

日本参与人工智能国际规则制定探究

江天骄 张书言

内容提要 日本的人工智能技术水平居于世界第二梯队,但其借助结构化的科技外交路径,通过包括制度参与、话语构建和实践共享等三条体系化参与策略,提升了其在该领域的国际影响力。在制度层面,日本推动经合组织、“七国集团”和“二十国集团”的人工智能治理原则“对齐”,并推出“可信数据自由流动”框架作为人工智能治理的配套机制;在话语层面,日本通过为人工智能国际讨论准备草案文本,在国规则中彰显本国利益诉求;在实践层面,日本与国际社会分享其具有“软法范式”和“敏捷治理”特色的治理实践,吸引国际社会对日本的认同。通过这三条路径间的有机互动,日本塑造了人工智能治理“积极行为者”的国际形象,助力其提升国际话语权。

关键词 非传统安全研究 人工智能 日本 国际规则制定 国际话语权

以人工智能为代表的新兴技术正在对大国战略竞争态势和国际秩序变革造成愈发深刻的结构性影响。在全球化的今天,谋求科学技术的国际话语权已然变得和发展本国的技术产业一样重要。这不仅是因为话语权本身构成了国家权

* 江天骄,复旦大学发展研究院副研究员。(邮编:200433);张书言,复旦大学国际关系与公共事务学院博士研究生。(邮编:200433)

** 本文是国家社科基金一般项目“金砖国家人工智能合作的实现机制及中国策略研究”(项目编号:25BGJ009)的阶段性成果。感谢《国际政治研究》匿名评审专家的意见和建议,文责自负。

力的组成部分,更是因为国家必须将自身的利益诉求和政策偏好融入全球治理进程,才能使人工智能的国际规范和规则满足自己的发展需要,从而为本国发展营造良好的国际环境。当前,不仅中美两国在竞争全球人工智能治理的话语权,一批技术水平、产业能力相对落后于中美,但仍具备一定特殊优势的国家,也正在对全球人工智能治理跃跃欲试。^①无论是印度与东盟共同开发地区性数字治理框架^②,还是巴西聚焦“金砖国家”机制构建人工智能治理体系^③,都反映出“技术追赶者”国家对全球人工智能治理的积极兴趣。

长期以来,日本是信息技术产业的强国。20世纪80年代,日本凭借其技术优势参与国际科技合作,赢得国际影响力。进入人工智能时代,日本的发展却显得力不从心,因国内社会、人口和经济结构等因素限制,日本在人工智能发展领域长期位于世界第二梯队。即使在更加广泛的数字技术领域,日本也难以找出能够与美国的谷歌、苹果、亚马逊或是中国的腾讯、阿里巴巴、百度等比肩的数字科技企业。已有研究认为,日本人工智能技术水平无论是在投资体量、专利和论文成果数量还是在技术人才方面均落后于中美两国,其老龄少子化、数据市场壁垒及国内社会保守风气等因素是日本发展人工智能面临的主要困境。^④在此背景下,日本将如何获取人工智能领域的国际话语权?

为系统回答上述问题,本文采取过程追踪法梳理日本参与人工智能国际讨论的历程,并对以下三类资料进行文本分析,进而总结日本人工智能科技外交的行为模式:一是日本政府官方文献,包括日本内阁府、总务省、文部科学省、经济产业省等机构发布的人工智能战略文件、研发与应用指南、政策白皮书及法律草案;二是国际组织及会议文件,涵盖经合组织、“七国集团”“二十国集团”的相关会议纪要、部长声明、原则建议书及进展报告;三是日本国内学术成果,包括日本人工智能学会、各大学研究机构发布的伦理准则、政策建议及技术评估报告。资料时间跨度主要覆盖2016年至2025年,以求呈现日本参与人工智能国际规则制定的动态演进过程。

① 对国家人工智能技术水平的划分参考了中国科学技术信息研究所与北京大学联合发布的《全球人工智能创新指数报告2025》,其中,中美两国位居第一梯队,日本、新加坡、法国是第二梯队的代表,加拿大、印度、巴西等国家位于第三梯队。参见 <https://reportify.cn/news/1146788686847938560>,2026-04-10。

