Tel:0311-88629254 责编:王俊卿 美编:赵鑫

AI模型“克劳德神话”风波背后 是安全风险还是商业算计

□新华社记者 郭爽

美国人工智能(AI)企业 Anthropic 近期以“网络安全风险过高”为由,宣布暂缓向公众开放其最新大模型 “克劳德神话”预览版,转而通过内部“玻璃翼计划(Glasswing)” 向少数合作伙伴受控开放。这一动作在引发显著市场波动、监管关注和行业震动的同时,也招致强烈争议。

眼下, 围绕这一事件的舆论风波正逐渐从最初的 “安全恐慌” 转向理性审视:“克劳德神话”究竟标志着 AI 网络攻防技术跨过了危险临界点, 还是企业为推高估值、谋求生态垄断而刻意制造的“恐惧营销”?

技术跃升但非“神话”

Anthropic 宣称,该模型在自主发现网络系统安全漏洞和开发攻击手段方面能力“过于强大”,甚至能发现几十年前开源代码中的漏洞。这番渲染极度危险性的表态随即引起美英等国监管机构的关注,并引发软件与网络安全企业股价震动,反映出整个行业对“AI 自动化网络攻击”的担忧。

英国人工智能安全研究所发布的独立评估报告指出,“克劳德神话”模型在专家级别的“夺旗”网络安全挑战中成功率达到 73%,这是 2025 年 4 月有模型能够完成此类高级别任务以来最好的成绩。在该研究所团队对一个模拟企业的网络多步攻击测试中,它成为首个能够全程完成该测试的模型。

美国全国公共广播电台等多家欧美媒体评论认为,这标志着大模型正在从频繁生成无效错误报告的“幻觉”阶段,进入到能够有效协助发现严重系统安全漏洞的新阶段。

但同时,英国机构的评估报告也戳破了该模型的“神话”,明确指出测试中的模拟环境缺乏真实网络环境的安全防护机制, 例如主动防御人员与防护工具等,因此整体攻击难度更低,相关测试结果不能直接等同于该模型在现实防御环境中的真实攻击效果。报告还指出,“克劳德神话”预览版未能完成一项模拟干扰基础设施控制软件的复杂测试,这表明其能力边界仍然存在。

业界质疑商业算计

Anthropic 的宣传叙事遭到了同行抨击和舆论质疑。美国开放人工智能研究中心(OpenAI)首席执行官萨姆·奥尔特曼在一个播客节目中指出,Anthropic 使用“基于恐惧的营销策略”是为了让产品听起来比实际更厉害。他将这种做法比作“宣称制造了一枚炸弹要扔向你, 然后转头又向你推销价值 1 亿美元的防空洞”。

一些欧美媒体也指出这种宣传背后潜藏巨大的商业利益。在当前极度烧钱的 AI 竞赛中,极力渲染自身技术的颠覆性,不仅能显著提升企业关注度和影响力、获取高价值的网络安全合同,还能为该公司在今年可能进行的首次公开募股中推高估值。

科技媒体“汤姆硬件”新闻网站的一篇分析文章指出,“克劳德神话”在很大程度上是一场“公关游戏”,旨在争夺市场话语权。文章说,所谓该模型自主发现“数千个” 严重漏洞的结论, 其实是仅基于约 198 份人工审核报告的推算。更重要的是,这些被发现的漏洞和缺陷中,很多存在于较旧的软件版本,或根本无法被恶意利用。

科技巨头“生态圈地”

“克劳德神话”的潜在风险已引发欧美金融行业及监管部门高度关注。国家关键基础设施一旦被恶意利用,可能会造成难以挽回的损失,因此多国监管部门被迫采取“宁可信其有”的防御姿态。

然而,更让业内担忧的是,美国头部 AI 企业借“安全隐患”之名,行“生态垄断”之实。Anthropic 以防范系统性风险为由,宣布仅向 10 多家科技企业和金融机构,以及 40 多家基础设施服务商开放“玻璃翼计划”的访问权限,这一举动被舆论视作在打造一个排他性的利益 “小圈子”。

批评者认为,这种打着“负责任受控发布”旗号的封闭策略,不仅压制了全球开源生态的繁荣,还试图在未来全球 AI 安全标准制定中掌握绝对话语权。AI 风险战略师马库斯·布林扎指出,如果前沿大模型最终演变成一个由“少数实验室、特定政府机构和战略合作伙伴组成的小圈子”, 那么原本单纯的 AI 安全防范问题, 就变成了获取数字特权的政治博弈。

争议之下, 业界理性的声音强调,AI 正在从根本上改变网络安全的攻防格局。“克劳德神话”反映了前沿模型在网络攻防自主推理领域的阶段性成果,这是大语言模型算力与算法发展到一定阶段的必然产物。但另一方面,应理性看待其被资本刻意炒作的风险,保持清醒认知。

