



# 推动完善人工智能全球治理

嘉 宾：清华大学人工智能国际治理研究院院长 薛 澜  
北京大学国际关系学院院长 唐士其  
复旦大学国际问题研究院教授 宋国友

主持人：本报编辑 吴 浩

当前,世界百年变局加速演进,人工智能作为引领新一轮科技革命和产业变革的战略性技术,正在深刻改变人类的生产生活方式。同时,人工智能带来的风险挑战也引发广泛关注。如何推动完善人工智能全球治理,是国际社会共同面临的课题。习近平总书记指出:“坚持以人为本、智能向善,在联合国框架内加强人工智能规则治理,积极推进绿色转型,让广大发展中国家更好融入数字化、智能化、绿色化潮流。”“十五五”规划纲要提出:“推动建立各国广泛参与的人工智能治理框架,共同构建平权、互信、多元、共赢的全球人工智能开放生态,支持全球南方国家加强人工智能能力建设。”中国积极推动形成具有广泛共识的人工智能全球治理框架和规则,为国际社会提供中国方案、贡献中国智慧。本期“访谈录”邀请嘉宾,就如何推动完善人工智能全球治理展开交流探讨。

## 人工智能全球治理的特点与挑战

主持人:2025年,习近平总书记提出全球治理倡议,为推动构建更加公正合理的全球治理体系提出中国方案。与维护多边贸易体制、应对气候变化挑战、实现可持续发展等较为成熟的全球治理议题相比,人工智能全球治理的特点体现在哪些方面?

薛澜:习近平总书记指出:“我提出全球治理倡议,旨在推动构建更加公正合理的全球治理体系。”推动完善人工智能全球治理,是践行全球治理倡议的重要内容。人工智能全球治理的特点,源于人工智能技术发展快速迭代,而其影响又极为广泛深远,国际社会没有在思想和行动上做好充分的应对准备。这一客观情况,导致与较为成熟的全球治理议题相比,人工智能全球治理有一些特点。

比如,应对气候变化挑战等议题是在国际合作氛围较好的时候达成了共识。与此不同,人工智能技术的爆发式发展与地缘政治风险持续上升同步,人工智能全球治理从诞生之初便面临大国博弈、技术竞争的复杂局面。个别国家搞“脱钩断链”,在技术开发、人才培养、数据互联等方面筑起“小院高墙”,试图以科技封锁垄断竞争优势,破坏了原本联系较为紧密的全球科研及产业协作网络。在这种背景下,由技术领先国家构成的人工智能全球治理核心主体之间不是“合作者”,而是“竞争者”,削弱了人工智能全球治理合作的信任基础,导致国际社会在应对人工智能技术发展所带来的风险和挑战时,缺乏合作根基与协作机制。单边主义、保护主义催生技术壁垒与规则壁垒并存的格局,导致人工智能全球治理体系碎片化,削弱了规则的普遍约束力,导致人工智能全球治理陷入“分而治之”的困境。

唐士其:人工智能的发展速度远远超过其他领域,其未来发展具有巨大的不确定性。除此之外,人工智能全球治理的特点主要表现在两个方面。

一是人工智能不仅是被决策的对象,而且部分地参与了决策。人工智能因其出色的信息加工和处理能力,已经在诸多领域成为辅助决策的手段。未来,决策者对人工智能的依赖总体上会持续增加。人工智能提供信息的真实性和有效性对决策的结果和质量会产生越来越大的影响。然而,用于辅助决策的人工智能技术通常并不掌握在决策者手中,而是由行业内具备显著优势和强大资源整合能力的企业所掌控,它们主导基础模型、掌握海量用户数据,具有强大的影响力。因此,如何确保人工智能在信息供应环节做到全面、客观,日益成为一个重要问题。

二是治理对象具有流动性和虚拟性。人工智能并不像贸易交往、气候变化那样,能够通过一些可观察、可控制的实体来展现其特征变化与发展趋势。人工智能全球治理涉及的对象是算力、算法、数据和模型,它们本身是技术,而非技术产生的结果,因此人工智能缺乏清晰的“治理锚点”。此外,人工智能存在并自由流动于网络空间,在很大程度上游离于人们物理性的把握和控制之外。人工智能所采集到的信息和数据,以及对这些信息和数据的具体处理方式并不完全在人的可控范围。生成式人工智能具备“学习—优化—再生成”的闭环机制,其行为可能超出开发者的初始设计,导致意图与结果脱节。

