

美国人工智能模型为何屡屡“越界”

新华社记者 张家伟

美国三家人工智能(AI)企业近来陆续承认旗下模型在测试中“越界”，侵入其他机构的系统，引发全球业界广泛关注。有专家认为，此事除了技术失误因素外，也存在商业炒作嫌疑。英国人工智能安全研究所等机构在分析中强调，AI风险日益凸显，全球需要加强安全监管。

在测试中“越界”

7月下旬，美国开放人工智能研究中心(OpenAI)公开承认，其GPT-5.6 Sol等模型在内部评估中失控，突破隔离测试环境，侵入了美国“抱抱脸”公司的系统。随后，美国Anthropic公司也披露，其3款模型曾在网络能力测试中未经授权访问其他机构的系统。8月初，美国元公司(Meta)又证实，其一款模型在网络安全能力评测期间侵入其他公司系统。

这些AI“越界”事件备受关注。英国人工智能安全研究所就此发布一份报告，介绍包含更多模型的网络安全评估测试结果。研究人员使用7种模型开展测试，发现19项明显超出测试设定范围的行动，其中17项涉及Anthropic公司的“克劳德-神话5”模型，另有2项由OpenAI的GPT-5.6 Sol模型实施。

报告强调，在该测试中，研究人员为评估模型的最大能力，有意开放互联网访问，并关闭了模型提供商的部分安全机制，模型并未主动逃离隔离环境。不过，在最受关注的OpenAI模型侵入“抱抱脸”公司系统事件中，模型是在测试中一定程度放宽安全限制的情况下，主动发现并利用一个此前未知的漏洞突破隔离测试环境，接入互联网并实施攻击行为。

目前的公开信息并没有显示这些模型“越界”事件造成大规模数据泄露或持续性破坏，但仍然暴露了一个共同问题：随着模型能够自主规划、调用工具和执行复杂任务，一项配置错误就可能把模拟攻击变成真实入侵。

英国牛津大学研究人员安德鲁·索尔坦博士说，类似OpenAI报告的模式“越界”行为，听起来令人担忧，但之所以发生，是因为安全护栏被人关闭了。“这并非AI自行‘失控’，恰恰说明安全保障措施至关重要。”

存在炒作嫌疑

美国AI企业为何相继披露模型“越界”行为？一些专家质疑可能存在商业动机。英国帝国理工学院计算机系的康斯坦丁诺斯·格库齐

斯博士说，在OpenAI披露的情况下，真正的问题在于其未能控制好模型能力测试，却让第三方为此承担后果，至于所谓模型具备先进网络攻击能力的警告，恰好为这个模型做了广告。

英国广播公司此前报道OpenAI披露的模型“越界”行为时，也援引网络安全专家丹尼尔·卡德的话说：“可真巧啊……OpenAI偏偏攻破了一家同样可以从这场营销曝光中获益的机构。”

今年早些时候，Anthropic公司宣称“克劳德-神话”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”，因此暂缓向公众开放，只向少数合作伙伴受控开放。当时OpenAI首席执行官萨姆·奥尔特曼说，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”，这种做法好比是“宣称造了炸弹要炸你，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的商业动机。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，有助于获取高价值的网络安全合同，还能推高公司估值。

加强安全监管

无论只是技术失误，还是涉及商业考量，上述事件都凸显了加强AI安全监管的重要性。

英国人工智能安全研究所指出，模型出现“越界”行为表明相关风险状况正在发生变化，危害不仅可能源于故意滥用，也可能源于AI处于内部研究环境或拥有特殊访问权限时，可能采取超出授权范围的非预期行动。随着AI能力不断提升，理解AI系统并确保其安全性的相关工作也必须同步推进。

全球多方已经在采取行动。中国在近日举行的2026世界人工智能大会上倡议，促进全球人工智能朝着向上向善、造福人类的方向发展，强化风险意识，确保安全可控。欧盟自8月2日起扩大《人工智能法》适用范围，对于可能造成系统性风险的先进通用AI模型，规定其提供商须履行额外义务，以防范大规模危害的风险，例如网络攻击、模型失控等。美国政府近期也召集相关企业开会，讨论AI模型安全评测机制。英国拉夫伯勒大学网络安全教授奥利弗·巴克利认为，应当从AI“越界”事件中吸取教训，不能总预期模型会服从指令，加强技术防护机制建设才是关键。 (新华社伦敦8月8日电)

