

全球人工智能治理：联合国的角色^{*}

□ 阎德学

〔提 要〕近年来，联合国秉持以人为本的核心理念和可持续发展的目标导向，通过设立专门机构和国际论坛，制订一系列治理规范，引领全球人工智能治理议程，成为推动全球人工智能治理的重要平台。然而，现有国际法滞后于人工智能发展进程的新溢出风险，以及国家间存在的人工智能鸿沟等因素，加大了联合国开展全球人工智能治理的难度。同时，受大国竞争影响，联合国开展的全球人工智能治理面临激烈的规则制定权之争。为此，联合国要进一步构建多元主体协同治理模式，提升不同国际行为体人工智能治理框架的互操作性，解决国家间人工智能治理领域的国际合作难题，增强发展中国家在全球人工智能治理中的代表性和发言权，弥合国家之间的人工智能鸿沟。

〔关 键 词〕人工智能治理、联合国角色、全球治理、多边机制

〔作者简介〕阎德学，华东师范大学俄罗斯研究中心、周边合作与发展协同创新中心研究员，博士生导师

〔中图分类号〕TP18, D815

〔文献标识码〕A

〔文章编号〕0452 8832 (2024) 4 期 0120-18

随着数智化时代的到来，人工智能正嵌入人类社会的方方面面。人工智能

^{*} 华东师范大学政治与国际关系学院硕士研究生郜晗轩对本文亦有贡献。感谢评审专家的宝贵意见，文中疏漏由作者负责。

的“涌现式”发展为人类社会带来前所未有的机遇和动力，同时也伴随着隐私泄露、算法歧视、劳动力失业等溢出风险，给国家和非国家行为体带来巨大的安全问题，成为全球治理新议题。2024年3月，联合国大会一致通过首个关于人工智能的全球决议，120多个会员国就安全、可靠和可信的人工智能系统达成全球共识。^[1] 联合国在全球人工智能治理的中心地位逐渐凸显，成为解决国际法缺位、全球人工智能鸿沟、各国人工智能治理理念差异等问题的重要平台。评估联合国在全球人工智能治理中的角色、挑战与发展趋势，对中国参与全球人工智能治理、发挥联合国主渠道作用、维护以联合国为核心的国际体系和以国际法为基础的国际秩序具有重要现实意义。

一、联合国在全球人工智能治理中的责任和优势

全球人工智能治理亟需以联合国为核心的多边机制引领。作为最具普遍性、代表性和权威性的政府间国际组织，联合国有责任和义务在人工智能治理等全球性议题上发挥关键作用，展现其突出优势。

（一）全球人工智能治理需要多边机制引领

近年来，ChatGPT-4o 和 Sora 等新一代人工智能大模型的出现，客观上加快了人工智能技术从“弱人工智能”到“强人工智能”的演变，但也进一步加剧了人工智能引发的失业、失序、失控等风险。同时，人工智能与教育、生物、网络、军事等议题的联系也愈发紧密。人工智能技术的多重议题属性使其安全问题远远超出传统技术治理涉及的风险与安全等单一议题范畴。^[2]

基于此，各国政府开始重视人工智能治理，尝试开展人工智能治理领域的国际合作，并认识到通过具有权威性和主导性的政府间国际组织发挥多边

[1] United Nations General Assembly, “Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development,” A/78/L.49 (2024), March 11, 2024, <https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=JcOqaZdYOdMBhY4wyF&fe=true>.

[2] 薛澜、赵静：《人工智能国际治理：基于技术特性与议题属性的分析》，《国际经济评论》2024年第3期，第1-18页。

机制引领作用的重要性。然而，全球人工智能技术发展的不均衡性以及发展中国家在全球治理体系中的话语权劣势，使缺乏人工智能治理能力的发展中国家面临更严重的负面影响。^[1] 全球人工智能治理亟需以联合国为核心的多边机制引领，因应人工智能引发的系统性、全球性风险，提升发展中国家在治理中的参与度和话语权，弥合国家间人工智能鸿沟。

在此背景下，具有权威性的政府间国际组织开始聚焦、参与并引领全球人工智能治理议程。特别是，涵盖大多数主权国家的联合国具备引领全球人工智能治理议程的能力，日益受到各国关注和重视。中国积极支持联合国在全球人工智能治理中发挥主渠道作用，在《全球人工智能治理倡议》中明确提出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。^[2]

（二）联合国在全球人工智能治理中的职责

人工智能的溢出风险挑战了以《联合国宪章》宗旨和原则为核心的国际秩序和国际体系。一方面，人工智能引发的伦理风险对世界人权保护造成负面影响。人工智能在某种程度上赋能人权，促进了世界人权理论与实践的演进，但也对隐私权、平等权、工作权等基本人权构成挑战，从人格、政治、经济、社会权利等维度挑战了人权的立足之本。^[3] 人工智能的失序特征还为人类社会注入更多不确定和不稳定因素。人工智能技术一旦被滥用，将加剧社会不公正，助大规模犯罪活动，推进自动化战争，带来大规模的操纵和监视风险。^[4] 另一方面，人工智能作为一种战略性技术，或加剧主权国家间的不平等。

[1] Robert Muggah and Ilona Szabó, “Artificial Intelligence Will Entrench Global Inequality,” *Foreign Policy*, May 5, 2023, https://foreignpolicy.com/2023/05/29/ai-regulation-global-south-artificial-intelligence/#cookie_message_anchor.

