

AI 技术每日分析：DeepMind 把 EVE Online 变成长程 Agent 试验场

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 8 月 22 日星期六

摘要

今天 AI 技术方向出现了几条很有代表性的变化：Google DeepMind 把 EVE Online 这样的长期、多人、持续演化虚拟世界进一步明确为研究连续学习、记忆、长程规划和多智能体协作的试验场；Google Research 则发布多智能体生物标志物发现框架，把生成式推理、确定性统计和对抗式验证组合到同一科研流程；Oracle 把企业 Agent 的持久记忆和生产数据库访问问题落到权限、审计和语义正确性上；轨道数据中心创业公司 Starcloud 完成 2.5 亿美元融资，AI 算力基础设施开始向太空延伸；开发者工具 Amp 则把“Token 花到哪里了”做成 Agent 使用分析能力。AI 竞争的重心继续从单次回答能力，转向长周期运行、数据可信、科研验证、算力位置与单位任务成本。

目录

摘要	3
一、DeepMind 把 EVE Online 做成长程 Agent 研究场：测试“按周、按月、按年”规划	3

二、Google Research 发布多智能体生物标志物框架：AI 科研开始把“质疑自己”写进流程	4
三、Oracle 讨论 Agent 进入生产数据库：真正危险的是“SQL 正确、业务含义错误”	5
四、Starcloud 融资 2.5 亿美元：轨道 AI 数据中心从概念进入制造扩张阶段	5
五、Amp 上线 Token 使用解释：Coding Agent 开始把成本可观测性做成产品能力	6
趋势观察	6
参考资料	7

摘要

今天 AI 技术方向出现了几条很有代表性的变化：Google DeepMind 把 EVE Online 这样的长期、多人、持续演化虚拟世界进一步明确为研究连续学习、记忆、长程规划和多智能体协作的试验场；Google Research 则发布多智能体生物标志物发现框架，把生成式推理、确定性统计和对抗式验证组合到同一科研流程；Oracle 把企业 Agent 的持久记忆和生产数据库访问问题落到权限、审计和语义正确性上；轨道数据中心创业公司 Starcloud 完成 2.5 亿美元融资，AI 算力基础设施开始向太空延伸；开发者工具 Amp 则把“Token 花到哪里了”做成 Agent 使用分析能力。AI 竞争的重心继续从单次回答能力，转向长周期运行、数据可信、科研验证、算力位置与单位任务成本。

一、DeepMind 把 EVE Online 做成长程 Agent 研究场：测试“按周、按月、按年”规划

Google DeepMind 8 月 21 日发布新的游戏 AI 研究路线，系统总结其从 Atari、AlphaGo、AlphaStar 到 SIMA 2 的演进，并重点介绍与 EVE Universe 开发团队 Fenris Creations 的合作。与传统游戏 Benchmark 不同，EVE Online 是一个已经持续运行二十多年、拥有真实供需、联盟、冲突、外交和大规模玩家互动的持续世界。

DeepMind 明确提出，这类环境适合研究四类前沿能力：连续学习、跨越当前上下文窗口的长期记忆、以周/月/年为尺度的长程规划，以及大规模多智能体协作、竞争、谈判和经济行为。研究会先从与真实玩家隔离的离线 EVE 实例开始，再进入 EVE Frontier 等更开放环境，只有能力足够成熟后才考虑进入正式在线世界。

这条路线的价值不在“AI 会玩游戏”。当前 Agent 在短任务 Benchmark

上表现已经很强,但一旦任务持续数天甚至数月,就容易出现目标漂移、遗忘、反复尝试和协作失效。一个规则复杂、持续变化、后果长期累积的虚拟世界,实际上提供了比静态 Benchmark 更接近真实组织和数字经济的 Agent 压力测试环境。

二、Google Research 发布多智能体生物标志物框架: AI 科研开始把“质疑自己”写进流程

Google Research 8 月 21 日发布 Biomarker Discovery Framework, 用于从可穿戴设备时间序列中筛选候选生物标志物。系统并非让一个大模型直接生成研究结论,而是由 Orchestrator、Scout、Literature、Hypotheses、Statistical、ML、Critic、Defender 等不同 Agent 分工,形成“提出假设—统计分析—模型训练—对抗验证—文献核验—报告”的闭环。

框架把确定性计算和生成式推理解耦: 数值统计由代码完成,模型负责提出假设、解释和组织证据; Critic 和 Defender 专门寻找数据泄漏、过拟合、混杂因素、不稳定性和生理学上不合理的结论。Google 在三个队列、合计 9279 个参与者观测样本中测试,最终筛出精神健康和代谢领域的多组候选数字生物标志物。

这里最值得关注的是科研 Agent 的“流程工程”。AI for Science 如果只追求自动生成更多假设,很容易制造更多伪相关。把不同 Agent 设置成互相质疑的角色,并要求数值结果回到结构化事实表核验,说明科研 Agent 正在从“自动写报告”向“可追踪的研究流程”发展。

三、Oracle 讨论 Agent 进入生产数据库：真正危险的是“SQL 正确、业务含义错误”

Oracle 开发者团队 8 月 21 日发布关于 Agent 记忆和生产数据库访问的工程总结。Oracle 计划在 AGNTCon + MCPCon Europe 展示如何把 Oracle AI Database 作为 Agent 的持久记忆，并让 Agent 查询企业数据而不自行生成 SQL。

