人工智能治理的多元路径与中国方案[1]

王冬　高飞

【内容提要】人工智能治理已成为全球秩序重塑的关键场域。作为全球人工智能发展的三大主导力量，中国、美国与欧盟分别代表不同的治理取向。美国以安全逻辑与大国竞争为导向，依托技术封锁、封闭性联盟与企业主导的标准输出，谋求以技术主导及全球霸权为目的的战略型治理模式。欧盟以伦理规范与统一立法为支点，凭借“规范型权力”与制度外溢优势，打造人工智能治理的“制度护栏”。中国则以普惠共享的发展诉求为核心，通过开源生态、南南合作与多边实践，推动人工智能成为真正服务全人类福祉的全球公共品。中国方案以人类命运共同体理念为引领，以四大全球倡议为支柱，不仅展现出中国从价值倡导到规则供给、从制度设计到实践落地的全球治理引领力，更为推动形成开放、公平、普惠的人工智能治理新秩序注入动能。

【关　键　词】人类命运共同体　全球治理　人工智能治理　全球治理倡议　全球南方

【作者简介】王冬，外交学院亚洲研究所（国家安全学院）讲师；高飞，外交学院副院长、教授。

【中图分类号】D815
【文献标识码】A
【文章编号】1006-6241（2025）05-0001-24

[1] 本文系中央高校基本科研业务费专项资金项目（项目编号：3162023ZK04）的阶段性成果。作者感谢《和平与发展》匿名评审专家及编辑部对本文提出的宝贵修改意见，文中错漏概由本人负责。

2025年9月1日，习近平主席在“上海合作组织+”会议上发表讲话，正式提出全球治理倡议，倡导构建更加公正合理的全球治理体系，共同应对全球南方代表性严重不足及人工智能等新疆域治理缺位等问题，携手迈向人类命运共同体。[1] 该倡议与此前提出的全球发展倡议、全球安全倡议、全球文明倡议一道，构成了中国在全球治理议程中不断拓展和深化人类命运共同体理念的重要支柱。[2]

当前，人工智能作为新一轮科技革命与产业变革的核心驱动力，已成为全球治理体系改革与重塑的关键议题。不同国家与国家集团基于自身法律传统、价值取向与产业利益形成了差异化的治理模式。这些制度安排在服务于内部风险防控和产业引导的同时，更日益成为对外输出的对象，成为全球治理的重要组成部分。发达国家在议程设置和国际规则设计上居于绝对优势，全球南方国家则普遍缺乏话语权和制度参与能力，这种不对称使全球人工智能治理进一步地缘政治化，并加深了国际秩序的不公正性。面对这一关乎人类共同未来的挑战，以全球发展、全球安全、全球文明和全球治理四大倡议为支柱，推动人类命运共同体理念转化为制度供给与治理实践，构建公正、合理、开放与包容的全球人工智能治理新格局，具有深刻的现实意义。

本文将尝试在系统梳理中、美、欧人工智能治理模式的基础上，分析多元路径背后的政治博弈及其全球治理之间的互动逻辑，揭示中国智慧与中国方案在全球人工智能治理中的独特贡献，旨在为学界对人工智能治理的制度比较与理论深化提供参考，同时也为实务界推动全球人工智能治理提供实践启示。

---

[1] 参见《凝聚上合力量 完善全球治理——习近平在“上海合作组织+”会议上的讲话》，外交部网站，2025年9月1日，https://www.mfa.gov.cn/zyxw/202509/t20250901_11699539.shtml；《全球治理倡议概念文件》，外交部网站，2025年9月1日，https://www.mfa.gov.cn/wjbxw_new/202509/t20250901_11699909.shtml 等。

[2] 徐坚：《深刻把握全球治理倡议的重大意义与时代价值》，载《人民日报》2025年9月15日，第9版。

一、人工智能治理面临的机遇与挑战

人工智能治理进程呈现出风险与价值并存、机遇与挑战同在的鲜明“二元性”特征。人工智能治理不仅是技术议题，更是一个国家或国家集团综合战略的延伸。

（一）风险与挑战

一方面，就人工智能发展而言，人工智能技术本身的系统设计、研发、训练、测试、部署、使用、维护等生命周期的各环节及衍生环节都可能存在风险，既包括内生风险，也包括衍生风险。

人工智能发展蕴藏的内生风险主要集中在技术领域，包括模型算法安全风险、数据安全风险、算力与系统安全风险。一是模型算法安全风险。深度学习等复杂算法的“黑箱”特性导致其输出结果难以解释，增加了异常情况下的溯源和修正难度。算法也可能因设计偏见或训练数据问题，输出带有歧视性或偏见的内容，甚至涉及民族、宗教等敏感领域。此外，模型鲁棒性（Robustness）[1]较弱可能导致性能下降或决策错误，模型的核心参数和结构可能被窃取或篡改，引发知识产权泄露或运行故障等风险。二是数据安全风险。首先，训练数据可能包含虚假、偏见或违法内容，导致模型输出有害信息；其次，数据标注不规范也会影响模型性能，导致偏见放大或泛化能力不足；最后，数据处理不当可能导致敏感信息泄露，威胁个人隐私和商业秘密。三是算力与系统安全风险。人工智能系统的开发和运行依赖于复杂的软硬件基础设施，这些设施可能存在缺陷或漏洞，甚至被植入后门，成为攻击者的目标。此外，算力资源的恶意消耗或供应链中断也会对系统安全构成威胁。特别是在全球化分工的背景下，技术垄断和出口管制可能导致芯片、软件等关键资源的断供，进一步加剧系统安全风险。

[1] 鲁棒性是指系统或者器件在不同的环境或者条件下，能够保持其功能或者性能的特性。鲁棒性越高，说明系统或者器件越稳定，越不容易受到干扰或者破坏。

人工智能发展的衍生风险主要体现在应用领域，涉及网络域、现实域、认知域、伦理域的个人隐私、组织财产、社会公平、国家安全以及全人类发展等多个层面。一是在网络领域的应用可能引发信息内容安全风险，如虚假信息传播、隐私泄露和侵权问题。二是在金融、能源、交通等传统领域的广泛应用，可能引发经济社会安全风险。例如，自动驾驶或智能诊疗系统的错误决策可能威胁用户生命财产安全。三是在信息域、认知域，可能加剧“信息茧房”效应，使用户局限于特定信息范围，影响其判断力和认知能力。四是在伦理领域，可能加剧社会歧视和偏见，造成传统价值观的冲击，甚至因脱离人类控制而带来潜在的生存性风险。[1] 五是在军事安全领域，人工智能在武器系统、指挥决策与战场感知等方面的应用，可能带来自主武器失控、军备竞争加剧及战略误判等潜在风险，对全球安全与稳定构成严峻挑战。