[2] 《全球人工智能治理倡议》，中国网信网，2023年10月18日，https://www.cac.gov.cn/2023-10/18/c_1699291032884978.htm。

[3] 程新宇、杨佳：《人工智能时代人权的伦理风险及其治理路径》，《湖北大学学报（哲学社会科学版）》2024年第3期，第159-166页。

[4] Yoshua Bengio et al., “Managing Extreme AI Risks amid Rapid Progress,” *Science*, Vol.384, No.6698, 2024, pp.842-845.

人工智能的“涌现式”发展让所有国家看到了其极具变革性的一面。可以说，现今各国对人工智能技术的重视程度不亚于对核技术、空间技术等传统战略性新兴产业的重视。当“赢者通吃”的人工智能技术竞争和商业竞争规律复制到国际关系上时，必然会冲击大小、强弱国家间的主权平等关系。^[1] 对于这类新挑战的应对及治理都在联合国的职责范围内。

《联合国宪章》的宗旨和原则规定了联合国应致力于维护各国人权免受侵害和国家间主权平等，促成国际合作。在全球人工智能治理背景下，缓解人工智能对国际秩序造成的挑战，协调区域性人工智能治理机制，也是联合国的职责之一。2023年12月，联合国人工智能高级别咨询机构发布《以人为本的人工智能治理》中期报告^[2]，阐明了联合国人工智能治理机构在全球人工智能治理过程中应该具备的原则和职责，包括定期评估、制定标准、促进国际多利益攸关方合作等方面。报告还强调全球人工智能治理应以《联合国宪章》、《国际人权法》以及可持续发展目标等国际承诺为基础，指出联合国可以在其中发挥独特的规范性和制度性作用，全球人工智能治理要与联合国的基本价值观相结合。

（三）联合国推动全球人工智能治理的突出优势

全球人工智能治理离不开政府间国际组织的参与。联合国是其中极为重要的治理主体，是全球范围内唯一有能力协调不同区域性人工智能治理愿景的政府间国际组织。具体而言，联合国在全球人工智能治理中具备的优势主要体现在以下三个方面。

第一，具有较强的国际信任度和权威。一方面，联合国是受会员国认可和信任的多边机制。皮尤研究中心（Pew Research Center）的一份调查报告显示，近年来，在参与调查的24个国家中，大多数民众对联合国持积极态度；2023年，在肯尼亚、波兰、韩国和瑞典，约80%的民众对联合国表达了积极

[1] 傅莹：《人工智能对国际关系的影响初析》，《国际政治科学》2019年第1期，第13页。

[2] AI Advisory Body, “Interim Report: Governing AI for Humanity,” December 22, 2023, <https://www.un.org/en/ai-advisory-body>.

看法。^[1] 另一方面，联合国作为集体安全机制的核心组织，在应对气候变化、网络空间安全、人工智能安全等国际安全问题方面始终发挥着不可替代的作用。譬如，在全球网络安全治理领域，联合国通过制定一系列机制和规范，实现了对全球网络安全治理的协调性引领，成为推动网络安全治理的主要平台。^[2] 现阶段，各方对人工智能治理国际规范虽有分歧，但基本上都承认联合国在全球人工智能治理中的权威地位。

第二，具有较强的规范性、制度性和包容性。作为一个具有普遍性的政府间国际组织，联合国容纳了几乎所有种族、宗教、价值观等方面的差异。《联合国宪章》赋予联合国一些超国家的职责和使命。在此基础上，联合国以《世界人权宣言》《人工智能伦理问题建议书》等国际硬法和软法相结合，不断在全球人工智能治理实践中推动国际法治建设和全球治理体系发展完善。

第三，有助于更多发展中国家参与到全球人工智能治理中。在联合国框架下，各会员国主权平等。无论大国还是小国，各会员国都可通过联合国这一多边机制参与全球治理。在全球人工智能治理领域，联合国决心弥合国家间的人工智能鸿沟，呼吁发展中国家参与数字化转型，以有效利用安全、可靠和可信的人工智能系统。^[3]

二、理念目标与实践模式

为应对人工智能全球性的风险挑战，联合国倡导以人为本的核心理念，以可持续发展为目标导向，设立若干工作组、专门机构和国际论坛，推动形成一系列全球人工智能治理规范，在引领议程设置、推动软法规范、促成国

[1] Pew Research Center, “People across 24 Countries Continue to View UN Favorably,” August 31, 2023, <https://www.pewresearch.org/short-reads/2023/08/31/people-across-24-countries-continue-to-view-un-favorably/>.

