人工智能全球治理的现状、困境和中国路径

韩永辉 周港隽 徐翠芬

摘 要：人工智能是引领新一轮科技革命和产业变革的战略性技术，人工智能全球治理是针对人工智能的研发和应用等全流程中涉及的多方面问题进行规范与管控，发挥着引导科技向善发展、塑造可持续的全球人工智能生态的关键作用。文章以人工智能全球治理现状为出发点，针对当前全球人工智能治理参与代表性不足、治理机制协调性不足、技术规则间协调难度较大、治理红利释放不足等困境，从深化多边合作机制建设、聚焦技术安全与伦理标准制定、注重数据跨境流动规则建设、加快落实人工智能治理承诺等方面提出人工智能全球治理的框架，以期为人工智能全球治理提供新思路，推动人工智能全球治理体系朝着更加开放包容的方向发展。

关键词：人工智能；全球治理；治理困境；中国路径

中图分类号：TP18;F49;F125 文献标识码：A 文章编号：1673-5706（2024）06-0094-09

党的二十届三中全会通过的《中共中央关于进一步全面深化改革、推进中国式现代化的决定》指出，要落实“三大全球倡议”，参与引领全球治理体系改革和建设，有效应对外部风险挑战，引领全球治理，要“建立人工智能安全监管制度”，鼓励支持中国推动人工智能国际交流合作。作为新兴技术，人工智能全球治理至关重要，在当今全球化时代，人工智能的迅猛发展超越了国界，合理的全球治理能够促进人工智能技术在全球范围内健康创新与有序扩散，推动全球共享技术红利，助力全球经济增长与可持续发展。随着人工智能的高速发展，发展中国家在治理对话中普遍缺位、规则竞争化武器化趋势凸显、现有治理机制之间协调性不足、治理实施效果与承诺存在差距等人工智能全球治理困境和挑战或阻碍人工智能的技术红利释放，不利于人工智能全球“善治”格局的构建。

作为人工智能全球治理的积极参与者和构建者，中国倡导开放与共享的精神，推动全球人工智能研究资源的交流与合作。2024 年 7 月 4 日，

基金项目：国家社科基金重大项目“全球产业链重构对全球经济治理体系的影响及中国应对研究”（21&ZD074）；国家自然科学基金项目“产业政策对参与全球价值链的影响研究：理论机制、实证识别与中国方案”（72073037）；广东省自然科学基金杰出青年项目“面向全球价值链的产业政策治理研究”（2022B1515020008）；深圳市建设中国特色社会主义先行示范区研究中心课题“新发展格局下深圳高水平对外开放研究”（SFQZD2402）资助。

上海世界人工智能大会向全球发布《人工智能全球治理上海宣言》，从人工智能发展、人工智能安全、人工智能治理体系、人工智能社会、社会福祉等五个方面提出改进人工智能全球治理的政策框架。在当今国际科技竞争愈发激烈、技术机制规则壁垒加强、科技发展鸿沟不断拉大的背景下，中国在人工智能全球治理上的愿景设想以及治理框架，为未来人工智能全球治理提供有益的政策借鉴。因此，梳理人工智能全球治理的现状与困境，结合中国参与全球人工治理的政策实践，提出全球人工智能治理的“中国方案”，是推动人工智能造福全人类、构建人类命运共同体的应有之义[1][2][3][4]。

在对人工智能全球治理的研究中，大多研究集中于围绕全球人工智能治理倡议展开分析，阐述人工智能治理倡议中的核心内涵，基于全球人工智能治理倡议挖掘中国智慧和中国方案[5]。从国际关系角度出发，主张建立人工智能全球性共识[6]，在此基础上针对其信任困境提出解决方案[7]。从全球经济角度出发，基于人工智能的特征解析其冲击全球经济治理秩序的具体机制[8][9][10]，在全球治理框架中剖析中国实施路径[11]。目前关于人工智能全球治理的相关研究大多从宏观的角度论述其理论重要性以及抽象内涵，而较少落实到微观层面。此外，针对人工智能全球治理提出的实践路径集中于国内视角，较少从全球角度提出对应的治理方案。本文结合宏观和微观视角，剖析人工智能全球治理面临的新兴挑战，从国际视角提出相对合理可行的中国路径。

一、人工智能全球治理的内涵与现状

（一）人工智能全球治理的内涵

近年来，人工智能已经成为推动人类步入智能时代的决定性力量。纵观历次工业革命的历史，新兴技术既能推动人类社会进入新阶段，也会对社会经济治理提出新挑战。在人工智能带来时代变革的同时，一系列社会问题和安全隐患也不断滋生，人工智能全球治理亟待完善。人工智能全球治理是指国际社会通过一系列的政策、规则、标准、机制和合作，对人工智能技术在全球范围内的开发、部署和应用进行协调、监督和管理的过程。各国应对人工智能带来的跨国界挑战，充分发挥人工智能的优势促进公平、安全、可持续的全球科技发展和社会进步成为人工智能全球治理的主导方向。

