全球治理

全球人工智能监管:规则建构与多重挑战*

程海烨

[内容摘要] 数据、算法与算力构成人工智能发展三大核心支柱,亦是引发人工智能潜在风险的关键因素。鉴于人工智能风险的全球性特征,国际组织、双多边机制以及非国家行为体等正积极建构人工智能监管规则并开展协调行动。当下,主要经济体制定的人工智能监管规则可划分为强化技术监管主导权、主张企业自我规制、采用风险分级和闭环监管、分领域管控与行业自律相结合等四种特征。然而,全球人工智能监管正面临多重挑战。全球监管结构的去中心化与碎片化以及协调机制本身存在的执行困难,导致国际层面的人工智能治理失效。经济体之间的监管理念与模式存在巨大差异,中美欧围绕不同目标争夺全球人工智能规则主导权的激烈斗争将延缓合作进度。此外,人工智能技术还将进一步加剧全球"数字鸿沟"。

[关键词] 人工智能监管 跨国监管协调 数字规则博弈 全球数据治理

[作者简介] 程海烨,上海社会科学院国际问题研究所助理研究员

[中图分类号]TP18 [文献标识码] A [文章编号]2095-5715(2025)03-0110-22

人工智能成为第四次工业革命下引领全球科技和产业变革的核心驱动力,正为全球产业发展带来巨大经济收益。② 据测算,2024 年全球人工智能产业收入规模达6421.8 亿美元,同比增长22.2%。预计2025 年人工智能市场规模将达到

* 本文系 2024 年国家社会科学基金青年项目"美欧跨境数据安全规制协调及我国对策研究"(项目编号:24CGJ039)的阶段性成果。

② 薛澜、王净宇:《人工智能发展的前沿趋势、治理挑战与应对策略》,《行政管理改革》2024 年第 8 期,第 4~13 页。

2437 亿美元,2030 年市场规模将达到 8267 亿美元。[①] 然而,尖端人工智能科技亦伴随着包括数字技术与政治危机、数据隐私保护等一系列风险与挑战,引起国际社会高度关注。对于全球人工智能监管研究,学界主要从以下三个视角展开:一是关注人工智能对国家安全、数据和隐私安全、军事政治、科技伦理等方面带来的风险特征与应对;二是聚焦美国、欧盟和中国等三大主要经济体的人工智能战略竞争、军事运用和监管规则等领域的合作与分歧;三是从宏观角度阐述人工智能对今后全球秩序、世界政治经济发展格局的重要影响,[②]但对全球人工智能监管规则和面临主要治理困境的系统梳理与深入分析尚不足。有鉴于此,本文将重点围绕全球人工智能监管规则建构的现状与主要困境展开讨论。首先分析人工智能技术的三大支柱及其风险,随后从国际层面与主要经济体两个视角探讨应对上述风险的规则建构、协调进展与监管特征,进而指出全球人工智能监管面临的多重挑战。

一、人工智能技术的三大支柱及其风险

数据、算法与算力是支撑人工智能发展与技术升级的创新三角,也是触发人工智能风险的核心因素。

---

① IDC, "Asia/Pacific AI Maturity Study 2024," May 2024, https://www.intel.com/content/dam/www/central-libraries/us/en/documents/2024-05/idc-infobrief-asia-pacific-ai-maturity-study-2024-australia.pdf; "Artificial Intelligence - Worldwide," https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide.

② 薛澜、劳拉·赞诺蒂、胡郁:《人工智能的全球治理:挑战与进路(笔谈)》,《探索与争鸣》2025 年第 1 期,第 108~121 页;贾开、俞晗之、薛澜:《人工智能全球治理新阶段的特征、赤字与改革方向》,《国际论坛》2024 年第 3 期,第 62~78 页;封帅:《从民族国家到全球秩序:人工智能时代的世界政治图景》,《外交评论》2020 年第 6 期,第 99~129 页;吴桐、刘宏松:《地缘经济转向、数字主权与欧盟人工智能治理》,《国际安全研究》2024 年第 5 期,第 81~108 页;Matthew U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies," *Harvard Journal of Law & Technology*, Vol. 29, No. 2, 2016, pp. 354~400; Johann Laux, Sandra Wachter and Brent Mittelstadt, "Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk," *Regulation & Governance*, Vol. 18, No. 1, 2024, pp. 3~32; Nathalie A. Smuha, "From a 'Race to AI' to a 'Race to AI Regulation': Regulatory Competition for Artificial Intelligence," *Law, Innovation & Technology*, Vol. 13, No. 1, 2021, pp. 1~26; Peter Cihon, Matthijs M. Maas and Luke Kemp, "Fragmentation and the Future: Investigating Architectures for International AI Governance," *Global Policy*, Vol. 11, No. 5, 2020, pp. 545~556。

(一)人工智能的数据风险

可持续高质量的数据供给是人工智能发展的不竭动力。[①] 庞大的数据是支撑算法和算力的重要基础,如何高效安全地获取数据是人工智能进步的关键。

数据获取过程中遇到的首要挑战是人工智能的数据标注,即对数据进行标记,以便人工智能模型能够解读数据。为减轻数据标注工作带来的高昂成本,企业通常会采用自动化技术和数据开源与共享方式,减少数据采集和处理过程中的重复劳动。然而,自动化技术中的无监督学习、半监督学习、主动学习和自我监督学习等,难以达到数据标注的精确度,从而产生数据偏见、失真等问题,进而影响数据标注的质量。同时,数据开源与共享也隐含数据滥用、知识产权侵犯、安全漏洞以及追踪与控制困难、规则不统一导致的操作难题等,这些都可能对人工智能模型的精确度产生负面影响。[②]

第二个核心挑战是数据安全。在人工智能开发的数据采集、数据处理、数据流动、数据调取以及数据存储等五个阶段,均存在显著的数据安全风险。[③] 在数据采集阶段,灰色交易和违规爬取现象频繁,用户知情权与授权机制缺失、无差别或过度的数据采集行为等,也可能侵犯个人隐私、甚至威胁国家安全与公共利益。在数据处理阶段,存在因数据质量低劣或缺乏标准化处理导致的数据污染、对人工智能大模型数据的恶意攻击和随意篡改等数据投毒行为、因训练数据样本或算法偏差导致的不公平与歧视结果等风险。在数据流动阶段,不同参与主体之间设置的标准或技术双重壁垒导致数据孤岛现象,限制了人工智能的发展,甚至催生了“人工智能黑色产业”。此外,数据跨境流动所涉及的隐私保护和国家安全等问题,亦是人工智能数据安全问题的一部分。在数据调取阶段,人工智能对离散数据的关联度分析和深度挖掘、逆向还原攻击、对抗样本攻击等,可能侵犯个人隐私、窃取商业和国家机密。在数据存储阶段,数据交互现象将揭露产业链安全薄弱环节,导致用户生物识别等个人数据信息的泄露,还涉及操作不当

① 张晓洁、严赋憬:《以高质量数据促进人工智能发展,国家数据局将开展四方面工作》,新华网,2025 年 3 月 25 日,https://www.xinhuanet.com/tech/20250325/5ddb17547ed94dc6a6b0a647bf3d1d4e/c.html。