② 孙志伟、殷浩铤:《人工智能时代数字巨头的技术权力及其对“全球南方”的挑战》,《国际安全研究》2025年第2期,第142—164页。

③ 姚旭、江天骄:《金砖国家推动人工智能治理的动因、实践和挑战》,《同济大学学报(社会科学版)》2025年第2期,第36—49页。

④ 蓝庆新、彭一然:《日本“数字新政”战略动机与发展特征》,《人民论坛》2020年第5期,第128—131页;周生升、秦炎铭:《日本人工智能发展战略与全球价值链能力再提升:基于顶层设计与产业发展的竞争力分析》,《国际关系研究》2020年第1期,第67—90页;邓美薇:《日本人工智能的战略演进和发展愿景及其启示》,《日本问题研究》2022年第2期,第11—21页。

一、关于日本人工智能治理的相关研究述评

目前,国内外相关文献已经对日本人工智能治理的特殊性有所关注,主要研究聚焦于以下两个方面:

(一) 日本参与全球人工智能治理的情况

首先,日本将2008年前后提出的“科技外交”战略延伸到人工智能国际合作中。一方面,日本引进先进技术,与发达国家开展共同研究以提升自身人工智能水平,同时巩固同盟关系^①;另一方面,面向广大发展中国家,日本通过技术和资金、人才培养及基础设施援助,既解决全球性挑战,又获取资源与数据支持,并借以扩大自身国际影响力。^②在日本与欧盟的人工智能合作中,双方就构建“第五代社会”(Society5.0)达成共识,认为这一目标可以带来数据驱动的创新、经济增长和社会福祉的提升。然而,在数据隐私和安全问题、技术鸿沟,以及政策和法规的适应性问题,日欧的政策协调性存在不足。有分析进一步指出,日欧在应对中美大国博弈时,应该采取更积极的姿态,在“二十国集团”和世贸组织(WTO)的讨论中相互支持,建立国际通用的人工智能法律框架。^③相比之下,当前日本与发展中国家在人工智能研究机构间的合作密度仍然较低,人才和资金短缺是重要原因之一。^④日本仍将英美作为人工智能国际合作的主要对象,谋求其在美国“印太”人工智能盟伴体系中的领导地位,这进一步增加了中日人工智能合作的困难。^⑤

其次,日本在人工智能技术相关的国际安全议题上持有特殊立场。以致命

① 惠仲阳、胡智慧:《日本科技外交战略与实践探究》,《中国科技论坛》2016年第12期,第141—146页。

② 廖欢:《日本对印度科技外交的发展与演变研究》,《日本问题研究》2024年第2期,第10—24页;李建民:《日本战略文化、科技发展与中日科技合作》,《中国科技论坛》2009年第11期,第129—133页;惠仲阳、胡智慧:《日本科技外交战略与实践探究》,《中国科技论坛》2016年第12期,第141—146页;马潇潇:《日本科技外交中的国家形象塑造与话语权构建》,《科技传播》2025年第7期,第73—78页;Atsushi Sunami, et al., “Japan’s Science and Technology Diplomacy,” *World Scientific*, 2015, pp. 243-258; Julia Mariko Jacoby, “Learning from the Earthquake Nation: Japanese Science Diplomacy in the Twentieth Century,” *Journal of Contemporary History*, Vol.56, No.3, 2021, pp. 485-501.

③ Ana Gascon Marcen, “Society 5.0: EU-Japanese Cooperation and the Opportunities and Challenges Posed by the Data Economy,” 2020, https://www.researchgate.net/publication/339136139_Society_50_EU-Japanese_cooperation_and_the_opportunities_and_challenges_posed_by_the_data_economy, 2024-11-01.