人工智能的发展并不局限于某一特定产业，使得人工智能全球治理存在技术治理尚未完善、负面影响扩散风险高、国际治理差距较大等负面特征。一是人工智能技术治理尚未完善。人工智能算法可能存在缺陷或被恶意篡改，大量个人信息、企业机密乃至国家敏感信息存在泄露风险。在人工智能的整个生命周期，从研发设计阶段的模型缺陷、训练阶段的数据偏差到部署运营后的适应性与更新维护问题，缺乏体系化、制度化、成熟化的解决方案和治理模式[12]。二是人工智能负面影响存在全球扩散传染的风险。人工智能的应用范围广泛且正处于高速发展阶段，在信息不对称的情况下，人工智能造成的负面溢出效应和下游影响是全球性的。人工智能技术成果在全球范围内迅速传播与应用，任何一个国家在人工智能领域的重大突破或失误会产生蝴蝶效应，通过信息交流等渠道扩散至其他国家。三是人工智能全球治理存在着巨大差距。人工智能引发的伦理道德和社会公平问题具有全球性，算法偏见等不公正、不均衡问题可能导致不同种族、性别在就业和教育等资源分配上的不公平[13][14]。人工智能倡议和机构之间的协调差距可能会将世界分裂成相对独立的人工智能治理体系[15]。各国治理政策与法规差异巨大，在跨国人工智能业务开展时面临监管套利与合规困境。因此，保障全球人工智能技术在合法、合规、符合伦理道德的框架内发展成为人工智能全球治理的必然要求。

在人工智能技术相关伦理和安全风险不断出现的形势下，人工智能全球治理以法律法规形式为主，行业标准和倡议、国际合作形式为辅。关于人工智能伦理和数据保护等的国际法律文件或共识性建议书是人工智能全球治理的必要工具[16]。2023年，美国、欧盟等主要经济体都在推进AI监管法规的制定。美国聚焦于人工智能社会治理层面，于2023年1月26日发布了《美国AI风险管理框架 1.0》（Artificial Intelligence Risk Management

Framework，AI RMF 1.0），由政府部门、私营企业、科研机构以及行业协会参与制定，通过立法对人工智能进行重点领域规制。AI RMF 1.0 的治理内容涵盖了人工智能系统全生命周期的风险，包括对人类、组织和生态系统的危害，同时关注到 AI 的可信度，确保其有效可靠。在政策框架上，AI RMF 1.0 提出了风险评估、信息共享、坚持问责“三位一体”的政策框架：建立了风险评估机制，组织评估 AI 系统风险并提出相应建议从而确定技术操作和管理方面的措施；执行了信息共享机制，加强各方交流合作，在组织内部与组织间共享风险和决策过程信息；坚持问责机制，明确各参与方责任，实施明晰权责划分机制，使 AI 系统风险可追究。与此同时，2024 年 9 月中国网络安全标准化技术委员会发布《人工智能安全治理框架》，参与制定主体包括政府部门、企业、科研院所、民间机构和社会公众，治理内容涵盖了人工智能内生安全风险和应用安全风险。相比于美国的政策框架，中国的人工治理框架则强调包容、协同发展等共享型发展理念。一是坚持包容审慎、确保安全的原则，鼓励创新的同时严守安全底线；二是风险导向、灵活治理，密切关注风险趋势，快速调整治理措施；三是技管结合、协同应对，综合运用技术和管理手段，明确各方责任。

（二）人工智能全球治理正处于秩序形成关键期

随着科技的迅猛发展与全球一体化进程持续推进，人工智能全球治理发展呈现出国际规制结构转换不断加速、多样化人工智能框架逐渐形成、政策重心向多维度治理转移、治理主体趋于多元化四方面特征，人工智能全球治理格局的重构进入关键阶段[17][18][19]。