② 杨学军、吴朝晖等著:《人工智能——重塑秩序的力量》,科学出版社 2023 年版,第 100~107 页。

③ 梁正主编:《前沿人工智能 发展与治理》,中国发展出版社 2024 年版。

或模型缺陷导致的商业机密或国家机密的泄露或被盗取等潜在风险。[①]

(二)人工智能的算法挑战

算法是人工智能发展的重要引擎。算法带来的风险源于其自身特质以及全球科技革命对算法风险性的扩散。

一方面,人工智能依赖以数据分析和认知技术软件为基础的算法,而算法通常存在人为偏见或歧视、技术缺陷、使用缺陷和安全缺陷等,这将重点导致三类算法风险:第一类是用于人工智能训练或输入的数据存在偏差,主要原因包括数据不完整、过时或不匹配;样本量不够大、缺少多样化;数据收集技术不成熟;以及用于训练算法的数据与操作过程中实际输入数据之间的不匹配等。第二类是算法设计中存在漏洞,比如逻辑缺陷、建模技术不适当、编码错误以及识别训练数据中存在的虚假模式等。第三类是对产出结果解释误差。[②]

另一方面,随着机器学习的不断发展,算法变得更加普遍与强大,缺乏透明度与公平性。其一,算法在后台的不透明运行规则,是造成人工智能"黑匣子"问题的主要原因。[③] 算法的可解释性构成了人类对算法进行控制的关键途径,也是评估人工智能模型决策行为、验证决策结果可靠性、安全性和问责制的重要依据。然而,当前以深度学习模型为基础的生成式人工智能大多属于经典的"黑箱算法"或"闭源模式",缺乏完整的全局解释性。这在高风险领域应用可能引发严重的安全问题,在中低风险场景中则可能导致验证困难和诊断缺陷等治理风险。其二,算法难以实现主观或法律意义上的公平性。算法存在分配和表征两种危害,人工智能的多模态模型还将增加算法歧视的风险。收集和标记训练数据、定义数据类标签以及自主决策程序等,都将影响算法决策程序,产生不公平的结果输出。其三,人工智能借助模型迭代实现多场景应用,却增加了以算法操纵为代

① 刘辉、雷崎山:《生成式人工智能的数据风险及其法律规制》,《重庆邮电大学学报》2024 年第 4 期,第 40 ~ 51 页。

② Digital Platform Regulators Forum, "Literature Summary: Harms and Risks of Algorithms," June 2023, https://dp-reg.gov.au/publications/working-paper-1-literature-summary-harms-and-risks-algorithms.

③ Dilip Krishna, Nancy Albinson and Yang Chu, "Managing Algorithmic Risks, Safeguarding the Use of Complex Algorithms and Machine Learning," 2017, https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-risk-algorithmic-machine-learning-risk-management.pdf.

表的算法妨碍弥散效应,通过虚假信息和心理诱导对个人或大规模人群进行操纵,实现算法推荐技术,加速群体圈层分化,从而产生算法偏见与不公平性。[①]

(三)人工智能的算力威胁

算力作为人工智能的核心动力与引擎,潜在风险主要源自运行数字基础设施的安全隐患以及算力监管或分配权的过度集中。

一是支撑计算的数字基础设施可能存在安全漏洞,导致隐私泄露、商业机密和国家机密外泄。人工智能计算依赖于数据中心、互联网和云计算等基础设施,网站和应用程序遭受攻击、分布式拒绝服务攻击、域名系统攻击以及凭证泄露等针对托管服务器的典型网络攻击形式,是威胁算力服务器安全性的主要因素。[②]此外,在云计算领域还存在数据完整性、数据信任等问题,行为主体在未经授权的前提下删除、更改或操作以及阻止访问有价值的信息或服务,亦或缺乏清晰度和损害用户敏感数据而产生的不信任,将造成数据损害、数据泄露、数据拦截和应用程序的不安全与不稳定。[③]

二是科技公司对算力进行大量集中部署并施加中央监管或分配权,加剧了控制权力的集中化态势。“没有大型科技公司就没有人工智能。”[④]数字科技公司正尝试争夺对算力基建的监管与控制权,借用“技术官僚权力”促成自身利益合法化,甚至左右国家行为体的相关决策。[⑤] 相较于抽象的算法与数据,大型科技公司直接掌握关键技术的供应渠道。人工智能管理依赖的算力产品与服务也大

---

① Digital Regulation Cooperation Forum, “The Benefits and Harm of Algorithms: A Shared Perspective from the Four Digital Regulators,” September 23, 2022, https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/the-benefits-and-harms-of-algorithms-a-shared-perspective-from-the-four-digital-regulators#current-and-potential-harms-of-algorithmic-processing;张欣:《生成式人工智能的算法治理挑战与治理型监管》,《现代法学》2023 年第 3 期,第 109 ~ 112 页;戴长征、刘浥晨:《数字技术爆发性增长对国家认知域安全的影响及其应对》,《政治学研究》2024 年第 5 期,第 44 ~ 53 页。

② Check Point, “Data Center Threats and Vulnerabilities,” https://www.checkpoint.com/cyber-hub/cyber-security/what-is-data-center/data-center-threats-and-vulnerabilities/.

③ Pinki Rani, Sukhdip Singh and Karanbir Singh, “Cloud Computing Security: A Taxonomy, Threat Detection and Mitigation Techniques,” *International Journal of Computers and Applications*, Vol. 46, No. 5, 2024, pp. 348 ~ 361.

④ Girish Sastry et al., “Computing Power and the Governance of Artificial Intelligence,” 2024, https://arxiv.org/abs/2402.08797.

⑤ Shaleen Khanal, Hongzhou Zhang and Araz Taeihagh, “Why and How is the Power of Big Tech Increasing in the Policy Process? The Case of Generative AI,” *Policy and Society*, Vol. 44, No. 1, 2025, pp. 52 ~ 69.

多由科技公司提供,如应用商店、支付系统、身份验证器、电子邮件和社交网络等。即便政府作出相关决策,执行亦需受到大型科技公司的技术权力制约。[①]

二、全球人工智能监管:规则协调与主要特征

为应对上述风险,国际社会正积极制定监管规则,对前沿人工智能技术发展设置“红绿灯”。当前,人工智能监管呈现碎片化与多中心特征,全球人工智能技术竞赛已转变为人工智能监管规则的较量。

(一)国际层面的人工智能监管:规则建构与协调情况

第一类是联合国、经济合作与发展组织、金砖国家等国际组织协调机制下的规则建构,核心工作是为人工智能三大支柱的技术风险制定原则性规则指南。

联合国为全球人工智能监管提供核心平台,促进人工智能向更加公正、安全和可持续的方向发展。一是发布针对人工智能发展方向的行动和倡议。二是提出系统性的人工智能治理方案与决议。三是发布人工智能治理白皮书,协调联合国框架下各部门,与其他国际组织加强监管方式沟通与对接。四是成立人工智能高级别咨询机构,提出具有一定约束力和实际影响力的人工智能规则。联合国正加强在国际层面制定人工智能监管最低标准中发挥的实际作用,尤其是注重平衡各国间制定统一的监管标准和基准,降低人工智能监管协调的复杂性。[②]

经合组织重点关注制定人工智能监管原则、降低潜在风险,并提升差异化监管模式的跨国互操作性。经合组织理事会部长会议上通过首个政府间人工智能标准《人工智能建议书》,重点包括两个实质性部分:一是建立可信赖的人工智能负责任5项管理原则,二是为可信赖的人工智能制定国家政策和国际合作。[③] 经

---

① “Tech Giants’ Race to Control Energy for AI Expansion,” October 24, 2024, https://linkdood.com/tech-giants-race-to-control-energy-for-ai-expansion/? utm_source = rss&utm_medium = rss&utm_campaign = tech-giants-race-to-control-energy-for-ai-expansion.

