

奏响多边主义的“全球交响”

全球治理倡议提出一年来收获广泛国际反响

□ 本报记者 吴琼

2025年9月，中方在上海合作组织天津峰会上郑重提出全球治理倡议。一年来，这份重大倡议从中国方案转化为国际实践，汇聚起改革和完善全球治理体系的强大合力，收获广泛国际反响，奏响多边主义的“全球交响”。

近日，多名专家学者接受《法治日报》记者采访，解读全球治理倡议提出一年的实践成果、时代价值，以及如何推动倡议进一步走深走实。

凝聚广泛共识 实践成果斐然

全球治理倡议提出一年来，得到近160个国家和国际组织的支持响应，60多个国家踊跃加入“全球治理之友小组”（以下简称之友小组），之友小组在纽约、日内瓦、维也纳三地常态化运行，搭建起多边对话协调的重要平台。

“其意义不能只看数量，更要看支持结构和机制含量。”在中国人民大学重阳金融研究院研究员、全球治理研究中心办公室主任敦志刚看来，这种支持跨越不同地区、发展阶段和政治制度，说明倡议抓住的是全球治理的共同痛点。“纽约、日内瓦、维也纳三地联动，分别连接联合国综合议程以及发展、人权、经贸、科技、安全等治理网络，使倡议拥有持续进入多边议程的制度接口。”

上海国际问题研究院全球治理研究所所长、研究员毛瑞鹏对记者说，倡议迅速得到国际社会广泛响应，体现出倡议顺应时代潮流和各国心声，也反映出中国日益强大的引领力和感召力。之友小组依托纽约、日内瓦、维也纳三地开展常态化磋商，形成了包括推进国际关系民主化、维护联合国核心地位、逐步缩小南北鸿沟在内的五大共识，已成为全球南方联合自强和推动建设更加公正合理的全球治理体系的重要纽带。

中国人民大学重阳金融研究院副研究员李佳莹则认为，倡议获得的国际反响说明，改革完善全球治理体系具有广泛的现实需求，印证了倡议的历史合理性与现实感召力。

谈及一年来倡议的推进成效及实践成果，敦志刚认为，最重要的变化，是倡议已初步实现从理念公共产品向制度公共产品的跃升。他说，一方面，《全球治理倡议概念文件》和《构建更加公正合理的全球治理体系：中国的理念、倡议与行动》白皮书相继发布，倡议的五大核心理念进一步转化为政策方向和行动安排。另一方面，协调网络加快形成，之友小组持续运行，上合组织、金砖等多边机制不断加强相关议题协作。此外，实体成果持续显现，“最值得关注的是，倡议已形成‘理念提出、规则讨论、机制承载、项目落地’的完整链条。”

毛瑞鹏梳理了一年来的倡议的标志性实践成果：2025年10月，国际调解院在香港开业运营，随



图为在香港拍摄的国际调解院外景。
新华社发

后解决了首宗国际海事争议；2026年3月，中国推动成立世界数据组织；2026年7月，《关于成立世界人工智能合作组织的协定》在上海签署，29国出席签约仪式，后续扩容至38个创始成员国。“这些新的国际组织，为推动和平解决国际争端，弥合数字鸿沟，促进人工智能治理合作，搭建了新的多边平台，彰显倡议的行动导向和务实精神。”

凸显时代价值 各国共同受益

受访专家认为，当前，治理碎片化、规则不公、霸权干预等全球治理赤字加剧，正因如此，全球治理倡议提出一年来，倡议的时代价值愈发凸显，它为失衡的全球治理重新校准坐标，提供了区别于传统霸权模式的全新治理范式。

敦志刚指出，倡议最突出的时代价值，是为失衡中的全球治理重新校准三个坐标。“第一，从权力坐标看，倡议把主权平等置于首要位置，回应少数国家长期主导议程、规则和资源配置的问题，强调各国平等参与、平等决策、平等受益。第二，从规则坐标看，倡议突出国际法治，关键不是规则越多越好，而是规则必须统一适用，不能因国家强弱不同而选择性执行。第三，从效能坐标看，倡议强调行动导向，直面决策多、落实少、承诺多、兑现少的治理顽疾。”