1. 国际规制性立法正逐步增强，扩张性政策相对减少。人工智能国际规制性立法朝着确保技术安全和伦理的方向逐步增强，以促进技术自由发展为主导的扩张性政策相对减少。以往全球的人工智能法案主要集中于扩张性法案，旨在增强国家的人工智能技术水平，现在愈发倾向于规制性立法，对人工智能的使用施加限制，为人工智能的发展设定明确的规则和界限。随着政策制定者越来越关注人工智能融入社会的潜在危害，意识到规范人工智能的必要性，各国针对人工智能规范性治理出台大量政策。2023 年在全球通过的 28 项 AI 相关法案中，18 项属于中等相关度的监管法规，体现了各国对 AI 风险管控的重视。各国大力发展人工智能，希望其成为国家经济科技实力乃至综合国力增长的重要驱动力量。随着人工智能技术逐渐成熟，与社会各领域的融合不断加深，诸如数据泄露、算法偏见和伦理道德等问题慢慢凸显。全球人工智能竞争愈发激烈，单纯依靠扩张性政策提升竞争能力已经不足以维持竞争优势地位，通过制定严格的规制性立法，提高人工智能产业的门槛和标准更加有利于国家利益的发展。欧盟于 2023 年 12 月颁布《人工智能法案》（Artificial Intelligence Act），建立了基于风险的监管框架，将人工智能系统划分不同风险等级，对高风险系统实施严格监管，要求其遵守多项要求，对低风险人工智能要求其遵守基本的透明度义务，同时明确禁止 AI 系统存在过高风险。各国政府和国际组织在人工智能全球治理领域渐渐朝着达成共识的方向前进，越来越重视对人工智能技术的发展和应用进行规范和监管以确保技术的安全、可靠和符合道德伦理标准。

2. 多边合作框架正在形成，“分而治之”趋势明显。人工智能多边合作框架正在构建，各国各地区治理需求上存在显著差异，呈现出“分而治之”的明显趋势。各国基于共同利益考量，积极寻求在人工智能治理方面的共识，推动人工智能多边合作框架逐步构建。以 2023 年 11 月的英国人工智能安全峰会（AI Safety Summit）为例，包括中国和美国在内的 28 个国家共同签署了布莱切利宣言，标志着国际社会在 AI 治理方面达成初步共识。联合国等相关国际组织已经将人工智能全球治理相关议题纳入工作层面，致力于推动人工智能全球治理朝着更加合理有序的方向发展。在人工智能多边合作框架形成中，不同的治理理念和利益诉求导致各国在人工智能治理上的侧重点和立场不同，各国内部治理模式渐渐分隔，呈现“分而治之”倾向。欧盟出台的《通用数据保护条例》（General Data Protection Regulation）

严格规定了企业使用个人数据的规则，对违反规定的企业处以高额罚款，而美国加利福尼亚州的《加州消费者隐私法案》（California Consumer Privacy Act）主要依靠行业自律，数据隐私法规相对宽松。各国发达程度不同，利益诉求也有差异，发达国家希望通过人工智能技术保持其在全球的科技领先地位和经济优势，主导人工智能国际规则的制定，在人工智能治理方面更强调保护知识产权、数据安全等，而发展中国家则更关注人工智能技术的普及和应用，希望通过人工智能技术提升国家的经济发展水平和社会治理能力，保障国家数据主权，在其治理方面更强调技术的可及性、公平性等。此外，国家利益和价值观念主导着人工智能治理模式的差异，如美西方担心中国在人工智能领域的崛起会挑战其长期以来的技术优势地位，影响其在全球经济和科技领域的主导权，进一步对本土企业施加更大压力，要求其限制向中国提供先进的人工智能与半导体技术[20][21][22][23]。在这种情况下各国可能会更加注重保护自身的技术优势和利益，在全球人工智能治理的关键问题上分歧加大，导致人工智能全球治理机制难以顺利建立。

3. 政策重心向多维度治理转移，治理覆盖领域不断扩大。人工智能政策导向发生显著转变，政策重心已逐步向多维度治理倾斜，治理覆盖领域呈现出持续扩大的态势。人工智能政策重心向多维度治理转移意味着政策关注点不再局限于单一维度，治理覆盖领域拓展到医疗、交通、金融、教育等多个行业，涵盖了从技术研发、数据管理到应用推广、社会影响评估等全生命周期环节。人工智能问世以来，不同发展阶段关注的问题逐渐增加，早期全球关注技术安全，保护信息数据等重要战略资源不外泄，如今随着人工智能运用到多个领域，人工智能治理覆盖范围也在不断扩大。在人工智能时代，高水平教育资源的稀缺性不断降低，获取知识的方式更加便捷[24][25]。从研发环节的数据使用规范，到应用环节的市场监管，人工智能几乎贯穿了人类生产生活的全过程，其发展的复杂性和全球性要求多维度、广覆盖的治理来保障人类利益和社会稳定。人工智能发展带来的外部性影响巨大，各国进行多维度治理的同时扩大治理领域，通过一系列落到实处的具体措施使得人工智能的研发者和经营者不得不考虑新增加的成本问题从而将外部性内部化。2023 年 7 月，中国发布《生成式人工智能服务管理暂行办法》，对生成式人工智能服务进行规范，覆盖技术研发、应用场景和数据管理等多个领域。国家制定人工智能规范条例和政策，企业需要调整自身业务模式和管理政策适应政策变化，企业合规成本增加，从生产端控制人工智能风险，将人工智能可能产生的负外部性影响内部化。