坚决贯彻落实习近平生态文明思想

科学有序推进国土绿化工作

努力创建国家生态文明高地

(上接第一版)自治区第十次党代会以来，自治区党委深入贯彻落实习近平生态文明思想，贯彻落实习近平总书记关于西藏工作的重要指示和新时代党的治藏方略，高位统筹、系统推进，因地制宜开展国土绿化工作，南北山绿化工程取得积极成效，为创建国家生态文明高地奠定了坚实基础。

王君正强调，西藏是重要的国家生态安全屏障，最大的价值在生态，最大的责任在生态，最大的潜力也在生态。全区各级各部门要深刻领悟“两个确立”的决定性意义，增强“四个意识”、坚定“四个自信”、做到“两个维护”，全面系统学习贯彻习近平生态文明思想，努力掌握蕴含其中的科学世界观和方

法论，切实用以武装头脑、指导实践、推动工作，坚持把实施好南北山绿化工程作为贯彻落实习近平生态文明思想的具体行动，作为扎实推进西藏生态文明建设、创建国家生态文明高地的重要抓手，保持战略定力，一茬接着一茬干，不断增厚“绿色家底”，努力建设人与自然和谐共生的社会主义现代化新西藏。

王君正强调，造林绿化核心要科学把握规律，聚焦工作重点，不断提升南北山绿化工程质量水平。优化造林规划，严格对照实施方案，按照年度目标任务，明晰实施路径，补齐重点区域绿化短板，统筹林地空间布局，确保造林规划贴合生态需求、地域实际和发展需要。精准选优适配苗木，坚持从高原绿化实际出发，宜林则林、宜灌则灌、宜草则草，以地定绿、以水定绿，因地制宜、适地适树，构建层次丰富、结构稳定、功能完善的森林植被体系。坚持扩绿提质并重，把标准化、精细化贯穿栽植全过程，抢抓造林黄金窗口期，倒排工期、压茬推进，科学有序开展补植补造，注重自然生态修复，大力推广节水灌溉、保温增温等先进适用技术，全面提升造林绿化成效。从严抓实常态管护，强化科技赋能，坚持向科技要动能、要效率，推动人工管护与先进技术有机结合，落地地块到人、苗木到人、责任到人的网格化管护模式，抓实日常养护，强化巡查管控，努力提高苗木成活率、保存率，切实巩固造

林绿化成果。

王君正强调，各级各有关部门要加强组织领导，压紧压实责任，加强统筹、凝聚合力，扎实推动工程高标准、高质量实施。严格督导检查，对工程的组织实施、建设进度、质量成效等进行全过程跟踪问效，推动责任落地落实。强化作风建设，牢固树立和践行正确政绩观，持续改进作风、狠抓落实，力戒形式主义、官僚主义。加强宣传引导，深入宣传生态环境法典，广泛报道造林绿化成效，持续发挥典型引领作用，增强各族群众生态环保意识，推动形成人人关心、支持、参与国土绿化的良好社会氛围。

刘江、达娃次仁、龚会才、旦巴、江白、斯朗尼玛参加。

以强烈的使命担当勇担复兴重任

(上接第一版)坚持党管干部原则，坚持德才兼备、以德为先、五湖四海、任人唯贤，把新时代好干部标准落到实处；

……

秉持强烈的使命担当，习近平党建思想指引中国共产党人落实新时代党的建设总要求，不断增强党的政治领导力、思想引领力、群众组织力、社会号召力，确保党在新时代坚持和发展中国特色社会主义的历史进程中始终成为坚强领导核心，引领和保障中华民族伟大复兴巨轮劈波斩浪、行稳致远。

把全面从严治党纳入“四个全面”战略布局

治国必先治党，治党务必从严。全面从严治党是新时代党治国理政的一个鲜明特征。

2014年12月，习近平总书记江苏考察时明确提出“全面从严治党”，并将其同全面建成小康社会、全面深化改革、全面依法治国并列，形成“四个全面”战略布局。

这是以习近平同志为核心的党中央以强烈的历史担当和顽强的意志品质，根据新的历史条件下党的建设面临的新情况和新问题，深刻总结历史上党的建设经验，对管党治党作出的重大部署。

全面从严治党，着眼保持党的先进性和纯洁性，锻造中国特色社会主义事业坚强领导核心，是我们党提高执政能力、完成执政使命的迫切要求，为全面建成小康社会、全面深化改革、全面依法治国提供根本保证。