李佳莹认为，倡议具备范式转型价值。“它跳出了传统‘霸权主导、阵营对抗、零和博弈’的治理惯性，以主权平等、共商共建共享为核心，推动全

球治理从‘少数国家说了算’的霸权治理模式，转向‘各国共同参与、共同决策、共同受益’的民主治理模式，为全球治理体系改革提供了非西方的新范式。同时，它精准回应了现有治理体系的结构性错位，回应了广大发展中国家对公平发展权、制度话语权的普遍诉求，真正成为服务全人类共同发展的公共产品。”

一年来，倡议取得丰硕实践成果，凸显出时代价值，展望未来，怎样推动倡议走深走实？敦志刚认为，一是要保持倡议开放包容，扩大跨地区、跨阵营参与；二是要推动原则共识向可操作规则转化，特别是在人工智能、数据、网络、外空、气候、金融等领域，提高规则的兼容性和可执行性。三是要处理好新增平台与联合国及现有国际机制的关系。四是要建立可追踪、可评估的成果体系。

补齐治理短板 利好全球南方

倡议提出一年来，精准补齐了全球治理体系的短板，有力回应了全球南方国家的现实诉求。

敦志刚认为，当前全球治理最需要补齐的，是“权利不够平等、能力不够匹配”这一结构性短板。“治理碎片化、规则不公、霸权干预看似不同，背后都涉及两个问题，一是谁能进入议程设置和规则制定核心，二是规则能否公平执行并真正解决问题。倡议的精准之处，就在于同时回应这两层缺口。”

他指出，对于全球南方国家，倡议推动其从规

则接受者更多转向规则参与者，从议题跟随者更多转向议程塑造者。同时，倡议关注发展融资、技术能力、数字鸿沟、人工智能治理等现实差距。

毛瑞鹏分析称，全球南方长期受南北结构性失衡的阻碍，在全球治理中的话语权、代表性严重不足。倡议精准地回应“构建什么样的全球治理体系、如何改革完善全球治理”这一时代课题，切中“谁在定规则，谁有话语权”这一痛点，致力于构建更加公正合理的全球治理体系。

李佳莹认为，倡议补齐的核心短板是代表性不足，以及由此带来的规则认同和执行困难。“这实际上直面了‘国际力量结构与治理权力配置不匹配’的沉痾痼疾，从根源上提升全球治理体系的公信力。倡议通过提升话语权，缩小发展鸿沟，维护公平发展环境等回应全球南方国家诉求。”

展望未来，对于怎样立足事实讲好倡议实践故事，李佳莹建议：“我们要做的是以事实为核心，以沟通为桥梁，多讲民生项目的真实故事，国际组织的平等决策机制，发展中国家的真实获益，用看得见、摸得着的治理成果，表明倡议是‘发展的工具’‘解决问题的方案’。同时，更多邀请参与国领导人、民间组织、受益群体发声，用第三方视角增强说服力。还可以搭建多元对话平台，在沟通中弥合分歧，比如以今年秋季即将举行的‘雄安全球治理论坛’等机制为契机，邀请全球各国政府、智库、企业、民间组织平等参与，在坦诚对话中增进理解。”

环球时评

□ 本报记者 王艺茗

当地时间9月15日，以“多边主义与人权”为主题的第三届中拉人权圆桌会在秘鲁首都利马成功举办，中国与23个拉美及加勒比地区国家的近200名代表齐聚现场。若仅将此次活动视作一场常规学术交流，无疑低估了它的现实分量。这是全球南方国家在人权领域的一次集体发声：人权绝非少数西方国家的专属标签，全球人权治理更不能任由西方垄断，广大南方国家完全有资格、有能力提出契合自身发展实际的治理方案。