② 钱亿亲:《联合国框架下的全球人工智能治理现状、挑战与展望》,《中国信息安全》2024年第4期,第62~66页。

③ 目前全球已有47个成员国承诺遵守该原则。参见OECD, “Recommendation of the Council on Artificial Intelligence,” March 5, 2024, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449。

合组织推动七国集团广岛峰会提出“先进人工智能系统开发组织国际行为准则”,提升成员国在人工智能规则制定中的影响力。[①]

金砖国家更重视“全球南方”在人工智能监管与规则制定中的参与度,要求在联合国框架内践行多边主义,建立全球人工智能监管框架,反对由少数发达经济体搞“小圈子”“小集团”,确保人工智能技术应用的公平与透明。在 2023 年约翰内斯堡峰会期间,成员国成立了“人工智能工作组”,重点聚焦人工智能领域的关键技术、算法、标准和应用场景。[②] 2024 年,喀山峰会上通过《金砖国家领导人第十六次会晤喀山宣言》,支持联合国在全球人工智能治理中发挥的重要作用,将缩小数字鸿沟、增进人类共同福祉作为推动全球人工智能监管的核心目标。[③]

第二类是全球人工智能峰会、七国集团和二十国集团等为代表的国际多边协调机制,聚焦因人工智能技术本身带来安全风险的规制行动。

从英国布莱切利、韩国首尔到法国巴黎连续举办的三届全球人工智能峰会,使世界各国实现了从规制人工智能潜在风险与确保“安全优先”的发展原则,到以实际行动推动“以人为本、开放、可信的人工智能生态系统”的重要转变。其中,首届国际人工智能峰会上,中美两国首次在前沿人工智能模型的风险应对上达成合作共识。[④] 在巴黎人工智能行动峰会上,包括中国、印度、法国、德国等在内的共 61 个国家和经济体达成《关于发展包容、可持续的人工智能造福人类与地球的声明》,但美国和英国并未签署。[⑤]

七国集团致力于提升成员国间相互信任和先进人工智能系统的标准互操作

---

① “AI Watch: Global Regulatory Tracker - OECD,” May 13, 2024, https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-oecd.

② 《金砖国家领导人第十五次会晤约翰内斯堡宣言》,2023 年 8 月 25 日,https://www.mfa.gov.cn/web/gjhdq_676201/gjhdqzz_681964/jzgj_682158/xgxw_682164/202308/t20230825_11132502.shtml。

③ 《金砖国家领导人第十六次会晤喀山宣言(全文)》,新华社,2024 年 10 月 24 日,https://www.news.cn/world/20241024/1d6664db75b142daaa5a2293a6ef0854/c.html。

④ 2024 年 10 月 23 日新西兰加入了该宣言,目前包括中国、美国等在内 29 个国和欧盟签署《布莱切利人工智能安全宣言》。参见“The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023,” February 13, 2025, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

⑤ “Artificial Intelligence Action Summit,” February 10-11, 2025, https://www.elysee.fr/en/sommet-pour-l-action-sur-l-ia.

性,并加强与经合组织等开展监管合作。例如,加拿大与法国在2018年的七国集团峰会上提出“全球人工智能伙伴关系”倡议,内容包括推动负责任的人工智能行动、数据治理等,汇集各行业专家共同应对人工智能带来的风险挑战,并交由经合组织负责召开定期会议。[①]《2024年七国集团科技与数字部长会议联合宣言》提出进一步加强关于开发安全与可信的公共部门人工智能系统、组织开发先进人工智能系统的国际行为准则、数字政府服务和数字身份方法等讨论。[②]

二十国集团在推动人工智能治理中更重视成员国的差异性,其协调工作目前仍处于发出倡议与制定基本原则的阶段。一是聚焦数字隐私、透明度、问责制和包容性增长等人工智能全球性挑战的核心议题。2019年二十国集团大阪峰会发布《大阪数字经济宣言》,追求以人为本的人工智能,承诺加强隐私和个人数据保护,促进人工智能能力建设和技能发展等。[③] 基于该项承诺,促进人工智能包容性可持续发展,加强人工智能监管透明度、稳定性与安全性等议题成为2024年二十国集团数字经济部长会议重点。[④] 二是倡导通过人工智能全球合作弥合“智能鸿沟”。在2024年二十国集团领导人里约峰会第二阶段会议上,中国强调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”。[⑤] 三是推动人工智能治理的法制化和标准化进行。二十国集团支持联合国《全球数字契约》中关于人工智能治理的目标,倡导建立国际人工智能科学小组和全球人工智能治理对话机制,促进国际合作。

第三类是美英、美欧和中美等主要经济体围绕人工智能风险规制开展的双边协调机制,关注人工智能在军事、科技等领域的应用风险。

---

① “The Global Partnership on Artificial Intelligence,” https://gpai.ai/projects/responsible-ai/.

② “Ministerial Meeting on Technology and Digital G7 Joint Statement,” October 15, 2024, https://www.g7italy.it/wp-content/uploads/1728987577-final-g7-digital-joint-ministerial-statement-15_10_24.pdf.

③ “Osaka Declaration on Digital Economy,” https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/special_event/en/special_event_01.pdf.

④ Anne Turner, “G20 Declarations for Digital Privacy and AI Regulation,” September 19, 2023, https://www.groundlabs.com/blog/g20-declarations-for-digital-privacy-and-ai-regulation/; “G20 Ministerial Declaration: September 13, 2024,” September 14, 2024, https://www.gov.uk/government/publications/g20-ministerial-declaration-maceio-13-september-2024/g20-ministerial-declaration-13-september-2024.

⑤ 《习近平在二十国集团领导人第十九次峰会第二阶段会议关于“全球治理机构改革”议题的讲话(全文)》,2024年11月18日,https://www.gov.cn/yaowen/liebiao/202411/content_6988048.htm。

2024 年 4 月 1 日,英美签署《美英人工智能安全科学领域谅解备忘录》,并以此为基础建立了全球首个人工智能安全科学领域双边合作框架,通过美英各自的人工智能安全研究所进一步推进双方在人工智能安全研究与评估等方面合作,共享人工智能安全治理能力。[①] 生成式人工智能带来的生物恐怖主义、核武器战争等高风险安全议题是美英合作的重点。[②] 目前,美国空军研究实验室和英国国防科学技术实验室已联合开发、选择、训练和部署最先进的机器学习算法能力,用于支持两国武装部队在未来战争中的合作。[③]

以美欧贸易与技术委员会为代表的跨大西洋人工智能科技联盟基于“民主价值观”和人权,正推动建立统一人工智能风险管理方法标准、制定人工智能联合研究计划等七大关键领域合作,并在能源优化、应急响应、城市重建等社会领域的各个风险节点合作中不断深化拓展。[④]

以中美人工智能政府间对话为代表的发达经济体与发展中经济体协调机制,重点围绕人工智能风险管控、治理模式与目标、技术出口管制等三个方面进行深入、专业和建设性讨论。美国更加关注中国对人工智能大模型的存在性风险监管与预防风险生成与扩散;中国坚持《全球人工智能治理倡议》的核心主张,强调要发挥联合国主渠道作用形成具有广泛共识的全球人工智能治理框架和标准规范,确保人工智能技术有益、安全与公平。[⑤] 尽管存在分歧,但双方仍积极寻求政府间合作与“二轨对话”的可能性。

第四类是以“人工智能伙伴关系”和人工智能无线接入网络联盟为代表的非

---

① “Collaboration on the Safety of AI: UK-US Memorandum of Understanding,” April 2, 2024, https://www.gov.uk/government/publications/collaboration-on-the-safety-of-ai-uk-us-memorandum-of-understanding/collaboration-on-the-safety-of-ai-uk-us-memorandum-of-understanding.

