

AI 技术每日分析：9750 亿参数开放权重模型登场，自动红队进入训练闭环

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 7 月 16 日星期四

摘要

7 月 15 日，AI 技术领域同时出现了模型开放、安全训练、智能体运行环境和行业数据接入方面的新进展。Thinking Machines 发布首个开放权重通用模型 Inkling，把竞争重点放在企业可定制性；OpenAI 披露自动化红队模型 GPT-Red，以自博弈方式持续生成新的提示注入攻击，并将攻击结果用于生产模型训练；LangChain 一边把无代码智能体直接部署到 Slack，一边强调为智能体配置与个人电脑、生产系统相隔离的执行环境。语音 AI 和医疗 AI 也在补齐评价与治理底座：Hume 推出多维真实语音基准，CData 则为受 HIPAA 监管的医疗场景提供受控的实时数据连接层。

Contents

一、Thinking Machines 发布 Inkling：开放权重竞争转向企业可定制性	1
二、OpenAI 披露 GPT-Red：自动红队进入模型训练闭环	2
三、LangSmith Fleet 接入 Slack：智能体进入团队协作入口	3

四、LangChain 强调“给智能体一台隔离电脑”：执行环境成为基础设施	3
五、Hume 推出 Real World VoiceEQ：语音 AI 不能只看一个总分	4
六、CData 扩展医疗 AI 数据层：敏感数据连接成为落地瓶颈	5
参考文献	5

一、Thinking Machines 发布 Inkling：开放权重竞争转向企业可定制性

Thinking Machines Lab 7 月 15 日发布首个开放权重通用模型 Inkling。官方资料显示，模型采用混合专家架构，共有 9750 亿参数，每次推理激活 410 亿参数，最长上下文达到 100 万 Token；预训练使用约 45 万亿 Token 的文本、图像、音频和视频数据。模型可接收文本、图像和音频输入，并配套发布了更轻量的 Inkling-Small 预览版本。

公司没有把 Inkling 描述成综合能力最强的模型，而是强调模型可以下载、部署、微调和改造。Reuters 报道指出，Inkling 在智能体相关任务上表现突出，但闭源前沿模型在综合能力上仍保持领先。Thinking Machines 同时把模型接入 Tinker 微调平台，希望企业能够围绕金融分析、编码、知识检索和专用智能体形成自己的版本。

这次发布的重要性在于，开放权重市场的竞争指标正在变化。企业不仅关心基础模型排行榜，也开始比较模型能否私有部署、能否用行业数据完成稳定微调、工具调用是否可靠，以及后续升级会不会破坏既有 workflow。Inkling 选择“基础模型 + 微调平台”的组合，意在把模型开放转化为持续的工程服务。

二、OpenAI 披露 GPT-Red：自动红队进入模型训练闭环

OpenAI 7 月 15 日发布 GPT-Red 技术说明。GPT-Red 是内部使用的自动化红队模型，通过自博弈强化学习与一组防守模型共同训练：攻击方尝试诱导模型违反原始任务或泄露数据，防守方则要抵御攻击并完成任务。攻击载体包括本地文件、网页横幅、电子邮件正文和工具返回结果等，重点覆盖智能体容易接触的间接提示注入场景。

OpenAI 称，在一套内部复现的间接提示注入测试中，GPT-Red 在 84% 的场景中找到有效攻击，而参与对照的人类红队人员为 13%。GPT-Red 还被用于测试自动售货机智能体和 Codex CLI 数据外泄场景。公司表示，自 GPT-5.3 以来已将逐步增强的自动红队模型用于后续生产模型训练；在其内部评测中，GPT-5.6 Sol 对 GPT-Red 直接提示注入的失败率降至 0.05%。这些数字来自 OpenAI 内部环境，仍需第三方评测验证其外部可复现性。

这项工作的价值不在于生成更多固定攻击样本，而在于形成“攻击者增强—防守者训练—再次攻击”的动态流程。智能体会访问网页、邮件、文件和外部工具，安全风险随着工具链扩展不断变化。自动红队能够提高测试覆盖率，但不能替代人工安全研究、第三方审计和生产环境监控。

三、LangSmith Fleet 接入 Slack：智能体进入团队协作入口

LangChain 7 月 15 日宣布，用户可以在 LangSmith Fleet 中无代码创建定制智能体，并一键部署到 Slack。每个智能体拥有独立的 Slack 身份，可以在对话线程中回答问题、处理附件、请求补充信息和发起人工批准。创建者负责配置技能、上下文和可访问系统，其他员工无需进入开发平台即可使用。

Fleet 采用“一人创建、团队使用”的模式。专业人员可以把企业流程和经验固化为智能体，员工通过原有沟通渠道调用。智能体遇到需要判断的环节时，可以留在同一 Slack 线程中请求批准、接受修改意见，再继续执行后续步骤。

Slack 只是一个分发入口，但它反映出企业智能体的产品形态已经发生变化。智能体不再局限于独立聊天页面，而是嵌入员工原有工作流，在同一线程里接收任务、补充信息和完成人机协作。今后的竞争将更多取决于渠道集成、权限管理和协作闭环。