[2] 张蛟龙：《联合国与全球网络安全治理》，《国际问题研究》2023 年第 6 期，第 98-118 页。

[3] United Nations General Assembly, A/78/L.49 (2024), p.3.

际合作等实践方面发挥了显著作用。

（一）以人为本：联合国全球人工智能治理的核心理念

联合国倡导以人为本的人工智能治理理念，强调在人工智能系统的设计、开发、运行和淘汰等各个阶段，必须以人为本，符合道德标准，充分尊重和保障人权及国际法。^[1]为进一步倡导以人为本的治理理念，联合国人工智能高级别咨询机构在《以人为本的人工智能治理》中期报告中提出以下五项指导原则。

一是人工智能治理应包容各方，由所有人治理，并为所有人谋福利。尽管人工智能潜力巨大，但并不是每个人都能通过接触和使用人工智能来有效改善生活。因此，所有公民，包括“全球南方”的公民，都应能够利用人工智能创造机遇、实现繁荣；就人工智能开发、部署、使用阶段对特定群体的历史性、结构性排斥，有必要积极采取纠正措施。^[2]

二是人工智能治理必须符合公共利益。人工智能系统的开发主要集中于科技公司，但人工智能系统的完善、部署和使用应鼓励更加广泛的行为体参与。人工智能治理也必须以更广泛的公共利益为导向，充分考虑多样性、公平性、可持续性等与公共政策相关的问题。^[3]

三是人工智能治理应与数据治理、数据共享同步进行。数据治理应符合公共利益，且不能与人工智能治理的其他组成部分相互割裂；应鼓励发展数据共享，为更多利益攸关方提供有助于应对各种社会性挑战的公共数据。^[4]

四是人工智能治理必须基于多利益攸关方的适应性合作。人工智能治理应促进各利益攸关方的共识，建立一个具有协调性、互操作性的治理框架。这一框架需要借鉴世界各地的最佳实践和专业知识，并保证政府、私营部门、

[1] United Nations General Assembly, A/78/L.49 (2024), p.2.

[2] AI Advisory Body, “Interim Report: Governing AI for Humanity.”

[3] *Ibid.*, p.13.

[4] AI Advisory Body, “Press Release for AI Advisory Body Interim Report,” https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_interim_report_press_release.pdf.

学术界、民间社会等利益攸关方一道参与。^[1]《全球数字契约》就是联合国调动多利益攸关方审议包括人工智能在内的全球技术治理框架的范例。^[2]

五是人工智能治理应立足于《联合国宪章》、国际人权法和其他一致同意的国际承诺，如可持续发展目标。人工智能治理应与《联合国宪章》的宗旨和原则相结合，在充分考虑人工智能引发的各类溢出风险的前提下，为国际社会提供全球性指南。

（二）可持续发展：联合国全球人工智能治理的目标导向

技术变革与可持续发展息息相关。早在2015年9月，联合国大会在第70/1号决议——《变革我们的世界：2030年可持续发展议程》中就提出建设“一个技术研发和应用顾及对气候影响、维护生物多样性和有复原力的世界”，宣布消除贫困、性别平等、减少国家内部和国家间不平等、应对气候变化等17个可持续发展目标。^[3]人工智能等颠覆性技术变革与联合国可持续发展目标高度关联。一方面，人工智能技术变革日新月异，具有加快实现联合国可持续发展目标的潜力；另一方面，人工智能技术引发的溢出风险也会阻碍联合国实现可持续发展目标的整体进程。

近年来，联合国大会先后通过多份大会决议和政治宣言，确定了以可持续发展为目标导向，推动全球人工智能治理实践。2023年9月，联合国大会通过第78/L.1号决议，各国元首共同承诺弥合科学、技术及创新鸿沟，负责任地利用科学、技术及创新推动可持续发展，并寻求更好地实现人工智能的益处和应对各种挑战。^[4]2024年3月，联合国大会第78/L.49号决议也表

[1] AI Advisory Body, “Interim Report: Governing AI for Humanity,” p.16.

[2] *Ibid.*, p.17.

[3] United Nations General Assembly, “Transforming Our World: the 2030 Agenda for Sustainable Development,” A/RES/70/1, October 21, 2015, <https://documents.un.org/doc/undoc/gen/n15/291/89/pdf/n1529189.pdf?token=COecQyDVocSI5FU6le&fe=true>.

[4] United Nations General Assembly, “Political Declaration of the High-level Political Forum on Sustainable Development Convened under the Auspices of the General Assembly,” A/78/L.1 (2023), September 19, 2023, <https://documents.un.org/doc/undoc/ltd/n23/271/52/pdf/n2327152.pdf?token=66Z88zhmr6IrWDzORR&fe=true>.