② U.S. Department of Commerce, “U.S. and UK Announce Partnership on Science of AI Safety,” April 1, 2024, https://www.commerce.gov/news/press-releases/2024/04/us-and-uk-announce-partnership-science-ai-safety.

③ Bryan S. Ripple, “U.S. and UK Rresearch Labs Collaborate on Autonomy and Artificial Intelligence (AI),” March 8, 2024, https://www.dvidshub.net/news/465738/us-and-uk-research-labs-collaborate-autonomy-and-artificial-intelligence-ai.

④ “AI for Public Good: EU-U.S. Research Alliance in AI for the Public Good,” https://digital-strategy.ec.europa.eu/en/library/ai-public-good-eu-us-research-alliance-ai-public-good.

⑤ Michael Martina and Trevor Hunnicutt, “US, China Meet in Geneva to Discuss AI Risks,” Reuters, May 14, 2024, https://www.reuters.com/technology/us-china-meet-geneva-discuss-ai-risks-2024-05-13/.

国家行为体制定人工智能技术规则,重点是解决人工智能技术操作风险问题。

“人工智能伙伴关系”由大型数字科技公司、高校与智库机构等组成,正致力于推动技术在国际社会中的公平与包容性,帮助各成员与经合组织、七国集团等人工智能监管标准形成准确对接,并重点从企业层面推动医疗、金融、交通等不同行业的人工智能监管行动,加强企业监管政策透明度,降低潜在或预测安全风险。该组织针对人工智能新形式的虚假信息、操纵和骚扰、以及有害数字内容在线传播等问题,提供了较多的指导性方案。“人工智能伙伴关系”还单独成立了“人工智能与共享繁荣倡议”指导委员会,专门应对人工智能技术可能造成的失业问题以及社会不稳定、经济动荡等难题,并建立人工智能安全风险事件的数据库,便于预测与防范人工智能风险滥用。[①] 人工智能无线接入网络联盟则由电信运营商、设备供应商、业界领先的人工智能芯片厂商和研发机构组成,更聚焦将人工智能融入蜂窝通信技术,进一步推进无线接入网络技术和移动网络的发展。目前,人工智能无线接入网络联盟的主要工作是利用人工智能和数字孪生的全部力量推动移动网络和电信行业转型。其中,重点利用人工智能提升无线接入网络能力,从而提高频谱效率,同时整合人工智能与无线接入网络流程,更高效地使用基础设施创造人工智能驱动的收入机会。此外,还将通过无线接入网络在网络边缘部署服务,提高运营效率并为移动用户提供新的服务。[②]

(二)主要经济体规制人工智能风险的四种特征

世界各国已深刻意识到人工智能技术实施监管的必要性。截至 2024 年,全球至少有 69 个国家和地区制定了多部人工智能相关规则,涵盖隐私保护、数据监管、伦理规范和人工智能治理等广泛议题。[③] 全球主要经济体制定的人工智能

① “人工智能伙伴关系”致力于负责任地使用人工智能,创始成员大多来自如苹果、微软、谷歌和脸书等美国大型数字科技公司,以及华盛顿大学法学院、布鲁金斯学会、清华—卡耐基全球政策研究中心、香港数字非洲中心、香港科技大学人工智能研究中心等高校与智库机构。目前,在全球已拥有 126 个合作伙伴。参见“PAI Brings Together a Diverse Community to Address Important Questions About Our Future with AI,” https://partnershiponai.org/。

② 人工智能无线接入网络联盟的创始成员包括亚马逊云科技、Arm、DeepSig、爱立信、微软、诺基亚、美国东北大学、英伟达、三星电子、软银公司和 T-Mobile 等。参见“Redefine What Networks Can Transform,” https://airan.org/。

③ “AI Laws Around the World,” https://www.aiprm.com/ai-laws-around-the-world/.

监管规则呈现以下四种特征：

一是强化对人工智能技术本身存在的数据安全、算法偏见、隐私侵犯等直接风险进行立法监管的主导权。以中国、印度等为代表。中国强调要“建立人工智能安全监管制度”“完善生成式人工智能发展和管理机制”，统筹发展与安全，积极应对人工智能安全风险。[①] 中国在人工智能监管方面的立法正不断完善，尤其注重数据安全、算法偏见等引发的网络安全等关键基础科技领域的法律规制。在数据安全领域，中国已构建以《网络安全法》《数据安全法》和《个人信息保护法》为核心的三部法律体系，旨在全面保护国家网络与数据安全、个人隐私以及数据跨境流动的合法性。中国还制定了《生成式人工智能服务管理暂行办法》等部门规章以及《国家人工智能产业综合标准化体系建设指南(2024)》等行业技术标准，并形成监管备案、伦理审查和安全框架三个维度相互依赖又紧密关联的制度保障体系。再如，印度首部综合性《2023 年数字个人数据保护法案》和《2023 年数字印度法案》凸显了印度对人工智能技术带来安全风险的高度重视，表明其削弱科技公司对印度数字市场的控制力，提升国家对人工智能等前沿技术监管权力的坚定决心。[②]

二是主张企业对人工智能技术的安全风险进行自我规制，并赋予监管机构一定的规则引导权。以沙特、新加坡等为代表。沙特政府强调创建一个由政府与私营部门实体构成的世界级人工智能生态系统，但并不是由政府层面开展监管，而是委托沙特数据与人工智能管理局引导人工智能企业合规发展和经营。该管理局颁布了《个人数据保护法实施条例》与《个人数据跨境传输条例》，指导人工智能相关企业在数据跨境传输、数据安全、隐私保护中完成自我合规，促进政府与企业之间能够安全、高效地共享数据，重点是为企业研发人工智能提供便

① 《中国共产党第二十届中央委员会第三次全体会议公报》，环球网，2024 年 7 月 18 日，https://china.huanqiu.com/article/4If8Pqm1MXX。

② “The Digital Personal Data Protection Act, 2023,” August 11, 2023, https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf; “Digital India Act 2023: Revolutionizing Internet Regulation in India,” October 7, 2024, https://www.nextias.com/blog/digital-india-act/.

