全国政协委员、奇安信集团董事长齐向东：

# 为AI预设“安全护栏”非常有必要

本报记者 卢梦琪

最近几年，AI技术发展日新月异，不论是大模型、智能体还是具身智能，安全问题都如影随形。全国政协委员、奇安信集团董事长齐向东在接受《中国电子报》记者采访时表示：“没有安全的创新是走不远，也走不稳的。从这两年大模型、智能体、具身智能等各种AI形态遭受的网络攻击来看，为AI预设‘安全护栏’非常有必要。”

在齐向东看来，安全和创新不对立，而是一体两翼。要实现“创新和安全的动态平衡”，关键有三点：一是把安全能力嵌入AI应用的全生命周期，做到纵深防御；二是明确合规红线，夯实安全主体责任，强化权限与内容管控；三是用AI对抗AI，让安全能力始终比安全风险快一步。

> AI技术发展日新月异，安全问题却如影随形。没有安全的创新是走不远，也走不稳的。为AI预设‘安全护栏’非常有必要。

对于具身智能而言，所面临的新安全挑战主要集中在三个层面。齐向东表示，一是决策层，数据投毒、数据库被污染、模型“幻觉”、框架缺陷等，会导致具身智能“大脑”混乱，造成决策混乱失真；二是执行层，具身智能的“四肢”由智能体和硬件“执行体”构成，软硬件供应链复杂，终端漏洞、云平台、API接口等处处都是攻击突破口，而这些口子一旦被攻破，会直接影响具身智能现实操作，让它做出危险动作、违规执行指令等；三是物理层，具身智能物理暴露的特性容易遭到临机攻击，可能被近距离劫持信号、植入程序，造成具身智能被非法控制。

解决具身智能可能面临的这些安全问题，齐向东提出了三个建议：第一是针对具身智能做专属安全防护。目前，奇安信已经探索构建了“端-网-云-机”一体化的防御体系，并聚焦攻击入口探测、威胁场景构建、CVE漏洞验证等核心环节，开展渗透测试，为具身智能的安全检测提供强有力的技术服务支撑。二是保障数据资产安全。关键是以数据资产为中心，形成集“事件监测、风险分析、策略调整、访问控制”为一体的全链条闭环体系。三是锻造高水平的实战能力。实时发现、精准研判、快速溯源、有效阻断各类网络攻击，做到全链路精准反制与立体式纵深防御。

齐向东也表示，这两年，奇安信大力推进AI安全防护和AI安全应用，围绕具身智能、Agent、大模型，构建了创新的防御体系，切实保障AI全场景安全。除此之外，奇安信还积极开展AI安全应用，推进全产品AI化，奇安信的智能运营平台，研判准确率已经达到95%以上，安全事件响应时间从传统“天”级缩短到了“分钟”级。

全国政协委员、天娱数科董事长贺晗：

# 构筑人工智能国家级技术护城河

本报记者 卢梦琪

过去几年，我国在AI大模型应用落地、算力基础设施建设及数据要素市场化方面取得了显著成效，构建了具有中国特色的AI应用生态。面对通用人工智能（AGI）的加速演进，全国政协委员、天娱数科董事长贺晗在接受《中国电子报》记者采访时表示，我国人工智能产业的优势更多集中在“从1到100”的应用创新与场景赋能上，我们要改变目前“跟随式迭代”的研发现状，在大力搞AI应用的同时，也要关注“从0到1”的基础创新，真正掌握核心技术，构筑人工智能国家级技术护城河。

在贺晗看来，我国人工智能产业底层原创创新能力不足，存在“跟随式迭代”的技术代差风险。虽然模型参数量和评测跑分不断攀升，但绝大多数底层框架和基础算法依然建立在海外开源生态的“底座”之上。贺晗坦言，这种“跟随式”研发模式导致我们始终存在半步到一步的技术代差，一旦海外顶尖开源模型闭源或对华限制使用，我国部分建立在海外底层架构上的应用生态将面临“釜底抽薪”的风险。

> 在大力搞AI应用的同时，也要关注‘从0到1’的基础创新，真正掌握核心技术，构筑人工智能国家级技术护城河。

贺晗表示，我国人工智能产业的算力生态“硬强软弱”，缺乏统一的异构计算底层软件栈。国内算力芯片呈现碎片化，各自为战，缺乏一个能够跨硬件、跨厂商统筹调度的“国产统一算力软件栈”（如高效的编译器和算子库），这导致开发者适配国产算力的成本极高，算力利用率低下，“有算力但不易用”。

我国人工智能“耐心资本”供给不足，制约了基础大模型的长周期探索。通用人工智能的研发是一场高投入、高风险、长周期的极限挑战，对此，贺晗表示：“我国资本市场相对偏好短期商业变现，即便是政府引导基金或国资背景的产业基金，也普遍面临严格的年度考核与保值增值压力，导致市场缺乏真正敢于押注基础技术突破、容忍高失败率的‘耐心资本’。”

此外，目前我国在全球开源生态的话语权较弱。“我国虽然也有优秀企业开源了模型，但在构建具有全球号召力的国际化开源社区方面仍显薄弱。”贺晗表示。

针对以上难题，贺晗提出了四项建议。一是改革科研评价与国资考核机制，大力培育“耐心资本”。贺晗建议出台针对人工智能基础研究的专项投融资政策。对主投AI底层算法、新型计算架构等“硬核科技”的政府引导基金和国有创投机构，大幅拉长考核周期，建立完善的“尽职免责”与“容错机制”。鼓励社保基金、保险资金等长期资本有序进入AI基础研发领域，为跨越周期的技术“无人区”探索提供充足的弹药保障。