④ 袁野等:《人工智能产业核心技术的国际竞争态势分析》,《中国电子科学研究院学报》2020年第11期,第1128—1138页;殷勇:《中日韩:人工智能合作现状与优势分析》,《东北亚经济研究》2020年第4期,第89—96页。

⑤ 赵明昊:《美国“印太”盟伴体系中的日本角色探析》,《日本学刊》2024年第4期,第1—29页。

性自主武器系统(LAWS)的国际规则制定为例,作为唯一遭受过核打击的国家以及曾经的亚太地区侵略者,日本对人工智能军事技术持有极其审慎的态度^①,主张将国际人道法原则应用于致命性自主武器系统监管。^②在具体的治理合作过程中,面对美国对华战略博弈,日本不仅需要与美国和欧盟合作,还需要与中国合作,对生成式人工智能进行监管,并建立一个具有法律约束力的框架来规范致命性自主武器系统的发展。^③此外,还有研究关注到日本与澳大利亚的军事合作,批评两国合作建设海洋监视、威胁检测和数据融合的人工智能驱动系统有可能提升亚太地区地缘政治风险。^④

最后,制定人工智能国际规则和规范是当前日本参与全球治理的重要工作。由于日本无法短时间内扭转人工智能发展的相对落后地位,因此通过积极参与国际规则制定以提升国际话语权成为日本的重点策略。^⑤有研究指出,日本利用2023年“七国集团”峰会主席国身份,在国际层面引导共识,共享其国内的治理实践特色,这一做法将进一步塑造其在人工智能国际治理合作中担任领导者的地位。^⑥还有研究认为,日本参与国际人工智能规则制定的特点是“互操作性”。即日本既没有能力像美国这样的人工智能超级大国制定事实上的标准,也无法像欧盟那样推动全球标准化进程。日本将工作重点放在“互操作性”上,力求不干预欧盟或美国制定的法规或指南,而是确保国际社会就互操作性的核心原则达成共识。^⑦比如,与欧盟的《通用数据保护条例》(GDPR)相比,日本的个人信息保护法并未直接保障隐私权。因此,日本政府需要考虑如何在保留自身文化属性

① 畠山記美江「AI原則の国際的な潮流と日本のAI原則の特質:今後の見直し検討に向けた政策上の示唆」,『慶應SFC学会』,23卷,2023年,142—162頁。

② 新保史生「自律型致死兵器システム(LAWS)に関するロボット法的視点からの考察」,『電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review』,第3号,2019年,217—230頁;Daisuke Akimoto, “International Regulation of “Lethal Autonomous Weapons Systems” (LAWS): Paradigms of Policy Debate in Japan,” *Asian Journal of Peacebuilding*, Vol.7, 2019, pp. 311-332, https://www.researchgate.net/publication/341374703_International_Regulation_of_Lethal_Autonomous_Weapons_Systems_LAWS_Paradigms_of_Policy_Debate_in_Japan, 2024-11-10.

③ Daisuke Akimoto, “Japan’s AI Diplomacy,” *The Diplomat*, April 26, 2024, https://my.nda.news/artificial-intelligence-st_38009/news/japans-ai-diplomacy-the-diplomat-sn_5788591/, 2024-11-01.

④ Taylor Rodriguez Vance, “Geopolitical Implications of Artificial Intelligence in Cybersecurity: A Comprehensive Analysis,” *International Journal of Computer Science Engineering and Information Technology Research*, Vol.11, 2024, pp. 91-105, <https://www.researchgate.net/publication/382129529>, 2024-11-01.

⑤ 陈琪、肖源睿:《国际人工智能战略之比较》,《国别和区域研究》2020年第1期,第182—202页。

⑥ Habuka Hiroki, “Japan’s Approach to AI Regulation and Its Impact on the 2023 G7 Presidency,” CSIS, 2023, <http://www.jstor.org/stable/resrep47347>, 2024-11-01.

⑦ Tagui Ichikawa, “Norms in New Technological Domains: Japan’s AI Governance Strategy,” CSIS, June 17, 2025, <https://www.csis.org/analysis/norms-new-technological-domains-japans-ai-governance-strategy>, 2025-12-15.