捷通道。[①] 又如,新加坡对人工智能的法律监管重点在于个人数据保护、在线安全、算法和数据泄露等,但新加坡采取了一种结合严格数据风险保护与企业自我规制的双重模式,其颁布的《用于生成式人工智能模型的治理框架》仅为企业提供了一套道德和透明度的开发准则,并不具有强制执行力。[②]

三是聚焦人工智能技术应用对人权、社会伦理与民主等带来的威胁,采用风险分级和闭环监管的方式划定前沿科技创新的"红线"。欧盟是典型代表。《欧盟人工智能法》已明确欧洲监管重点是针对人工智能系统与应用的安全风险,采纳基于风险分级的差异化监管策略。该法律设有代价高昂的罚款与惩戒措施,并赋予监管机构执法权力。[③] 目前,欧盟还面向国际社会发布了《人工智能与人权、民主和法治框架公约》,第 70 ~ 79 条规定每个成员国设置至少一个通知机构和市场监督机构作为国家级别的主管部门,确保各成员国能够从国家层面应对人工智能带来的潜在风险。这是全球第一部对人工智能技术使用要符合人权、民主和法治进行规制且具备国际法律约束力的条约。[④]

四是将分领域管控与制定行业自我监管的软法相结合,为提升人工智能技术创新扫除障碍。美国是典型代表。一方面,美国对涉及人工智能军事武器化、威胁国家安全、生物安全等相关领域进行严格管制。《促进国家安全中人工智能治理与风险管理的框架》等指导性文件,已对人工智能在维护国家安全的前提下发挥作用提出严格要求。[⑤] 美国现行的对人工智能风险规范的立法政策集中在州政府层面,大约 42 个州及华盛顿特区都提出人工智能风险监管相关法案,主

---

① "Reimagining Saudi Arabia's Economy," https://www.accenture.com/us-en/case-studies/artificial-intelligence/reimagining-saudi-arabia-economy; "The Implementing Regulation of the Personal Data Protection Law," https://sdaia.gov.sa/en/SDAIA/about/Documents/ImplementingRegulation.pdf.

② "Model Artificial Intelligence Governance Framework," https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf; "AI for the Public Good for Singapore and the World," https://file.go.gov.sg/nais2023.pdf.

③ Nathalie A. Smuha, "The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence," *Computer Law Review International*, Vol. 20, Vol. 4, 2019, pp. 97 ~ 106.

④ Council of Europe, "Council of Europe Opens First Ever Global Treaty on AI for Signature," September 5, 2024, https://www.coe.int/en/web/portal/-/council-of-europe-opens-first-ever-global-treaty-on-ai-for-signature.

⑤ The White House, "Framework to Advance AI Governance and Risk Management in National Security," October 24, 2024, https://ai.gov/wp-content/uploads/2024/10/NSM-Framework-to-Advance-AI-Governance-and-Risk-Management-in-National-Security.pdf.

要倾向于对人工智能应用施加限制，特别是涉及军事武器化、威胁国家安全、金融危机、生物安全等领域的严格管制。[①] 另一方面，自我监管是美国人工智能监管模式的核心原则。美国基于《反垄断法》鼓励人工智能行业实施自我监督与行业监督机制，认为从事人工智能行业或领域相关的企业、私营机构和其他利益群体应共同努力规范和实践人工智能监管规则，国家仅作为监管方或管理者对其相关政策举措进行观察或认证，并通过如“行为准则、技术标准”等软法进行协商，而非直接参与监管。这符合自由与民主的美国数字市场竞争原则，也为美国数字科技巨头赢得更灵活的市场环境。[②] 特朗普上任第一天重新签署的《消除美国在人工智能领域领导地位的障碍行政命令》，旨在通过大幅削减监管成本，为美国人工智能产业创造一个更加宽松的创新环境。[③]

然而，鉴于人工智能发展的复杂与多样性特质，上述分类难以完全论及主要国家的人工智能监管特征。部分国家将采取多种监管方式相结合，以便获得最优监管效果，并在国际社会中赢得监管规则制定的话语权。

三、全球人工智能监管面临的多重挑战

目前，人工智能技术的迭代速度已远超国际社会的监管与立法行动，全球人工智能监管规制仍然存在一定的局限性，主要面临以下挑战。

（一）国际层面的人工智能治理失效

一方面，在监管规则制定过程中，全球监管结构的去中心化与碎片化、横向或纵向立法规制的差异，提升了国际社会开展人工智能监管协调的难度。在多边机制内，联合国及其下属各机构已经围绕人工智能开展自上而下的监管行动，

① “Artificial Intelligence 2024 Legislation,” September 9, 2024, https://www.ncsl.org/technology-and-communication/artificial-intelligence-2024-legislation.

② Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology*, New York: Oxford University Press, 2023; Roger Clarke, “Regulatory Alternatives for AI,” *Computer Law & Security Review*, Vol. 35, No. 4, 2019, pp. 398 ~ 409.

③ The White House, “Removing Barriers to American Leadership in Artificial Intelligence,” January 23, 2025, https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

涉及人工智能监管议题的参与者还有联合国教科文组织、国际劳工组织、国际电信联盟等,经合组织、二十国集团、亚太经合组织、金砖国家等国际组织也正在积极推动开展人工智能监管,但均未形成权威性。此外,围绕人工智能开展的双多边行动还分散在区域多边机制中。比如,美国为主导的国家正与盟友建立人工智能技术联盟的“小圈子”,包括美欧贸易与技术委员会、“印太经济繁荣框架”、七国集团以及美印日澳四方安全对话等机制下的人工智能监管与技术合作等。《数字经济伙伴关系协定》等区域数字贸易协议也包含了具体的人工智能监管承诺。[①] 国际层面的人工智能监管协调行动呈现碎片化态势,缺乏中心权威机构引导规则制定方向,各成员国的监管规则也需要与不同的分散机制相匹配,才能拥有获得双边或多边合作的可行性。同时,不同法域间横向或纵向立法规制的差异,大幅增加了各国达成跨国监管共识性协议的难度。

另一方面,国际层面协调机制本身存在规则执行困难。一是国际组织自身的执行能力较弱,技术监管能力与治理需求不匹配,国际层面监管效能依然受制于自身的制度矛盾和成员国地缘政治博弈。“全球南方”在追求技术进步中缺乏与发达经济体讨价还价的能力。例如,尽管金砖国家等强调在人工智能技术中应兼顾发展中国家,帮助发展中国家摆脱数字鸿沟,但在具体执行过程中,国际组织需要应对美国霸权的压力,难以帮助发展中国家跨越技术鸿沟。二是多边机制制定的全球人工智能监管规则不具约束力、无强制性监督机制。例如,二十国集团的大多数倡议和行动方案停留在原则上,缺少有效监督机制,各成员国的执行情况较为空洞。七国集团则以“西方民主标准”的“小圈子”规则抵制其他国家,实际意图是制定有利于西方人工智能技术垄断、推行符合自身监管模式与利益诉求的监管规则,并迫使其他国家选边站队,因此相关规则无法获得发展中国家的认同。三是部分多边机制沦为美西方技术联盟与科技排华的“政治俱乐部”。比如,经合组织和七国集团制定的规则大多代表发达经济体利益,成为美西方国家数字规则“扩张”的工具。再如“人工智能伙伴关系”和人工智能无线接

① Joshua P. Meltzer, “Developing Global AI Governance for Foundational AI,” Hinrich Foundation, May 21, 2024, https://www.hinrichfoundation.com/research/article/tech/developing-global-ai-governance-for-foundational-ai/.