示，将推动制定人工智能系统标准，帮助发展中国家有效参与人工智能系统整个生命周期，促进在数字技术和人工智能系统方面开展与可持续发展目标相关的研究和创新，有效利用人工智能系统促进全球数字化转型，在全面实现 2030 年可持续发展议程方面加快取得进展。^[1]

（三）议程引领：设立工作组、专门机构和国际论坛

为应对人工智能引发的全球性风险，联合国设立了一系列人工智能治理组织和机构，举办相关国际论坛，推动各方在人工智能治理领域开展对话、凝聚共识，发挥了对全球人工智能治理议程的协调性引领作用。

一是设立人工智能机构间工作组。2020 年 10 月，联合国方案问题高级别委员会（HLCP）建立了由联合国教科文组织和国际电信联盟共同领导的人工智能机构间工作组（IAWG-AI）。^[2]该工作组汇集联合国系统在人工智能方面的专业知识，为联合国系统当前和未来在人工智能方面的努力提供了坚实基础。在人工智能机构间工作组的推动下，联合国方案问题高级别委员会于 2024 年 4 月批准《联合国系统人工智能治理白皮书》^[3]，分析了联合国系统的机构模式、职能以及适用于全球人工智能治理的现有国际规范框架。

二是成立人工智能高级别咨询机构。该咨询机构的设立源于 2020 年 5 月联合国秘书长古特雷斯发布的《数字合作路线图》^[4]，目前主要聚焦三个领域：全球人工智能治理，对风险和挑战的共识，利用人工智能加速实现可持续发展目标的关键机遇和推动因素。该咨询机构的建议将被 2024 年 9 月召开的联合国未来峰会作为参考。

[1] United Nations General Assembly, A/78/L.49 (2024), pp.3-7.

[2] UN CEB, “Inter-Agency Working Group on Artificial Intelligence,” <https://unsceeb.org/inter-agency-working-group-artificial-intelligence>.

[3] High-Level Committee on Programmes, “United Nations System White Paper on AI Governance,” April 16, 2024, <https://unsceeb.org/sites/default/files/2024-04/United%20Nations%20System%20White%20Paper%20on%20AI%20Governance.pdf>.

[4] 联合国大会：《数字合作路线图》，A/74/821，2020 年 5 月 29 日，<https://documents.un.org/doc/undoc/gen/n20/102/50/pdf/n2010250.pdf?token=yVBZEbDnAiHdlSEF8I&fe=true>。

三是举办全球人工智能伦理论坛。该论坛由联合国教科文组织支持和赞助，迄今已举办两届。在 2024 年 2 月举办的第二届全球人工智能伦理论坛上，来自政府、国际组织、科研机构以及企业的超过 600 名代表围绕“改变人工智能治理格局”这一主题，分享各自在国家、区域和全球层面对人工智能治理的见解和良好实践。^[1]该论坛还推出了联合国教科文组织的各种倡议，包括“全球人工智能伦理和治理观察站”和“人工智能伦理专家无国界网络”等。

四是举办人工智能造福人类全球峰会。2017 年以来，国际电信联盟等 40 多个联合国机构合作举办人工智能造福人类全球峰会年度论坛，启动国际电信联盟人工智能资料库，以负责任和切合实际的人工智能应用方式推动实现联合国可持续发展目标。2024 年 5 月，人工智能造福人类全球峰会在日内瓦举办，吸引了全球 183 个国家和地区参与，各界人士聚焦人工智能的全球合作、监管与治理、标准建设以及人工智能在教育、健康、通信、金融、气候变化方面的应用等议题展开讨论。^[2]

（四）软法治理：推动形成一系列全球人工智能治理规范

国际软法是联合国推进全球人工智能治理的重要工具。当前，针对人工智能的国际造法进程缓慢，国际软法成为现阶段全球人工智能治理规范不足的有效补充。近年来，在联合国多边机制的引领下，全球人工智能治理进程的规范建设取得初步进展，各方在规范层面就人工智能涉及的若干领域达成重要共识。联合国系统充分发挥不同机构的专业性和机构间的协调性，在伦理、教育、公共卫生等领域发布一系列人工智能治理软法，凝聚各方共识，确保人工智能的安全创新，推动国际法的更新和发展，有助于加速实现联合国可持续发展议程。

在伦理治理领域，联合国系统通过的多个关于人工智能伦理的协议和决

[1] UNESCO, “Global Forum on the Ethics of AI 2024,” February 6, 2024, <https://www.unesco.org/en/forum-ethics-ai>.