4. 政企合作治理模式初现，治理主体趋于多元化。人工智能领域政企合作的治理模式初步显现，治理主体正朝着多元化方向发展。随着政府和企业合作共同推动人工智能治理的模式逐渐兴起，人工智能的治理不再只依靠单一的政府或国际组织，企业、科研学术机构和社会组织渐渐参与其中[26][27][28]。政府凭借其政策制定和监管优势，企业依靠技术创新和实践经验，二者实现互联互通，携手为人工智能治理注入新活力。2023 年 7 月，包括谷歌、微软、Meta 等七大 AI 企业与白宫达成自愿承诺，承诺在 AI 系统发布前进行安全评估、分享风险信息等，体现了多方参与的治理趋势①。人工智能技术具有复杂性，企业作为人工智能的研发者，比政府更加了解人工智能，与企业合作政府能够更好地理解其内在逻辑，制定出切实可行的技术标准和监管措施。对于企业而言，人工智能产业蕴含巨大的经济潜力，积极参与人工智能治理有助于其获得市场竞争优势地位。此外，人工智能产品和服务的应用是全球性的，跨国企业的业务遍及世界，单一国家难以对跨国的人工智能业务进行有效监管，需要多国政府和企业共同协作，才能够保证数据存储、处理等环节符合各国法律规定和安全标准。在人工智能跨国企业业务发展的推动下，形成了政企合作治理、治理

① 数据来源：《Artificial Intelligence Index Report 2024》。

主体趋于多元化的格局。

二、人工智能全球治理的困境

（一）治理参与的代表性不足，发展中国家在治理对话中普遍缺位

人工智能全球治理对话中发展中国家普遍缺位，代表性的严重缺失致使治理对话呈现失衡格局，难以构建全面且公正的人工智能全球治理框架。当前人工智能领域的规则制定与战略研讨被少数发达国家主导，国际会议、标准设定组织以及相关决策平台上来自发展中国家的声音微弱。发展中国家由于在人工智能技术研发、基础设施建设和专业人才储备等方面相对薄弱，缺乏深度参与人工智能全球治理所需的技术话语权。受限于经济实力，发展中国家难以大规模投入资源开展相关研究和国际交流合作，在治理体系中的影响力自然不足。此外，美国为代表的西方发达国家限制人工智能技术向其他发展中国家扩散，如动用政策和法律打压中国智能科技企业和研究机构，遏制中国在人工智能等高科技领域的发展，使得人工智能技术的发展方向和创新路径主要由其掌控，其他国家很难参与其中。在 7 个主要的非联合国 AI 治理倡议中，只有 7 个国家（加拿大、法国、德国、意大利、日本、英国和美国）参与了全部倡议，而有 118 个国家（主要是全球南方国家）完全未参与任何倡议[①]。发展中国家拥有庞大的人口基数与多样的应用场景，其独特需求与视角若被忽视，诸如贫困等特殊社会问题得不到解决，易引发数据隐私、伦理道德标准在不同发展水平国家间的差异与冲突。这种治理代表性的失衡导致了 AI 发展红利分配不均，扩大科技鸿沟，各国综合实力差距增大，阻碍了国际政治经济格局向公平包容方向发展。

（二）现有治理机制之间协调性不足，规则竞争化武器化趋势凸显

人工智能全球治理机制之间协调性不足，规则的竞争化与武器化趋向显著。国际组织在调和各国利益、促进协同治理方面进展缓慢，各组织间也缺乏沟通协调机制，在联合国系统内部，虽然许多机构都涉及 AI 治理，但由于具体职责划分，没有具体机构能够全面处理或主导 AI 治理问题。例如联合国教科文组织主要侧重于从伦理和教育文化等角度对人工智能进行规范和引导，但对于人工智能的技术标准制定和市场监管等方面则较少涉及[②]。联合国下属机构世界知识产权组织主要关注人工智能相关的知识产权问题，但对于人工智能的伦理道德问题则无法处理[③]。此外，各国将人工智能视为关键技术竞争领域，发达国家掌握主导权，就会倾向于选择或者创建符合自身利益偏好的排他性全球或地区治理机制。美国、欧盟等掌握相对技术优势的经济体通过频繁出台措施限制 AI 技术及相关产品出口，颁布法案限制人工智能人才交流，在国际人工智能合作中设置障碍，企图主导全球 AI 治理规则制定，使得规则竞争大于规则对接合作，规则武器化趋势明显。这种碎片化趋势可能导致世界分裂为互不兼容的 AI 治理体系，安全问题更加突出。技术标准的差异化以及对接难度的加大，增加了数据传输过程中的安全风险，监管政策不同导致企业为规避监管将业务转移至监管宽松地区，造成安全监管空白地带，增加了安全隐患。由于人工智能治理机制不协调，国家之间在人工智能技术交流方面受到诸多限制，阻碍了技术交流和创新扩散。规则竞争化武器化趋势凸显，导致各国建立起自身的人工智能市场准入标准从而形成贸易壁垒，引发不正当的市场竞争。