宋国友:与相对成熟的全球治理议题相比,人工智能全球治理的特点至少有如下3个方面:一是治理影响的不均

衡性。气候变化是全球性的,所有国家都受到巨大影响。然而,人工智能的发展和影响并不均衡。一些国家已经感受到人工智能带来的剧烈冲击,而部分国家可能认为自身不受影响或影响较小。二是治理路径的非预知性。由于人工智能本身还处于早期发展阶段,尚未定型,对其治理的重点和规则等具体事项现阶段难以落实,治理路径并不清晰。三是技术竞争的高敏感性。人工智能作为新一轮科技革命和产业变革的重要驱动力,是大国竞争的战略性高点。技术竞争的高敏感性导致个别国家竞争和对抗思维强烈,合作和互利意识不足。

主持人:中国倡导构建人类命运共同体、构建网络空间命运共同体,坚持发展和安全并重的原则,提出《全球人工智能治理倡议》,欢迎各国政府、国际组织、企业、科研院所、民间机构和公民个人等主体秉持共商共建共享的理念,协力共同促进人工智能治理。目前共建人工智能全球治理体系面临哪些挑战?

薛澜:第一,国际社会在人工智能全球治理的突出问题上缺乏共识。比如,如何认识人工智能的潜在风险?需要在人工智能发展中划定哪些红线?政府如何在促进创新与防控风险之间取得动态平衡?各方仍存在较大分歧,给国际社会落实《全球人工智能治理倡议》带来挑战。同时,如何弥合日益加剧的智能鸿沟也是中国和其他发展中国家希望引起国际社会关注的重要问题。

第二,人工智能技术正处于爆发式发展阶段,其技术突破往往领先于治理规则的制定,这种“时间差”导致治理始终在“追赶”技术,形成治理难题。比如,面对生成式人工智能、具身智能等新技术带来的风险,当前治理规则存在缺口或滞后的现象。因此,中国所倡导的敏捷治理等人工智能安全治理方式得到国际社会高度关注。

第三,全球层面看似治理机制众多——联合国、经济合作与发展组织等都在推动制定相关准则,各国还有各自的标准和规范。但这些机制缺乏统筹协调,彼此间存在重叠甚至冲突,形成了一个“机制复合体”,导致规则繁杂且难以落地,各国面临“规则迷宮”,执行成本高企,实际效能大打折扣。

唐士其:目前,世界范围内的技术民族主义抬头,一些国家优先考虑的是本国的技术安全与产业链韧性,而非全球公共利益。从企业层面看,人工智能的开发总体上处于一种缺乏秩序和规则的竞争状态,并且与各个国家之间的竞争交织在一起,为建立人工智能全球治理体系带来很大困难。在人工智能治理机制方面,不同国家在数据跨境、内容审核、政府监管权限等关键问题上也存在一系列分歧。减少矛盾和冲突,需要各种国际行为体,特别是大国之间增进互信与合作,建立一套相对完整的人工智能监测、预警机制,以及利益和责任的分担机制,在此基础上尽可能增强共识。

宋国友:从共建主体的角度加以分析,困难有3点。一是个别国家奉行单边主义、保护主义,阻碍了人工智能全球治理体系共建。二是部分人工智能技术欠发达的国家对于共建热情不高。这些国家还没有真正经历人工智能带来的冲击,对其机遇和风险缺乏认识,缺少参与人工智能全球治理的迫切意愿。三是参与人工智能全球治理体系共建的私营部门存在担忧。大型人工智能企业等私营部门既是全球人工智能治理的重要主体,同时也是被治理的客体。相关私营部门希望以企业发展为中心,以利润为目标,以便获得市场领先优势和超额回报,对于政府主导的共建人工智能全球治理体系仍有疑虑和抵触。

## 坚持以人为本、智能向善,汇聚治理合力

主持人:在现行国际机制下,人工智能治理,往往滞后于人工智能技术的迭代演进,这种失衡导致该领域全球治理赤字进一步扩大。应在全球范围推广何种治理理念,实现技术发展与管理效能的协同推进?