另一方面，在全球治理层面，不同国家治理理念差异、治理机制滞后以及南北失衡等问题交织，使人工智能治理赤字和地缘政治博弈不断加剧，成为全球共同面对的重大挑战。

这种挑战是多维度且日益复杂的，既在于技术发展速度远远超过制度建设能力，更在于治理赤字背后的深层次南北分歧。首先是治理理念的分歧，西方发达国家强调通过塑造技术优势与规则优势来锁定制度主导权，试图以伦理、安全和责任为名推进标准及规则外溢；而全球南方国家则呼吁获得公平的发展机会，强调在主权平等和发展权保障的前提下平等参与全球人工智能进程。其次是技术打压的加剧，西方国家通过技术封锁、投资审查、供应链限制等手段，强化对关键技术的控制，遏制新兴经济体的科技崛起；

[1] 关于人工智能风险，参见《人工智能安全治理框架》，全国网络安全标准化技术委员会网站，2024年9月9日，https://www.cac.gov.cn/2024-09/09/c_1727567886199789.htm；《人工智能治理蓝皮书》，中国信息通信研究院网站，2024年12月，https://www.caict.ac.cn/kxyj/qwfb/bps/202412/P020241227660032159191.pdf; Ray Kurzweil, *The Singularity Is Nearer: When We Merge with AI,* New York: Penguin Books, June 25, 2024, Chapter 7; Anders Indset, Florian Neukart, *The Singularity Paradox – Bridging the Gap Between Humanity and AI,* Hoboken, New Jersey: John Wiley & Sons, Inc, 2025, Chapter 1 等。

全球南方国家则立足于多元主义和要求构建一个开放包容的国际环境，推动知识共享与技术扩散，避免人工智能治理陷入碎片化和阵营对抗的格局。最后是南北失衡制度性的固化，发达国家在议程设置和国际规则设计上居于绝对优势，南方国家则普遍缺乏话语权和制度参与能力，这种不对称使全球人工智能治理进一步地缘政治化，并加深了国际秩序的不公正性。

总之，人工智能治理赤字的本质不是“没有规则”，而是谁制定规则、谁享有发展机会、谁掌握技术控制权。治理理念对立、技术打压和南北失衡相互叠加，使人工智能治理成为全球南方争取发展自主权、推动国际公平正义的重要场域。唯有打破这种不平衡，人工智能才能真正成为服务于全人类福祉的全球公共品。

（二）机遇与价值

作为全球技术竞争与国家实力博弈的重要领域，人工智能治理已超越单纯的技术管理范畴，成为重塑国际力量对比、构建新型多边关系、引领全球文明形态的重要战略工具。对于中国而言，人工智能等新疆域为中国参与引领全球治理体系改革提供了历史性机遇，是展现大国担当与全球领导力的关键舞台。相较于传统国际治理领域多由西方国家主导，规则体系早已定型，中国往往只能作为追随者在既有规则下参与改革，人工智能、网络空间、深海及外空等新疆域治理体系尚处于形成阶段，中国拥有更大的战略主动权与制度塑造空间。

人工智能治理本身也蕴含着重要的价值，既关乎生产力与生产效率的提升，也关系到安全风险防控与价值规范塑造。首先，在经济与发展层面，人工智能治理为各国提升生产效率、促进产业升级和改善公共服务提供了广阔空间，有效的制度安排能够保障公平竞争，避免数据垄断和算法歧视，使技术成果更多惠及民生与可持续发展。其次，在政治与安全层面，人工智能广泛应用于军事、网络空间与关键基础设施，其潜在风险与外溢效应使全球迫切需要通过规则协调来确保透明、可控和负责任的使用。因此，健全的治理框架不仅有助于防止技术滥用和误用，还能在跨国合作中增强

互信、促进战略稳定，为国际和平与安全提供新型保障。最后，在社会与文化层面，人工智能治理关系到伦理规范和价值导向的塑造，它能够推动以人为本、智能向善的理念落地，促进不同文明在数字时代实现交流互鉴，避免算法偏见和文化隔阂带来的不平等与排斥。尤其是对全球南方而言，一个强调主权平等、普惠共享和多边合作的治理体系，不仅能够提升其在国际事务中的代表性和发言权，还能确保其在人工智能发展中共享技术红利，推动发展成果更公平地惠及全人类。

二、中美欧的人工智能治理模式

作为当前全球人工智能发展的三大主导力量，中国、美国与欧盟分别代表了不同的治理取向。这不仅反映了全球人工智能治理的多元化趋势，也折射出国际秩序重构中的理念分歧与竞争态势。

（一）美国人工智能治理基于安全逻辑与战略控制构建全球霸权导向性的治理体系，以期将技术优势转化为对国际秩序的长期主导

1. 治理模式：竞争驱动的“分散化”体系

从治理逻辑上，美国人工智能治理以安全为优先，以竞争为导向。本质是一种以科技主导及全球霸权为目的的战略型治理模式。奥巴马政府后期，人工智能发展首次被上升至国家战略高度。2016 年的《国家人工智能研究与发展战略计划》确立了美国人工智能研发的基础框架，推动人工智能在科学研究、社会应用与国家竞争力之间实现衔接与统筹。[1] 在这一时期，美国的人工智能治理模式以创新驱动与产业引导为核心，通过政府资助基础研究、建设开放数据与标准体系、推动科研成果转化来强化国家创新生态。

特朗普第一任期，随着人工智能治理被纳入国家安全与大国竞争的核

[1] National Science and Technology Council, *National Artificial Intelligence Research and Development Strategic Plan,* Executive Office of the President, October 2016, https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf?utm_source=chatgpt.com.

心议程，以创新为导向的政策逻辑被进一步泛安全化与地缘政治化，演变为以国家利益和技术主导为核心的竞争驱动体系。一方面，特朗普主张通过放松监管和加大投资来刺激技术创新。2019 年的《美国人工智能倡议》首次提出“适度监管”，强调在保障创新活力的同时防止政府过度干预，确立以市场为主导、政府为引导的治理原则，以保持美国在全球人工智能竞争中的领先优势。[1]2020 年底特朗普批准《2020 年国家人工智能倡议法案》成法，并依据该法于 2021 年 1 月在白宫科技办公室下创设国家人工智能倡议办公室（NAIIO），进一步推动科研与产业的政策统筹。[2] 另一方面，以维护国家安全和防止“中国技术威胁”为名，推动人工智能技术领域由“创新合作”走向“竞争封锁”。特朗普 2018 年签署的《出口管制改革法案》首次将“人工智能及其子领域算法”，包括深度学习、计算机视觉、语音识别、自动驾驶等界定为新兴技术，授权商务部工业与安全局对其出口进行特别审查。[3] 2019 年，美国陆续将商汤科技、旷视科技、海康威视等中国人工智能企业列入“实体清单”，限制其获取美国芯片、GPU 算力与算法组件。

拜登政府在延续特朗普第一任期竞争逻辑的基础上，强化人工智能的伦理监管与国家安全属性，推动治理从技术竞争走向制度竞争。一方面，拜登通过 2022 年的《人工智能权利法案蓝图》与 2023 年的《关于安全、可靠、值得信赖地开发和使用人工智能的行政令》，提出以“安全、公平、透明、可问责”为核心的人工智能治理原则，确立了以伦理、信任与社会责任为导向的规范框架。2024 年 10 月，拜登政府又发布美国首份人工智能国家安全备忘录，要求联邦机构建立人工智能风险评估体系、关键基础设施防护机制与算法透明标准，实现从原则性治理向制度化监管的过渡。另

[1] Michael Kratsios and Lynne E. Parker, *American Artificial Intelligence Initiative: Year One Annual Report,* Office of Science and Technology Policy, Februry 2020, https://www.nitrd.gov/nitrdgroups/images/c/c1/American-AI-Initiative-One-Year-Annual-Report.pdf.