（三）技术规范呈现多元化发展态势，技术规则间协调难度较大

在人工智能全球治理进程中逐渐形成治理标准和技术规范多元化发展格局，各规则体系间的协调统一面临着巨大挑战。全球多个国家和地区以及国际组织等多种治理主体参与人工智能全球治理规则制定，不同国家和地区对于人工智能的认识和发展需求以及人工智能应用行业特点不同，

① 数据来源：《Governing AI for Humanity》。

② 资料来源：联合国教科文组织官网，http://www.unesco.org/。

③ 资料来源：《WIPO Technology Trends 2019–Artificial Intelligence》。

导致治理标准和规范存在差异，加之国际上缺乏权威、统一的协调机制来整合各国治理标准和技术规范，难以达成广泛有约束力的国际协议。国际电信联盟、国际标准化组织、国际电工委员会和电气与电子工程师协会等机构都在制定 AI 标准[①]，但这些标准之间缺乏共同语言，许多关键概念如公平性、安全性、透明度等没有统一定义。各国际组织的主要职责、价值观、治理需求差异化，使得所制定的技术规范和行业标准具有一定的局限性和偏向性，如国际电信联盟作为联合国的专门机构，在人工智能治理标准制定时会侧重考虑数据传输等技术规范问题[②]，而电器与电子工程师协会作为一个非政府组织，其治理标准往往反映的是行业内技术领先企业和专家群体的利益和技术倾向[③]。在缺乏具有强制执行力的国际组织决议的情况下，人工智能全球治理标准无法有效整合。在数据隐私方面，中国颁布的《中华人民共和国数据安全法》构建了数据安全和个人信息保护的法律框架，而美国的数据隐私规则则相对碎片化，在联邦层面没有统一的法律。不同治理标准反映不同的社会价值观和伦理观念，不同国家基于自身的文化、社会结构等因素有不同的定义和衡量标准，技术规范以及技术标准的多元化使得全球难以达成统一的人工智能治理共识，加剧了国家间的竞争和战略博弈，促使地缘政治格局围绕人工智能治理进一步分化重组，影响全球治理体系的和谐稳定与有效运行。

（四）治理实施效果与承诺存在差距，治理红利释放不足

人工智能全球治理在愿景与实操之间失衡，承诺未能有效转化为实际成果，治理效能的短板致使其红利难以充分释放。经济全球化、科技全球化、信息全球化乃至治理机制的全球化，不可避免地带来人工智能风险的全球性。国际组织提出的治理原则往往停留在理论宣示阶段，缺乏具体的实施细则与强制执行机制，难以对各国及各类主体的行为形成有效约束，加之各国在人工智能治理上的步调不一，导致治理实施效果与承诺存在差距。目前很多治理措施仍停留在自愿性遵守层面，如日本发布的《人工智能原则》，其中多为倡导性建议，缺乏强制执行力，企业和机构在开发应用人工智能时可选择性遵守，使得治理效果受限，实践与承诺之间存在明显差距。各国政府和国际组织都强调要确保人工智能系统的安全性和可靠性，但尽管有安全标准和测试要求，在复杂的现实场景中人工智能系统仍可能出现故障。由于人工智能治理的滞后，人工智能在众多应用领域面临诸多伦理与法律争议，企业在人工智能研发与应用投入上存在顾虑，技术创新的规模效应难以形成，治理红利无法充分释放。例如在面部识别技术领域，176 个国家正在使用人工智能监控系统[④]，但是由于其应用涉及隐私权问题，国际商业机器公司（International Business Machines Corporation，IBM）停止提供用于大规模监控的技术，使得面部识别技术在安防等领域的推广应用受限，在提升公共安全等方面的治理红利也无法充分释放。人工智能涉及多个领域知识，具备专业素养的人才短缺，同时由于人工智能系统的复杂性需要先进的监测和评估技术，然而当前的监管技术难以对人工智能进行有效监控。治理红利释放不足使企业难以充分利用人工智能进行创新，导致人工智能行业的发展在不同国家或地区出现失衡，促进人工智能社会层面利好的效应持续减弱，助长人工智能的恶意应用，甚至对国际安全秩序构成潜在威胁。

