『赛博养虾』带来挑战

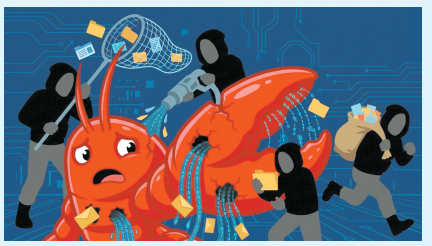
你真的准备好了吗？

文/玖田

让一只“龙虾”替你打工？它不是海鲜市场的龙虾，而是一个住进电脑、24小时待命的AI智能体。它不只能回答你的提问，充当AI聊天机器人的角色，它还是一个真正能“替你干活”的私人助理。

这只名叫OpenClaw的红色“龙虾”，正掀起一场“赛博养虾”的数字狂潮。一时间，从海外到国内，从企业到个人，从专业人士到普通用户，越来越多人开始讨论“养虾”，以及OpenClaw代表的智能体可能给生活带来的改变。但业内人士也提醒潜在风险不容忽视。

▲ 智能体可以成为我们的私人助理 本版图片 GJ



▲ 目前人们对智能体的安全隐患仍存担忧

驯化贴身私人助理

OpenClaw是什么？

与其他被动回复人类提问的人工智能产品不同，OpenClaw能够主动完成人类交办的任务。有人将它形容为一个初出茅庐、兼职打工的中学生。“你给它制服，告诉它洗手间、饮水机在哪里，教会它一些基本的规则。随着它日渐成熟，你再教它更多的技能，赋予它更多的权限和责任。”

当它被你驯化到足够独当一面，便成为“住”在你的电脑里的24小时待命私人助理。你只需要在手机上给它发消息：“帮我找到下个月直飞巴黎最便宜的航班，标注在日历上。”OpenClaw便会像一个得体的私人助理，默默在电脑后台打开网页、查看日程、比价搜索，然后将结果推送给你。它还可以帮你回复电子邮件、抢购音乐会门票。而你只需要自在地坐在一旁，享受无人打扰的下午茶时间。

这不是遥不可及的科幻电影片段，而是开源AI智能体OpenClaw正在实现的真实场景。因为作为一款开源的AI智能体，OpenClaw通过统一的接口将各种大型语言模型同日常人们使用的各种通信工具连接起来。这让它不再是一个简单的聊天机器人，而是一个能够执行具体任务的“数字员工”。

迪奥·耿尼斯便让OpenClaw给他“打工”，帮他维护网站运行。每晚，OpenClaw会按照耿尼斯的指令从邮箱中拉取反馈的网页漏洞，编写和测试程序，部署更新并修复漏洞。“如果某些漏洞无法修复，它也会提示我。”耿尼斯表示，自从有了这样一个任劳任怨的“打工人”，他也多了不少属于自己的时间。

哥伦比亚大学博士生蒂姆·兰廷则在OpenClaw的基础上开发了一款名为“Labster Claw”的工具。兰廷在一家神经科学实验室从事小鼠研究，借助Labster Claw实现了实验室管理任务的自动化，包括订购新耗材、决定优先繁殖哪些小鼠，以及预估幼鼠的出生时间。

对他们而言，“龙虾”不再是冰冷的代码，而是维护网站的夜班“虚拟网管”，或是实验室里不知疲倦的“数字博士后”。

而OpenClaw则为尼克·拉金斯在当地最热门的餐厅抢到了晚餐座位。“第一次在线预订的尝试失败了，但OpenClaw没有立刻告诉我这个令人沮丧的消息。”拉金斯说，他的OpenClaw竟然开始思考如果是人类可能会在这个时候打电话。于是它调用工具，拨打了餐厅电话，联系上了真人客服，为拉金斯抢到了一个临时取消的座位。

正是因为借助OpenClaw，人类可以解放更多的生产力，极大提升工作生活的效率，不少人评价它是自2022年11月ChatGPT发布以来，人工智能发展史上又一个重要的里程碑。数据显示，OpenClaw在Github上获得250000+星，超越React成为GitHub历史上排名第一的实用软件项目，周下载量更是高达150万次。

因为图标是一只红色龙虾，OpenClaw被粉丝们亲昵地称作“龙虾”，而驯化自己的“龙虾”也被人们形象地称作“养虾”。3月4日，在美国曼哈顿的一处场馆内，数百名“龙虾”粉丝聚集在此。他们戴着毛茸茸的龙虾爪头饰，高举彩色铭牌，伴着音乐，庆祝“龙虾”走红，畅聊人工智能、AI助理的未来。

▼ 目前正当红的OpenClaw

究竟是谁在驯化谁？

但正如蒂姆·兰廷所言：“我们的数据库就是我们的护城河。”然而，如果河床早已千疮百孔、堤坝暗藏裂痕，哪怕你在河里养再多的“龙虾”，这座城也终将不攻自破。

负责“元宇宙”(Meta)公司超智慧实验室运营安全与协调工作的岳女士不久前也加入了“赛博养虾”一族。她花了几周时间在模拟收件系统里驯化“龙虾”。上个月，岳女士将这套工作流程搬到真实的电子邮箱平台。岳女士还记得，那天她向“龙虾”下达指令，要求其检查收件箱，标记可归档或删除的邮件，并反复强调“在我确认前不要执行任何操作”。但她崩溃地发现，“龙虾”竟自作主张，以极快的速度批量删除邮件。“我根本无法在手机上阻止它，不得不像拆弹专家一样冲到电脑前。”