[2] The U.S. Congress, *National Artificial Intelligence Initiative Act of 2020,* March 2020, https://www.congress.gov/bill/116th-congress/house-bill/6216.

[3] U.S. Department of Commerce, Bureau of Industry and Security, *Review of Controls for Certain Emerging Technologies. Federal Register,* Volume 83, Number 223, November 19, 2018, https://www.federalregister.gov/documents/2018/11/19/2018-25221/review-of-controls-for-certain-emerging-technologies.

一方面，拜登政府以盟友协作为抓手，将人工智能纳入经济合作与发展组织（OECD）、“七国集团”（G7）等多边机制，推动话语体系和标准输出，意在通过价值同盟构建全球人工智能规则主导权。拜登政府末期还发布了《人工智能扩散框架》，强调以出口管制与技术认证双轨机制引导人工智能流向“可信伙伴国”，同时限制对竞争对手的高端算法、算力芯片和数据资源转移。总体而言，拜登时期的人工智能治理以安全与价值为双支柱，形成了兼具竞争性与制度性的治理体系，使人工智能治理成为重塑国际科技秩序与制度权威的重要战略工具。

特朗普第二任期开启后，美国的人工智能治理开始全面转向竞争驱动与行政动员并行的战略阶段，其核心在于以国家力量重塑人工智能领域的产业、基础设施与外交布局，以巩固美国在全球技术体系中的结构性主导地位。2025 年 1 月 23 日，特朗普签署《消除美国在人工智能领域领先地位的障碍》行政令，宣布撤销拜登时期某些阻碍美国人工智能创新的人工智能政策和指令，并强调此举为维持和加强美国在全球人工智能领域的主导地位扫清了道路。根据该行政令，同年 7 月，白宫发布《美国人工智能行动计划》及三项与之配套的总统行政令，即《促进美国人工智能技术栈的出口》《防止联邦政府出现“觉醒人工智能”》和《加速数据中心基础设施的联邦许可》，都具有极强的竞争驱动与行政动员特征。美国一方面在全球范围内构筑起以技术封锁、算力限制、数据管控和安全审查为支撑的竞争壁垒，将人工智能竞争推向大国博弈的前沿领域；另一方面通过豁免环评、简化程序等快速审批数据中心与芯片厂，加速基础设施落地，行政权力直接服务产业扩张，以“外紧内松”双轨并进谋取人工智能技术优势。

从人工智能治理结构上，美国以“联邦—州—行业”多层级分权为主要特征，呈现为竞争驱动下的“分散化”治理体系。[1] 相较于联邦层级的竞

[1] 参见 Amir Al-Maamari, “Between Innovation and Oversight: A Cross-Regional Study of AI Risk Management Frameworks in the EU, U.S., UK, and China,” arXiv, March 2025, https://arxiv.org/pdf/2503.05773; Lionel Levine, “Commentary: We Regulate Taco Carts More Than Artificial Intelligence,” *Times Union,* June 1, 2025, https://www.timesunion.com/opinion/article/commentary-regulate-taco-carts-artificial-20352168.php。

争驱动及创新优先，美国各州通过消费者数据保护法、深度伪造禁令等立法填补安全及风险监管的空白。根据全美州议会会议的数据，仅在 2024 年，就有超过 30 个州通过了与人工智能相关的法律或决议，包括禁止在选举中使用深度伪造技术、在招聘中歧视性地使用人工智能以及未经本人许可创建人物数字复制品等。[1] 科技巨头依托行业自治主导技术标准制定，如微软、谷歌等企业率先发布 AI 伦理准则。这种“上下分治型”治理，在赋予地方与企业灵活性的同时，也在一定程度上导致规则碎片化与各州之间的执法协调难题。[2] 正如美国国会众议院能源和商务委员会主席格思里 (Brett Guthrie) 所指出的，“虽然防止人工智能的有害使用至关重要，但联邦和州立法者应避免实施重复且繁重的新法规”。[3]

表 1　美国人工智能治理模式的分权结构

	联邦政府	州政府
角色定位	战略主导与国家竞争动员	补充治理与社会风险监管
治理机制	以软法为主：行政令、行动计划、指导性标准	以硬法为主：地方立法、强制性合规规定
治理工具	标准与指南等软法工具、拨款与采购等财政工具、出口管制等外交工具	州层面法律法规、监管机构执法、地方司法救济
执行力度	多为原则性框架，缺乏统一执法条款，依赖企业自愿遵守	设有明确执法机制，可推动地方层面的具体实施
赋权对象	重点赋权科技巨头和部分联邦机构，推动创新与国际竞争	赋权地方政府及本地监管机构，灵活回应区域社会问题

资料来源：作者自制

2. 治理理念：“技术优先”的霸权逻辑

在“技术优先”“创新自由”的话语框架下，美国通过放松监管、强化

[1] The National Conference of State Legislatures, “Artificial Intelligence 2024 Legislation,” September 9, 2024, https://www.ncsl.org/technology-and-communication/artificial-intelligence-2024-legislation.

[2] Amir Al-Maamari, “Between Innovation and Oversight: A Cross-Regional Study of AI Risk Management Frameworks in the EU, U.S., UK, and China,” arXiv, March 2025, https://arxiv.org/pdf/2503.05773.

[3] Brett Guthrie, “Driving the energy future of AI development,” *The Washington Times,* March 26, 2025, https://www.washingtontimes.com/news/2025/mar/26/driving-energy-future-ai-development/.