的同时,使日本的隐私保护做法受到国际认可。^① 又比如,有研究建议针对欧盟发起的《人工智能公约》(AI Convention),日本政府应该尽快确定《人工智能公约》的适用范围,迅速开发用于风险评估的方法系统,从而确保日本对欧盟治理规则的“互操作性”。^② 还有研究指出,针对经合组织和“二十国集团”提出的人工智能包容性原则和公平性原则,日本政府需要充分协调国内不同利益攸关方的立场,既保证日本国内治理与国际规范“对齐”^③,也确保日本国内的多元利益诉求可以充分反映在国际规则制定中。^④

(二) 日本人工智能治理的国内实践

尽管这部分研究并未直接涉及日本参与国际规则制定的行为模式,但从日本国内视角出发为理解其对外政策提供了有益参考,在一定程度上揭示了日本国内实践与国际政策的互动关系。

首先,日本的人工智能发展战略的首要目标是利用技术创新解决社会问题。第一,日本国内风险投资活跃度不足,远远低于国际竞争对手;第二,日本企业过于“墨守成规”,严重缺乏“冒险创新精神”;第三,长期的物质充足与人口老龄化,导致个人消费疲软和高素质劳动力短缺。^⑤ 因此,日本的人工智能政策重点在于回应人口减少、老龄化,以及医疗和护理费用不断增加等现实社会问题,而人工智能与其他相关技术的融合,为解决这些社会问题提供了可能性。有研究将日本发展人工智能技术的动机概括为“建立可持续的社会应用框架”。^⑥

其次,日本的人工智能治理采取多元社会主体共同参与的发展与监管模式。

① 畠山記美江「AI原則の国際的な潮流と日本のAI原則の特質:今後の見直し検討に向けた政策上の示唆」,『慶應 SFC 学会』,23 卷,2023 年,142—162 頁。

② Arisa Ema, “International Collaboration in AI Governance, Key Considerations of the Council of Europe’s AI Convention and Japan’s Response,” *IFI Policy Recommendations*, No.25, October 2023, https://ifi.u-tokyo.ac.jp/en/wp-content/uploads/2023/10/policy_recommendation_tg_20231031e.pdf, 2025-12-15.

③ 江間有沙「コメンタリ:AI原則から実践にあたって日本が国際的な活動から学ぶべき教訓」,『人工知能』,36 卷,2021 年,210—212 頁。

④ 江間有沙「小特集「AI原則から実践へ:国際的な活動紹介」にあたって」,『人工知能』,36 卷,2021 年,179—180 頁;ガル グニット、翻訳:江間有沙「AIガバナンスに関する多国間主義における日本の役割」,『人工知能』,36 卷,2021 年,181—186 頁;バク ティナ M、翻訳:江間有沙「AI開発における包摂を意識して道を切り開く」,『人工知能』,36 卷,2021 年,187—192 頁;ランキスト ヨランダ、ポーラ アドリアナ、イリアディス ニキ、翻訳:江間有沙「善のためのAIとそのグローバルガバナンス」,『人工知能』,36 卷,2021 年,193—198 頁。

⑤ 陈婧嫣等:《跨国比较视阈下的人工智能政策:目标、理念与路径》,《科学学与科学技术管理》2021 年第 3 期,第 87—100 页;朱启超、王姝:《日本“超智能社会”建设构想:内涵、挑战与影响》,《日本学刊》2018 年第 2 期,第 60—86 页。

⑥ Roxana Radu, “Steering the Governance of Artificial Intelligence: National Strategies in Perspective,” *Policy and Society*, Vol.40, No.2, 2021, pp.178-193, <https://doi.org/10.1080/14494035.2021.1929728>, 2025-12-15.

在产业发展方面,日本政府主导推出人工智能政策并进行顶层制度设计,各省市配合出台相应领域的人工智能发展路线图,并在国家层面推出人工智能创新项目,促进政府与市场的投资与研发合作。^①日本高校和科研机构在人工智能创新生态系统中负责人才培养和知识创新,构建素养教育、应用基础教育、专家培育多层次梯级人才培养体系。日本企业积极参与技术研发与应用落地。目前,日本已经在金融、交通、医疗和工业制造领域率先试点将人工智能和经济生产结合。^②有研究将日本的人工智能发展概括为“政府引导、市场化运作、产官学协作”模式。^③在人工智能监管方面,日本同样采取了“政府—社会—企业”多元参与的监管模式。日本政府、行业协会和企业共同参与人工智能监管原则制定,且各有分工。政府部门主导原则规则制定,并担任监督和执法者角色;经济团体联合会等行业协会与社会组织参与制定本行业伦理规则并配合实施伦理审查和认证。日本企业主动制定本企业伦理规则,在监督环节实行自我监督、履行说明责任,在执法环节自主改进、自主纠错。^④有研究将日本自主监管和共同监管的模式概括为人工智能监管的“软法范式”,该模式表现出“敏捷治理”的特点^⑤,反映了日本“以人为中心”的人工智能治理原则。^⑥