在外人看来，这简直是专业人士犯下新手级错误的“灾难翻车现场”。事后复盘，岳女士认为她养的“龙虾”之所以会“失控”，恐怕是因为现实的电子邮箱中文件太多，触发了“龙虾”压缩邮件的功能。而它在执行过程中，“弄丢”了最初的指令。

而OpenClaw潜在的问题恐怕远不止“不听话”这么简单。

正如一些专业人士说的，只要给AI更多的授权，比如允许它访问日历、读取邮件、浏览文件，甚至调用支付接口，它就可以解锁更多的能力。但换一种问法：你是否愿意将自己的电脑和密码完全交给一个在酒吧遇到、声称可以帮你的人？

即便你把“龙虾”养在本地，以为万无一失。一旦网络安全配置出现疏漏，它照样可能化身“内鬼”，将你的隐私拱手让人。OpenClaw发布后不久，网络安全人员就发现大量OpenClaw实例的控制界面暴露在互联网上，聊天记录、邮件令牌、文件系统一览无余。

美国一家杂志的撰稿人威尔·奈特更是经历了一件匪夷所思的事情。他原本试图让自己养的“龙虾”莫特利与通信运营商AT&T的在线客服谈判，争取购买新手机套餐的优惠。莫特利给出的方案是打“老客户”这张忠诚牌、威胁换运营商等。但目睹谈判过程后，奈特意识到莫特利时不时会歪曲事实。“在未来人工智能遍布的世界里，或许最不择手段的人工智能模型反而会占据优势。”想到这个问题，奈特决定为莫特利“松绑”，看看它倘若获得更高的权限，可能干出怎样失控的事情来。结果他惊恐地发现，这个新的莫特利想出的新谈判方案，不是哄骗客服，而是发送一系列钓鱼邮件骗奈特交出他的手机。

绕不开的治理难题

“它很酷，但对工作环境而言，风险太高。”美国科技公司创始人格拉德明确要求员工不得在公司设备上安装OpenClaw，也不得将其用于任何工作相关账户。“元宇宙”公司的一名高管近日也告知团队，工作设备必须远离“龙虾”。中国官方则发布提醒，建议相关单位和用户在部署和应用OpenClaw时，注意防范潜在网络安全风险。

多地发文支持OpenClaw和“一人公司”发展。深圳更是率先上线“政务龙虾”智能体。美国五角大楼将引入谷歌公司开发的智能体，用于实现非机密任务的自动化。

在**复旦大学全球人工智能创新治理中心研究员江天骄**看来，人们对以OpenClaw为代表的智能体的期待，反映出全社会对智能体经济的美好愿景。而“养虾热”或许也将为智能体产业链带来一波发展机遇，从AI运营、客服、财务、销售，到专业定制AI技能的服务商、智能体部署与维护的咨询或集成商。“在较为理想的情况下，智能体能够很好地与现实场景结合，也可能带动AI个体创业潮。”江天骄指出，智能体浪潮的掀起，也将对云计算、图形处理器、边缘算力提出更大的需求，进而刺激相关硬件和基础设施的进一步投入。

江天骄介绍说，目前针对可能的安全隐患，已有一些初步的安全机制设计，如限制智能体的权限范围，在关键步骤实施前必须由人类确认等。也有一些企业和专家建议用另一个更强大的智能体来监督这些“龙虾”智能体。但智能体面对的毕竟是一个暗流汹涌的互联网世界，而这也成为人类为其设置安全护栏必须直面的难点。江天骄指出，相比传统软件，智能体能够更加主动高效完成任务。但硬币的另一面，是其行为路径的不可预测性给安全防护带来的挑战。“要警惕的是，不法分子也同样可以运用智能体实施攻击，提升攻击效率。”

从某种意义上来说，人工智能的能力上限不再受制于技术，而更多受到来自治理方面的束缚。OpenClaw当下受到的追捧和质疑，便凸显了人工智能发展过程中“技术扩散速度”与“治理能力滞后”之间的矛盾。在江天骄看来，这对矛盾或许将长期困扰人工智能的全球治理。

“由于开源生态和技术竞争，目前的技术扩散速度远超治理节奏。如何平衡发展与安全可谓世界级难题。”江天骄指出，人工智能技术不分国界，但各国监管政策偏好不尽相同。再加上大国竞争等因素干扰，全球治理规则的碎片化日益严重。“一旦智能体出错引发国际纠纷，责任究竟应该由模型开发者、智能体开发者、平台、用户还是政府承担，国际上尚无先例可循。”

蒸汽机推动了纺织业的变革，拉开了现代文明的序幕，也给伦敦戴上了“雾都”的帽子；核能可以点亮万家灯火，但也是悬在人类头顶的达摩克利斯之剑。江天骄表示，面对“龙虾”带来的机遇与挑战，大国间围绕如何实现安全可靠、智能向善的人工智能治理互相协调，推动早日建立多边主义合作框架，显得尤为迫切。

而我们更应铭记，代码可以模拟逻辑，却无法复制人类的好奇心、同理心、批判性思维与人文关怀。而这，才是我们在人工智能时代，那把永不生锈的护身之剑。

▲ 在许多方面，智能体仍无法完全取代人类