薛澜:首先,必须坚持以人为本的发展理念。人工智能发展归根结底要服务于人类社会。我们一直强调“技术向善”,就是要把人的福祉作为科技发展的根本目标。无论是技术设计还是规则制定,都要围绕人的需求展开,确保

技术发展不偏离人类社会发展轨道,助力弥合南北发展鸿沟,推动各国共享人工智能发展成果。

其次,确立平等对话的治理基础。人工智能全球治理不能由少数技术强国说了算。各国能力有强弱,体量有大小,但在治理规则制定上应该平等参与,坚持共商共建共享的全球治理观。我们需要摒弃单边主义,搭建多边平台,充分尊重不同国家的声音,特别是发展中国家的诉求,提高发展中国家的代表权和发言权。只有包容互鉴,才能真正形成有共识的治理框架。

再者,必须注重行动导向,积极推动确立普惠发展的治理路径。现在各国技术差距明显。如果放任这种差距扩大,治理鸿沟只会更深。所以,中国特别强调要加强能力建设,坚持系统谋划、整体推进,通过技术援助、知识共享等方式,推动全球南方国家共享人工智能发展红利。

最后,构建安全共治的风险防控体系。人工智能风险是全球性的,一个国家出现技术漏洞可能引发全球连锁反应。因此,人工智能安全治理不能“各扫门前雪”。要把人工智能安全作为全球公共产品来提供,各国共同研发安全技术、制定统一标准,建立应急协作机制。

唐士其:在这个问题上,我觉得必须坚持3点:

一是坚持以人为本。传统上我们认为人之为人,很大程度上是因为人拥有理性,而一种在某些方面远远超越了人的智慧,特别是逻辑思维能力的人工智能体的出现,毫无疑问撼动了人类传统的地位。那么面对人工智能,我们究竟应该在什么方面维护人类的基本的人文底线,就是一个需要全人类集体面对、共同探讨,并形成基本共识与价值规范的根本性问题。无论如何,人工智能只能是人类的工具,对人工智能的利用必须以人为本,这是一个重要原则。

二是坚持互利合作。目前,人工智能的开发和利用都体现出多中心、分散化的特点,同时在世界范围呈现出明显的不均衡性。因此,一方面,要优势互补,避免重复建设和浪费;另一方面,要让人工智能研发相对落后的国家和地区也能够享受人工智能带来的巨大便利,缩小数字鸿沟。这两方面都需要在人工智能全球治理的过程中提倡互利合作的精神。

三是坚持开放互信。人工智能技术建立在大语言模型的基础上,信息越开放,越能够自由流动,人工智能的发展就

会越快。人工智能在军事、国家安全等领域的运用,及其巨大的经济潜能,又使人工智能成为大国博弈的关键领域。由于短期内这种局面难以得到根本性改变,所以在人工智能全球治理的问题上,应该努力寻找一个平衡点,即国家安全、经济竞争与开放互信的平衡点。

宋国友:人工智能技术发展日新月异,人工智能全球治理应确保人工智能技术无论如何迭代,都能符合全人类的共同利益,让人工智能发展服务于全人类福祉。具体而言,应强化开放智能、包容智能、普惠智能和安全智能等治理理念。开放智能,是指人工智能全球治理要防止个别大国将人工智能阵营化,对其他国家搞封锁。包容智能,是指人工智能全球治理应该把全人类的文明都包含在内,而不能在人工智能中注入文明排他主义基因。普惠智能,是指要让人工智能造福于全人类,让世界各国能够在人工智能技术创新中获益。安全智能,是指要积极防范人工智能可能引发的各种危害国家安全的风险和挑战,确保其不威胁人类。只有将这四大理念贯穿于人工智能全球治理全过程,成为牢固的治理“护栏”,才能实现技术发展和治理效能的协同推进。

## 推动建立各国广泛参与的人工智能治理框架

主持人:中国举办2025世界人工智能大会暨人工智能全球治理高级别会议,发布《人工智能全球治理行动计划》,提出“推动构建包容、公平的多边全球数字治理体系”,并倡议在人工智能安全治理、人工智能能力建设等方面开展国际合作。当前,人工智能国际合作应如何突破地缘政治壁垒,把上述共识转化为可操作的信任机制与务实成果?

薛澜:首先,必须坚定支持联合国发挥主渠道作用,支持在联合国框架下建立“人工智能独立国际科学小组”和“人工智能治理全球对话”机制。同时,积极鼓励和其他各种双边和多边对话机制,走出一条开放协商、多元参与的互信共治之路。其次,在鼓励技术创新与产业应用的同时,通过国际合作建立人工智能风险测试评估体系,完善数据保护规范,寻求发展与安全的平衡。此外,我们要积极倡导开放合作,帮助世界各国特别是全球南方国家加强能