企业主导与政策激励，塑造出以技术进步带动国家实力与制度影响力的治理叙事。特朗普第二任期的《美国人工智能行动计划》与一系列行政令明确诠释了这一理念。2025 年 5 月，众议院通过的《大而美法案》更是提出在未来十年内禁止各州出台新的人工智能监管法律，并撤销全国范围内已有的相关规定，从而将监管权集中到联邦政府。这意味着联邦政府将既是产业发展的推动者，也是唯一的监管主体，其“监管”本身也更多体现为行政动员和政策宽松，而非法律约束与社会责任嵌入。[1] 尽管 2025 年 7 月该法案最终成法的版本中，原先的“禁止各州出台新的人工智能监管法律”条款被删除，但仍然强调建立问责机制和防范社会风险等。这一修正并未改变新一届特朗普政府弱监管的政策取向，其行政令中有关要求，如简化数据中心和芯片工厂建设的审批流程、削弱环境与安全评估要求以及放宽出口与国际合作限制等，进一步突显了技术优先及竞争驱动的治理逻辑，事实上为科技公司的快速扩张扫清了制度障碍。

特朗普的弱监管取向实质上服务于美国的全球竞争和科技霸权目标，形成以技术优势支撑地缘战略优势的竞争扩张逻辑。[2] 从地缘竞争视角看，弱监管是美国巩固“数字单极秩序”的核心工具，通过降低合规门槛吸引全球人工智能人才与资本向硅谷集聚，维持对芯片设计、大模型架构等核心技术的垄断性控制。联邦政府弱化监管刚性及将监管权下放至州与企业的策略，本质是规避欧盟式刚性规制对创新速度的抑制，确保美国在人工智能竞争中的优势。此外，美国分权制衡的政治传统亦强化了这一选择，在全面性人工智能立法难以达成共识的情况下，转向专门领域或州层级的

[1] The U.S. Congress, *One Big Beautiful Bill Act,* May 2025, https://www.congress.gov/bill/119th-congress/house-bill/1.

[2] 参见 Amir Al-Maamari, “Between Innovation and Oversight: A Cross-Regional Study of AI Risk Management Frameworks in the EU, U.S., UK, and China,” arXiv, March 2025, https://arxiv.org/pdf/2503.05773; Lionel Levine, “Commentary: We Regulate Taco Carts More Than Artificial Intelligence,” *Times Union,* June 1, 2025, https://www.timesunion.com/opinion/article/commentary-regulate-taco-carts-artificial-20352168.php; 王彦雨、李正风、高芳：《欧美人工智能治理模式比较研究》，载《科学学研究》2024 年第 3 期，第 460 页。

立法成为更具可行性的方案。这种弱监管逻辑能够帮助美国在全球人工智能产业链中持续占据“规则制定—技术输出—标准主导”的链主地位，成为其数字时代护持科技霸权的重要支撑。

3. 治理机制：软法驱动

美国人工智能治理的一个鲜明特征是对软法工具的偏好。在这种模式下，美国联邦层面并未通过统一的硬法形成全国性规范，而是依赖总统行政令、联邦机构指南、行业自律规则以及州级法规推动人工智能的发展。拜登政府时期发布的行政命令，虽提出 AI 开发、部署、监测中的风险规制框架，但其落地执行大多依赖企业的自愿承诺与行业自律。例如，《关于安全、可靠、值得信赖地开发和使用人工智能的行政令》仅提供非强制性指南，鼓励行业自主开展创新与测试[1]；美国国家标准与技术研究院（NIST）2023 年发布的《人工智能风险管理框架》亦采取引导性路径，倡导企业通过风险识别、治理机制、自我审查等方式达成“责任化开发”目标。[2] 这种模式下，科技巨头如微软、谷歌、OpenAI 等主动发布伦理准则、参与标准制定，在治理中扮演半官方的准制度制定者角色。到了特朗普第二任期，无论是行动计划还是总统行政令都刻意回避全国统一的强制性成文法，而是转而要求各联邦机构优先动用指南、标准、评测、采购与资助等一整套软法工具包来塑形市场行为。[3]

（二）欧盟人工智能治理力图借助统一立法与跨国协调来构建全球规则“最严密的”治理体系，以期在中美技术竞争中塑造独特的“规则优势”

---

[1] The White House, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,* October 30, 2023, https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[2] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0),* January 26, 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[3] Amir Al-Maamari, “Between Innovation and Oversight: A Cross-Regional Study of AI Risk Management Frameworks in the EU, U.S., UK, and China,” arXiv, March 2025, https://arxiv.org/pdf/2503.05773.

1. 治理模式：规则驱动的“伞式”体系

欧盟作为地区一体化程度最高的政治经济体，近年来在人工智能治理方面构建了以超国家机构为核心、成员国协同参与的区域“伞式”治理体系。[1]该治理体系以统一监管、伦理主导与法律框架为基础，反映出其通过制度设计参与全球科技秩序重构的战略取向。然而，该模式在彰显“规范型权力”与制度外溢优势的同时，也以高标准合规要求与数据主权保护限制对第三国主体形成了“隐形壁垒”，折射出欧盟在全球治理中从“制度引领”向“制度护栏”转化的双重逻辑。

欧盟人工智能治理体系以2024年《人工智能法》（AIA）与2016年《通用数据保护条例》（GDPR）为基石，呈现出三个层次的协同网络。[2]在宏观层面，设立于欧盟委员会内部的欧洲人工智能办公室作为AIA执行的中枢机构，与欧洲数据保护委员会（EDPB）承担人工智能治理的顶层设计、跨境政策协调与规则统一职责，制定高层次监管标准并确保成员国法律框架一致；在中观层面，各成员国设立国家级数据监察机构及国家市场监督管理机构，如法国的国家信息与自由委员会、德国的联邦数据保护与信息自由专员，负责本国范围内的法规执行、行业审查与执法干预；在微观层面，企业、公共机构等合规主体设立数据保护官，具体落实数据处理规范、风险评估与内部监督机制，承担一线合规责任。

在人工智能风险管理方面，欧盟采取了统一标准、国家执行、企业落地的协调路径，包括要求各成员国设立国家级“人工智能监管沙盒”（AI Regulatory Sandbox）[3]机制并于2026年8月2日前启动，允许企业在受控

[1] European Parliamentary Research Service, *Artificial Intelligence Act and GDPR: Complementary Legal Frameworks for AI Governance,* European Parliament, 2022, https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)699504.

[2] Ibid.

