

AI早报：Agent workflows 前移，治理与安全同步升温

中文内部早报 · 5分钟版

今日总览

今天的AI新闻主线集中在智能体工作流、企业部署和治理安全。头条里最重要的几条分别是微软开源 Orchard 智能体训练框架、AirLLM 实现单块 4GB GPU 运行 70B 模型推理、Google Agent Skills 幕后：如何构建、测试与规模化，整体呈现出从模型能力展示转向真实工作流和企业环境落地的趋势。

智能体、编程与工作流：最重要的是微软开源 Orchard 智能体训练框架。这类新闻的共同点是，AI 工具正在进入任务拆解、跨系统协作、版本控制和长期执行等环节，不再只是停留在代码补全或单轮问答。同板块还包括Cloudflare 推出 @cloudflare/computer 预览版：为智能体提供虚拟文件系统与多执行环境、Cloudflare Workers 与 Containers 现已支持入站 TCP 连接和 gRPC。

前沿模型与多模态：最重要的是Paramount CTO Phil Wiser：AI 属于史上最伟大技术趋势，但“iPhone 时刻”尚未到来。模型能力的亮点更多体现在产品嵌入和场景化调用上，视频编辑、旅行对话、模型聚合入口等消息都指向同一个方向：能力需要被放进具体工作流，读者更应关注实际可用性和成本。

大厂与产业链：最重要的是AirLLM 实现单块 4GB GPU 运行 70B 模型推理。企业部署、云上交付、内容合作和基础设施投入仍是产业侧重点；算力、水电、版权和企业采购会继续影响 AI 服务成本与可落地范围。

监管、安全与研究：监管安全侧最重要的是Google Agent Skills 幕后：如何构建、测试与规模化。今天的信号集中在漏洞修复、治理控制面和数据驻留等主题，说明企业在采用模型时需要同步处理审计、权限和供应商边界。研究评测方面，OpenRouter 推出 Ori Eval 简化评估流程值得放入方法论更新。开发者工具方面，Qwen3.8-Max 发布：开源最强编码与协作模型，2.4T 参数可作为近期试用入口。

头条今日头条

1. 微软开源 Orchard 智能体训练框架

X: Microsoft Research (@MSFTResearch) · 08/04 08:00 · 智能体、编程与工作流

Orchard 是一个面向研究社区的开源框架，用于跨任务类型训练和评估 AI 智能体。它降低了复杂性，同时通过让研究人员复用同一套基础设施，支持较小模型也能实现强劲性能。<https://msft.it/6019a8fqP>

2. AirLLM 实现单块 4GB GPU 运行 70B 模型推理

Hacker News · 08/04 08:00 · 大厂与产业链

AirLLM 项目支持在单块 4GB 显存 GPU 上运行 70B 参数大模型推理，无需多卡或大规模显存配置。该项目已开源，相关讨论在 Hacker News 上获得 103 点热度，引发社区关注。

3. Google Agent Skills 幕后：如何构建、测试与规模化

Google AI: DEV 作者专属 · 08/04 08:00 · 监管、安全与版权

Google Agent Skills 团队详解其开源技能库的构建与治理流程：项目始于 Google Cloud Next 2026 前的“swarm”冲刺，发布后 GitHub 星标超 15,000。为保证规模化下的质量，每个技能须通过标准化目录结构、CI/CD 流水线（含 linter、链接检查、AI 辅助清单）及提交时与每周的持续评估，并优先引用远程 MCP 工具。

4. Cloudflare 推出 @cloudflare/computer 预览版：为智能体提供虚拟文件系统与多执行环境

Cloudflare Blog · 08/04 08:00 · 智能体、编程与工作流

Cloudflare 发布 @cloudflare/computer 早期预览版，这是一个开源智能体运行时，为每个智能体提供虚拟文件系统，并支持在 isolate、容器沙箱或浏览器中执行代码。

5. Cloudflare Workers 与 Containers 现已支持入站 TCP 连接和 gRPC

Cloudflare Blog · 08/04 08:00 · 智能体、编程与工作流

据Cloudflare Blog，Cloudflare Workers 与 Containers 现已支持入站 TCP 连接和 gRPC。这类进展主要影响智能体编排、代码生成和企业工作流落地。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

模型前沿模型与多模态

1. Paramount CTO Phil Wiser：AI 属于史上最伟大技术趋势，但“iPhone 时刻”尚未到来

Runway 新闻 · 08/04 08:00

Paramount 首席技术官 Phil Wiser 认为 AI 应跻身人类史上最伟大技术趋势前五，其影响堪比火的使用。他主张等待依赖条件成熟、以“aha 时刻”引导采用，而非以技术为先导；并警告行业巨头过度分析将错失窗口期，这一窗口可能仅 5 至 10 年。Paramount 两年前已通过 Runway 向全员开放 AI 工具，并以业务成果而非 token 消耗量衡量成效。