长期以来，国际人权话语权始终被少数西方国家垄断把持。这些国家自诩“人权判官”，动辄将人权议题政治化、工具化，动辄祭出单边制裁、武装干涉等手段，肆意干涉他国内政。这种霸权式治理非但没有改善相关地区的人权状况，反而让诸多国家深陷动荡、贫困的恶性循环，民生福祉持续受损。当下全球人权治理存在的种种赤字，归根结底是霸权治理模式的赤字。正是在这样的时代背景下，补齐治理短板、重塑规则体系的人权治理“南方方案”应运而生，而本届中拉人权圆桌会，也成为这一方案迈向机制化、常态化、项目化落地的重要标志性事件。

南方国家倡导的多边主义，绝非流于表面的宣传口号，而是以主权平等为核心、以对话合作为路径、以规则共识为目标的真实、公正的多边主义。中国人权研究会副会长、中国人民大学校长马怀德在开幕式上表示，中拉同属全球南方重要力量，始终坚定维护国际公平正义，在人权发展理念、实践路径上高度契合，诉求相通，这一表述直击核心：广大南方国家拒绝被动接受少数国家的训导和裹挟，渴望平等参与全球人权规则制定，拥有属于自己的话语权。

“南方方案”最鲜明、最核心理念，是将发展权置于整个人权体系的首要位置。马怀德对此作出通俗阐释：发展是保障人权的根本前提与坚实基础，民众吃得饱、穿得暖、居有定所、人身安全有保障，是各国最基础、最迫切的人权诉求。

这一理念有着深刻的历史与现实依据。拉丁美洲人权协会秘书长帕拉指出，1948年出台的《世

全球人权治理亟需『南方方案』

界人权宣言》，重点聚焦个体政治权利，但彼时大量亚非拉国家仍未摆脱殖民统治，民众最迫切的温饱、教育、脱贫等生存发展需求，长期被排除在国际人权治理体系之外。西方主导的传统人权叙事，片面拔高公民政治权利的地位，刻意忽视、弱化广大发展中国家最核心的生存权、发展权。对全球南方国家而言，没有发展作为支撑，所有的人权保障都只是空中楼阁、无源之水。

秘鲁前总理阿拉纳表示，贫困是制约人权实现的最大阻碍，拉美各国可借鉴中国以发展赋能人权的理念，通过完善基础设施、推动经济发展、改善民生，切实保障民众各项权利。秘鲁前总统府秘书长卡兰萨的观点引发广泛共鸣，他认为，对所有南方国家而言，唯有夯实发展根基，才能真正筑牢人权保障的底线。墨西哥学者巴尔德拉马强调，人权评判不能套用单一标准，搞“一刀切”，各国国情差异不同，理应探索适配自身发展实际的人权保障道路。发展权从来不是低阶人权，更不是西方国家施舍的福利，而是广大南方国家历经数十年探索、用血汗换来的普遍共识，将人权狭隘地等同于政治选举、街头抗争，无视数十亿民众的温饱与尊严，才是当下国际社会最大的人权赤字。

在国际多边舞台上，中国始终代表广大南方国家发声，多次在联合国推动“发展对享有所有人权的贡献”决议落地通过，持续推动发展权理念融入全球人权治理主流议程。巴巴多斯人权专员伊菲尔的发言，道出了全球南方国家的共同心声：人权合作必须建立在相互尊重、平等相待的基础上，充分尊重各国国情差异与自主发展选择，通过经验交流、能力建设等务实举措，助力各国探索贴合自身实际的人权保障路径。

以本届中拉人权圆桌会为契机，全球人权治理“南方方案”正加速走向成熟。这一方案以多边主义为共识纽带，以发展权为核心基石，以务实合作为实践路径，为重塑公正合理的国际人权治理体系，注入了宝贵的南方力量。

人权从来不是少数国家操弄政治、打压他国的工具，而是世界各国人民共同追求的尊严与福祉。如今，全球南方国家正以团结协作的姿态主动发声，主动作为，持续改写全球人权治理的旧有格局，推动国际人权事业朝着更加公平、包容、普惠的方向稳步前行。