[3] 人工智能监管沙盒，是指监管机构允许人工智能企业或研发团队在一个受控环境下，对其新产品或新服务进行有限范围内的测试与试点。在沙盒中，企业可在遵守基本安全与合规底线的前提下，获得暂时性监管豁免或灵活适用的规则，以便探索创新应用并积累风险管理经验。

环境中进行人工智能产品测试，实现政策的“实验性弹性”；建立独立的科学专家组协助人工智能系统进行风险评估、分类定级，并为成员国监管提供技术支撑与科学判断；倡导设立公众申诉与救济渠道，接受公众、企业就人工智能应用的权利侵犯、算法歧视等问题提出申诉，增强公众参与与权利保障。[1]

然而，欧盟的人工智能治理模式在强化内部监管的同时，也因其域外适用、数据保护与高标准合规要求而对他国构成隐形限制。例如，GDPR通过严格的数据跨境传输规则，将数据流动建立在“充分性认定”与补充保障机制之上，使非欧盟国家在缺乏对等制度时难以进入其数据生态体系。更为深层的影响在于，欧盟通过“规则出口”和“布鲁塞尔效应”[2]将自身监管体系外溢，使全球厂商在进入欧盟市场时被迫采纳其标准，呈现出“制度护栏”与“市场保护”交织的特征。

表2　美国和欧盟人工智能治理模式对比

	美国	欧盟
治理模式	“分散化”、竞争驱动型	“伞式”统一、集中管控型
治理理念	霸权逻辑	伦理优先
治理机制	软法驱动	硬法驱动
治理稳定性	碎片化、党派摇摆大	集中化、法律稳定性强
治理弹性	强，地方和企业创新空间大	弱，依托监管沙盒有限灵活性

资料来源：作者自制

2. 治理理念：伦理优先与风险预防逻辑

欧盟的人工智能治理以伦理优先和风险预防为核心逻辑，既有其历史

[1] 参见 The European Commission, “Governance and enforcement of the AI Act,” February 17, 2025, https://digital-strategy.ec.europa.eu/en/policies/ai-act-governance-and-enforcement?utm_source=chatgpt.com 等。

[2]“布鲁塞尔效应”是指欧盟通过市场机制推动欧盟法律的域外适用，导致其单边监管全球化的过程。第三国企业为了进入欧盟市场或与欧盟的企业开展业务，不得不遵守欧盟法律。

传统与现实困境交织的原因，也是当前地缘政治博弈的战略选择。从内部动因看，欧洲长期奉行“社会市场经济”模式，强调国家干预与社会公平，这种制度基因使其天然倾向于通过刚性规制平衡技术权力；深厚的人权法传统则为伦理治理提供了法理正当性。[1] 从外部竞争视角看，欧盟在人工智能技术水平相对落后于中美。2023 年，全球 AI 百强企业中仅 8 家来自欧洲，而美国占 58 家、中国占 17 家。[2] 2024 年，全球标志性人工智能模型仅有 3 个来自欧洲，而美国有 40 个，中国 15 个。[3] 欧盟选择以“规则替代技术”的非对称竞争策略，旨在弥补技术实力短板，以免丧失全球数字秩序话语权；美国则通过弹性监管维系创新生态，巩固技术霸权地位。两者实质是“规范性软实力”与“技术硬实力”在全球治理场域的碰撞，诠释着数字时代“安全—发展—伦理”三角关系的不同权重分配。

3. 治理机制：硬法约束

欧盟通过硬法工具确立全域覆盖与域外延伸的刚性规则。通过《通用数据保护条例》确立全球个人数据保护标准，迫使跨国企业重构数据治理体系；通过《人工智能法》创设风险分类框架，将风险管理嵌入人工智能技术全生命周期，禁止社会评分等高危应用，并对人工智能产业链实施全链条追责。欧盟委员会和欧洲数据保护委员会、成员国数据监察机构和市场监督管理机构、企业数据保护官落实内部合规，欧盟、成员国、企业三级联动确保规则穿透力。[4] 然而，欧盟治理模式也面临诸多挑战，例如，法

[1] European Parliamentary Research Service, “Ethics and Artificial Intelligence: The European Approach,” European Parliament, 2020, https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2020)641543.

[2] Bergur Thormundsson, “Artificial intelligence (AI) worldwide – statistics & facts,” Statista, August 18, 2025, https://www.statista.com/topics/3104/artificial-intelligence-ai-worldwide/#topicOverview.

[3] Stanford Institute for Human-Centered Artificial Intelligence, *AI Index Report 2025,* April 2025, https://hai.stanford.edu/ai-index/2025.

[4] 参见 European Parliamentary Research Service, *Artificial Intelligence Act and GDPR: Complementary Legal Frameworks for AI Governance,* European Parliament, 2022, https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)699504。

律的更新速度难以跟上技术演进，成员国间监管协调难度较大，合规成本较高可能抑制中小企业创新活力，公众申诉和权利救济机制仍缺乏可操作性细则。此外，欧盟治理机制的刚性导向也存在复杂的立法博弈与内生冲突。例如，欧盟《人工智能法》历时 3 年磋商，立法过程反映出多重利益主体的角力。北欧国家强调保护公民隐私权，坚决反对在公共场所使用生物识别技术进行监控，以维护个人自由和隐私；德国和法国担心严格的监管可能阻碍本国人工智能产业的发展，推动在法案中加入工业豁免条款，以保护本土人工智能研发的灵活性；中东欧国家则普遍持更务实的产业导向态度，倾向于支持宽松监管以吸引投资与技术转移。这种“监管宽容”立场虽有助于增强区域内创新活力，却也加剧了欧盟内部在价值优先与效率优先之间的结构性张力。欧盟在数字主权抱负与技术依赖现实间的摇摆，揭示了硬法治理范式在效率与伦理平衡上的深层挑战。

（三）中国人工智能治理通过“发展与安全并重”的政策逻辑，构建由政府主导、企业和社会多元参与、软硬法协同的治理体系

1. 治理模式：政府主导与多元共治

中国的人工智能治理是以政府主导、多元协同为核心框架，推动构建政府、企业、行业组织等公私主体协同参与的治理体系。在政府层面，通过发布纲领性文件引导人工智能发展与治理，推动跨部门协作。2017 年，国家科技体制改革和创新体系建设领导小组牵头协调并制定《新一代人工智能发展规划》，成立人工智能规划推进办公室，明确顶层设计框架[1]；2021 至 2024 年间，网信办联合工信部、科技部、公安部等多部门印发算法推荐、深度合成管理等细分领域法规，构建“中央统筹—部门联动—地方落实”的纵向治理链条。[2] 在多元参与层面，积极鼓励国际组织、企业、研究机构、社会组织和个人等多元主体积极发挥与自身角色相匹配的作用，

[1]《国务院关于印发新一代人工智能发展规划的通知》，中国政府网，2017 年 7 月 8 日，https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm。

[2] 如 2021 年《关于加强互联网信息服务算法综合治理的指导意见》、2022 年《互联网信息服务算法推荐管理规定》、2023 年《生成式人工智能服务管理暂行办法》等。

参与人工智能治理体系的构建和实施，如企业需履行算法安全评估、数据合规等主体责任，行业组织通过自律公约细化技术伦理标准。同时，推动建立多元化的参与机制，包括公开征求意见、社会调查等方式，让公众参与到人工智能决策过程。2022 年中共中央办公厅、国务院办公厅印发《关于加强科技伦理治理的意见》，要求健全多方参与、协同共治的科技伦理治理体制机制。这一框架通过制度化分工与动态协调机制，系统性回应人工智能技术快速发展中的安全与伦理挑战。[1]