入网络联盟的成员大多来自美国、欧盟等西方发达经济体,尤其是人工智能无线接入网络联盟的成员排除华为等中资通信企业,实质上为美西方技术联盟体系服务,成为提升美国在全球通信技术发展中的主导权以及对华开展 6G 竞争的工具。[①]

(二)经济体之间的监管规则差异难以协调

国家层面的人工智能监管模式差异化为形成统一的监管规则增添了诸多困难,这体现在经济体关注不同的监管风险类型、选择不同的监管目标,却为全球监管协调带来交叉缺陷与深层矛盾。

首先,监管理念存在差异。是以人为本的善治为先、亦或以促进科技创新为首要目标、还是坚持国家数据安全高于一切的监管初衷,在全球层面仍然存在较大差异。例如,欧盟监管规制的重点是以保障个人权利为基本出发点,更强调对公民个人数据隐私的绝对保护。这与欧盟在数据跨境流动领域的监管规则理念是一致的,体现了欧盟数字治理路径的惯性特点。[②] 欧盟重视人工智能的伦理设计,还提出包括指导人工智能可信任、责任指令等相关的伦理指南。澳大利亚则以促进科技创新为主,支持侧重于科技发展的灵活监管政策。澳大利亚制定了 10 项自愿标准,促进其与经合组织人工智能监管规则对接,在企业参与人工智能研发和促进经济发展的过程中,推动其对人工智能安全、透明性的自觉保护意识。[③] 美国却更多强调创新的重要性,更在意的是人工智能应用带来的一系列挑战。例如,当“深度求索”公司旗下的开源模型 DeepSeek-R1 引发全球广泛关注时,特朗普首先判断其是否对美国国家安全、能源安全造成影响。面对人工智能模型风险时,美国表示愿意和其他国家合作应对一切可能的安全风险。[④] 中国则秉持“以人为本、智能向善”的人工智能发展原则,主张不断提升人工智能技术的

① U.S. Department of State, “Joint Statement Endorsing Principles for 6G: Secure, Open, and Resilient by Design,” February 28, 2024, https://2021-2025.state.gov/joint-statement-endorsing-principles-for-6g-secure-open-and-resilient-by-design/.

② 严少华、杨昭:《欧美人工智能治理模式比较及启示》,《战略决策研究》2024 年第 3 期,第 41 ~ 66 页。

③ Tim Lyons and Olivia Newbold, “Shaping the Future: Australia's Approach to AI Regulation,” September 9, 2024, https://www.lexology.com/library/detail.aspx?g=46db3cb4-e14b-4ef0-971f-3a5e594318cf.

④ Matt Sheehan and Scott Singer, “What DeepSeek Revealed About the Future of U.S.-China Competition,” February 3, 2025, https://foreignpolicy.com/2025/02/03/deepseek-china-ai-artificial-intelligence-united-states-tech-competition/.

安全性、可靠性、可控性和公平性等。

其次,监管力度与技术竞争力之间的冲突。人工智能监管力度正呈现“逐顶竞争”或“逐底竞争”的规则博弈态势,导致各国间的监管协调成本增加,全球不同的监管规则难以衔接。从跨境数据到人工智能,美欧一直在“强监管与弱创新”和“弱监管与强创新”之间纷争不断。强监管政策将扼杀人工智能的创新与竞争力。过度监管增加本地企业和海外企业的合规成本,严重阻碍欧洲人工智能技术创新。[①]《欧盟人工智能法案》是欧盟在全球人工智能监管竞赛中引领规则“逐顶竞争”的代表,该法案的第 99 条处罚措施中,对违背禁止人工智能实践的行为将处以最高 3500 万欧元的行政罚款,对企业的最高罚款为其上一财年全球年营业总额的 7%。[②] 这是继《一般数据保护条例》后的另一项专门针对人工智能的严苛惩罚措施。苹果、谷歌等美国科技巨头反对欧洲过于严苛的人工智能监管环境。美国副总统万斯认为,欧盟对人工智能的大规模严苛与过度监管将扼杀技术创新与产业变革,甚至沦为官僚机构与独裁者的审查工具。[③] 特朗普政府正制定人工智能行动方案,以放松监管和强化产业主导为核心,更加鼓励科技企业实施自我监督和行业自律,为美国数字科技巨头赢得更灵活的市场环境。以创新驱动为主要特征的行为体也更加偏好人工智能监管的“逐底竞争”。比如,新加坡便没有单独为人工智能制定相关的监管规则条例,个人数据保护委员会、金融管理局和卫生部等监管机构采取推广软法的方式,引导各行业负责任地使用人工智能,提倡人工智能数据分析过程中的公平、道德、可问责和透明度原则。[④] 过于严格的人工智能监管标准将限制技术的进步和产业的效益,但过于宽松的监管环境则无法管控人工智能带来的潜在风险挑战。目前,由“深度求索”公

① Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology*, New York: Oxford University Press, 2023, pp. 105 ~ 146.

② “AI Act,” June 23, 2024, https://eur-lex. europa. eu/legal-content/EN/TXT/PDF/? uri = OJ: L _ 202401689.

③ Jeffrey Dastin and Ingrid Melander, “Vance Tells Europeans that Heavy Regulation Could Kill AI,” Reuters, February 12, 2025, https://www. reuters. com/technology/artificial-intelligence/europe-looks-embrace-ai-paris-summits-2nd-day-while-global-consensus-unclear-2025-02-11/.

④ “AI Watch: Global Regulatory Tracker - Singapore,” May 13, 2024, https://www. whitecase. com/insight-our-thinking/ai-watch-global-regulatory-tracker-singapore.

司推出的开源模型引发各国对人工智能应该开源还是闭源的争议,即开源更便于技术审查和监管,而闭源则有助于提升科技企业创新与商业效益。[①] 这一争议背后,是各国寻求在提升人工智能监管与提升科技创新竞争力之间的一种平衡。

(三)中美欧围绕不同目标争夺全球人工智能规则主导权

对于人工智能规则制定权的争夺已成为全球地缘政治竞争的核心领域之一。美国、欧盟等发达经济体与以中国为代表的“全球南方”,正围绕技术标准、伦理规范和数据主权等展开激烈博弈,争夺人工智能规则制定的主导权。

特朗普第二任期在人工智能规则制定行动中的核心目标是维持“美国优先”、强化美国在全球人工智能领域的科技霸权和规则主导权,并对中国采取单边主义制裁。[②] 一方面,美国自身采取的市场驱动型模式,从制度上为人工智能科技企业在尖端科技领域的创新与竞争力保驾护航。美国正抢抓人工智能发展机遇、推动更宽松的监管政策、塑造包容的监管环境,并优先将美国人工智能技术推广到全球。[③] 特朗普政府还宣布了“星际之门”人工智能基础设施投资项目,计划在未来四年投资 5000 亿美元,确保人工智能在美国的全球领先地位。[④] 另一方面,特朗普政府对中国人工智能技术从产业链到供应链展开全方位、多领域的限制,并联合盟友共同对中国人工智能技术进行打压。例如,限制关键技术出口、限制关键数字基建投资、限制使用相关应用程序等。在半导体出口管制方面,特朗普政府正在制定更加严厉的半导体限制措施,对日本、荷兰施压,让盟友效仿美国芯片设备公司实施对中国出口限制。同时,美国将进一步限制向中国出口英伟达公司芯片的类型、阻止中国内存芯片制造商长鑫存储技术有限公司

① George Lawton, “Attributes of Open vs. Closed AI Explained,” July 8, 2024, https://www.techtarget.com/searchenterpriseai/feature/Attributes-of-open-vs-closed-AI-explained.