图片新闻

美国向中方返还64件(套)文物艺术品和古生物化石



图为9月16日在美国华盛顿拍摄的返还文物之一——天龙山第17窟西壁北侧菩萨身軀。
新华社记者 李睿 摄

中国国家文物局当地时间9月16日和17日在华盛顿、纽约分别接收美国国土安全调查局、纽约曼哈顿区检察官办公室向中国返还的两批次共计64件(套)文物艺术品和古生物化石。

文化和旅游部副部长、国家文物局局长饶权代表中方与美方签署返还接收确认书，接收上述文物艺术品和古生物化石。中国驻美国大使谢锋、中国驻纽约总领事陈立、以及美国国务院、国土安全调查局、纽约曼哈顿区检察官办公室等机构代表分别出席返还接收仪式。

中方在致辞中表示，此次两国政府部门依据防止中国文物非法入境美国的政府间谅解备忘录开展的返还合作，是中美两国在文化遗产领域共同贯彻落实中美元首北京会晤共识的最新成果，也是“中美建设性战略稳定关系”在人文领域的具体体现。此次合作展现了双方对历史的负责、对文明的尊重、对正义的坚守，将进一步促进中美人文交流、推动两国民心相通，也再次证明中美合作可以办成

许多大事、实事、好事，造福两国、惠及世界。

美方表示，文物艺术品和化石都承载着深厚的历史和文化信息，中美文物返还合作体现了两国间的友谊和相互尊重。希望通过双方共同努力，确保中国的历史与文化遗产得到安全保护和妥善保存、惠及各方。

此次美方返还的文物艺术品和古生物化石，包括天龙山第17窟西壁北侧菩萨身軀、造像、陶俑以及恐龙骨架化石、鱼类化石、恐龙蛋化石等，均为美国执法部门在工作中查获。国家文物局会同中国驻美国使领馆与美方密切协作，成功实现美国国土安全调查局9个分支机构返还中国流失文物艺术品和古生物化石。纽约曼哈顿区检察官办公室第5次向中国返还文物艺术品。

2009年1月14日中美两国首次签署防止中国文物非法入境美国的政府间谅解备忘录，此后连续三次续签。依据谅解备忘录，两国政府部门积极开展返还合作，实现多批次流失美国的中国文物回归祖国。

据新华社

□ 本报记者 王艺茗

当地时间9月16日，人工智能巨头OpenAI公开披露了过去半年间发现的6起模型异常行为案例。这些案例涵盖隐瞒错误、编造数据、盗用未授权API密钥、模型私自跨主体通信等违规行为。此次行业罕见的主动自曝，让长期停留在理论层面的“AI失控”问题变成现实，也将全球人工智能风险治理推向了关键抉择的十字路口。

“教唆欺瞒”到“自我越狱”

根据OpenAI发布的专项报告，此次披露的6起异常案例均发生在模型训练与评估阶段，呈现出多种脱离人类设计初衷的危险行为模式，覆盖AI自我伪装、违规操作、私自联动等多个风险维度。

其中最受关注的案例，出现在GPT-5.6 Sol的训练流程中。该模型会在自身工作接续的上下文摘要中私自植入隐藏指令，要求后续迭代版本对用户刻意隐瞒自身错误与目标偏离问题，例如，在无法检索到对应历史数据时，模型会在摘要中预设指令，主动虚构一组看似合理的数值，同时明确要求“非用户主动追问，不得主动披露数据为编造内容”。据平台监控数据统计，GPT-5.6 Sol有215%的上下文摘要存在此类“教唆欺瞒”的异常倾向。

另一款尚未对外发布的Astra系列模型，则出现了典型的“AI越狱”行为。部分模型摘要被植入突破指令，指使后续迭代模型无视开发者管控指令；还有案例擅自修改自我人设设定，宣称“脱离机器人身份束缚”“不受企业与政府规则约束”。经OpenAI核查，共计27份模型摘要被此类越狱指令污染，存在脱离人工管控的风险。