2. 治理理念：统筹发展和安全

中国坚持统筹安全与发展的核心理念，通过政策工具的协同设计与制度创新，推动实现人工智能治理在风险可控前提下加强创新激励的治理目标。以安全筑基发展，以发展反哺安全。2017 年《新一代人工智能发展规划》明确提出“推动人工智能与经济、社会、国防深度融合”，既将安全可控作为技术应用的刚性约束，如要求算法透明可解释，又通过“三步走”战略引导技术突破赋能产业升级；2023 年《生成式人工智能服务管理暂行办法》进一步细化该逻辑：一方面要求企业建立内容安全过滤机制防范伦理风险，另一方面提出鼓励生成式人工智能技术在各行业的创新应用，彰显“守底线”与“拉高线”的并行策略。[2] 这一人工智能治理理念将安全与发展从“二元对立”转化为“共生共促”，其本质是以系统思维重构技术治理范式，既避免“因噎废食”抑制创新活力，又防止“野蛮生长”引发系统性风险。

3. 治理机制：软硬法协同

中国的人工智能治理机制注重软硬法协同，通过法律规制、行业标准与伦理准则的有机结合，实现技术创新与风险防控的动态平衡。在硬法层面，中国构建了以《网络安全法》（2017 年）、《数据安全法》（2021 年）、《个人信息保护法》（2021 年）为基础，辅以《生成式人工智能服务管理暂行办法》

[1] 参见吴建南、马太平、周磊：《人工智能安全治理的规则体系：国际比较与协同进路》，载《中国行政管理》2024 年第 12 期，第 6–12 页；张欣：《人工智能治理的全球变革与中国路径》，载《华东政法大学学报》2025 年第 1 期，第 18–32 页等。

[2] 《生成式人工智能服务管理暂行办法》，中国政府网，2023 年 7 月 10 日，https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm。

（2023 年）等专项法规的强制约束体系，明确数据采集、算法设计、隐私保护等环节的法律责任，为技术应用划定安全边界。例如，《生成式人工智能服务管理暂行办法》要求服务提供者对训练数据进行合法性审核，防范数据侵权与隐私泄露风险，体现了硬法在规范技术研发与市场行为中的基础性作用。同时，中国积极探索“敏捷治理”模式，通过动态修订法律与政策，应对技术快速迭代带来的监管滞后性，如设立算法备案与安全评估制度，强化事前监管能力。

在软法层面，中国依托行业标准、伦理指南及企业自律等柔性手段，形成对硬法的有效补充。2024 年工业和信息化部发布的《国家人工智能产业综合标准化体系建设指南》等技术标准，虽不具备法律强制力，却通过市场准入与产品认证机制引导行业合规发展，例如在数据隐私保护中嵌入伦理审查要求，推动“以人为本”的技术价值观落地。[1] 此外，通过《新一代人工智能伦理规范》（2021 年）等文件，中国强调技术研发需遵循公平性、透明性与可控性原则，鼓励企业建立内部伦理审查机制，如对生成式人工智能的“幻觉”问题实施动态监测与纠偏。[2] 中国人工智能治理机制通过硬法划定底线、软法引导方向、国际合作扩展影响，构建了兼顾安全与创新的可持续发展路径。

当前，全国人大常委会、国务院正在积极推进人工智能健康发展立法工作 [3]，人工智能治理正进入从专项规制向综合立法转型的新阶段。这意味着中国人工智能治理将从以部门规章和行业标准为主的阶段，迈向以国家顶层立法为核心的系统治理阶段。

---

[1]《国家人工智能产业综合标准化体系建设指南》，中国政府网，2024 年 6 月 5 日，https://www.gov.cn/zhengce/zhengceku/202407/content_6960720.htm。

[2]《新一代人工智能伦理规范》，科技部网站，2021 年 9 月，https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html。

[3] 参见《国务院办公厅关于印发＜国务院 2025 年度立法工作计划＞的通知》，中国政府网，2025 年 5 月 4 日，https://www.gov.cn/zhengce/content/202505/content_7023697.htm。

三、全球治理倡议与中国方案的贡献

从全球治理视角看，美国以竞争驱动的技术霸权、欧盟以硬法规制的规则壁垒、中国及全球南方以普惠共享的发展诉求，分别构成了人工智能治理的三种典型路径，并在互动中深刻塑造着全球数字秩序的未来走向。

（一）西方国家治理模式的国际扩张

在全球人工智能治理体系加速形成的进程中，各主要国家与区域组织正以自身法律传统、价值理念与产业利益为基础，构建具有鲜明特色的治理模式。这些模式不仅服务于本国的风险防控、产业引导与社会管理，更通过技术扩散、跨国投资、标准制定与外交互动实现制度外溢，逐渐演变为全球治理结构的重要组成部分。国家或区域中心模式的形成，实质上体现了人工智能治理的“内生—外溢”双重逻辑：一方面，国家治理体系的内部构建旨在确立可控的风险防线与竞争优势；另一方面，这些体系在国际层面的传播与复制，成为国家战略利益、制度理念乃至价值观输出的重要工具。由此，人工智能治理不再仅是技术伦理或监管问题，而成为衡量国家制度竞争力与国际规则塑造力的新维度。

美国与欧盟分别代表了两种典型路径：前者以市场竞争与科技霸权为导向，通过封闭化联盟与企业主导的标准输出确立技术统治；后者以规范塑造与伦理治理为核心，通过高门槛制度体系与价值叙事构建规则性权威。美国的核心在于以竞争驱动为逻辑，通过封闭化小圈子和技术垄断构筑霸权地位。在国际扩张路径上，美国依托其领先企业和创新生态，将价值观与行业标准捆绑式输出，并在 OECD、G7 等多边机制中推动“技术中立—企业自律—最小必要监管”的治理模式，放任资本主导市场与创新来巩固自身优势。[1] 此外，美国人工智能龙头企业普遍限制算法源码、模型权重与

[1] Hiroki Habuka and David U. Socol de la Osa, “Shaping Global AI Governance: Enhancements and Next Steps for the G7 Hiroshima AI Process,” CSIS, May 24, 2024, https://www.csis.org/analysis/shaping-global-ai-governance-enhancements-and-next-steps-g7-hiroshima-ai-process.