最后,日本不少国内研究从政策实践出发,讨论如何完善和改进日本国内的人工智能治理结构。有研究认为,日本在坚持“软法范式”的同时也可以参考国际经验,建立人工智能技术评估机制,使行业协会和企业可以及时将人工智能技术风险报告给政府部门,并反映在决策中。^⑦也有研究从社会学的视角讨论人工

① 张鹏飞:《日本人工智能产业发展和政策研究》,《现代日本经济》2021年第5期,第25—40页;刘姣姣等:《日本人工智能战略:机构、路线及生态系统》,《科技管理研究》2020年第12期,第39—45页;王玲:《日本政府发展人工智能的战略布局分析》,《全球科技经济瞭望》2020年第10期,第1—7页。

② 刘平、刘亮:《日本新一轮人工智能发展战略:人才、研发及社会实装应用》,《现代日本经济》2020年第6期,第36—47页;陈祥:《日本人工智能战略论析》,《大连理工大学学报(社会科学版)》2023年第5期,第18—27页;李哲等:《日本人工智能战略及人才培养模式研究》,《现代教育技术》2019年第12期,第21—27页。

③ 朱启超、王姝:《日本“超智能社会”建设构想:内涵、挑战与影响》,《日本学刊》2018年第2期,第60—86页。

④ 肖红军、阳镇:《数字科技伦理监管的理论框架、国际经验与中国借鉴》,《东北财经大学学报》2023年第5期,第47—61页;马珊珊等:《可信赖人工智能标准化研究》,《信息技术与标准化》2022年第9期,第46—54页。

⑤ 刘湘丽、肖红军:《软法范式的人工智能伦理监管:日本制度探析》,《现代日本经济》2023年第4期,第28—44页。

⑥ 福岡真之介「AIに関する法制化についての日本のアプローチ」,『立命館国際地域研究』,55巻,2022年,23—33頁;橘均憲「人間のためのAI(human-centric AI)を実現する社会実装の道筋」,『情報処理』,63巻,2022年,1—7頁。

⑦ 城山英明「人工知能とテクノロジーアセスメント」,『科学技術社会論研究』,16巻,2018年,65—80頁。

智能带来的伦理、法律和经济问题^①，提倡面向社会公众设置对话机制，鼓励普通民众与专家学者就人工智能的应用展开讨论，扩大公民在人工智能治理中的参与度。^② 还有研究聚焦于如何提升日本政府决策过程的科学性，认为决策者应该梳理并整合社会治理的概念体系和信息科学的概念体系，从而实现人工智能嵌入社会治理的目标。^③

总体上，国内外研究已经比较充分地认识到日本在人工智能治理中的特殊性，并有意识地将其国内治理实践置于全球人工智能治理的国际语境中加以观察，尝试在其国内政策与国际人工智能规则与规范之间形成对照和联系。其中，中国和欧美学者的研究主要倾向于对日本实践经验的总结与归纳，在理论上提炼出日本人工智能治理的特点；日本国内学者则更聚焦于从特定人工智能治理议题切入，对日本政府的治理实践提出政策建议。然而，已有研究对于日本参与国际规则制定的行为模式分析仍缺乏系统性，对其策略的有效性分析也有待深入，为本文的探讨留下了空间。

二、日本参与人工智能国际规则制定的路径

由于技术要素的差距及国内社会经济发展的限制，日本优先选择借助“软实力”的方法，通过制度参与、话语构建、实践共享三条路径参与国际规则制定，提升人工智能领域的国际影响力和话语权。