三、加快构建人工智能全球治理的中国逻辑和路径

（一）中国逻辑：一个愿景、三大原则、四项行动

在人工智能全球治理秩序变革的关键期，中

① 数据来源：《Governing AI for Humanity》。
② 资料来源：经济日报。
③ 资料来源：《对企业使用人工智能的呼吁》。
④ 数据来源：《人工智能全球监控指数》。

国在全球人工智能治理变革过程中的角色愈发重要。因此，应以中国人工治理的政策经验和实践为基础，为人工智能全球治理改革提出“一个愿景、三大原则、四项行动”的变革框架。一个愿景指构建平等包容、共建共享的人工智能全球治理体系并惠及全人类的伟大愿景；三大原则指发展与安全并重、包容合作和公平正义；四项行动指制度建设、技术创新、能力建设和伦理规范。

2023 年习近平主席出席第三届“一带一路”国际合作高峰论坛开幕式，在主旨演讲中提出了《全球人工智能治理倡议》，表达了构建包容、公平、共享的人工智能全球治理体系，促进人工智能技术惠及全人类的伟大愿景。旨在打破国家和地区之间在人工智能发展进程中的壁垒，让不同经济水平、科技实力的国家都能参与人工智能的全球合作网络，无论是发达国家的前沿技术研发成果，还是发展中国家的应用场景探索经验，均能自由交流和融合。在人工智能治理中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。包含了务实、合作与公平等新质治理力的中国倡议，既是对中外治理经验的系统总结，也是一国战略与全球定位融合的积极尝试 [29]。

发展与安全并重、包容合作和公平正义三大原则是建设人工智能全球治理秩序的核心逻辑。一是秉持发展与安全并重原则，推动人工智能发展的同时，高度重视安全问题。在国家战略层面，积极推进 AI 发展，通过强化顶层设计和统筹规划，力争在新一轮国际科技竞争中掌握主导权。在国家安全层面，加强管控 AI 所带来的安全风险，不断提升 AI 的安全性、可靠性、可控性、公平性，并确保其始终朝着有利于人类文明进步的方向发展。二是秉持包容合作原则，充分尊重各国政策和实践差异性，通过对话与合作凝聚共识。全人类共同价值深谙人类社会的多样性、差异性，它虽然倡导“共同”，但并不否定矛盾和分歧，而是始终尊重各国自主选择的制度模式和发展道路，始终主张各国在全球治理中承担“共同但有区别的责任”[30][31]。倡导各国摒弃技术保护主义与地缘政治竞争思维，以开放包容的心态接纳不同国家在人工智能领域的多元发展路径与特色模式，推动多利益攸关方积极参与，在人工智能全球治理领域形成广泛共识。三是秉持公平正义原则，确保 AI 发展红利惠及发展中国家，缩小数字鸿沟，推动全球 AI 治理民主化。增强发展中国家在人工智能全球治理中的代表性和发言权，确保各国人工智能发展与治理的权利平等、机会平等、规则平等，开展面向发展中国家的国际合作与援助，不断弥合智能鸿沟和治理能力差距，让全球人工智能产业的发展红利在各国间合理分配。

制度建设、技术创新、能力建设和伦理规范四项行动是建设人工智能全球治理秩序的基本方案。一是制度建设为人工智能全球治理筑牢根基。应积极参与国际人工智能治理规则制定，推动建立科学合理的全球 AI 治理框架，完善国内 AI 监管制度体系。二是技术创新成为驱动人工智能蓬勃发展的核心引擎。应加强人工智能核心技术研发，推动 AI 产业健康发展，为全球 AI 治理贡献中国方案和中国智慧。三是能力建设是提升人工智能全球治理整体参与水平的有力支撑。应完善人工智能人才培养体系，促进国际交流合作，帮助发展中国家提升 AI 治理能力。四是伦理规范为人工智能安全治理的设定底线。应推动建立人类命运共同体理念下的 AI 伦理框架，构建符合人类价值观与国际规范的伦理原则，分享中国在伦理规范制定与实践方面的经验，引导全球人工智能可持续发展。