四、LangChain 强调“给智能体一台隔离电脑”：执行环境成为基础设施

LangChain 同日发布智能体执行环境指南，提出安全边界应从提示词层面提升到机器层面。智能体可以获得能够运行代码、安装依赖和处理文件的完整环境，但该环境应与用户电脑、生产系统及其他智能体工作区隔离。

其建议包括把智能体生成或下载的代码默认视为不可信内容，为每个任务提供独立内核、文件系统和网络边界，并限制 CPU、内存、网络访问和凭据使用。智能体运行过程还应记录执行命令、修改文件、安装软件包和外部网络请求，以便审计和复现。

智能体能力越强，执行环境越不能被视为附属功能。工具调用失败通常可以重试，但凭据泄露、生产数据误删和横向访问会造成不可逆影响。企业部署编码智能体或数据智能体时，需要把沙箱、最小权限、人工批准和审计日志作为统一运行时的一部分。

五、Hume 推出 Real World VoiceEQ：语音 AI 不能只看一个总分

Hume AI 7 月 14 日推出 Real World VoiceEQ Bench，用四类榜单分别评价自动语音识别、语音理解、文本转语音和实时语音智能体。官方称，基准建立在超过 100 万人次的人类评分基础上，其中包括约 78.5 万次 TTS 评分和 4.8 万次语音到语音评分。

研究结果显示，语音 AI 表现具有明显的维度差异。一个系统可能在自然度上领先，却在身份稳定、情绪理解、噪声环境或实时互动上表现一般；接收音频输入也不代表模型会利用语调、停顿和犹豫等副语言信息，一些语音智能体仍主要依赖转写文本。

这类多维评测对客服、陪伴、教育和语音助手更有参考价值。实际部署需要同时考虑内容识别、情绪理解、表达稳定性、抗噪能力和延迟，单一平均分容易掩盖产品在具体场景中的短板。

六、CData 扩展医疗 AI 数据层：敏感数据连接成为落地瓶颈

CData 7 月 15 日宣布，Connect AI 面向受 HIPAA 监管的医疗机构、支付方、生命科学公司和医疗科技企业提供受治理的数据访问层。产品允许 AI 应用和智能体连接电子病历、支付系统、实验室平台及企业应用中的实时数据，而不必先复制或集中暴露受保护健康信息。

平台提供身份感知访问控制、加密、审计日志、统一治理和业务伙伴协议支持。每次 AI 查询和用户交互均可记录，模型按照既有用户权限访问数据，医疗机构无需为每个智能体单独建设一套数据复制管道。

医疗 AI 的难点已经从演示模型能力转向受控访问业务数据。没有清晰的数据权限、访问记录和责任边界，临床、运营和支付智能体很难离开

试点。CData 这类连接层代表了新的基础设施方向：模型保持可替换，数据治理和企业系统连接则作为长期能力沉淀。

参考文献

- Thinking Machines Lab: 《Inkling: Our open-weights model》, 2026-07-15; 用途: 核验模型架构、参数规模、上下文、训练数据和产品定位。
<https://thinkingmachines.ai/news/introducing-inkling/>
- Thinking Machines Lab: 《Inkling Model Card》, 2026-07-15; 用途: 核验输入模态、数值格式、部署与模型限制。
<https://thinkingmachines.ai/model-card/inkling/>
- Reuters: 《AI startup Thinking Machines launches an open-weight AI model》, 2026-07-15; 用途: 核验市场定位及与闭源模型的比较。
<https://www.reuters.com/business/ai-startup-thinking-machines-launches-an-open-weight-ai-model-2026-07-15/>
- OpenAI: 《GPT-Red: Unlocking Self-Improvement for Robustness》, 2026-07-15; 用途: 核验自博弈训练、提示注入评测、案例和模型加固数据。
<https://openai.com/index/unlocking-self-improvement-gpt-red/>
- LangChain: 《New in LangSmith Fleet: Bring agents into Slack in one click》, 2026-07-15; 用途: 核验 Slack 部署、团队共享和人工批准机制。
<https://www.langchain.com/blog/new-in-langsmith-fleet-bring-agents-into-slack-in-one-click>
- LangChain: 《Agents need their own computer. Here's how to give them one safely》, 2026-07-15; 用途: 核验机器级隔离和运行环境建议。
<https://www.langchain.com/blog/agents-need-their-own-computer>
- LangChain: 《Introducing LangSmith Sandboxes: Secure Code Execu-

tion for Agents》，2026-03；用途：补充沙箱的临时环境和资源控制机制。

<https://www.langchain.com/blog/introducing-langsmith-sandboxes-secure-code-execution-for-agents>

- Hume AI: 《Introducing Real World VoiceEQ: Measuring the Human Quality of Voice AI》，2026-07-14；用途：核验人类评分规模和评价维度。
<https://www.hume.ai/blog/introducing-real-world-voiceeq-measuring-the-human-quality-of-voice-ai>
- Hume AI: 《Real World VoiceEQ Bench》，2026-07；用途：核验四类榜单及多维评价结论。
<https://www.hume.ai/rw-voice-eq>
- CData: 《CData Extends Connect AI to HIPAA-Regulated Healthcare》，2026-07-15；用途：核验医疗 AI 实时数据连接、治理和 HIPAA 控制。
<https://www.cdata.com/company/press/cdata-connect-ai-hipaa-regulated-healthcare/>

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>