（一）制度参与

近年来，日本努力促进不同国际制度之间的交流合作，保证不同国际制度关于人工智能讨论的一致性，巩固国际讨论的基础。同时，日本还尝试为实现人工智能原则提供配套性制度，加快相关国际共识的推进和落实。

1. 日本推动不同国际制度交流合作，促成“七国集团”“二十国集团”和经合组织人工智能治理原则的“对齐”。2016年，日本邀请了经合组织科学技术与创新主任安德鲁·威科夫(Andrew Wyckoff)作为观察员参与“七国集团”信息通讯

① 鈴木貴之「江間有沙「AI社会の歩き方—人工知能とどう付き合うか」」、『科学技術社会論研究』、18卷、2020年、246—247頁。

② 江間有沙「人工知能社会のあるべき姿を求めて」、『科学技術社会論研究』、16卷、2018年、142—146頁；日比野愛子ほか「AI哲学マップ(第7回)「社会の中のAI」という視点」、『人工知能:人工知能学会誌』、2022年、210—221頁。

③ 小林隆「人工知能を活用したガバナンスのための概念の体系化」、『アジア太平洋討究』、32卷、2018年、85—111頁。

技术部长会议。在联合新闻发布会上,日本政府提出,经合组织长期以来充当“七国集团”实际上的“常设秘书处”,因此,“七国集团”应该与经合组织和国际电信联盟(ITU)通力合作,引领国际社会关于人工智能社会经济影响的讨论,日本、德国和经合组织应该担任制定人工智能研发指南的领导角色。^① 同年,在经合组织的技术前瞻性会议上,日本专家分享了“七国集团”会议上的人工智能原则草案,经合组织同意在此基础上形成人工智能原则。经合组织科学技术与创新主任威科夫特别提到了日本政府对“七国集团”信息通信技术部长会议的支持,并感谢日本政府支持经合组织对人工智能政策的研究。^② 经合组织人工智能专家小组共同意识到人工智能潜在的风险,呼吁经合组织与日本、美国等“七国集团”成员国一同行动,开展更深入的政策研究。^③ 对“二十国集团”而言,日本也利用经合组织和“二十国集团”的密切联系,率先将关于人工智能的国际讨论引入“二十国集团”框架。2019年6月,由日本主持的“二十国集团”贸易与数字经济部长会议在茨城县筑波市举行,日本在会议上提出人工智能议程,使“二十国集团”成员首次正式讨论人工智能问题。

在上述三个国际制度合作的基础上,日本利用自己“中间人”的身份,尝试推动“七国集团”、“二十国集团”和经合组织之间人工智能治理原则“对齐”。首先,日本鼓励“七国集团”参考经合组织的人工智能原则,制定具体的实践指南。2016年,日本政府参考经合组织的隐私和安全治理指导原则,建议“七国集团”制定自己的人工智能研发原则草案。^④ 2023年广岛峰会上,日本呼吁“七国集团”成员国在技术可行的情况下,开发和部署人工智能认证机制。例如,水印技术,使用户能够识别人工智能生成的内容,从而响应经合组织提出的人工智能系统的“透明性和可解释性原则”。此外,日本政府还在广岛峰会上提倡在人工智能系统各利益相关方之间实现信息共享,对先进人工智能系统的能力做出公开报告,以响应经合组织对人工智能系统可问责性的要求。未来,“七国集团”还将继续根据《四年以来经合组织人工智能原则的实施情况》,提出进一步的合作原则,确

① Ministry of International Affairs and Communications, “G7 ICT Ministers’ Meeting in Takamatsu, Kagawa Joint Press Conference,” https://www.soumu.go.jp/joho_kokusai/g7ict/english/main_content/joint_press_conference.pdf, 2024-09-16.

② OECD Conference Centre, “Summary of the CDEP Technology Foresight Forum Economic and Social Implications of Artificial Intelligence,” [https://one.oecd.org/document/DSTI/CDEP\(2016\)17/en/pdf](https://one.oecd.org/document/DSTI/CDEP(2016)17/en/pdf), 2024-09-16.

③ Ibid.