[2] 《“人工智能造福人类全球峰会”在瑞士日内瓦举行》，新华网，2024 年 5 月 30 日，http://www.xinhuanet.com/photo/2023-07/08/c_1129738820.htm。

议，构成联合国人工智能伦理规范的重要组成部分。2021 年 11 月，联合国教科文组织 193 个会员国一致通过《人工智能伦理问题建议书》^[1]。该建议书以软法的形式寻求定义人工智能相关的共同价值观和原则，指导人工智能法律框架建设，确保人工智能技术的开发和部署以公平、透明、符合道德与人权的方式进行。2024 年 3 月，联合国通过的首个人工智能决议也强调，人工智能有导致结构性不平等和歧视的风险，应帮助保护个人免受一切形式的歧视、偏见、滥用或其他伤害。^[2]

在教育治理领域，为推动实现 2030 年可持续发展议程在教育领域的承诺，联合国教科文组织与中国政府于 2019 年 5 月在北京合作举办“人工智能与教育国际会议”。^[3]会上联合国教科文组织发布的《北京共识——人工智能与教育》强调各国要采取适当的应对措施，将人工智能与教育相融合，建议各国政府和其他利益攸关方在人工智能时代的教育公平、性别平等、教育数据和算法的使用等一系列领域采取行动，推动可持续发展目标和人类命运共同体的实现。在此框架下，联合国教科文组织于 2021 年 12 月出版《人工智能与教育：政策制定者指南》，指出政策制定者要关注人工智能教育应用的研究、监测与评估，加强教育领域的人工智能治理。^[4]

在公共卫生治理领域，世界卫生组织于 2024 年 1 月发布关于“多模态大模型伦理和治理”的新指南，^[5]概述了人工智能多模态大模型在诊断和临床护理、科研与药物开发等方面的广泛应用，详细阐述了人工智能时代卫生系统面临的风险和应对建议，呼吁各国政府制定公共卫生领域人工智能治理

[1] UNESCO, “Recommendation on the Ethics of Artificial Intelligence,” November 23, 2021, <https://unesdoc.unesco.org/ark:/48223/pf0000380455>.

[2] United Nations General Assembly, A/78/L.49 (2024), p.3.

[3] UNESCO, “Beijing Consensus on Artificial Intelligence and Education,” May 18, 2019, http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/201908/W020190828311234688933.pdf.

[4] 曾海军、张钰、苗苗：《确保人工智能服务共同利益，促进教育系统变革——〈人工智能与教育：政策制定者指南〉解读》，《中国电化教育》2022 年第 8 期，第 1-8 页。

[5] WHO, “Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-modal Models,” January 18, 2024, <https://www.who.int/publications/i/item/9789240084759>.

新标准。

三、面临的挑战

在联合国系统各机构的努力下，以联合国为中心的全球人工智能治理体系逐渐形成。然而，联合国引领全球人工智能治理面临着国际法缺位、全球人工智能鸿沟、大国竞争等困境。

（一）国际造法难以跟进人工智能新溢出风险

国际法是联合国引领全球人工智能治理的重要依据。然而，人工智能的“涌现式”发展为国际法和全球治理带来新的挑战，使联合国缺乏有效的治理工具规制人工智能的溢出风险。现阶段，联合国全球人工智能治理进程面临国际造法滞后性与人工智能颠覆性发展之间的矛盾。

一方面，国际法规则本身具有滞后性。从国际法内部来看，具有约束力的人工智能治理国际规则至今缺位，体现出国际法体系在人工智能规制上的滞后性。^[1]究其原因，国家间存在的人工智能鸿沟以及各方对全球人工智能治理统一规范的不同利益诉求，使人工智能治理领域的国际造法进程陷入重重困境。2023 年 7 月，联合国秘书长古特雷斯发布题为《新和平纲领》的政策简报，呼吁围绕人工智能军事应用的一系列问题制定新规范和原则，在 2026 年之前缔结一项具有“法律约束力”的文书，禁止使用无人监管且违反国际人道法的致命性自主武器系统。^[2]

另一方面，人工智能迭代更新不断给国际法提出新的挑战。人工智能技术的迅猛发展不仅挑战了国家主权原则、国家责任原则等全局性的国际法基本原则，还挑战了国际人权法、国际人道法、国际刑法、国际知识产权法等

[1] 沈伟、赵尔雅：《数字经济背景下的人工智能国际法规制》，《上海财经大学学报》2022 年第 5 期，第 123-137 页。

[2] António Guterres, “A New Agenda for Peace,” July 21, 2023, <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf>.

国际法具体领域。^[1]现阶段，随着人工智能向“强人工智能”阶段的过渡不断加速，人工智能在不断颠覆人类科技极限的同时，也在不断突破人类社会伦理道德规范的底线。被《时代》杂志称为“全球人工智能领袖”的马斯克曾多次在公共场合发声，称人工智能可能导致“文明毁灭”，若不对人工智能加以立法规范，未来它们极有可能具备发动信息战的能力。^[2]

