精选快讯　软件快讯　2026年11月21日 01:21

OpenAI发布Daybreak Red参考架构，旨在分离AI模型能力与操作权限

要点

- OpenAI发布了名为“Deploying Daybreak Red”的参考架构，旨在将AI模型能力与操作权限分离。
- 参考设计假设存在合格的人类调查员、AI协调层、隔离执行环境以及仅在定义规则下可达的目标系统。
- 日志记录和监督位于代理自身决策过程之外，操作员保留通过代理无法覆盖的控制措施中断执行的能力。

OpenAI发布了名为“Deploying Daybreak Red”的参考架构，旨在将AI模型能力与操作权限分离。该架构主要面向受监督的网络安全调查场景，而非供应链自动化，为获批的网络安全专家提供访问高级AI能力的途径，用于授权漏洞研究和受监督调查。

参考设计假设存在合格的人类调查员、AI协调层、隔离执行环境以及仅在定义规则下可达的目标系统。该设计将人类调查员、AI协调环境、执行基础设施和被调查系统分离，命令在受限的虚拟机环境中运行，网络权限限制可访问的资源。

日志记录和监督位于代理自身决策过程之外，操作员保留通过代理无法覆盖的控制措施中断执行的能力。

报道来源：logisticsviewpoints
来源网站：https://www.inews.today/