

AI 技术每日分析：OpenAI Presence 推动企业智能体运营化，Codex 规模部署验证生产效率

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 26 日星期日

摘要

近期人工智能产品的变化集中在三个环节：个人智能体开始连接邮件、日历等常用服务；企业智能体转向权限、评测和持续运营；大规模组织应用开始积累可量化的生产效率数据。与此同时，多轮对话攻击暴露出传统单轮安全测试的局限，医疗等高风险领域也开始采用大样本真实交互研究评估 AI 系统。

这些动态表明，智能体的评价对象已经从单个模型扩展到“模型、工具、权限、运行环境和人工接管机制”组成的完整系统。模型能力仍然重要，企业部署的主要难点则逐渐落在任务定义、权限边界、过程监控和结果验证上。

目录

一、Meta 扩展个人智能体能力，邮件和日历成为任务入口	3
二、OpenAI Presence 把企业智能体做成持续运营系统	3
三、NTT DATA 将 Codex 扩大到约 9000 名员工	3
四、多轮攻击使智能体安全评测面临新的缺口	4

五、Google 用近 1.4 万人研究症状评估智能体	4
趋势判断	5
参考资料	5

一、Meta 扩展个人智能体能力，邮件和日历成为任务入口

Meta 近期升级 Muse Spark 1.1，推动 Meta AI 从对话助手向个人任务代理延伸。新版本可连接 Gmail、Google Calendar 等服务，用于整理每日信息、辅助研究、安排计划和推进个人项目。与独立智能体应用相比，Meta 的优势来自社交产品入口和用户上下文，短期内更适合高频、低风险、需要持续了解个人偏好的任务。

这一方向的关键不在一次完成多少任务，而在能否长期管理个人信息。邮件、日历、联系人和社交关系构成个人智能体最有价值的上下文，同时也是权限风险最集中的区域。接下来需要观察 Meta 是否提供细粒度授权、操作预览、任务撤销、数据隔离和跨应用审计能力。

二、OpenAI Presence 把企业智能体做成持续运营系统

OpenAI 发布 Presence，面向客服、保险理赔、账单处理和企业 IT 服务等业务，支持语音和文字智能体调用企业系统、执行批准范围内的操作，并在必要时转交人工。企业可以限定智能体使用哪些知识、访问哪些系统、执行哪些动作，以及在什么条件下必须获得批准。

Presence 的产品重点包括政策规则、护栏、升级机制、生产会话分析和持续评测。智能体上线后，企业可根据真实会话和人工接管记录识别能力缺口，再更新知识、流程和评测集。这类闭环意味着企业智能体将逐步形成类似软件运维的持续运营体系，采购对象也会从单纯的模型接口转向完整的智能体生产平台。

三、NTT DATA 将 Codex 扩大到约 9000 名员工

NTT DATA 披露，Codex 已扩展到约 9000 名技术和非技术岗位员工。一个关键系统的事故分析任务，过去需要 5 名经验丰富的工程师工作

3 天，使用 Codex 后在 30 分钟内完成。内部调查中，超过 96% 的受访者对 ChatGPT Enterprise 表示满意，超过 95% 报告了生产效率提升。

非技术人员也开始利用 Codex 处理多文件数据、生成分析报告、填写费用表单和编写轻量工具。NTT DATA 同步制定安全指南，明确允许使用的数据、可连接的系统、网络流量管理方式、沙箱模式、自动化等级和必须人工复核的环节。大规模部署的基础由模型能力、员工培训、内部卓越中心和治理规范共同构成。

四、多轮攻击使智能体安全评测面临新的缺口

Cisco 开展的研究对 15 个主流模型实施 6986 次多轮攻击，部分攻击的突破率达到 88.3%。攻击者并不依赖一条明显违规的指令，而是在连续对话中逐步改变上下文、试探边界并调整表达方式。

这一结果说明，单轮拒答率不足以代表智能体的实际安全水平。具备长期记忆、工具调用和自主规划能力的系统，需要记录风险在多轮交互中的累积过程，并针对权限提升、目标漂移、提示注入和跨工具数据泄露开展连续测试。企业评测集也应从静态问题库升级为包含状态变化和对抗反馈的任务轨迹。

五、Google 用近 1.4 万人研究症状评估智能体

Google Research 公布 SymptomAI 研究，通过一项包含 13917 名参与者的随机化研究，测试五种基于 Gemini Flash 2.0 构建的症状评估智能体。研究关注智能体如何在连续询问中收集信息、形成鉴别判断并向用户提供下一步建议。

这类研究的价值在于把 AI 评估从标准答案扩展到真实用户交互。对于医疗、教育、金融等高风险应用，结果质量与提问顺序、表达清晰度、用户理解程度、风险提示和转交专业人员的时机有关。未来高风险智能体的

评测将更多采用真实任务、大样本用户研究和长期结果跟踪。

趋势判断

第一，智能体竞争开始形成两条产品路线。个人智能体依靠用户入口、生活数据和跨应用连接提高使用频率；企业智能体围绕岗位任务、系统权限、责任边界和持续评测建立生产能力。

第二，模型正在成为智能体系统的一项可替换组件。企业更关注任务完成率、过程可追溯性、人工接管比例、单位任务成本和错误后果，而不只比较公开基准成绩。

第三，安全评测的基本单位将从一条提示词升级为完整任务轨迹。多轮上下文、记忆、外部工具和权限变化会产生新的风险组合，持续监控能力将成为智能体平台的重要基础设施。

参考资料

- Meta: 《Meta AI Doesn't Just Think, It Acts》; 2026 年 7 月 24 日。用于核实 Muse Spark 1.1 及个人智能体能力。
- Axios: 《Meta inches toward its agentic future》; 2026 年 7 月 24 日。用于了解 Meta 个人智能体的产品定位及竞争情况。
- OpenAI: 《Introducing OpenAI Presence》; 2026 年 7 月 22 日。用于核实企业智能体产品、业务范围和治理设计。
- OpenAI: Presence 生产部署说明; 2026 年 7 月 22 日。用于分析权限、护栏、人工接管和持续评测机制。
- OpenAI: 《NTT DATA Group cuts incident analysis to 30 minutes with Codex》; 2026 年 7 月 22 日。用于核实 9000 名用户及事故分析案例。

- OpenAI: NTT DATA 企业应用成效; 2026 年 7 月 22 日。用于核实满意度和生产效率调查结果。
- OpenAI: NTT DATA 安全治理实践; 2026 年 7 月 22 日。用于分析沙箱、数据、网络和人工复核要求。
- VentureBeat: 《Multi-turn attacks broke AI models 88% of the time》; 2026 年 7 月 24 日。用于分析多轮对话攻击风险。
- Google Research: 《SymptomAI》; 2026 年 7 月 22 日。用于核实大样本症状评估智能体研究。
- Hugging Face: 《Measuring the human quality of voice AI》; 2026 年 7 月 16 日。用于补充真实场景评测与传统基准差异。

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>