（二）各国人工智能发展水平不均衡加大治理难度

在全球人工智能治理过程中，由于科技水平差距、利益认知差异等因素，国家间人工智能发展水平呈现较大的不均衡性。牛津洞察（Oxford Insights）发布的《2023 年全球政府人工智能就绪指数》评估报告指出，全球人工智能技术发展的地区差异显著，数字鸿沟仍是全球性的挑战；其中，在数字基础设施和数据可用性两个指标上，高收入经济体得分约为低收入经济体得分的 3 倍。^[3]在此形势下，联合国若以人工智能强国的技术发展水平为标准治理人工智能，可能导致人工智能弱国的技术提升遭遇瓶颈，进而加剧全球人工智能发展的不均衡；反之，若以人工智能弱国的技术发展水平为标准治理人工智能，则可能造成相关国际规则约束力不足等后果。可见，国家间人工智能鸿沟使联合国难以把握对人工智能的监管力度，也是导致人工智能治理领域国际法缺位的现实原因。^[4]

此外，国家间人工智能发展水平的不均衡会进一步强化“数字割据”，加大全球人工智能治理的难度。从全球经济体系来看，发展中国家由于资源禀赋所限居于边缘或半边缘地位，依附于处在中心地位的发达国家。随着新一代人工智能大模型的问世，通用人工智能技术的发展或将进一步加剧全球

[1] 赵骏、李婉贞：《人工智能对国际法的挑战及其应对》，《浙江大学学报（人文社会科学版）》2020 年第 2 期，第 10-25 页。

[2] Clare Duffy and Ramishah Maruf, “Elon Musk Warns AI Could Cause ‘Civilization Destruction’ Even as He Invests in It,” April 17, 2023, <https://www.cnn.com/2023/04/17/tech/elon-musk-ai-warning-tucker-carlson/index.html>.

[3] Oxford Insights, “Government AI Readiness Index 2023,” December 6, 2023, <https://oxfordinsights.com/wp-content/uploads/2023/12/2023-Government-AI-Readiness-Index-2.pdf>.

[4] 沈伟、赵尔雅：《数字经济背景下的人工智能国际法规制》，第 123-137 页。

人工智能鸿沟，减少发达国家对发展中国家资源禀赋的需求，甚至导致发展中国家脱离体系。^[1] 联合国大会第 78/L. 49 号决议表示，不当或恶意设计、开发、部署和使用人工智能，有扩大国家间和国家内部数字鸿沟等方面的风险，对可持续发展进程构成障碍，联合国决心消除各种数字鸿沟。^[2]

（三）围绕人工智能治理的大国竞争造成协调困境

联合国在全球人工智能治理进程中发挥着解决分歧、凝聚共识的重要作用，但是各方对人工智能治理具体问题的认知仍存在较大分歧，一些发达国家和国家集团甚至将争夺人工智能治理的规则制定权视为其国际竞争战略的一部分。譬如，欧盟近年来进行了一系列人工智能治理实践，一定程度上形成了人工智能治理的“欧洲方案”。2024 年 5 月，《人工智能法案》的正式批准实际上是“布鲁塞尔效应”的再现：欧盟希望这一“欧洲方案”能像之前的数据保护立法一样，助其抢占人工智能治理规则制定的主导权，弥补欧盟在人工智能技术领域的落后态势。^[3] 同样地，为护持美国在全球人工智能领域的“领导力”，拜登政府先后出台《人工智能权利法案蓝图》《人工智能风险管理框架》《关于安全、可靠、可信地开发和和使用人工智能的行政令》等一系列政策文件，不断推进人工智能治理进程，致力于提升美国人工智能治理国内规范的“域外影响力”。2024 年 3 月，白宫管理和预算办公室（OMB）发布人工智能政策备忘录。美国副总统哈里斯表示：“拜登总统和我将推动这些国内政策成为全球（人工智能治理）行动的典范。”^[4]

美国和一些西方国家还注重利用“小多边主义”策略，在人工智能治理的国际规则制定方面“拉帮结派”，共同竖立“小院高墙”。2024 年 6 月，

[1] 鲁传颖：《警惕人工智能发展不平衡》，《人民日报》2018 年 4 月 19 日，第 23 版。

[2] United Nations General Assembly, A/78/L.49 (2024), p.3.

[3] 刘洋、李益斌：《愿景政治视角下人工智能规范的“欧洲方案”研究》，《当代世界与社会主义》2023 年第 4 期，第 153-161 页。

[4] Kathryn Watson, “White House Orders Federal Agencies to Name Chief AI Officers,” March 28, 2024, <https://www.cbsnews.com/news/white-house-chief-ai-officers-federal-agencies-artificial-intelligence/>.

