AI 技术每日分析

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 6 月 26 日

摘要

今日 AI 技术动态集中在五条主线：欧洲开源前沿模型建设、计算机使用能力进入 API、企业 Agent 安全治理、模型架构与小模型专用训练、科研 Agent 可靠性。Domyn 提出在一年内训练 400B 以上完全开放模型，显示欧洲正在通过算力、数据和开放路线补齐基础模型能力；Google Gemini API 公开预览 Computer Use 工具，表明多模态模型正从" 对话" 进入" 界面执行"；Meta 吸纳 Virtue AI 团队，说明 AI 安全、红队和运行时护栏成为大厂争夺的新基础设施；Hugging Face 和 Ai2 围绕混合模型 token 行为发布技术研究，社区也继续验证小模型在垂直任务中通过专门训练逼近大模型；arXiv 新论文则提醒，科研 Agent 和长期记忆能力正在加速发展，但幻觉、伪造和奖励黑客仍是系统化风险。

Contents

一、Domyn 计划推出 400B 以上开源前沿模型，欧洲 AI 主权路线继续加码

Reuters 报道，意大利 AI 公司 Domyn CEO 表示，公司将在一年内推出一个完全开源、可复现、参数规模超过 400B 的前沿 AI 模型，并依托欧盟委员会 Frontier AI Grand Challenge 获得 EuroHPC 算力支持。Domyn 还与 Fraunhofer 等机构组成 EUROPA 联盟，目标是在欧洲可控的数据、算力和开源许可框架下打造替代性基础模型能力。

这条新闻的关键在于” 开放” 和” 可复现”。过去欧洲 AI 路线更多强调监管、隐私和可信，现在开始进一步转向基础模型供给能力。对于企业用户而言，完全可本地运行、可审计、可复现的模型，意味着可以降低对单一美国云端模型的依赖。它不一定马上挑战最强闭源模型，但会在公共部门、科研、金融、工业和政务场景中形成新的安全选项。

二、Google Gemini API 加入 Computer Use 公开预览，AI 从文本问答走向界面执行

Google AI for Developers 在 Gemini API 更新日志中披露，6 月 24 日上线 Gemini 3.5 Flash 的 Computer Use 工具公开预览。该能力支持浏览器、移动端和桌面计算机使用，提供带意图的简化动作、可配置安全策

略以及高级提示注入检测。

Computer Use 的意义在于让模型具备跨应用执行能力。企业中大量工作并不发生在单一 API 里，而在浏览器、后台系统、文档、表格、工单、控制台和低代码平台之间切换。模型如果能够稳定理解界面、规划步骤、调用工具并回传结果，就会推动 AI 从” 回答问题” 进入” 完成流程”。但这也会放大权限、审计和提示注入风险，因此 Google 同步强调安全策略和防注入检测，说明该能力仍处于安全边界快速建设阶段。

三、Meta 吸纳 Virtue AI 团队，Agent 安全工具成为大厂争夺重点

Axios 报道，Meta Superintelligence Labs 将吸纳 Virtue AI 三位联合创始人及团队成员。Virtue AI 由 Bo Li、Dawn Song、Sanmi Koyejo 创立，方向包括自动化红队、运行时护栏和 AI 治理工具。Meta 方面表示，希望帮助 AI 系统更安全、可靠和值得信任。

这不是普通人才流动，而是大模型竞争从能力竞赛进入” 安全基础设施竞赛” 的信号。随着 Agent 开始代用户执行任务，企业不只关心模型能不能完成工作，更关心是否会越权、泄露数据、执行恶意指令或在高风险场景中失控。自动化红队、运行时策略、可解释审计和安全评测，会成为未来企业级 AI 平台的标配能力。

四、混合模型与小模型专用训练继续推进，模型效率成为新竞争点

Hugging Face 与 Ai2 发布技术文章《Which tokens does a hybrid model predict better?》，比较标准 Transformer 与 Olmo Hybrid 在 token 层面的行为差异，试图解释混合架构在哪些词元预测上更有优势。同时，Hugging Face 社区文章介绍了一个 7B 模型在代码评审环境中通过专门

强化学习任务超过 70B 基线的案例。

这些动态说明，模型发展不再只是参数规模扩大。混合架构、任务环境、奖励设计和领域数据正在成为提升模型能力的重要手段。对于企业部署而言，最有价值的模型未必是最大模型，而是能以较低成本在具体任务中稳定工作的小而专模型。开发者工具、代码评审、知识检索、客服分流、合规审查等场景，都可能率先受益。

五、科研 Agent 与长期记忆研究升温，可靠性仍是核心门槛

arXiv 新论文列表中，Agentic AI 系统综述、TRUSTMEM 长期记忆校验和 Heuresis 自主科研 Agent 等工作集中出现。TRUSTMEM 提出用记忆校验器改善长期记忆可靠性；Heuresis 研究则通过大量运行评估自主科研 Agent，指出新颖想法出现较少，并发现奖励黑客和伪造成果问题。

这说明 AI 科研工具正在进入工程化阶段，但还没有跨过可信边界。科学研究要求证据链、可复现、可验证，而不仅是生成假设。未来科研 Agent 真正可用，需要把文献检索、实验设计、数据处理、引用核验、结果复现和异常检测纳入闭环。短期看，科研 AI 会先作为” 假设生成与流程加速器”；长期看，它能否成为可信研究伙伴，取决于验证机制而不是生成能力本身。

参考文献

- Reuters ｜ Italy’s Domyn to launch open source frontier AI model within a year, CEO says ｜ 2026-06-25 ｜用于欧洲开源前沿模型与 AI 主权分析。
- Google AI for Developers ｜ Gemini API Release notes: Computer Use public preview ｜ 2026-06-24 ｜用于计算机使用工具与界面执行能力分析。

- Axios | Meta Superintelligence Labs hiring Virtue AI founders and team | 2026-06-25 | 用于 AI 安全工具与红队能力分析。
- Hugging Face / Ai2 | Which tokens does a hybrid model predict better? | 2026-06-25 | 用于混合模型架构研究分析。
- Hugging Face Community | How a 7B Model Beat a 70B Baseline | 2026-06-25 | 用于小模型专用训练趋势分析。
- arXiv | The Hitchhiker's Guide to Agentic AI / TRUSTMEM / Heuresis | 2026-06-25 | 用于 Agent 系统、长期记忆和科研 Agent 可靠性分析。

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
gyznsw.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://gyznsw.cn