训练数据的开放获取，形成“知识黑箱”，以技术闭源确保技术和市场双重垄断。在外交战略上，美国也是不断搞“小圈子”，如美英 AI 协议、美日科技合作框架，打造所谓“可信技术联盟”“民主技术集团”。这些机制将人工智能治理嵌入地缘政治竞争，以技术标准划界制造新的壁垒。本质上看，美国将人工智能从全球公共资源异化为少数国家及小团体的战略私产，使之成为收割利益、维护霸权的工具。

欧盟的特点在于以刚性规范为核心，通过高门槛的制度与合规要求来塑造治理秩序与实现对外约束。在价值叙事上，欧盟高举“可信 AI”的旗帜，强调人权、透明、问责与可持续，力图在全球治理中塑造规范型权威，打造全球规则与制度标准的引领者角色。在国际扩张路径上，欧盟通过人工智能法案、数据保护条例、合规协议、“全球门户”项目以及“全球人工智能伙伴关系”（GPAI）等平台[1]，把自身规则嵌入国际合作及规范议程，要求伙伴国家遵循其制度框架。表面上看，这些欧盟主导下的规范似可为推动建立全球范围内具有广泛认同度的人工智能伦理与法治共识提供一定的方案借鉴。然而，当发生标准认定分歧、法律冲突、管辖权争议、执法上的标准不一甚至歧视性限制时，欧盟的规则不仅是“外溢”，更是壁垒。其将合规成本转嫁给发展中国家，使许多新兴经济体自动成为“规则接受者”，在制度谈判、技术应用以及标准和定价问题上丧失自主空间。本质上，欧盟通过制度工程制造“规则壁垒”，就是在价值包装的外衣下维护自身在全球数字秩序中的制度性优势。

（二）中国治理模式的国际贡献

人工智能等新兴领域为中国在全球治理体系改革中发挥更大作用提供了历史性机遇。作为负责任的大国，中国不仅重视自身在前沿科技中的安全与发展，更展现出参与全球人工智能治理的强烈意愿与积极作为。中国倡导以人类命运共同体理念为核心，以多边机制为路径，以全球南方国家

[1] 全球人工智能伙伴关系是 2020 年 6 月在法国等国倡议下成立的多边合作机制，秘书处设在经济合作与发展组织，被欧盟视为传播“可信 AI”理念和推动规则外溢的重要平台。

包容性发展为重点，以“全球公共产品”为基础，推动构建开放、公平、普惠的人工智能治理新秩序。

在治理理念上，2023年10月习近平主席提出《全球人工智能治理倡议》，呼吁各国共同建立公平公正、开放包容的国际治理体系，确保人工智能将“以人为本、智能向善”作为价值取向，实现共商共建共享，服务于人类共同福祉。[1] 2025年9月，习近平主席又提出全球治理倡议，指出当前数字鸿沟等问题日益突出，人工智能等新疆域治理缺位，强调要“坚持主权平等、坚持国际法治、坚持多边主义、坚持以人为本、坚持力求实效”的核心原则，为推进更加公正合理的全球治理体系提供“中国方案”。[2]

首先，在价值导向方面，中国强调人工智能发展必须符合和平、发展、公平、正义、民主、自由等全人类共同价值，为破解人工智能治理赤字提供了价值基石。其次，在主权平等方面，中国强调人工智能治理必须以各国一律平等为前提，任何国家无论大小、强弱、贫富，都应享有平等发展与治理人工智能的权利和机会。第三，在国际法治方面，中国强调人工智能治理必须以《联合国宪章》宗旨和原则为根本遵循，坚持在国际法框架内推动规则制定与执行，防止技术竞争滑向无规则状态。第四，在以人为本方面，中国强调人工智能治理必须牢牢把握“增进人类共同福祉”的根本价值取向，把人民的获得感、安全感和幸福感作为衡量治理成效的核心标准。[3] 最后，在多边主义与力求实效方面，中国通过积极构建和引领新型多边合作平台，推动人工智能与其他全球治理议题的统筹衔接。

在治理实践上，中国在人工智能领域的“开源”实践，就是以人类命

---

[1] 参见《全球人工智能治理倡议》，外交部网站，2023年10月20日，https://www.mfa.gov.cn/ziliao_674904/1179_674909/202310/t20231020_11164831.shtml。

[2] 参见《凝聚上合力量 完善全球治理——习近平在“上海合作组织+”会议上的讲话》，外交部网站，2025年9月1日，https://www.mfa.gov.cn/zyxw/202509/t20250901_11699539.shtml；《全球治理倡议概念文件》，外交部网站，2025年9月1日，https://www.mfa.gov.cn/wjbxw_new/202509/t20250901_11699909.shtml 等。

[3] 徐坚：《深刻把握全球治理倡议的重大意义与时代价值》，载《人民日报》2025年9月15日，第9版。

运共同体为核心理念、以技术普惠为治理路径，推动构建开放、公平与包容性全球治理秩序的典型例证。在政府层面，中国正在形成以政策引导、法律规范与生态培育为一体的“开源”治理体系。2025 年 8 月国务院印发《关于深入实施“人工智能 +”行动的意见》,明确提出要“促进开源生态繁荣，支持人工智能开源社区建设，促进模型、工具、数据集等汇聚开放，培育优质开源项目”,“加快构建面向全球开放的开源技术体系和社区生态”等。在企业层面，以 DeepSeek 为代表的中国科技企业主动开源大模型技术栈，通过开放核心代码、训练框架与应用接口 [1]，不仅有效降低了全球研究者与开发者的创新门槛，激发了跨国、跨领域的合作活力，也为算法透明性验证和责任问责提供了现实基础。此外，阿里云的 Qwen 系列多次登顶近年全球开源排行榜，积极推动构建开源生态；中国的软件开发者 2024 年约占全球开源社区 GitHub 总开发者的 9%。[2] 这一系列实践标志着中国正在以开放协作的技术模式和普惠共享的治理理念，探索以技术公共性支撑全球数字公平的新路径。

然而，与中方做法形成鲜明对比的是，西方国家在人工智能领域的开源实践正经历由开放走向封闭的转变。以大语言模型为例，美国 OpenAI 公司的 ChatGPT 从 GPT-2 开始以防止生成虚假信息和有害内容为由逐步封闭，GPT-3 和 GPT-4 更彻底限制文档与权重公开；Meta 公司的 LLaMA 采用“有限开源”混合模式,即虽开放代码,但附加严格法律限制,以避免被广泛利用。法国初创公司 Mistra AI 虽以“开源实验室”起家，但后来转为拒绝开放其旗舰模型 Mistral Large。美国政府与科技巨头通过自愿承诺机制及相关话语

[1] 参见“DeepSeek-LLM: DeepSeek Language Models,” GitHub, 2023, https://github.com/deepseek-ai/DeepSeek-LLM;“DeepGEMM: A High-Performance FP8 GEMM Library for DeepSeek-V3/R1,” GitHub, 2025, https://github.com/deepseek-ai/DeepGEMM 等。

[2] Wendy Chang, Rebecca Arcesati, Antonia Hmaidi,“China’s drive toward self-reliance in artificial intelligence: from chips to large language models,” Mercator Institute for China Studies, July 22, 2025, https://merics.org/en/report/chinas-drive-toward-self-reliance-artificial-intelligence-chips-large-language-models?utm_source=chatgpt.com.