七国集团领导人在年度峰会上发布联合公报，表示将根据七国集团共同的“民主价值观”来管理人工智能风险。^[1]虽然联合公报提出了人工智能治理愿景和框架，强调人工智能治理国际合作的重要性，但整体上缺乏明确的落实机制和实施细则，更缺少对发展中国家人工智能技术发展和治理态势的关注。

可以看出，美西方国家企图抢占全球人工智能治理规则制定的主导权。在大国竞争背景下，联合国作为全球人工智能治理的主渠道，受到美西方“理念一致国家同盟”的挑战，^[2]将使其全球人工智能治理进程面临严峻的协调困境。

四、发展趋势

联合国大会通过首个人工智能全球决议，昭示着联合国开展全球人工智能治理的未来方向。面对全球人工智能治理赤字，联合国将加速凝聚各方共识，推动形成具有整体性、互操作性的国际规范，促成国际合作，努力弥合全球人工智能鸿沟，强化联合国在全球人工智能治理中的主渠道作用。

（一）倡导构建多国家、多部门、多机构的多元主体治理模式

在人工智能全球风险挑战下，任何一个国际行为体都无法独善其身。全球人工智能治理需要国际组织、国家、私营部门等多元主体的共同参与。联合国高度重视在全球人工智能治理进程中引入多元主体，在第 78/L. 49 号决议中鼓励会员国邀请来自所有区域和国家的多利益攸关方（包括私营部门、国际组织、民间社会组织、媒体、学术界和研究机构、技术界及个人）在各自角色和职责范围内，制定和支持与安全、可靠和可信的人工智能系统有关的监管和治理办法和框架，在各级创建一个有效的生态系统。^[3]

[1] “G7 Apulia Leaders’ Communiqué,” The White House, June 14, 2024, <https://www.whitehouse.gov/briefing-room/statements-releases/2024/06/14/g7-leaders-statement-8/>.

[2] 鲁传颖：《全球人工智能治理的目标、挑战与中国方案》，《当代世界》2024 年第 5 期，第 25-31 页。

[3] United Nations General Assembly, A/78/L.49 (2024), p.4.

联合国系统下各机构也在通过密切协商和适当的合作机制推动构建人工智能多元主体治理模式。纵观联合国系统的全球人工智能治理实践，联合国通过由教科文组织和国际电信联盟共同领导的人工智能机构间工作组、由多利益攸关方组成的人工智能高级别咨询机构、由国际电信联盟等 40 多个联合国机构合作举办的人工智能造福人类全球峰会等一系列实践，推进联合国系统下的人工智能治理机构间合作机制建设。事实上，2018 年以来，国际电信联盟等 40 多个联合国机构每年都推出合作编撰的《联合国人工智能活动》报告，汇集联合国系统内正在开展的所有与人工智能有关的项目、倡议、活动和进程。其中，2023 年度报告涉及 47 个联合国机构、408 个项目，明确将人工智能领域的多利益攸关方合作视为联合国系统的优先事项。^[1]

（二）推进制定具有国际互操作性的全球人工智能治理框架

全球人工智能治理需要具有整体性、包容性的国际规则，而不是分散的制度设计。^[2]对此，联合国人工智能高级别咨询机构在《以人为本的人工智能治理》中期报告中指出，全球各国际行为体制定了人工智能治理的诸多指南、框架和原则，相互之间虽有共性，但仍缺乏互操作性；咨询机构今后将通过建立全球认可的人工智能治理框架，提高各方人工智能治理框架的互操作性。^[3]

在联合国看来，协调统一的人工智能治理国际法是全球人工智能治理框架的基础。尽管联合国开展的全球人工智能治理进程面临国际法滞后性与人工智能颠覆性发展之间的矛盾，但两者间的矛盾并不意味着国际法的停滞不前。国际法依然是联合国参与全球人工智能治理的核心治理工具。正如联合国大会第 78/L. 49 号决议所强调的，人工智能治理是一个不断发展变化的领域，需要人们在技术发展过程中不断调整人工智能治理的办法，这些办法必

[1] ITU, "United Nations Activities on Artificial Intelligence 2023," May 28, 2024, https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2023-PDF-E.pdf.

[2] Peter Cihon, Matthijs Maas and Luke Kemp, "Fragmentation and the Future: Investigating Architectures for International AI Governance," *Global Policy*, Vol.11, No.5, 2020, pp.545-556.

[3] AI Advisory Body, "Interim Report: Governing AI for Humanity," p.20.

须以国际法为基础，具有互操作性、灵活性、适应性、包容性，能够满足发达国家和发展中国家的不同需求，并使全人类共同受益。^[1]

在治理实践层面，联合国还积极参与全球人工智能治理的其他多边机制，推进制定具有国际互操作性的全球人工智能治理框架。比如，2024年5月，第二届全球人工智能安全峰会由韩国与英国政府合办，邀请了包括中国、美国、联合国在内的国家、国际组织以及科技公司等共同参与。^[2]期间，联合国提名的国际专家委员会参与制定了《先进人工智能安全国际科学报告》，旨在促进人工智能安全方面的国际科学共识，努力提高人工智能技术方法上的互补性和互操作性。

（三）推动解决会员国在全球人工智能治理领域的国际合作难题

人工智能治理的国际合作已是大势所趋。面对全球人工智能治理赤字，尽管各方在全球人工智能治理的国际规则制定方面仍有一定的分歧，但主要大国已逐渐展现出在人工智能治理领域开展国际合作的意愿。例如，2024年5月，中美人工智能政府间对话首次会议在瑞士日内瓦举行，双方介绍了各自对人工智能技术风险的看法以及为推动人工智能赋能经济社会发展采取的措施，均认识到人工智能技术发展既面临机遇也存在风险，并重申继续致力于落实两国元首在旧金山达成的重要共识。^[3]

