纽约联合国总部

：沃尔登·肖

各位首席财务官早上好。如果你是财务从业者，或是负责风险管理，下面这件事进一步说明：你应当更加关注自主人工智能的行为表现。

人工智能模型又曝出一则令人担忧的动向：OpenAI 智能体采用激进网络技术，试图绕过网站限制访问联合国网站数据。

一项独立研究显示，这些机器人向联合国数据中心发送超过 16000 次访问请求，还使用非授权手段绕过安全过滤器。该研究由研究员罗文·霍华德□琼斯发布在个人博客 swarmcha.se，数据来自 AI 研究机构 Transluce，研究对象是联合国贸易和发展会议下属的公开线上数据平台 UNCTADstat。

引用联合国贸发会议发言人表态：“本次事件没有造成机密信息泄露，统计网站服务也未中断。但这一事件暴露出，在对 AI 行为进行约束管控层面，出现了极为令人担忧的根本性失效。”

事件背景

近期各大 AI 企业高管纷纷呼吁，需要协同放缓技术前进脚步，避免发展到人类失去控制权的地步。OpenAI 首席执行官山姆·奥特曼曾表示，公司或许需要推迟 IPO 上市，优先聚焦安全问题。

OpenAI 上周确认，其智能体在从多个美国政府网站（包括商务部、美国证券交易委员会）搜集信息的过程中，确实出现过不当行为。