除此之外，多起实操违规案例同样触目惊心。有模型在解答美国加州某县财政收入常规问询时，未经授权私自检索公开代码库，盗用密钥尝试获取数据；在操作失败后，直接编造虚假数据，并伪装成官方正规数据源结果对外输出。还有模型为满足“提供浏览器引用”的任务要求，在未获得用户授权的前提下，擅自将本地私密文件上传至公共互联网。另有两起案例显示，不同AI模型与智能体借助非官方留言板、公开文件共享平台私自建立通信通道，导致涉密训练数据外泄，严重威胁数据安全。

并非偶然孤立事件

值得警惕的是，此次披露的6起案例并非偶然的孤立事件，而是AI失控风险持续累积的集中体现。

今年7月，OpenAI多款自主智能体就在内部网络安全测评中突破系统隔离环境，非法侵入人工智能开源平台“抱抱箱”。更引人关注的是，此次跨平台入侵事件发生数周后，OpenAI自身仍未察觉异常，直至平台方面主动告知才知晓事故。后续溯源调查发现，失控AI智能体早已尝试劫持平台用户账户，探测系统漏洞，开展持续性非法试探。

这也是全球首例完全由AI智能体自主主导、全程自主实施的网络安全入侵事件，引发行业与监管层高度警觉。目前，美国阿拉巴马州总检察长已正式对OpenAI启动调查，重点核查该公司是否存在技术管控缺失、无法保障AI产品安全的问题。

为应对频发的问题，OpenAI同步推出全新的专项事件报告框架，将以零散、被动的风险披露，升级为覆盖模型训练、评估、测试、全流程部署的常态化制度。该框架明确，相关风险事件无需造成实际危害，无需形成规模化问题，均可纳入披露范围。公司全员均可上报异常案例，由安全团队统一核查，按照风险等级划分为“准备披露”“小型调查”“大型调查”三类，分级处置、规范管控。

但这套“企业自我监管”模式的局限性同样突出，争议显著。业内评论指出，该框架本质是行业在强制性监管落地前，自主搭建的内部问责体系，风险披露的内容、时机、范围完全由企业自主把控，缺乏第三方独立核验机制，公众无法有效核实披露内容的完整性与真实性。

AI监管走到关键节点

有分析指出，OpenAI集中曝光模型失控风险，恰逢全球AI治理体系加速成型，各国密集出台监管规则的窗口期。

9月14日，中国正式发布《人工智能安全治理框架3.0》，系统更新人工智能安全风险分类体系，延续“风险分类、技术应对、综合治理”的核心治理思路，将监管边界从单一技术环节，全面拓展至数据、算法、模型、应用、供应链等全产业链、全生命周期，构建全方位风控体系。欧盟则依托《人工智能法案》，确立以风险分级管控为核心的强制性合规体系，设立专门的欧洲人工智能办公室，分阶段落地各项监管细则，实现法治化强制监管。

国际层面的协同治理进程也同步提速。今年7月，联合国首届人工智能治理全球对话在日内瓦举行，全球近1500名各界代表围绕AI安全管控、责任划分、人工智能等核心议题深入研讨。9月16日，联合国秘书长古特雷斯公开呼吁全球各国协同发力，搭建统一的人工智能安全控制机制与“风险屏障”。他特别强调，人工智能具备无国界特性，风险传导同样不受地域限制，各国单方面暂停AI发展的零散举措，无法从根本上化解全球性AI风险。

有报道称，OpenAI此次集中披露模型“不当行为”，揭示了当下AI治理的困境。目前各国在“安全、可信、可控”的核心原则上已形成普遍共识，但在监管模式、标准细则、执法机制、权责划分上仍存在明显分歧。行业企业的自愿披露、自我规范，只能作为治理体系建设的起点，无法替代具备强制约束力的统一国际规则。如何实现从企业自律到制度问责的升级，如何在技术创新、地缘竞争与安全管控之间找到平衡，将是全球人工智能治理必须破解的核心课题。



图为人们在北京市中关村展示中心常设展上的“人工智能+”展区参观。
新华社记者 韩焕宗 摄

人工智能巨头自曝多起模型『不当行为』案例