（二）推动人工智能全球治理改革的中国路径

1. 深化多边合作机制建设，借力“AI 热”完善国内治理体系。积极投身国际多边合作框架下的人工智能议题探讨和规则共商，借全球“AI 热”之势，从多维度发力构建并完善国内人工智能治理体系。在涉及人工智能的国际组织中积极倡导改革成员参与机制，增加发展中国家的投票权和话语权。积极参与国际标准制定，定期举办人工智能全球治理主题论坛，邀请发展中国家政府官员、企业界人士和专家学者代表共同探讨人工智能全球治理关键议题，鼓励发展中国家代表分享观点和经验，促进不同国家立场和观点的交流与

融合。此外，进一步加强与发达国家的对话与合作，特别是在 AI 安全、伦理等领域，学习发达国家在人工智能治理领域的先进方案。在“AI 热”进程中加快完善国内监管框架，推动制度监管、提升技术智治，以建立强大和适应性强的监管实践来增进全球人工智能安全治理信任。可设立人工智能全球治理专项基金，用于支持发展中国家的相关项目和研究，基于自身经验帮助发展中国家合理配置人工智能资源，完善其国内人工智能治理体系。

2. 聚焦技术安全与伦理标准制定，推动国际 AI 产业创新合作。以构建兼顾安全与创新的技术伦理标准体系为基，积极引领国际合作平台搭建与规则塑造，汇聚全球智慧与资源，推动 AI 产业持续创新。主张广泛征求全球各界关于人工智能安全规范的意见，确保标准的科学性和合理性。将中国传统哲学思想中的有益元素融入人工智能伦理标准之中，如“仁智合一”思想可启发对人工智能智能水平与道德责任匹配性的思考，从而为国际标准制定提供独特的中国视角与智慧。应推动人工智能的国家治理与全球治理的统一，顺应全球技术与产业革命发展大势，在不断强化人工智能技术发展的既有优势的基础上，制定和完善扶持优势智能产业的政策工具箱，助力企业推进数字化转型与智能化改革。鼓励企业加大在深度合成和大模型技术上的研发投入，引导企业良性竞争，增强人工智能知识外溢效应，建立人工智能创新合作园或产业集聚，在深度合成、大模型等领域形成技术优势，推动 AI 产业创新合作。

3. 注重数据跨境流动规则建设，保障国家数据安全稳定。构建多层次数据跨境流动监管框架，平衡数据开放与安全，确保国家数据主权不受侵害，实现数据跨境有序流动与安全保障的协同共进。与主要贸易伙伴和数据大国签订双边或多边数据跨境流动协议，积极参与全球数据治理规则制定，推动建立公平合理的数据跨境流动机制。此外，应建立必要的数据安全审查制度，确保关键数据和核心技术安全。以分类分级、风险评估为核心手段，制定分层级的数据跨境流动分类管理规则，根据数据敏感程度将数据划分为不同类别，分别设定相应的跨境流动审批程序和安全评估标准。应建立国家层面的跨境数据流动监管机制，构建灵活性与全面性结合的跨境数据流动统一架构。加强国内数据安全基础设施建设，同时建立数据备份与恢复中心，确保在面临自然灾害、网络攻击等意外事件时数据的完整性和可用性，保障国家数据安全平稳运行。

4. 加快落实人工智能治理承诺，推动多边治理监督机制完善。制定具体的实施计划和时间表，确保国际组织和各国政府落实人工智能治理承诺。推动各国政府和国际组织共同制定和实施人工智能治理标准和规范，减少企业在全球范围内开展人工智能业务的障碍和风险。确保人工智能系统的安全性和可靠性，强化监管技术的研发和应用，提高监管效能。支持各国政府和国际组织开展人工智能治理宣传和教育活动，提高公众对人工智能治理的认识和理解。同时，加强对企业和机构的人工智能治理实施情况的监督和评估，确保其遵守相关标准和规范。加快培养具备专业素养的人才，推动人工智能监管技术的研发和应用，推动人工智能治理的国际合作和交流，加强人工智能治理的宣传和教育。

参考文献：

[1] 高奇琦 . 全球善智与全球合智：人工智能全球治理的未来 [J]. 世界经济与政治，2019，(7).

[2] 孙伟平 . 人工智能与人的“新异化”[J]. 中国社会科学，2020，(12).

[3] 武琼 . 美国人工智能反恐：路径、动因与挑战 [J]. 新疆社会科学，2022，(3).

[4] 武琼 . 科技向善：中国全球人工智能治理倡议的核心要义与理论价值 [J]. 和平与发展，2024，(5).

[5] 部彦君 . 智引未来：全球人工智能治理的现实困境与中国方案 [J/OL]. 当代经济管理，1-19[2025-01-07].http://kns.cnki.net/kcms/detail/13.1356.F.20240929.1658.004.html.

[6] 高奇琦 . 人工智能、四次工业革命与国际政治经济格局 [J]. 当代世界与社会主义，2019，(6).

