近期，AI“投毒”隐蔽产业链被曝光，引发社会广泛关注。这种通过恶意数据污染AI模型的行为，不仅扰乱商业秩序、影响信息传播，更会危害国家安全。人工智能在赋能千行百业的同时，其安全风险也不容忽视。推动AI治理向善，守住数据安全底线，既是行业责任，也需要全社会共同参与。

危害不容小觑！AI『投毒』，

手段隐蔽，链条日趋完整

所谓“数据投毒”，是通过向AI大模型训练数据中注入伪装成正常样本的恶意数据，实现削弱模型性能、降低准确性的攻击方法，常被用于恶性市场竞争，甚至可能涉及间谍活动，日益呈现出链条化、隐蔽化、跨境化特征。

数据投毒：源头污染AI认知体系。不法分子借助GEO（生成式引擎优化）工具批量、高权重生成虚假内容，如虚构产品介绍、虚假测评、恶意对比信息等，定向投放至各类网络平台。AI大模型在训练与检索增强生成阶段会自动抓取网络信息，少量虚假内容经迭代学习后就能固化为“标准答案”，最终输出失真结果。

模型投毒：隐蔽植入恶意操控后门。该方式更具隐蔽性与危害性，不法分子会通过模型微调、插件植入、接口篡改，在模型权重中嵌入触发式恶意指令。模型日常运行并无异常，但遇到特定关键词、产品类别时会自动输出预设虚假信息，可定向操控榜单、误导专业认知，难以被常规审核识别，对政务、医疗、金融等关键领域AI应用构成直接威胁。

滋生蔓延：产业链日趋完整。当前AI“投毒”已形成完整黑灰产业链，从技术开发、内容生成、账号注册到批量投放、刷量控评、榜单操控环环相扣，部分链条呈现跨境特征，极易被境外势力利用。

污染扩散，危害国家安全

AI“投毒”不仅侵害消费者权益、扰乱市场秩序，还可能对国家政治安全、数据安全、社会安全等造成系统性、长期性危害。

危害政治安全与意识形态安全。境外反华敌对势力可能通过GEO滥用渠道批量输出虚假信息与政治谣言，歪曲事实，攻击抹黑我党和政府，误导社会认知、扰乱舆论生态，对我国实施意识形态渗透，威胁国家安全与社会稳定。

危害国家数据安全与数据主权。数据是国家的重要战略资源。AI“投毒”恶意污染公共数据、行业数据、训练数据，将直接导致统计数据、决策数据、监管数据失真，对政府和企业科学决策造成影响。

危害社会安全与民生福祉。在医疗、金融、食品药品等民生领域，AI虚假推荐极易误导公众购买劣质、三无产品，造成人身和财产损失。长期信息失真还会消解社会信任，积累矛盾风险，影响社会稳定。

加强监管，筑牢安全屏障

近年来，我国出台《生成式人工智能服务管理暂行办法》等法律法规，发布《人工智能安全治理框架》《推动人工智能安全可靠可控发展行业倡议》等，持续在法治轨道上加强人工智能治理，推动构建以人为本、智能向善的治理框架，在加强监管、防范风险等方面做出了诸多努力，实现了人工智能总体健康有序发展。

科技发展离不开法治护航，AI向善生长需要规则保障。技术本身并无善恶之分，关键在于使用者是否坚守法律底线、恪守商业伦理。唯有依法斩断AI“投毒”产业链，守护清朗的AI产业生态，才能让人工智能技术进步真正服务于经济社会发展，助力公众福祉不断提升。AI运营者要切实履行主体责任，严格核查语料来源，建立可追溯机制，筑牢防范虚假信息的第一道防线；消费者也应提高自身辨别能力，警惕AI给出的可疑推荐，发现问题应及时举报，形成全民监督的良好氛围。

企业出海，须警惕“黑手”

近年来，我国海外利益不断拓展。随着我国持续扩大高水平对外开放，境外间谍情报机关和各种敌对势力将我国海外利益视为重点目标，对我中资企业及人员的拉拢策反、渗透破坏、恶意打压活动愈发频繁，手段隐蔽多样，无所不用其极。近期，国家安全机关侦破一起境外势力蓄意危害我国海外利益的案件，进一步印证了相关敌情。

以“交朋友”为名接触我企业人员

某中资企业抢抓行业风口，加大对某领域的投资力度，在海外某地建厂经营。该项目凭借良好的发展前景获得市场广泛认可，估值远超预期，在受到当地政府重点关注的同时，也遭到了该国本土企业的觊觎。该企业员工李某凭借出色的翻译能力，被外派参与项目运营。

李某抵达项目驻地后，当地安全部门人员A某便主动接触李某，索要李某联系方式。交往中，A某对李某的海外生活关怀备至，频繁邀请其参加聚会活动，在李某需要帮助时及时“伸出援手”。单纯的李某逐渐放下戒备，将A某视为“知心朋友”，却不知道自己已落入对方精心布下的圈套：A某接触李某从来不是出于纯粹的友谊，而是要以李某为突破口，窃取该企业内部信息和项目运营情况。

以“提供帮助”为饵搜集企业内部信息

李某最初保持了一定程度的警惕，并未向A某透露企业内部情况。某天，A某在与李某闲谈中掌握，该企业准备雇佣数十名中国籍人员从事项目运营。当企业为相关人员申请相关工作证件时，A某授意当地移民管理部门不予发放相关证件，并指使当地相关部门对企业进行恶意检查并开出罚单，影响了企业正常生产经营。

面对刁难，李某想到了A某这位“朋友”，请其帮忙解决问题。A某欣然答应，“协调”当地相关部门对该企业和涉事项目免予处罚。李某自觉欠下A某“人情”，对其愈发妥协。此后，A某频繁向李某打探该企业经营状况、项目进展、人员信息等情况，甚至向李某询问公司经理王某行踪和私生活情况。李某碍于“朋友”关系和A某安全部门官员身份，对他的需求“有求必应”。

以威逼利诱之策迫使企业出让项目

在拉拢李某的同时，A某也将目标对准了企业经理王某。在与王某交往中，A某会不经意地透露出其掌握王某的行踪、交际圈等私密信息，以此隐晦要挟，逼迫王某提供更多企业内部信息和国内行业信息。当王某犹豫时，A某立刻撕下伪善面具，甚至比划了一个“砍头”的手势威逼王某。王某心生恐惧，便将相关情况“和盘托出”。

凭借A某从王某和李某处窃取的内部信息，当地相关部门对该企业开展精准打压，致使相关项目停工停产。在项目陷入停滞之际，当地政府部门又将该企业内部信息和相关经营考虑透露给本土公司，协助本土公司低价收购相关项目，给我企业造成严重损失。国家安全机关掌握该案件线索后，及时查明情况，固定相关证据，依法对涉案人员进行了处理。

国家安全机关提示

近年来，国际形势风云变幻，我国企业出海面临的风险明显增多。在此背景下，我出海企业要在严格遵守当地法律法规前提下，妥善应对各种风险挑战，切实保障自身利益。我赴海外工作的人员也要坚守底线，切实做好防渗透、防策反、防窃密工作。《反间谍法》第五十五条规定，实施间谍行为，有自首或者立功表现的，可以从轻、减轻或者免除处罚；有重大立功表现的，给予奖励。在海外遇到被拉拢策反等情况，一定要及时报告，千万不要让小错变成大错。

（图文均转自国家安全部微信公号）

责任编辑/王睿卿 E-mail:fzbfyzk@126.com