2. OpenAI 更新模型能力与产品方案

OpenAI 新闻 · 08/03 15:00

据OpenAI 新闻，OpenAI 更新模型能力与产品方案。这类进展主要影响模型能力、API 选型和产品可用性。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

3. 欧盟《人工智能法案》透明度规则生效，违规最高罚 1500 万欧元

The Verge: AI · 08/04 08:00

欧盟《人工智能法案》下的新透明度义务于 8 月 2 日生效，要求公司披露用户何时与 AI 模型互动，并为合成音视频和文本添加机器可读标记。欧盟还推出了一套可选的 AI 披露标签供平台采用，但标注要求本身是强制性的。违规公司面临最高 1500 万欧元（约 1720 万美元）或全球年营业额 3% 的罚款，8 月 2 日前推出的模型有 4 个月宽限期。

4. Databricks 推出 Variant 通用可用版本，加速半结构化数据摄取

Databricks: Blog · 08/04 08:00

Databricks 宣布 Variant 正式全面可用（GA），用于更快速、高效地摄取 JSON、XML、CSV 等半结构化数据。该功能旨在简化传统上复杂的数据导入流程，提升处理效率。

智能体 智能体、编程与工作流

1. China 更新 AI 工作流与智能体方案

The Decoder · 08/03 21:52

据The Decoder，China 更新 AI 工作流与智能体方案。这类进展主要影响智能体编排、代码生成和企业工作流落地。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

2. Launch 更新 AI 工作流与智能体方案

Hacker News · 08/04 00:32

据Hacker News，Launch 更新 AI 工作流与智能体方案。这类进展主要影响智能体编排、代码生成和企业工作流落地。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

3. Kimi Work 幻灯片制作教程发布

X: Kimi.ai (@Kimi_Moonshot) · 08/04 08:00

使用 Kimi Work 制作幻灯片 - 教程 #1。Kimi Slides 处理整个幻灯片制作流程：- 清晰的结构与研究，由 Kimi K3 驱动 - 连贯的设计，包括精美的图表和 SmartArts - 可编辑并可直接下载 欢迎在评论区告诉我们你还想看什么内容！

4. Alibaba 更新 AI 工作流与智能体方案

The Decoder · 08/04 01:12

据The Decoder，Alibaba 更新 AI 工作流与智能体方案。这类进展主要影响智能体编排、代码生成和企业工作流落地。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

产业 大厂与产业链

1. Cloudflare 推出 Billable Usage API：为自助账户提供按产品与计费周期的程序化成本可见性

Cloudflare Blog · 08/04 08:00

Cloudflare 发布 Billable Usage API，为自助账户提供单一端点，一次调用即可返回按产品和计费周期拆分的用量与成本，覆盖 Workers、R2、D1、Workers AI、Vectorize、Images 和 Stream。

2. Data Commons on Spanner Graph 正式可用：统一公共与私有数据构建知识图谱

Google Cloud: Databases · 08/04 08:00

据Google Cloud：Databases，Data Commons on Spanner Graph 正式可用：统一公共与私有数据构建知识图谱。这类进展主要影响云算力、硬件投入、企业采购和服务成本。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

3. Palantir 强劲季度后，CEO Alex Karp 称 AI 行业“马克思主义”

TechCrunch AI · 08/04 08:00

Palantir CEO Alex Karp 在季度股东信中警告，前沿 AI 实验室对企业过于不可信，并称其意图“占有所谓合作伙伴的生产资料”，带有“马克思主义色彩”。该公司第二季度营收 19 亿美元，同比增长 93%，利润 11 亿美元。Karp 主张 Palantir 提供模型无关的 AI 与分析软件，让企业掌控自身数据与 AI “废气”（提示词、编排、上下文）。

4. TechCrunch 推出 AI 产业部署新进展

TechCrunch AI · 08/04 04:00

据TechCrunch AI，TechCrunch 推出 AI 产业部署新进展。这类进展主要影响云算力、硬件投入、企业采购和服务成本。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

开源 开源项目与开发者工具

1. Qwen3.8-Max 发布：开源最强编码与协作模型，2.4T 参数

Qwen: Blog Retrieval (API) · 08/04 08:00

Qwen 正式发布 Qwen3.8-Max，这是 Qwen 家族迄今最强的模型，拥有 2.4T 参数（95B 激活），并首次开源 Qwen-Max 级权重，开放权重将于下周发布。