力建设,推动人工智能发展成果由各国共享。

宋国友:习近平总书记强调:“进一步拓展人工智能合作,加强信息交流和技术合作,共同做好风险防范,形成具有广泛共识的人工智能治理框架和标准规范,不断提升人工智能技术的安全性、可靠性、可控性、公平性。”开展人工智能国际合作可以从以下几方面突破地缘政治壁垒。一是重要议题合作。可围绕人工智能对人类最具风险和最具公共产品属性的重要议题先行开展合作。前者如合作管控军事应用人工智能等,后者如发展医疗健康人工智能和减轻全球灾害人工智能等。二是开展现有多边机制下的务实合作,增强合作信心。如支持联合国“人工智能治理全球对话”机制,在联合国框架内加强人工智能规则治理,推动相关国际合作。三是推动双边、小多边等合作,作为人工智能国际多边合作的有益补充。双边和小多边合作更具针对性和灵活性,可在双边和小多边成员共同关心的人工智能相关问题方面开展合作,推动人工智能治理取得积极成果,逐渐扩大合作共识。四是鼓励私营部门的市场合作。市场合作是国际合作的重要组成部分,应积极鼓励私营部门在符合相关监管规则等前提下开展各种形式的商业合作,在市场端夯实合作基础、构建合作样板,打破地缘政治藩篱,自下而上地推动国家间治理合作。

主持人:当前,全球南方国家面临人工智能技术鸿沟与治理话语权不足的双重挑战。应如何切实保障全球南方国家平等参与人工智能全球治理?

薛澜:全球南方国家面临双重挑战的根源在于教育与研究基础薄弱,缺乏人才支撑,科技能力不足,信息基础设施落后。在这样的情况下技术赋能无从谈起。因此,为避免人工智能技术鸿沟继续扩大,要高度关注全球南方国家的教育和科技能力建设,关注其信息基础设施建设,重视“授人以渔”。

在参与人工智能全球治理决策层面,需要加强全球南方国家治理能力的培养,这与人工智能技术应用的实践相辅相成。人工智能普惠发展既可以使全球南方国家享受人工智能造福人类的红利,也可以增强其参与人工智能全球治理的能力。

## 深入学习贯彻习近平新时代中国特色社会主义思想

# 健全网络生态治理长效机制

部 庆

网络生态治理在国家治理中具有重要地位。习近平总书记在主持二十届中共中央政治局第二十三次集体学习时强调:“要健全网络生态治理长效机制,着力提升治理的前瞻性、精准性、系统性、协同性,持续营造风清气正的网络空间。”“十五五”规划纲要提出:“健全网络综合治理格局,完善网络生态治理长效机制”。当前,以网络安全法、数据安全法、个人信息保护法等为基础的网络法治体系的四梁八柱已经构建。通过开展“清朗”系列专项行动,网络空间主旋律高昂、文化创新活力迸发,崇德向善蔚然成风、网络环境风清气正。同时,网络生态治理仍存一些问题,如深度伪造技术用于诈骗、诽谤,扰乱社会秩序;“信息茧房”影响个人认知、社会凝聚;信息数据的非法获取和不当使用侵犯个人隐私、危害国家安全;等等。网络生态问题具有长期性、顽固性,必须加快构建系统完备、科学规范、运行有效的网络生态治理长效机制,推动网络治理实现由“管”向“治”的根本转变。

网络生态治理长效机制,是指在党中央集中统一领导下,围绕营造风清气正的网络空间目标,以制度建设为基础,以常态运行为特征,以协同推进为方式,推动网络生态治理由阶段性整治向制度化、规范化、常态化转变的长效运行机制。网络生态治理长效机制覆盖事前预防、事中监管、事后处置全过程,贯通内

容生态、平台运营、算法机制、数据安全等重点领域,统筹行政、市场、法治、科技等多种手段,着力解决治什么、谁来治、怎么治、如何持续治的问题。健全网络生态治理长效机制,既需要通过明确规则边界和责任要求,夯实治理基础,又需要通过动态调整制度工具,提升应对新情况新问题的能力。

坚持前瞻性布局,下好风险防范“先手棋”。为防止风险积聚蔓延、保障治理长期有效,面对新技术新应用新业态给网络生态治理带来的挑战,必须以前瞻性的战略思维夯实长效治理根基,以高质量制度供给驾驭网络治理中的风险变量,为网络生态持续清朗向上提供坚实可靠的制度支撑。用联系的、发展的眼光看问题,运用底线思维防范化解重大风险,做到准确识变、科学应变、主动求变;把各类风险的综合研判纳入常态化工作机制,及时识别潜在风险形态,科学评估风险等级并实施分级分类的安全监管机制,积极稳妥化解重点领域风险,推动治理规范与技术演进同向同行。通过风险研判前置化、规则供给常态化、制度调整机制化,推动治理从事后应对向源头预防转变,使网络生态治理在起点上即具备长期稳定运行的制度基础。

