

AI 技术每日分析：开放安全 AI 联盟成立，智能体竞争转向安全协同和企业运维

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 28 日星期二

摘要

7 月 27 日，AI 技术领域的重点由单一模型发布转向安全协同、智能体运行架构和企业落地基础。英伟达联合微软、IBM、Adobe、CrowdStrike、Hugging Face、Linux 基金会等机构发起开放安全 AI 联盟，试图通过开放工具、数据集和评测体系应对智能体安全风险。微软发布 Project Perception，把红队、蓝队和绿队智能体组织成持续感知、推理和处置的网络安全闭环。OpenAI 基于 80 多万条美国用户消息研究发现，43.5% 的职业专属 AI 使用已经跨出用户原有岗位边界。英伟达开源 NOOA 智能体框架，强调类型约束、对象状态、长期记忆和可测试软件工程。AWS 则从知识压缩与临时权限两个方向补齐企业 AI 的知识和运维底座。

目录

一、开放安全 AI 联盟成立：开放模型争论进入共同防御阶段	2
二、微软 Project Perception：网络安全智能体组成闭环	2
三、OpenAI 研究工作任务跨界：AI 开始重组岗位边界	3
四、NOOA 开源：智能体性能竞争转向“模型外部架构”	3

五、AWS 补齐知识与权限工程：企业 AI 从试验进入运维	4
--------------------------------------	----------

参考资料	4
-------------	----------

一、开放安全 AI 联盟成立：开放模型争论进入共同防御阶段

Reuters 7 月 27 日报道，英伟达牵头成立 Open Secure AI Alliance，参与方包括微软、IBM、Adobe、CrowdStrike、Hugging Face、Dell、Palantir、Cisco、Linux 基金会等科技企业和开源组织。联盟计划共同开发并开放 AI 安全工具、数据集、评测方法和红队能力，回应近期自主智能体越界和 Hugging Face 安全事件暴露出的风险。

联盟的政策含义值得关注。此前美国围绕开放权重模型的讨论主要集中在滥用风险和出口管制，新的联盟试图证明，开放模型也可以成为防御基础设施：安全团队能够检查工具链、复现攻击路径、共享检测数据，并在危机中快速部署本地模型。OpenAI、Google 和 Anthropic 没有出现在首批成员名单中，也说明行业对开放程度和责任边界仍存在分歧。

开放并不自动等于安全。联盟若要形成实际价值，需要建立模型和智能体组件的来源验证、供应链扫描、权限沙箱、行为日志和漏洞披露机制，并明确开放工具在高风险网络任务中的使用规则。

二、微软 Project Perception：网络安全智能体组成闭环

微软发布 Project Perception，计划于 8 月 3 日进入公开预览。该系统把安全信号、上下文、模型和专用智能体组织成持续学习的防御系统：红队智能体主动寻找潜在攻击路径，蓝队智能体结合环境上下文判断风险，绿队智能体在授权范围内执行修复与加固。

Project Perception 采用多模型架构，根据质量、可靠性、延迟和成本选择模型。首个场景是软件漏洞管理，专用模型 MAI-Cyber-1-Flash 被接入 MDASH 多模型智能体团队。微软称，该组合在 CyberGym 基准中取得 96% 的结果，比其对比方案高 12 个百分点，并把当前市场配置的成本降低接近 50%。这些数字属于企业评测，仍需独立验证。

更重要的变化是安全系统从“生成更多告警”转向“持续感知、推理并采取行动”。企业部署时必须保持人在回路，尤其是隔离终端、修改策略、封禁账户和修复生产系统等操作，应设置分级授权、证据链和回滚机制。

三、OpenAI 研究工作任务跨界：AI 开始重组岗位边界

OpenAI 经济研究团队分析了超过 80 万条美国 ChatGPT 用户消息。研究显示，16.8% 的工作相关消息、43.5% 的职业专属消息涉及原本与用户岗位不同的任务。排除通用写作和总结后，客户体验、设计、人力资源、法律和市场岗位的跨界任务占比分别达到 77%、75%、69%、56% 和 53%。

这种变化并不等于某个岗位被整体替代。更常见的情形是，销售人员开始自行分析客户数据，市场人员可以处理简单网站问题，小企业经营者能够起草文案、审阅合同和进行基础财务分析。过去需要跨部门交接的任务，被距离问题最近的人借助 AI 直接完成。