2. 新开源项目 patchy631/time-to-first-token 上线

GitHub · 08/02 21:02

GitHub 新仓库 patchy631/time-to-first-token 在本次覆盖窗口内创建。仓库简介显示，该项目面向 AI 开发者工具或自动化场景。当前公开数据约为 205 Stars、19 Forks，主要语言为 HTML。该条目代表新项目出现，不把普通代码提交描述为版本发布。

3. 新开源项目 FareedKhan-dev/kimi-k3-in-c 上线

GitHub · 08/01 17:29

GitHub 新仓库 FareedKhan-dev/kimi-k3-in-c 在本次覆盖窗口内创建。仓库简介显示，该项目面向 AI 开发者工具或自动化场景。当前公开数据约为 1,233 Stars、189 Forks，主要语言为 C。该条目代表新项目出现，不把普通代码提交描述为版本发布。

4. 新开源项目 Anionex/codex-vision-proxy 上线

GitHub · 08/01 18:00

GitHub 新仓库 Anionex/codex-vision-proxy 在本次覆盖窗口内创建。让纯文本模型在 Codex 中无障碍调用内置看图工具（view_image）的方案，附为纯文本 LLM 设计的视觉工具包&skill | Let text-only。当前公开数据约为 252 Stars、10 Forks，主要语言为 Python。该条目代表新项目出现，不把普通代码提交描述为版本发布。

观点

KOL 与巨头言论

1. EA 首席战略官谈生成式 AI 如何进入可游玩的实时游戏世界

Runway 新闻 · 08/04 08:00

EA 首席战略官 Mihir Vaidya 认为，游戏是 AI 的试验场，但生成式 AI 进入游戏面临 60 帧/秒、数千玩家同步和低延迟等严苛约束，不能只追求“看起来真实”，而必须“行为正确”。他主张采用神经符号架构，在生成能力之外保留确定性与可控性，并称“控制是下一个前沿”。EA 将 AI 影响分为效率、扩展和转型三个层面，其中《模拟人生》已服务超 5 亿玩家，拥有近万亿种游玩排列组合。

2. 关于 Astra 与数学的两则重要更新：Anthropic 数学家 24 小时复现 OpenAI 半数结果

Gary Marcus: The Road to AI We Can Trust · 08/04 08:00

Gary Marcus 称 OpenAI 的 Astra 可能并非其宣传的突破，Anthropic 数学家 Levent Alpöge 用已公开的 Fable 模型在 24 小时内复现了 OpenAI 半数结果，质疑其真实进展。

3. Show 发布 AI 行业观点与判断

Hacker News · 08/03 20:41

据Hacker News，Show 发布 AI 行业观点与判断。这类信息反映行业参与者对技术路线和商业节奏的判断。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

监管

监管、安全与版权

1. 德国法院裁定AI音乐生成器Suno侵犯版权，驳回合理使用抗辩

The Decoder: AI News · 08/01 18:40

慕尼黑法院裁定，AI音乐生成器Suno在训练过程及输出结果中均侵犯版权，并驳回其合理使用抗辩。法院认定Suno 3.5和4版本模型可复现六首知名歌曲的原创元素，构成“记忆化”侵权，且责任归于Suno而非用户。该判决还认定美国版权法下的合理使用不适用于此案，目前尚未最终生效。

2. OpenAI 新模型 Astra 数学表现出色，但被过度吹捧

Gary Marcus: The Road to AI We Can Trust · 08/03 05:25

OpenAI 内部测试的新模型 Astra 在数学问题上表现惊艳，但 Gary Marcus 指出相关讨论犯了“合成谬误”：擅长某类数学不等于擅长所有数学、科学乃至一切认知任务。数学之所以成为突破口，是因为它便于用符号工具验证且能廉价生成海量合成数据，而开放世界问题无法如此模拟。此外，OpenAI 未公布方法细节，尚无法评估其真实意义。

3. Databricks 完成对 Panther 的收购，加速安全湖仓时代

Databricks: Blog · 08/04 08:00

Databricks 宣布正式完成对安全数据平台 Panther 的收购，旨在加速安全湖仓（Security Lakehouse）时代。此次收购将 Panther 的安全分析能力整合进 Databricks 平台，帮助企业在统一的数据架构上运行安全运营与威胁检测工作负载。具体交易条款未披露。

研究

研究、评测与方法论

1. OpenRouter 推出 Ori Eval 简化评估流程

X: OpenRouter (@OpenRouter) · 08/04 08:00

推出 Ori Eval：编写首个评估的最简单方式。没有绝对最好的模型，只有最适合每项任务的模型。Ori Eval 利用 OpenRouter 的 API 处理代码库中的每项任务，然后评估结果。curl -fsSL https://openrouter.ai/skills/spawn-ori-eval