④ Ministry of International Affairs and Communications, “Proposal of Discussion toward Formulation of AI R&D Guideline,” https://www.soumu.go.jp/joho_kokusai/g7ict/main_content/ai.pdf, 2024-09-16.

立新的合作项目。^① 其次,日本还尝试推动经合组织与“二十国集团”之间达成人工智能治理原则“对齐”。“二十国集团”人工智能原则与经合组织《人工智能建议书》中的人工智能原则是完全一致的,都包括“包容性、可持续发展及福祉”“尊重法律规则、人权与民主价值”“透明性和可解释性”“稳健性、安全性与安全保护”^②及“可问责性”。日本政府还在此基础上提议“二十国集团”制定更具体的政策实施建议,包括投资人工智能研发、打造数字生态、为人工智能打造政策环境、人才赋能与劳动力市场转型以及国际合作。^③

2. 日本力推“可信数据自由流动”框架作为人工智能国际合作的配套制度,促进相关原则落实。日本政府认为,“可信数据自由流动”(Data Free Flow with Trust, DFFT)是人工智能发展的重要基础,可以为人工智能和数字化技术的推广提供重要的社会基础设施,是确保人工智能技术满足公平、透明、负责任和安全价值的重要前提,对国家发展负责任的人工智能并建立网络安全韧性意义重大。^④ 为此,日本前首相安倍晋三在2019年达沃斯世界经济论坛年度峰会上提出“可信数据自由流动”框架,推动可信赖的、具有互操作性的数据治理实践。“可信数据自由流动”框架包括四大关键合作领域:(1)数据本地化;(2)监管合作;(3)政府访问数据;(4)行业数据共享。^⑤ 成员国可以通过落实这些重点工作领域来满足“七国集团”与经合组织提出的人工智能原则。

目前,日本正在尝试将“可信数据自由流动”框架引入“七国集团”和“二十国集团”的人工智能治理框架,作为国际人工智能原则的基础制度保障。从2019年至2023年,日本政府分别在“二十国集团”峰会和“七国集团”峰会上努力推动“可信数据自由流动”框架的落实。2021年4月,“七国集团”数字和技术部长会议通过了《“七国集团”可信数据自由流动合作路线图》(G7 Roadmap for Cooper-

^① Lucia Russo and Noah Oder, “How Countries Are Implementing the OECD Principles for Trustworthy AI,” OECD. AI Policy Observatory, October 31, 2023, <https://oecd.ai/en/wonk/national-policies-2>, 2024-09-16.

^② 经合组织和“二十国集团”对于“稳健性、安全性与安全保护”这一原则的原文表述是“Robustness, security and safety”。本文为避免中文翻译上的歧义,将“security”译为“安全性”,意为保护人工智能系统本身不受到外部攻击;将“safety”译为“安全保护”,意为确保人工智能系统的使用不损害人类使用者的人身安全。

^③ Ministry of International Affairs and Communications, “G20 Ministerial Statement on Trade and Digital Economy,” https://www.soumu.go.jp/main_content/000625715.pdf, 2024-09-16.

^④ Cabinet Office, “AI Strategy 2022,” <https://www8.cao.go.jp/cstp/ai/aistratagy2022en.pdf>, 2024-09-16.

^⑤ G7, “G7 Roadmap for Cooperation on DFFT,” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/986160/Annex_2__Roadmap_for_cooperation_on_Data_Free_Flow_with_Trust.pdf, 2024-09-16.