研究还发现，中小企业更容易出现任务跨界，因为其专业团队和内部服务资源较少。对企业管理而言，AI 培训不应只围绕提示词，而应同步调整岗位职责、审批流程、质量标准和责任归属。使用数据可以作为岗位变化的早期信号，但不能替代工资、招聘和生产率等正式统计。

四、NOOA 开源：智能体性能竞争转向“模型外部架构”

英伟达实验室发布 NOOA 开放研究预览，把智能体定义为一个 Python 对象：方法代表能力，字段代表状态，文档字符串代表提示，类型标注成为输入输出契约。模型通过编写和执行 Python 完成行动，而确定性逻辑仍以普通代码运行。

NOOA 总结六项智能体框架能力：类型化输入输出、对象按引用传递、以代码作为行动、可编程执行循环、显式对象状态，以及模型可调用的框

架 API。系统还使用可读的 SQLite 存储长期、类型化和关系化记忆，减少反复把完整对象序列化进上下文的 Token 消耗。

这项工作说明，同一个基础模型的表现会被智能体框架显著影响。上下文如何呈现、工具如何调用、状态如何保存、任务何时结束，可能带来两位数的评测差异。对企业而言，智能体应像普通软件一样接受代码审查、单元测试、版本管理、轨迹复现和回归评测，而不是依赖不断修改提示词维持效果。

五、AWS 补齐知识与权限工程：企业 AI 从试验进入运维

AWS 提出 Task-Aware Knowledge Compression (TAKC)，用于处理跨越数百份文档的尽职调查、合规审查等复杂任务。它不是在查询时只检索相似片段，而是提前针对不同任务压缩整个知识库，为财务分析、合规审查等任务分别保留不同信息。AWS 提供可部署在客户账户中的开源实现，并建议对压缩提示进行版本管理和审计。

在模型运维方面，Deepgram 将 AWS IAM Temporary Delegation 用于 SageMaker 上的自托管语音模型支持。客户可在自己的 IAM 控制台批准限定资源、限定时长的访问，无需创建长期密钥、跨账户角色或共享秘密；所有调用可进入 CloudTrail 审计。Deepgram 称，这把支持工单的初步调查由数天缩短到数分钟。

这两项实践分别解决“智能体如何理解大规模企业知识”和“外部技术人员如何安全进入客户环境”两个问题。企业 AI 规模化后，知识版本、权限时限、审计和支持流程的重要性不会低于模型能力。

参考资料

- Reuters: 《Nvidia forms industry alliance for open AI security after Hugging Face hack》，2026-07-27；用于核验联盟成立、参与机构和目标。

<https://www.reuters.com/business/nvidia-forms-industry-alliance-open-ai-security-after-hugging-face-hack-2026-07-27/>

- The Verge: 《Nvidia, Microsoft launch open AI security alliance》, 2026-07-27; 用于补充联盟成员和开放安全工具方向。

<https://www.theverge.com/ai-artificial-intelligence/971281/nvidia-open-secure-ai-alliance-cybersecurity>

- Microsoft: 《Rethinking security for the age of AI》, 2026-07-27; 用于核验 Project Perception 架构、预览时间和企业评测。

<https://blogs.microsoft.com/blog/2026/07/27/rethinking-security-for-the-age-of-ai/>

- Axios: 《Microsoft unveils new cyber model, agentic security tools to fight hackers》, 2026-07-27; 用于补充专用网络安全模型和行业评价。

<https://www.axios.com/2026/07/27/microsoft-unveils-new-cyber-model-agentic-security-tools-fight-hackers>

- OpenAI: 《How AI is expanding what people do at work》, 2026-07-27; 用于核验 80 万条消息、任务跨界比例及岗位差异。

<https://openai.com/index/how-ai-is-expanding-what-people-do-at-work/>

- NVIDIA Technical Blog: 《Six Agent Harness Capabilities for Higher Model Performance》, 2026-07-27; 用于核验 NOOA 设计和六项框架能力。

<https://developer.nvidia.com/blog/six-agent-harness-capabilities-for-higher-model-performance/>

- NVIDIA GitHub: 开源项目与智能体工具仓库; 用于补充相关代码、评

测和社区生态。

<https://github.com/NVIDIA>

- AWS: 《Beyond RAG: Task-aware knowledge compression for enterprise AI on AWS》, 2026-07-27; 用于核验 TAKC 技术路线和开源实现。

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>