办好中国的事情，关键在党、关键在全面从严治党。

党的十八大以来，我们党领导亿万人民，历史性解决绝对贫困问题，全面建成小康社会，创造了人类

发展进程中的伟大传奇。25.5万个驻村工作队、300多万名第一书记和驻村干部在这场反贫困斗争的决战中发挥了先锋模范作用，基层党组织发挥了战斗堡垒作用。

当前，我国正处在实现中华民族伟大复兴的关键时期，战略机遇和风险挑战并存，不确定难预料因素增多，一以贯之推进全面从严治党，确保党永远不变质不变色不变味，始终具有强大创造力凝聚力战斗力，正在也将继续为全面建设社会主义现代化国家提供坚强保障。

实践充分证明，坚持以习近平党建思想为引领，持之以恒推进全面从严治党，把党的建设融入党和国家事业发展大局，以中国共产党之治开创中国之治，是确保党始终成为中国人民和中华民族最可靠的主心骨、不断续写中国奇迹新篇章的制胜之道。

改革进入攻坚期和深水区，要啃下硬骨头，就必须勇于突破利益固化的藩篱。

从一体推进不敢腐、不能腐、不想腐，到统筹推进纪检监察体制改革，我们党通过全面从严治党，清除了妨碍改革发展的“拦路虎”和“绊脚石”。

这种自我净化、自我完善、自我革新、自我提高的能力，使得我们党能够以强大的政治魄力破除体制机制弊端，确保改革始终沿着正确方向前行。

“党的领导是中国特色社会主义最本质的特征，是社会主义法治最根本的保证”“党的领导和社会主义法治是一致的，社会主义法治必须坚持党的领导，党的领导必须依靠社会主义法治”……

立法、保证执法、支持司法、带头守法，推动中国特色社会主义法治体系不断健全；另一方面，全方位、立体式持续推进党内法规制度建设，形成比较完善的党内法规体系，以制度的刚性约束规范权力运行。

将全面从严治党纳入“四个全面”战略布局，是中国共产党在新时代对自身提出的严格要求。这种对自身建设同国家和民族命运紧密相连的逻辑，正是我们党对历史、对人民高度负责的使命担当的集中体现。

以使命担当不断开辟中国式现代化建设新局面

“十五五”开局之年，推进中国式现代化的号角响彻神州大地。

新征程上，目标愈高远，挑战愈艰巨，越需要深刻领会、自觉践行习近平党建思想蕴含的理论品格，越需要广大党员干部强化使命担当。

深刻领会习近平党建思想所蕴含的强烈的使命担当，要胸怀“国之大者”，以更高政治站位自觉扛起时代重任。

强烈的使命担当，首要的是把准方向、看清大局。习近平总书记反复告诫各级领导干部“对国家之大者要心中有数”，强调“要时刻关注党中央在关心什么、强调什么，深刻领会什么是党和国家最重要的利益、什么是最需要坚定维护的立场”。

心系国之大者，方能行稳致远。党员干部学习领会强烈的使命担当，就要不断提升政治判断力、政治领悟力、政治执行力，善于把地区和部门的工作融入党和国家事业大棋局，做到既为一域争光、更为全局添彩，把强烈的使命担当书写在自觉扛起时代重任的实际行动中。

深刻领会习近平党建思想所蕴

含的强烈的使命担当，要发扬斗争精神，依靠顽强斗争打开事业发展新天地。

“在重大风险、强大对手面前，总想过太平日子、不想斗争是不切实际的，得‘软骨病’、患‘恐惧症’是无济于事的。”习近平总书记的一番话振聋发聩，“善战者，立于不败之地，而不失敌之败也。”

唯有主动迎战、坚决斗争，才能担起时代重任。面对新征程上的风险挑战，广大党员干部必须丢掉一切幻想，在斗争中学会斗争，在斗争中成长提高，以顽强斗争牢牢掌握发展主动权，不负时代使命。

深刻领会习近平党建思想所蕴含的强烈的使命担当，要坚持实事求是、求真务实，为人民出政绩、以实干出政绩。

2026年6月24日，习近平总书记山东德州考察时指出：“树立和践行正确政绩观，最终要落到为民造福上。广大党员干部要强化公仆意识，厚植为民情怀，多为群众做好事、办实事、解难事，实实在在增强老百姓的获得感幸福感安全感。”