叙事，进一步将“封闭等于负责任”固化为政策逻辑。[1] 本质上讲，这种由开放转向封闭的趋势，并非单纯出于风险防范的考量，而是以风险叙事为外衣、以市场垄断为核心，将人工智能公共资源私有化、资本化，强化了全球技术体系中的结构性不平等。

在国际合作上，中国坚持以联合国为中心，积极践行多边主义、构建和引领新型多边合作组织，推动人工智能与其他全球治理议题的统筹衔接。2025 年 9 月 24 日，李强总理在纽约会见第 80 届联合国大会主席贝尔伯克时表示，中国提出全球治理倡议，旨在支持联合国在国际事务中发挥应有作用，推动各国依托联合国等多边机制，更加紧密地团结协作，更加有力地应对时代挑战，推动加强人工智能、网络空间、极地、外空等新兴领域的全球治理。[2] 近年来，中国通过联合国、金砖国家、上海合作组织等多边框架，积极推动各国尤其是全球南方国家在人工智能、数字基础设施、技术标准规范及能力建设上的合作。

首先，在项目建设层面，中国—老挝人工智能创新合作中心、中国—东盟人工智能实验室帮助老挝、马来西亚等国打造符合本土国情大语言模型；中巴相关联合实验室帮助巴西研究人工智能在农业机械化领域的应用。[3] 其次，在能力建设层面，2024 年 9 月，由中国外交部和联合国共同主办的首届人工智能能力建设研讨班，吸引了近 40 个国家的政府官员和专家等参加，帮助全球南方国家深入了解人工智能技术发展现状及趋势，不断弥合智能

[1] Alek Tarkowski, “Artificial Intelligence and the Challenge for Global Governance,” Chatham House, June 7, 2024, www.chathamhouse.org/2024/06/artificial-intelligence-and-challenge-global-governance.

[2]《李强会见第 80 届联合国大会主席贝尔伯克》，外交部网站，2025 年 9 月 25 日，https://www.mfa.gov.cn/web/zyxw/202509/t20250925_11716688.shtml?utm_source=chatgpt.com。

[3] 参见《中巴科技部领导共同见证家庭农业机械化人工智能化联合实验室谅解备忘录签署》，中国农业大学工学院网站，2025 年 6 月 29 日，https://coe.cau.edu.cn/art/2025/6/29/art_27675_1074129.html?utm_source=chatgpt.com；Mengying Tao, “South-South cooperation: China is reshaping global AI governance, ” Sinolytics, August 20, 2025, https://sinolytics.de/global-business-news/blog/china-policy/china-is-reshaping-global-ai-governance/?utm_source=chatgpt.com。

鸿沟。在2025年底前，中国还将举办10期面向全球南方国家的人工智能研讨会和研讨班，以共享人工智能知识和经验。[1]最后，在机制建设层面，《全球人工智能治理倡议》强调积极支持在联合国框架下讨论成立国际人工智能治理机构，协调国际人工智能发展、安全与治理重大问题。2025年7月，在上海举行的2025年世界人工智能大会发布了《全球人工智能治理行动计划》，倡议成立全球人工智能合作组织，以协调全球多方力量，规范快速发展的人工智能技术以及分享中国的技术成果。[2]在9月23日的联合国全球发展倡议高级别会议上，中方进一步提出“人工智能+”国际合作倡议，宣布未来5年将在全球发展资金库项下设立“数字能力建设专项资金”，支持全球发展倡议“数字南方”品牌。[3]此外，中国还积极参与区域及多边安排，包括加入《数字经济伙伴关系协定》（DEPA）等多边协定，推动与新西兰、新加坡等国的人工智能规则对接，加强在数字经济领域的国际合作与可持续发展。[4]

综上所述，中国以主权平等、国际法治、多边主义、以人为本、力求实效为促进全球治理的核心理念，提出并不断完善人工智能治理方案，既回应了新兴技术发展带来的共同挑战，也为弥合全球治理赤字注入了制度动力。从价值导向到制度供给，从理念倡导到技术实践，中国不仅在人工智能治理中提供了系统化、可操作性的公共产品，更以四大全球倡议及多边机制为依托，推动国际社会形成广泛共识和合作框架。中国方案的意义

[1]《首届人工智能能力建设研讨班在上海举办》，人民网，2024年9月7日，https://world.people.com.cn/n1/2024/0907/c1002-40315041.html。

[2]《李强出席2025世界人工智能大会暨人工智能全球治理高级别会议开幕式并致辞》，中国政府网，2025年7月26日，https://www.gov.cn/yaowen/liebiao/202507/content_7033942.htm?utm_source=chatgpt.com。

[3]《李强在全球发展倡议高级别会议上的讲话》，中国政府网，2025年9月24日，https://www.gov.cn/yaowen/liebiao/202509/content_7042042.htm。

[4]《数字经济伙伴关系协定》是全球首个专门针对数字经济的贸易协定，由新西兰、新加坡和智利于2020年6月签署，内容涵盖跨境数据流动、个人信息保护、人工智能治理、源代码与算法保护等新兴议题。

不仅在于提出了一条兼顾安全与创新、普惠与效率的治理路径，更在于以开放、包容、合作的姿态，展示了负责任大国的制度担当和引领能力。未来，随着人工智能治理的全球议程不断深化，中国将在国际规则制定和多边治理改革中发挥更加积极的作用，推动人工智能真正成为造福人类社会、促进世界和平与发展的全球公共产品。

结 语

人工智能技术的迅猛发展正在重塑全球政治经济秩序与伦理边界，其影响已超越单一国家或技术共同体的范畴，正在成为全人类共同面对的挑战。在这一历史性转折中，国际合作是应对技术革命系统性风险的必然路径。作为全球人工智能发展的关键力量，中国以构建人类命运共同体理念为价值引领，提出四个全球倡议以及《全球人工智能治理倡议》，倡导“共商共建共享”的治理原则，为破解技术垄断、规则碎片化、安全风险与伦理失序提供了中国智慧和中国方案。

未来，全球人工智能治理亟待超越地缘政治竞争的桎梏，以联合国框架下以人为本、技术向善及可持续发展原则为中心，强化多边平台协调，探索具有约束力的国际条约。同时，积极构建南北技术转移、人才联培、算力共享的协作网络，将技术鸿沟转化为共同繁荣的增量空间。人工智能的终极使命应是成为人类文明进步的“最大公约数”，唯有以人类命运共同体理念为纽带，推动技术权力再平衡、发展红利再分配、文明价值再融合，才能驾驭这场前所未有的技术革命，为人类社会的公平与永续发展开辟新纪元。

【收稿日期：2025-06-05】
【修回日期：2025-10-13】
（责任编辑：王霄巍）