第二，要集中力量攻坚“国产算力统一软件生态”，打破软硬解耦壁垒。建议通过重大科技专项，组织国内头部大模型企业、算力芯片厂商及顶尖高校联合攻关，打造国家级的“统一异构算力软件底座”。强制推进国产硬件和国产大模型框架的深度适配与标准化，降低开发者向国产算力迁移的门槛。通过政府首购、服务券补贴等方式，强力推行全栈国产化算力方案在关键行业中的应用，以规模化市场反哺国产软件栈的迭代。

第三，要实施“前沿探索特区”计划，强化“从0到1”的原创技术策源。贺晗表示，要改变过去资源过度集中于“打榜跑分”的现状，建议设立“AGI基础架构先导特区”，定向资助非Transformer架构、类脑计算、物理世界具身大模型等具有颠覆性潜力的前瞻性研究。赋予顶尖科学家团队绝对的技术路线决定权和资金使用自主权，在科研体制上给予最大程度的松绑，全力孵化属于中国本土的底层原创技术。

第四，以国家力量背书，打造具有全球影响力的开源AI基础设施。贺晗表示，建议由相关部门统筹指导，联合国内科技巨头与开源基金会，建设具有国际顶级规格的开源代码托管与多模态数据集托管平台。不仅要开源国内最顶尖的基础模型与海量中文、多语种高质量语料，更要通过举办全球性的AI开发者大赛、设立高额的国际开源贡献奖金等方式，主动出击，将该平台打造为全球AI开发者的“第二极”，在国际规则制定中抢夺话语权。

全国政协委员、新大陆科技集团CEO王晶：

# 将善治理念贯穿人工智能产业发展全过程

> 让AI技术发展与伦理建设同频共振，坚守科技向善的价值底线，健全符合我国国情、适配全球发展的AI伦理治理体系。

本报记者 卢梦琪

当前，我国人工智能技术研发与产业应用进入提速期，但AI发展的伦理约束与善治体系建设滞后于技术创新步伐，导致AI技术应用面临数据失真、算法歧视、技术滥用等多重风险。全国政协委员、新大陆科技集团CEO王晶在接受《中国电子报》记者采访时表示，以AI善治筑牢伦理底线，核心是将善治理念贯穿人工智能技术研发、应用落地、产业发展的全过程，打通伦理规范与技术创新的融合堵点，让AI技术发展与伦理建设同频共振，坚守科技向善的价值底线，健全符合我国国情、适配全球发展的AI伦理治理体系。

王晶向记者详述了人工智能发展过程中伦理建设与善治体系层面的痛点。一是伦理建设滞后，与技术创新发展脱节。人工智能技术研发呈现高速迭代态势，但与之匹配的伦理规范体系建设相对滞后，缺乏覆盖AI研发、设计、应用、迭代全链条的统一伦理标准；部分研发机构存在“重技术突破、轻伦理考量”的倾向，将伦理建设视为技术发展的“附属品”，导致AI技术从研发源头就缺乏伦理约束，为后续应用埋下风险隐患。

二是治理机制梗阻，专业服务与协同体系缺失。AI伦理治理的专业人才队伍建设滞后，缺乏既懂人工智能技术，又通伦理规范和法律法规的复合型人才，导致AI伦理审查、风险研判的专业服务链条断裂；AI伦理治理的多方协同机制未有效建立，科研机构、科技企业的主体责任不明确，监管部门的监管体系不完善，行业协会的自律作用未充分发挥，形成“各自为战”的治理格局。王晶强调：“企业在AI技术应用中面临伦理界定模糊、治理规则不明的问题，部分企业因缺乏明确指引而出现‘不敢用、不会用’或‘随意用、滥用’的两极现象。”

三是伦理实践缺位，发展风险与全球协同不足。王晶坦言，当前科技向善的理念未充分贯穿人工智能发展全过程，部分AI技术应用存在数据采集不规范、算法设计有偏见、技术使用超边界等问题，数据失真、算法歧视、隐私泄露等风险逐步凸显，不仅会影响人工智能产业的健康发展，更可能危及数据安全、社会公平与国家安全。

针对以上痛点，王晶提出，推动AI善治与人工智能产业深度融合，打通伦理规范、技术研发、产业应用的全链条，构建人工智能健康发展体系。

一是强化AI伦理顶层设计。以国家数字经济发展战略和全球AI发展趋势为导向，依托新型举国体制凝聚科研机构、高校、行业协会的协同合力，加快制定覆盖AI研发、设计、应用、迭代全链条的统一伦理准则，明确人工智能发展的伦理红线与价值底线。在王晶看来，需要深化AI研发评价改革，破除“唯技术、唯成果”的桎梏，将伦理建设成效、风险防控能力纳入科研人员与研发机构的核心考核指标，激励研发主体从源头融入伦理考量。

二是健全AI伦理治理机制。构建科研机构、科技企业、监管部门、行业协会多方联动、责任共担、协同共治的AI伦理治理机制，推动监管部门做好规则制定与监督管理，科研机构做好伦理研究与技术融合，科技企业履行伦理应用主体责任，行业协会发挥自律引导作用。王晶强调，要培育专业化的AI伦理治理人才队伍和服务机构，搭建AI伦理审查、风险研判、技术咨询的公共服务平台，打通AI伦理治理的专业服务链条。

三是推动AI伦理实践落地。坚持“道引领、法保障、术赋能、器支撑”的原则，推动建立健全AI伦理审查机制、数据管理制度和算法监管体系，规范数据采集、存储、使用、流转全流程，加强对算法设计、训练、应用的全链条监管，严厉打击AI技术滥用行为。王晶表示：“要建立AI伦理失范的容错纠错机制和风险防控机制，从制度层面降低AI技术应用的伦理风险、数据风险和社会风险。”