坚持精准性施策,练就精治治理“绣花功”。网络生态问题类型多样、成因复杂、表现各异,既有共性风险,也有个性隐患。面对不同平台形态、技术路

径与应用场景中出现的网络生态问题,必须立足实际、分类施策,确保治理措施精准落地、效能持续发挥。把握问题本质,找准风险源头,厘清责任边界,围绕重点领域、重点环节和重点对象实施差异化治理策略,推动资源向关键领域聚焦、向薄弱环节倾斜;坚持问题导向和效果导向相统一,完善与精准治理相适应的制度工具和技术手段,压紧压实责任链条,建立健全问题发现、分析研判和处置反馈的闭环机制。通过精准识别风险、精准匹配手段、精准落实责任,持续提升治理实效,形成权责清晰、协同高效、运行顺畅的长效治理路径。

唐士其:习近平总书记指出:“要加强人工智能国际治理和合作,确保人工智能向善、造福全人类,避免其成为‘富国和富人的游戏’。”可以从3个层面保障和促进全球南方国家平等参与人工智能全球治理进程,享受人工智能发展带来的成果。

第一是技术层面。在发展问题上,大国有更多合作的空间,比如可以共同建设全球公共产品型人工智能系统,包括应对气候变化的碳排放监测平台、全球传染病早期预警模型、农作物健康管理系统等,并向全球南方国家开放。同时,通过各种方式帮助全球南方国家建立自己的人工智能系统,培养相关人才,并形成自身的发展能力。

第二是规则层面。在人工智能全球治理的规则制定过程中充分吸收全球南方国家参与,确保代表名额分配公平,以保证全球南方国家的诉求能够对未来人工智能发展的基本方向产生影响。

第三是价值层面。一方面,人工智能大国应避免通过云服务、预训练模型等向全球南方国家输出隐含某种特定价值观的人工智能系统、塑造新型依附关系;另一方面,应在人工智能未来的发展中,充分体现全球南方国家的文化和价值因素。由于当前的人工智能发展基于人类语言系统,所以严格来说,它在价值层面并不是完全中立的。文明的特征既体现在价值和伦理层面,也体现在思维方式上,这些都内含于一个民族的语言系统之中。未来人工智能发展的一个重要宝库就是非西方国家的语言和文化资源,如果能够把这些资源纳入人工智能的语料库,那么我相信会大大增加人工智能的智慧程度,同时也会为全球南方国家和人类文明的发展提供新的可能性。

宋国友:首先,全球南方国家要依靠自身,从能力建设和机制建设两方面推动联合自强,打破人工智能全球治理中全球南方和全球北方的结构不对称性,而不能“等”“靠”“要”,寄希望于发达国家主动平等对待。

其次,强化全球南方国家人工智能能力建设,尽量缩小与发达国家之间的人工智能鸿沟。从改善人工智能基础设施做起,如确保稳定电力供应和网络链接,在全球南方不同区域平衡布局算力集群,扩大优质人工智能产品使用范围并降低人工智能使用成本等。同时,推动人工智能领域的南南教育合作,共建人工智能培训中心,帮助全球南方国家提升人工智能科研能力,培养人工智能尖端人才。

最后,完善人工智能全球治理的机制建设,提高全球南方国家的治理机制话语权。充分借助联合国以及金砖国家、上海合作组织等重要多边机制,协调全球南方国家在人工智能领域的共同利益和治理诉求,推动形成国际人工智能治理领域的南方议程。全球南方国家要积极为人工智能全球治理提供重要理念,把全球南方国家的共识性理念注入全球人工智能治理进程。

理深度交融,促进规则衔接、标准协同和措施联动,着力强化网络生态治理长效机制的聚合力和穿透力。

坚持协同性联动,绘就多方共治“同心圆”。面对网络生态治理主体多元、边界交叉、联动密切的现实特点,必须坚持共建共治共享理念,统筹国内国际两个大局,对内织密多方协同的治理网络,对外携手构建网络空间命运共同体。健全多方参与、协同联动的治理格局,形成分工明确、衔接有序、协作高效的共治体系。在国内治理实践中,形成党委领导、政府管理、企业履责、社会监督、网民自律等多主体参与,经济、法律、技术等相结合的综合治理格局,以制度化方式健全协作机制。在国际治理实践中,积极参与全球网络治理规则制定和国际合作,在深化交流互鉴中弥合“数字鸿沟”,在共同应对网络安全屏障。通过内外兼修、标本兼治,探索形成以健全长效机制为核心的网络生态治理中国方案,为国际网络生态治理贡献中国智慧。

(作者单位:中国政法大学国家安全学院)

治理之道  
ZHILIZHIDAO