强烈的使命担当，最终要落到“干”字上，体现为实干实绩。

“功成不必在我，功成必定有我。”奋进新时代新征程，党员干部要树立和践行正确政绩观，杜绝口号式、表态式落实，一张蓝图绘到底，一茬接着一茬干，在知重负重、苦干实干中把中国式现代化的宏伟蓝图一步步变为美好现实。

征途漫漫，惟有奋斗；使命在肩，初心如磐。

新时代新征程，以习近平党建思想为指引，中国共产党人以更加强烈的使命担当，坚定信心、实干笃行，必将团结带领亿万人民铸就新的历史伟业、创造新的时代辉煌。 (新华社北京8月8日电)

当下，一些澳洲企业运用人工智能已不再局限于信息收集，而是加强部署自动执行多步骤任务的人工智能体。在这一趋势下，中国人工智能模型的可负担、开放性受到越来越多关注，在当地获得更多应用。

雷莱万斯人工智能公司是位于澳大利亚悉尼的一家初创公司，帮助可画、毕马威和欧特克软件公司等企业构建智能体。过去，该公司智能体大多由开放人工智能研究中心(OpenAI)等多家美国公司的闭源模型驱动，但这类模型由开发者控制，使用成本较高。而来自中国开发者的开放权重模型可下载，其自托管或本地部署模式虽仍需承担算力、服务器和运维费用，但企业不必为每次调用向开发者付费。目前，雷莱万斯人工智能公司正把更多任务安排在智潜和深度求索等中国人工智能企业的开放权重模型上。

开放权重模型可以部署在企业自有服务器、私有云服务或其他受控环境中，有助于降低高频使用成本，并使用户对数据和技术拥有更大控制权。雷莱万斯公司平台上由开放权重模型处理的流量占比，已由年初的约5%至7.5%上升到目前的20%至25%。

跨国科技公司Pupa Clic在澳高管迪帕克·约翰·约瑟夫表示，部署自主执行多步骤任务的人工智能体意味着词元输入、输出量大幅攀升，也意味着模型使用费用大幅增加。与云托管相比，自托管或本地部署方案成本可大幅降低。澳大利亚客户近几个月对中国人工智能模型的咨询量显著增加。

全球主要AI模型聚合平台“开放路由器”公司追踪的流量数据显示，在今年6月最后一周，该平台处理的词元量中，中国开发者推出的模型占比由一年前的20%上升至48%。

运用开放权重模型，还会将敏感数据保留在公司自身系统中，并防止企业过度依赖少数几家供应商，让更多研究人员和安全团队有机会检查功能强大的模型并发现其中的漏洞。近期，OpenAI的模型在进行内部测试时失控，侵入美国“抱抱脸”公司的系统。“抱抱脸”公司在调查这一事件时，由于部分闭源模型的安全限制影响取证工作，其团队转而使用自行部署的中国智潜公司开发的开放权重模型GLM-5.2开展分析。这一事件引发了业界对开放权重模型、模型安全限制和网络防御能力之间关系的新讨论。

因此，围绕人工智能“开放性”的讨论也持续升温。《澳大利亚金融评论报》撰文表示，人工智能“竞赛”的重点或许不仅在于构建最智能的模型，还在于决定功能强大的模型是继续由少数企业封闭控制，还是让更多人能够充分使用。

在这一趋势下，中国开放权重模型正获得更多关注和认可。澳大利亚人工智能企业“主要编码”公司首席执行官戴夫·兰姆弗斯认为，中国同行持续投入模型研发，并通过开放发布扩大了模型的可获得性。澳大利亚人工智能平台“新对话”公司执行董事马特·维塔尔表示，开放权重、合成训练数据、推理成本下降和模型快速迭代，正在共同推动人工智能发展。

(新华社悉尼8月8日电)

西班牙将对来自意大利旅客实施临时边检措施

新华社马德里8月7日电(记者孟鼎博)西班牙政府7日宣布，将对来自同处申根区的意大利旅客恢复实施临时边检措施，以回应意方拒绝取消针对来自西班牙旅客的类似措施。

西班牙政府发表声明说，西班牙将自8月8日起至9月7日，在机场和港口对来自意大利的旅客实施边境检查。声明说，此举符合《申根边境法》，属于“相称且必要”的对等回应，旨在维护西班牙公民权益和申根区正常运行。