② Charlotte Yuan, “What Does Trump Have in Store for China's AI Export Controls?” February 22, 2025, https://www.techpolicy.press/closing-the-loopholes-options-for-the-trump-administration-to-strengthen-ai-chip-export-controls/.

③ Aamer Madhani, “Vance Offers an ‘America First’ Argument on AI Deregulation in His First Foreign Policy Speech,” AP News, February 12, 2025, https://apnews.com/article/vance-artificial-intelligence-summit-paris-b3c90fe7fae1cabac07f87aa3a077826.

④ Bernard Marr, “What Does Trump's $500 Billion Stargate Mean for the World of AI?” Forbes, January 23, 2025, https://www.forbes.com/sites/bernardmarr/2025/01/23/what-does-trumps-500-billion-stargate-mean-for-the-world-of-ai/.

购买美国技术等,还可能考虑对中芯国际进行限制。[①] 特朗普签署的《美国优先投资政策》和《鼓励外国投资、保护国家安全》备忘录等,体现美国外国投资审查委员会已开始阻止中国在人工智能等与国家安全有关领域的对美投资。[②] 在人工智能程序使用方面,美国共和党参议员提出《2025 美中人工智能能力脱钩法案》,包含下载使用"深度求索"模型的个人或将监禁 20 年、企业最高罚款 1 亿美元等极端严苛的规定,该法案现已提交司法委员会。[③]

围绕"欧洲数字主权"战略目标,欧盟在人工智能领域的政策实践路径从兼顾人工智能创新与监管双重目标,转向确保产业发展优先、抢抓全球标准制定的话语权。"欧盟数字主权愿景"强调欧盟应争取在人工智能发展过程中获得主动权和控制权,倡导成为全球人工智能技术开发与风险监管的领导者。然而,鉴于缺乏领先的人工智能企业和连贯的创新激励战略,除了提升监管能力和监管标准,欧盟几乎没有创新企业可以推动其进一步引领人工智能标准。[④]《人工智能法案》在很大程度上标志着欧盟在人工智能技术发展与监管上的重大思路转折,即对人工智能风险采用分级监管,取代了与《一般数据保护条例》相似的"一刀切"做法,进一步降低具有低风险和通用的人工智能应用开发门槛,并建立沙盒监管机制激活创新。2025 年巴黎人工智能行动峰会后,欧盟还宣布撤回《人工智能责任指令》和《电子隐私条例》,凸显欧盟正在调整监管与创新的优先项,寻求在人工智能竞争中"破局"。[⑤] 欧盟委员会主席冯德莱恩在巴黎人工智能行动峰

① Mackenzie Hawkins, Cagan Koc and Jenny Leonard, "Trump Team Seeking to Toughen Biden's Chip Controls over China," Bloomberg, 25, 2025, https://news.bloomberglaw.com/ip-law/trump-team-seeks-to-toughen-bidens-chip-controls-over-china.

② The White House, "American First Investment Policy," February 21, 2025, https://www.whitehouse.gov/presidential-actions/2025/02/america-first-investment-policy/; The White House, "Fact Sheet: President Donald J. Trump Encourages Foreign Investment While Protecting National Security," February 21, 2025, https://www.whitehouse.gov/fact-sheets/2025/02/fact-sheet-president-donald-j-trump-encourages-foreign-investment-while-protecting-national-security/.

③ U.S. Congress, "S.321 - Decoupling America's Artificial Intelligence Capabilities from China Act of 2025," January 29, 2025, https://www.congress.gov/bill/119th-congress/senate-bill/321/text.

④ Daniel Mügge, "EU AI Sovereignty: for Whom, to What End, and to Whose Benefit?" *Journal of European Public Policy*, Vol.31, No.8, 2024, pp.2200~2225; Andrea Calderaro and Stella Blumfelde, "Artificial Intelligence and EU Security: the False Promise of Digital Sovereignty," *European Security*, Vol.31, Vol.3, 2022, pp.415~434.

⑤ "EU Commission Withdraws AI Liability Directive," February 25, 2025, https://www.lexology.com/library/detail.aspx?g=7f70826f-8480-4557-8583-3dafa1217922.

会上还宣布启动“投资人工智能”倡议,计划筹集 2000 亿欧元用于人工智能投资,包括 200 亿欧元的欧洲基金专门支持建设人工智能超级工厂,以提升超算能力。[①] 另一方面,欧盟正借助双多边国际合作平台构建基于“欧标版”人工智能国际行动和标准融合,进一步释放与扩散“布鲁塞尔效应”辐射范围。[②] 欧盟人工智能办公室在美欧贸易与技术委员会双边机制中与美国人工智能安全研究所直接开展对话,向美国、七国集团推介基于风险的人工智能监管方法。以法国为代表的欧洲国家借助 2025 年巴黎人工智能行动峰会,积极通过“伦理优先”和“规则输出”推动其在区块链、量子计算等前沿科技领域实现技术主权,达成以欧盟人工智能监管“风险分级分类”为标准的全球行动与国际共识。[③]

以中国为代表的“全球南方”希望借助国际组织与多边机制有更多机会参与人工智能领域的标准制定。中国提出的《全球人工智能治理倡议》呼吁建立风险等级测试评估体系、逐步建立健全法律和规章制度,以应对人工智能在数据安全、个人隐私保护、知识产权等问题上带来的潜在风险。[④] 中国还通过与金砖国家启动人工智能研究组、与阿拉伯国家共建 10 家人工智能联合实验室等方式,与世界各国相互合作、共同应对风险。同时,中国正积极促进人工智能技术更加开放与普惠,鼓励各国广泛参与人工智能监管,尤其是增加发展中国家在人工智能监管与规则制定中的代表性与发言权,避免人工智能技术成为“富国和富人游戏”。[⑤] DeepSeek-R1 模型最重要的意义之一就是从闭源到开源,揭开了人工智能技术背后的神秘面纱,帮助更多发展中国家有机会参与全球科技竞争,从而推动核心技术由少数国家垄断走向更多国家共享,打破高科技领域“赢者通吃”的叙

① European Commission, “EU Launches InvestAI Initiative to Mobilise 200 Billion of Investment in Artificial Intelligence,” February 11, 2025, https://ec.europa.eu/commission/presscorner/detail/en/ip_25_467.

② 苏可桢、沈伟:《欧盟人工智能治理方案会产生“布鲁塞尔效应”吗?——基于欧盟《人工智能法》的分析》,《德国研究》2024 年第 2 期。

③ Robert D. Atkinson, “A Transatlantic G2 Against Chinese Technology Dominance,” April 5, 2024, https://itif.org/publications/2024/04/05/a-transatlantic-g2-against-chinese-technology/.