[7] 韩娜，董小宇 . 全球人工智能安全治理的

信任困境与破解路径 [J]. 国际论坛，2024，（6）.

[8] 贾开，蒋余浩 . 人工智能治理的三个基本问题：技术逻辑、风险挑战与公共政策选择 [J]. 中国行政管理，2017，（10）.

[9] 韩永辉，张帆，彭嘉成 . 秩序重构：人工智能冲击下的全球经济治理 [J]. 世界经济与政治，2023，（1）.

[10] 郑煌杰 . 可信的人工智能：技术伦理风险下 AIGC 的治理基点 [J/OL]. 科技进步与对策，1-11[2025-01-07].http://hfffg3e1e79c5cd824acbh50o9qxfo9xqp6nfx.fxyh.librra.gdufs.edu.cn/kcms/detail/42.1224.G3.20241101.1042.002.html.

[11] 郭小东 . 大模型时代全球人工智能治理的挑战与中国方案 [J/OL]. 科学学研究，1-15[2025-01-07].https://doi.org/10.16192/j.cnki.-2053.20240724.001.

[12] 罗有成 . 数字时代主权的嬗变与国际安全秩序重塑 [J]. 国际展望，2024，（6）.

[13] Calvano E, Calzolari G, Denicolò V, et al. Artificial Intelligence, Algorithmic Pricing, and Collusion[J]. American Economic Review，2020，110，（10）.

[14] PINHEIRO LG. Protocols of Production: The Absent Factories of Digital Capitalism[J/OL]. American Political Science Review. Published online 2024:1-14. doi:10.1017/S0003055424000911.

[15] 闫广，忻华 . 权力互动与平衡性竞争——中美欧数字权力竞争的国际政治经济学分析 [J]. 国际展望，2024，（6）.

[16] 赵骏 . 国际法的守正与创新——以全球治理体系变革的规范需求为视角 [J]. 中国社会科学，2021，（5）.

[17] 吴汉东 . 人工智能时代的制度安排与法律规制 [J]. 法律科学（西北政法大学学报），2017，（5）.

[18] 阙天舒，张纪腾 . 美国人工智能战略新动向及其全球影响 [J]. 外交评论（外交学院学报），2020，（3）.

[19] 马欢，李磊，盛斌，等 . 全球生产智能化对中国贸易与福利的影响 [J]. 世界经济，2024，（9）.

[20] 刘蕊，熊炜 . 国际竞争、标准化战略与欧盟的技术权力 [J]. 太平洋学报，2024，（9）.

[21] 吴桐，刘宏松 . 地缘经济转向、数字主权与欧盟人工智能治理 [J]. 国际安全研究，2024，（5）.

[22] 杨楠 . 人工智能武器化与网络威慑战略的未来 [J]. 国际安全研究，2024，（5）.

[23] 陈小鼎，刘洋 . 数字外交对国际话语权博弈的影响及中国应对 [J]. 吉林大学社会科学学报，2024，（5）.

[24] 李益斌，李浩洋 . 欧美中人工智能监管规范比较研究 [J]. 当代世界与社会主义，2024，（5）.

[25] 汝鹏，苏竣，韩志弘，等 . 智能引领未来：生成式人工智能的社会影响与标准化治理 [J/OL]. 电子政务，1-13[2025-01-07].http://hfffg3e1e79c5cd824acbh50o9qxfo9xqp6nfx.fxyh.librra.gdufs.edu.cn/kcms/detail/11.5181.TP.20241121.1330.006.html.

[26] 支振锋 . 为互联网的国际治理贡献中国方案 [J]. 红旗文稿，2016，（2）.

[27] 吴志成，董柞壮 . 国际体系转型与全球治理变革 [J]. 南开学报（哲学社会科学版），2018，（1）.

[28] 张发林，杨佳伟 . 统筹兼治或分而治之——全球治理的体系分析框架 [J]. 世界经济与政治，2021，（3）.

[29] 全燕，张入迁 . 新质治理力：全球人工智能治理张力下中国倡议的机制创新 [J]. 传媒观察，2024，（11）.

[30] 殷文贵 . 批判与重塑：全球治理体系的内在缺陷及其变革转向 [J]. 社会主义研究，2021，（5）.

[31]Acemoglu D, Restrepo P. Demographics and Automation[J]. The Review of Economic Studies, 2022, 89(1).

作者：韩永辉，广东外语外贸大学广东国际战略研究院副院长、教授、博导、中组部“万人计划”青年拔尖人才、珠江学者

周港隽，广东外语外贸大学经济贸易学院助教

徐翠芬（通讯作者），广东外语外贸大学经济贸易学院助教

责任编辑：钟晓媚