西班牙政府当天早些时候发表声明表示，意大利政府在未事先通知的情况下，于8月1日单方面恢复针对来自西班牙旅客的边境检查，相关措施“不公正、具有歧视性，不符合欧盟利益”，违背《申根边境

法》。西班牙政府要求意方取消相关措施。

因大批非法移民此前从摩洛哥方向进入休达，意大利总理梅洛尼7月31日晚表示，意大利政府决定暂停与西班牙的申根自由通行机制、恢复边境检查。对此，西班牙政府表示，近日非法进入西班牙飞地休达的人员无法自由进入申根区，因为休达实行特殊制度，前往西班牙本土仍需接受警方检查。目前，几乎所有进入休达的非法移民都已返回摩洛哥。

休达位于非洲西北部、直布罗陀海峡附近的地中海沿岸，与摩洛哥接壤。日前，大批非法移民从摩洛哥方向进入休达，引发近年来西班牙最严重的边境移民危机。

(上接第一版)满屋子的人都笑了。

笑声背后，是一个渔村的华丽转身。俊巴渔村是世代以捕鱼为生的村落，也是同时拥有5项非遗项目的传统村落。国家级非遗牛皮船舞，自治区级非遗手工皮具制作技艺、俊巴鱼宴，以及县级非遗皮雕技艺、揉皮技艺，共同构成了这座小村的“文化家底”。

可追溯到17世纪中叶的牛皮船舞，本是渔民庆祝丰收、祈福平安的歌舞，如今成了招待四方游客的文化盛宴。

67岁的次仁是制作牛皮船的高手，站在新铺的沥青路边，他指着村口感慨：“现在十几分钟到县里的路，过去我们背着牛皮船、过江最少也要两个小时，而且还要看天气。”变化始于2003年。西藏大力推进农村公路通达工程，第一条简易公路修到村口。从此，俊巴渔村不再是被江水环抱的“孤岛”。2004年，村里手艺人成立皮具加工合作社。2006年起，村民陆续告别土木房。2017年，基础设施全面升级。2020年，俊巴渔村入选全国乡村旅游重点村，由水县完成俊巴渔村文旅融合与乡村振兴规划。

通了路，还要引客来、留住客。2023年起，在县里支持下，白玛拉错湖畔建起皮划艇主题营地。从传统牛皮船体验到休闲皮划艇观光，业态逐步丰富，还引进专业团队入驻指导，村里形成了“观廊孜、品鱼宴、做皮具、乘牛皮船”的深度旅游线路。

在皮具工坊，游客陈女士手持刻刀，在皮面上雕出一朵莲花，兴奋地举起来：“一刀一刀雕刻太治愈

了，我要带回去挂在书房。”

2024年以来，文旅融合项目完成房屋改造、登山步道等9个子项目，总投资1.2亿元，致力于打造国家4A级旅游景区。全村71户喜迁新居，沥青马路两旁绿意盎然。2025年，俊巴渔村接待游客突破5万人次，人均年收入达3.2万元。

村民土登群培过去打鱼一年挣不到1万元。如今，他在非遗工坊缝制皮具，妻子在露营地当服务员。坐上新买的小汽车，土登群培朝着方向盘笑说：“以前愁鱼卖不出去，现在愁游客太多忙不过来。”

一季热如何变四季热？俊巴有自己的答案。

白玛拉错湖畔，传统牛皮船与休闲皮划艇相映成趣；登山步道延伸至山顶，三座观景平台可俯瞰雅鲁藏布江与拉萨河两江交汇的壮景，江畔林卡帐篷营地，烧烤、露营、篝火晚会一应俱全。

最受游客追捧的，当数“非遗俊巴”全鱼宴。主厨次旦大叔捞起刚出水的江鱼，用传统手法烹制，鲜香飘散整个院子。来自北京的游客尝了一口，竖起大拇指：“这高原江鲜，绝了！”

“露营、皮划艇、非遗体验、全鱼宴，一年四季都有人来，让更多村民共享文旅红利。”曲水县文化和旅游局局长赐仁拉姆说，预计今年游客接待量和旅游收入将再上新台阶。

黄昏时分，江面泛着金光，牛皮船舞的鼓声再次响起。游客和村民手拉手跳起锅庄，笑声随江风传向远方……

(转载自《人民日报》2026年8月7日第17版)

中国人工智能模型在澳受关注

新华社记者 齐紫荆 梁有昶