④ 《全球人工智能治理倡议》,2023 年 10 月 20 日,https://www.fmprc.gov.cn/web/ziliao_674904/1179_674909/202310/t20231020_11164831.shtml。

⑤ 习近平:《携手构建公正合理的全球治理体系——在二十国集团领导人第十九次峰会第二阶段会议关于“全球治理机构改革”议题的讲话》,2024 年 11 月 18 日,https://www.gov.cn/gongbao/2024/issue_11746/202412/content_6991662.html。

事,真正实现人工智能技术的普惠性。

(四)人工智能技术将进一步加剧全球“数字鸿沟”

人工智能技术快速发展将进一步扩大全球“数字鸿沟”,制约“全球南方”共享全球数字科技红利。

首先,“全球南方”在人工智能技术、算力、数据中心等提升人工智能竞争力的关键领域不具备优势。据《全球人工智能实力排行榜》,2023 年排在全球前 10 的国家中仅有中国、印度、阿联酋等主要发展中国家,分别位列第 2、4、5 名;在 36 个经济体排名倒数 10 名中,就有沙特(27)、俄罗斯(29)、土耳其(31)、墨西哥(33)、巴西(34)和南非(36)等 6 个发展中国家。[①] 此外,数字基础设施是决定各国算力竞争的核心之一,但发展中国家仍缺乏获得高速宽带、稳定电力和电信网络的条件,进一步限制了其参与全球数字化进程。2024 年世界数字竞争力排名中,博茨瓦纳、哈萨克斯坦和尼日利亚的 5G 市场移动宽带用户仅占 3.15%、2.44% 和 1.04%。在互联网宽带速度方面,冰岛、新加坡和法国均达到 200 兆比特/秒以上,但委内瑞拉、印度尼西亚等发展中国家基本在 30 兆比特/秒以下,博茨瓦纳甚至仅为 10.61 兆比特/秒。[②] 近年来,谷歌、亚马逊、微软等美国科技巨头以及法国的科技企业正积极在国内甚至全球布局建设数据中心。在全球前 10 大数据中心排名中,除了中国电信和中国移动,其余均为美国、英国和葡萄牙等发达经济体的企业。[③] 到 2027 年,美国数据中心建设预计达到 250 亿美元,未来 5 年将增加约 2825 兆瓦电力容量等。[④] 较高的数字基建要求与昂贵的投资成本等,使得很多发展中国家难以承受。大部分发展中国家并不具备充足谈判优势。

---

① “Global AI Power Rankings: Stanford HAI Tool Ranks 36 Countries in AI,” November 21, 2024, https://hai.stanford.edu/news/global-ai-power-rankings-stanford-hai-tool-ranks-36-countries-ai.

② “IMD World Digital Competitiveness Ranking 2024,” November 2024, https://imd.widen.net/s/xvhld-krrkw/20241111-wcc-digital-report-2024-wip.

③ Amber Jackson, “Top 10: Biggest Data Centres,” Data Center, October 23, 2024, https://datacentremagazine.com/top10/top-10-biggest-data-centres.

④ “U.S. Data Center Construction Market to Reach Around $25 Billion by 2027. Around 2,825 MW Power Capacity will be Added in the Next 5 Year - Arizton,” August 29, 2022, https://www.globenewswire.com/news-release/2022/08/29/2506279/0/en/U-S-Data-Center-Construction-Market-to-Reach-Around-25-Billion-by-2027-Around-2-825-MW-Power-Capacity-will-be-Added-in-the-Next-5-Year-Arizton.html.

其次,“全球南方”在人工智能相关监管规制体系建设仍存在制度缺失。相较美欧等发达经济体,“全球南方”更需要对人工智能技术进行必要的监管。一是应对数据安全、隐私保护、算法偏见、不透明决策等威胁,甚至是对国家数据安全、公民生物数据特征滥用等更严重侵犯的重要保护与规制。二是保护本国科技企业创新,快速提升技术竞争力。三是吸引匹配的外国科技企业投资,并促进知识交流、技术转让、人才引进等。然而,亚非拉等多数国家还没有出台人工智能全面且具有约束力的法律或法规,也没有形成统一的立法体系。

再次,“全球南方”在国际组织、双多边机制和非国家行为体主导的联盟机制中参与全球人工智能规则制定的话语权较弱。其一,在多边机制中的代表性不足,与发达经济体对话普遍缺位。以美国、欧盟为代表的发达经济体在人工智能国际规则制定中占据主导地位。在人工智能多边机制谈判中,中国等诸多发展中国家并未参加第二届首尔人工智能安全峰会,七国集团制定的人工智能相关规则直接向二十国集团和经合组织输出,并未征求发展中国家的意愿。全球首份《人工智能与人权、民主和法治框架公约》由美国、挪威、加拿大、澳大利亚和日本等 50 多个国家签署,中国和许多发展中国家也并未参与其中。[①] 其二,多边规则协调能力不足,缺乏牵头制定规则或协调发展中国家与发达经济体规则差异的专门机构。由于美欧掌握人工智能关键核心技术主导权,经合组织、七国集团等成为美西方实施技术垄断、制定有利于发达经济体规则诉求的国际组织,而二十国集团等则成为发达经济体输出规则、促使规则武器化、胁迫发展中国家的场域。其三,“全球南方”难以形成统一的立场和行动,将进一步削弱其在国际规则制定中的话语权。比如,金砖国家成员国的人工智能技术发展差异较大,规则协调与整合机制匮乏,难以形成具有约束力的“全球南方”人工智能监管规则。此外,“全球南方”不具备顶尖的科技企业参与全球人工智能科技竞争,几乎很难有资格参与人工智能科技跨国联盟。

① Coucil of Europe, “The Framework Convention on Artificial Intelligence,” https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence.

四、结　　论

2025 年是世界各国正式迈入人工智能时代的“元年”,也是全球积极探索人工智能监管、应对科技变革的关键之年。国际社会已围绕人工智能监管规则制定与协调开展行动,但仍然面临国际层面的人工智能治理失效、经济体之间的监管理念与模式巨大差异、中美欧围绕不同目标争夺全球人工智能规则主导权、以及全球“数字鸿沟”加剧等多重挑战。

在此情形下,中国应强化多层次高水平国际交流合作,实现从人工智能监管规则“跟随者”向“引领者”的跨越。一是积极加入全球与区域层面现有国际制度框架下的人工智能监管规则制定工作。基于联合国、世界贸易组织、二十国集团、金砖国家为代表的多边机制,建构兼顾科技创新与国家安全的全球人工智能监管规则。在亚太经合组织、《数字经济伙伴关系协定》等区域性合作机制中,增加与人工智能安全与风险规则相关的监管合作,提升监管规则在缔约方司法管辖区内进行有效对接。二是探索构建新的人工智能监管国际合作机制。依托“一带一路”倡议和“数字丝绸之路”建设推动与合作伙伴国家共享人工智能领域知识、经验和资源时,重视监管规则的输出。同时,进一步推动金砖国家加强人工智能规则制定与跨国监管协调等方面的合作。三是在美欧人工智能治理规则博弈中提升中国影响力。除了与美国、欧盟分别建立人工智能安全稳定的“二轨对话机制”,中国还应在美欧因监管模式差异产生的难以调和分歧中开辟一条提升中国人工智能规则制定话语权的新路径:一方面,掌握人工智能核心技术是中国在美欧人工智能监管博弈中寻求自身安全与发展的必由之路。“深度求索”公司的成功向世界证明:尖端的技术是快速提升中国赢得人工智能监管话语权的核心竞争力。另一方面,在监管风险重点和国家干预力量中努力寻求一种平衡的监管张力,而非陷入二元对立,是中国在突破美欧人工智能规则冲突中寻求双边合作的关键,也是今后学界研究的主要方向。