文章强调，语言模型天然无状态，而企业 Agent 如果要连续运行数周，需要工作记忆、长期事实记忆和流程记忆；这些记忆一旦包含真实企业信息，就需要与业务数据使用相同的访问控制、备份和审计机制。更重要的是，Agent 访问数据库时最危险的失败未必是 SQL 语法报错，而是 SQL 完全可以执行，却由于对“月份”“客户”“收入”等业务语义理解错误而返回一个看似合理的错误结果。

Oracle 引用 8 月 21 日 BIRD Text-to-SQL 榜单称，最佳公开模型执行准确率为 80.04%，而人类为 92.96%。这提醒企业，Agent 接入生产数据库之后，治理重点会从“能不能连上数据库”进一步转向语义层、权限层和结果验证层。

四、Starcloud 融资 2.5 亿美元：轨道 AI 数据中心从概念进入制造扩张阶段

Reuters 8 月 21 日报道，轨道数据中心创业公司 Starcloud 完成 2.5 亿美元融资，估值 23 亿美元。本轮由 Manhattan West 领投，Nvidia 和 Cisco Investments 成为新投资方；公司自 2024 年成立以来累计融资 4.5 亿美元。

Starcloud 正在与 Nvidia 合作开发面向辐射和极端太空环境的 Space-1 Vera Rubin Module。新资金将用于扩大制造能力、推进与 Nvidia 的工程

协作以及采购新产品。公司提出的长期目标非常激进：建设 8.8 万颗卫星、形成 20GW 轨道算力，并在华盛顿州建设约 10 万平方英尺的 Starcloud-3 制造设施。

轨道数据中心仍面临发射成本、热管理、维护、通信时延和商业模型等大量不确定性，但这笔融资说明 AI 基础设施竞争正在推动新的“算力位置”实验。地面数据中心受到电力、土地、冷却和审批约束后，资本开始愿意为更极端的基础设施路径下注。

五、Amp 上线 Token 使用解释：Coding Agent 开始把成本可观测性做成产品能力

AI 编程工具 Amp 在 8 月 21 日更新中上线“Explain Usage”，允许用户向 Puck 询问 Token 到底消耗在什么地方。更新本身很小，却对应 Agent 进入团队后的一个实际问题：长时间自主运行、反复调用工具、打开大量上下文后，开发者很难仅通过 API 单价判断一次任务为什么变贵。

随着 Coding Agent 从单轮补全走向后台任务、远程执行和多 Agent 协作，成本管理会逐渐类似云计算的可观测性：不仅看模型每百万 Token 多少钱，还要看到一次任务中检索、上下文、工具调用、重试和长链推理分别消耗多少资源。Agent 平台能否解释“钱花在哪里”，会直接影响企业是否敢于扩大自主运行范围。

趋势观察

今天几条信息共同把“模型之外”的价值进一步放大：DeepMind 用持续世界测试长期智能，Google Research 用多 Agent 互审提高科研严谨性，Oracle 把记忆和业务语义交给数据库治理，Starcloud 尝试重构算力部署位置，Amp 开始追踪 Agent 单位任务成本。与此同时，OWASP 在

Agentic Skills Top 10 中把 Skill 层的权限、来源和跨平台复用风险单独列出，Nvidia 近期研究也显示，Harness 和监督 Agent 可能显著改变同一底层模型在长程任务中的表现。AI 系统正在从“一个模型”变成由环境、记忆、工具、数据、运行时和成本控制共同构成的工程系统。

参考资料

1. Google DeepMind | 《From Atari to EVE Online: Building on 15 Years of AI Research in Games》 | 2026-08-21
<https://deepmind.google/blog/from-atari-to-eve-online-building-on-15-years-of-ai-research-in-games/>
2. Google Research | 《An AI tool for prioritizing candidate biomarkers from wearable sensor data》 | 2026-08-21
<https://research.google/blog/an-ai-tool-for-prioritizing-candidate-biomarkers-from-wearable-sensor-data/>
3. Oracle Developers | 《What we learned about letting agents into a production database》 | 2026-08-21
<https://blogs.oracle.com/developers/what-we-learned-about-letting-agents-into-a-production-database>
4. Reuters | 《Orbital data center startup Starcloud valued at \$2.3 billion in latest funding》 | 2026-08-21
<https://www.reuters.com/business/media-telecom/orbital-data-center-startup-starcloud-valued-23-billion-latest-funding-2026-08-21/>
5. Amp | 《Chronicle — Explain Usage》 | 2026-08-21
<https://ampcode.com/chronicle>

6. Dark Reading | 《OWASP Flags Top AI Skill Risks in New Security Blueprint》 | 2026-08-06
<https://www.darkreading.com/application-security/owasp-flags-top-ai-skill-risks-security-blueprint>
7. OWASP Foundation | 《OWASP Agentic Skills Top 10》
<https://owasp.org/www-project-agentic-skills-top-10/>
8. TechCrunch | 《Nvidia just showed that the harness, not the AI model, is now the real hero》 | 2026-08-21
<https://techcrunch.com/2026/08/21/nvidia-just-showed-that-the-harness-not-the-ai-model-is-now-the-real-hero/>
9. Google | 《Inside the Gemmaverse: Celebrating one billion Gemma downloads》 | 2026-08-20
<https://blog.google/innovation-and-ai/technology/developers-tools/gemma-one-billion-downloads/>
10. Reuters | 《How a Texas student blew the whistle on a rogue AI hacking attempt》 | 2026-08-20
<https://www.reuters.com/world/how-texas-student-blew-whistle-rogue-ai-hacking-attempt-2026-08-20/>

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>