <https://radicaldatascience.wordpress.com/2026/04/07/ai-news-briefs-bulletin-board-for-april-2026/>
 6. Mean CEO Blog, *Open AI News | April, 2026 (STARTUP EDITION)*. <https://blog.mean.ceo/open-ai-news-april-2026/>
 7. CNBC, *Meta debuts new AI model, attempting to catch Google, OpenAI after spending billions*. <https://www.cnbc.com/2026/04/08/meta-debuts-first-ai-model-attempting-to-catch-google-openai-after-spending-billions.html>
 8. NVIDIA Blog, *National Robotics Week, Latest Physical AI Research, Breakthroughs and Resources*. <https://blogs.nvidia.com/blog/national-robotics-week-2026/>
 9. Relvai, *Best AI News April 2026: 5 Game-Changing Developments Ranked*. <https://www.relvai.com/blog/best-ai-news-april-2026-developments-ranked/>
 10. CNBC, *'Silent killers': How AI start-ups are trying to solve one of the retail industry's biggest problems*. <https://www.cnbc.com/2026/04/05/ai-retail-start-ups-virtual-try-on-tech-margins.html>

联系我们，请扫描二维码



新质生产力工作委员会
官方公众号



工业智能算网
gyznswn.cn

新质生产力工作委员会：

中国高技术产业发展促进会新质生产力工作委员会，专注于推动工业人工智能、智能制造、数字化转型等前沿技术发展，为企业提供政策解读、技术咨询和产业对接服务。

工业智能算网：

专注于工业人工智能、新质生产力、工业软件 CAE、智能制造等前沿技术。提供每日动态分析、技术趋势解读、解决方案分享，推动工业智能化转型。

网站地址：<https://gyznswn.cn>