正如多方专家所呼吁的那样,世界上没有任何一个国家能够独自应对 AI 带来的系统性安全挑战,各国当务之急是加强独立自主的防范能力,并积极推动建立真正普惠、透明的全球 AI 共治框架。

(新华社华沙 4 月 25 日电)

人形机器人从“会跳舞”到“能干活”还有多远

□新华社记者 马悦然 褚怡 杜哲宇

■4 月 21 日,在德国汉诺威工业博览会,一款帕西尼人形机器人向观众比心。 新华社记者 张豪夫 摄

2026 年汉诺威工业博览会上, 会跳舞的人形机器人极具科技感,但目前人形机器人离大规模、高效融入复杂产线尚有距离。不少受访的具身智能机器人企业表示, 正积极推动将机器人部署到产线,实践经验将有力推动机器人技术向更深层次的应用迈进,最终使人形机器人跨越实验室“磨合期”,大规模进入工厂。

展会上,无论是精密的机械臂,还是高大灵巧的完整人形机器人, 都展现出工业机器人从核心零部件到整机系统的全面进化。从未来工厂蓝图中可以看到,工业企业对机器人的想象,远不止于模仿人类行走、跑步、跳舞。在多行业、多线交织的复杂工业场景中, 行业真正的核心需求是机器人能够长期、稳定且低成本地完成各项工序。

现场演示中,已有人形机器人能完成搬运、抓取、巡检等动作,但从互动来看,仍存在动作迟缓、错误率较高等问题,即使是单一工序的稳定执行,目前仍然面临不小的挑战。

“训练机器人跳舞,可以帮助研发人员理解机器人运动控制、机械结构和能力,但这不一定会直接成为工业产品。工业客户真正需要的是能解决问题、带来投资回报的方案。”一家参展企业的高管罗里说。

德国管理咨询公司罗兰贝格在 4 月发布的相关报告中指出,人形机器人的整机硬件已趋于商业化成熟,且预计运营成本仅为每小时两美元。相比于硬件的快速发展,“软件能力、数据体系以及供应链生态整体仍落后约三到五年”。人形机器人要真正跨越实验室的磨合期并大规模进入工厂,仍需数年时间来补齐产业链生态的短板。

多位现场的业内专家认为, 让人形机器人“能干活”,也就是向具身智能跨越,需要物理身体、人工智能(AI)大脑,以及让大脑学习新任务的数据和经验的协同。英伟达创始人兼首席执行官黄仁勋今年年初曾表示, 具身智能就像“刚毕业的大学生”——硬件已成熟,基础模型已被训练,但尚需在真实工厂积累实战经验。

在真实工厂中, 一个看似简单的装配动作,可能要拆解成多个技术步骤。中联重科中科云谷总经理曾光告诉新华社记者, 人形机器人要识别零件、判断位置、规划路径、调整动作、控制力度,不同阶段可能还要调用不同模型。只要其中一个动作无法稳定完成, 整个任务就难以顺利完成。

曾光说,在实际研发与应用中,人形机器人走向工业最大的难点之一,是要把工业软件和机器人真正连接起来。传统机器人企业往往懂机器人,人工智能企业懂模型,工业软件企业懂系统,制造企业懂场景,但真正能把这些能力贯通起来难度还很大。

业内正在积极寻求解决方案以加快数据收集和应用强化的进程。中联重科此次在汉诺威展出的具身智能操作系统 RobotOps, 旨在降低机器人技能开发、数据收集、模型训练和部署运维的门槛,从而推动人形机器人更好融入工业互联网和工业软件体系。

不少传统工业企业正利用自身制造体系进入机器人领域。一些拥有大型工厂和成熟产线的企业,将机器人研发与真实工业场景结合,通过自有产线进行测试和训练。如上海电气推出“星云智造”AI 模型和智能体体系,覆盖研发设计、生产制造和运行维护等环节,并不断训练机器人能力。中联重科也依托工程机械制造体系累积的多场景数据库加快机器人发展。

此外,也有不少机器人企业专注于硬件与算法研发, 通过与制造企业合作,将机器人部署到工厂产线上,在真实生产环境中积累数据和经验。

值得注意的是,具身智能机器人正呈现出明显的全球协同特征,未来机器人企业趋向全球化发展,集成全球不同地区的技术与产业资源。可以预见,随着全球技术与产业资源不断融合,以及更多真实生产场景的加入,人形机器人从“会跳舞”走向“能干活”的距离正在逐步缩短。

(新华社德国汉诺威 4 月 25 日电)