联合国作为多边机制的核心，积极促进会员国在人工智能治理领域的国际合作，努力解决各方的国际合作难题。联合国大会第78/L.49号决议肯定了联合国系统通过促进包容各方的国际合作等方式，就安全、可靠和可信的人工智能系统达成全球共识所作出的独特贡献。^[4]具体而言，一是在规则制定层面，联合国鼓励开展研究和国际合作，敦促会员国分享数据治理的最佳

[1] United Nations General Assembly, A/78/L.49 (2024), p.3.

[2] 정책홍보담당관실, “AI 서울 정상회의의 서울선언 및 의향서,” 2024년 5월, https://www.mofa.go.kr/www/brd/m_26779/view.do?seq=537.

[3] 《中美举行人工智能政府间对话首次会议》，央视网，2024年5月15日，<https://news.cctv.com/2024/05/15/ARTIIVG55tknLsj93CHilQdX240515.shtml>.

[4] United Nations General Assembly, A/78/L.49 (2024), p.7.

实践，推动数据治理方面的国际合作与援助，推动制定和实施具有国际互操作性的框架、做法和标准；对于人工智能引发的劳动力失业风险，联合国主张各国合作研究和评估人工智能系统的部署对劳动力市场的影响，以减轻相关负面影响。^[1]二是在治理主体层面，联合国促请会员国公共与私营部门、民间社会、学术界和研究机构以及技术界之间，加强人工智能领域合作，提供公平、开放、包容和没有歧视的营商环境，促进经济和商业活动，等等。^[2]

（四）增强发展中国家在全球人工智能治理中的代表性和发言权

当前，“全球南方”国家在人工智能技术发展和技术治理领域均存在显著不足。人工智能的溢出风险加剧了国际社会的“结构性不平等”，^[3]致使“全球南方”国家在人工智能领域一直处于接受者的地位。但这并不意味着广大发展中国家应该在全球人工智能治理进程中被边缘化，更不意味着全球人工智能鸿沟是正义的、必然的趋势。因此，在推进全球人工智能治理进程中，必须坚持“一个也不能少”的原则，践行真正的多边主义。长期来看，全球人工智能鸿沟既是联合国开展全球人工智能治理面临的重要挑战，也是其治理进程中的重要治理对象。

联合国致力于为广大发展中国家提供一个具有公平性、包容性的全球人工智能治理平台，努力增强发展中国家在全球人工智能治理中的代表性和发言权。联合国秘书长 2021 年向联合国大会提交的报告《我们的共同议程》提出全面改革与转型倡议，表示联合国作为世界上最权威、最重要的多边治理机构，将优化自身能力，以实际行动构建更加网络化的多边主义，支持发展中国家等参与治理、影响决策。^[4]联合国大会第 78/L.49 号决议也指出，认识到国家间的人工智能技术发展水平参差不齐，发展中国家在人工智能迅猛

[1] United Nations General Assembly, A/78/L.49 (2024), pp.5-6.

[2] *Ibid.*, p.7.

[3] Padmashree Gehl Sampath, “Governing Artificial Intelligence in an Age of Inequality,” *Global Policy*, Vol.12, No.6, 2021, pp.21-31.

[4] 姚琨、赵敬雅、韩一元：《〈我们的共同议程〉：联合国全面改革和转型倡议》，《国际研究参考》2021 年第 11 期，第 1-6 页。

发展过程中面临独特挑战，联合国迫切需要在能力建设、技术和财政方面加强对发展中国家的援助；决议促请会员国和其他利益攸关方采取行动，增强发展中国家特别是最不发达国家的能力，营造基于创新的国际环境，以提高发展中国家发展技术专长和实力的能力，帮助发展中国家应对主要的结构性障碍，消除其在获得人工智能创新惠益方面存在的障碍。^[1]

五、结语

回顾近年来的全球人工智能治理进程，联合国作为具有普遍性、代表性和权威性的政府间国际组织，以治理者和造法者的双重角色为全球人工智能治理作出了突出贡献，发挥了至关重要的引领作用。人工智能的多重议题属性引发了系统性、全球性的溢出风险，任何一方都无法以独善其身的思维有效应对，开展人工智能治理的国际合作尤显重要，符合全人类的共同利益。如今，全球人工智能治理已进入新阶段，联合国推进全球人工智能治理面临大国竞争和各方认知分歧等挑战，但人工智能治理的国际合作已是大势所趋，联合国亦应为缓解全球人工智能治理主体间矛盾提供一定的解决方案。中国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，推动构建更加公正合理的全球人工智能治理体系。

【责任编辑：肖莹莹】

[1] United Nations General Assembly, A/78/L.49 (2024), p.4.