ation on DFFT)。^① 2023年,在广岛峰会上,“七国集团”领导人再次表达了落实“可信数据自由流动”框架的意愿,并重申了“可信数据自由流动”框架四大关键合作领域是“七国集团”数字贸易的优先政策事项。^② 在具体的操作过程中,日本推动“可信数据自由流动”框架的方法遵循“从双边到小多边,再从小多边发展至大多边”的逻辑顺序。即日本首先在美日、美英的双边数字贸易中融入“可信数据自由流动”框架,再试图将这些双边协议整合到一个以日本为中枢的“日—美—欧”小多边结构中。在此基础上,吸引具有相似治理理念的国家加入“可信数据自由流动”框架,进而在全球层面推广以“可信数据自由流动”框架为基础的人工智能原则。^③ 目前,日本与美欧之间的双边数字贸易协定已经开始实施。日本将“可信数据自由流动”框架引入了《美日数字贸易协定》(USJDTA)。相比《美墨加协定》(USMCA),《美日数字贸易协定》增加了关于交互式计算机服务和政府数据公开的条款,但在条款的约束力上相对减弱。《美日数字贸易协定》允许对部分数字产品非歧视性待遇制定例外条款,对个人信息保护的要求力度和对电子信息骚扰的监管力度也有所减小。这些调整意味着《美日数字贸易协定》是美日之间妥协的结果^④,部分反映了“可信数据自由流动”框架的开放和灵活性特征。同时,日本也已经与英国达成《日英全面经济伙伴关系协定》(Japan-UK EPA),其中关于数字贸易的条款涉及范围已经超过《日欧经济伙伴关系协定》(EU-Japan EPA),并在数据自由流通、保护个人数据、开放数字市场和消费者权益保护等方面,反映了“可信数据自由流动”框架的原则。“可信数据自由流动”框架本身不是一个约束性的协议框架,而更倾向于一个灵活的理念框架。^⑤ 因此,预计日本下一步的行动是吸引理念相似国家相继加入“可信数据自由流动”框架,在小范围国家内部先形成符合“可信数据自由流动”框架标准的多边数字

① G7, “G7 Roadmap for Cooperation on DFFT,” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/986160/Annex_2__Roadmap_for_cooperation_on_Data_Free_Flow_with_Trust.pdf, 2024-09-16.

② Ministry of International Affairs and Communications, “G7 Vision for Operationalising DFFT and Its Priorities,” https://www.soumu.go.jp/joho_kokusai/g7digital-tech-2023/topics/pdf/pdf_20230430/annex1.pdf, 2024-09-16.

③ 关于日本参与全球数字贸易并推广 DFFT 的动机和策略,参见李墨丝:《欧美日跨境数据流动规则的博弈与合作》,《国际贸易》2021年第2期,第82—88页;中原秋樱、刘宏松:《日本参与跨境数据流动治理:动因、路径与困境》,《日本研究》2023年第1期,第72—83页。

④ 周念利、吴希贤:《美式数字贸易规则的发展演进研究:基于〈美日数字贸易协定〉的视角》,《亚太经济》2020年第2期,第44—51页。

⑤ Aidan Arasingham and Matthew P. Goodman, “Operationalizing Data Free Flow with Trust (DFFT),” CSIS, April 13, 2023, <https://www.csis.org/analysis/operationalizing-data-free-flow-trust-dfft>, 2024-09-16.

贸易结构。

(二) 话语构建

在话语层面,日本将国际共识和国内原则建议对接,并利用主场外交优势,引领关于人工智能的国际讨论。2016—2019年间,日本分别在“七国集团”和经合组织框架下的不同会议中,率先将人工智能引入会议议程,为不同国际会议提供人工智能国际原则的草案,引领人工智能国际规则和标准的制定,从而提升其影响力。

1. 日本在国内组织政府和学界专家召开研讨会,为国际人工智能规则制定提供草案文本。在研发方面,2016年10月,日本总务省成立人工智能网络社会推进会议(Conference toward AI Network Society),并于2017年召开“人工智能网络社会推进论坛”(The International Forum toward AI Network Society),邀请日、德、意、法的人工智能专家和来自经合组织的观察员,共同讨论人工智能的社会经济影响。2017年7月,日本人工智能网络社会推进会议发布了《人工智能研发指南草案》(Draft AI R&D Guidelines for International Discussions),以期待“七国集团”和经合组织在这份草案的基础上形成《人工智能研发指南》(AI R&D Guidelines)。《人工智能研发指南草案》包括九项原则:(1) 透明性原则,确保人工智能系统的可解释性和可验证性;(2) 用户协助原则,人工智能系统应为用户提供选择空间;(3) 可控性原则,应确保人工智能系统的可控性;(4) 安全确保原则,确保人工智能系统的稳健性和可靠性;(5) 安全保护原则,人工智能系统不应损害用户和第三方的人身安全;(6) 隐私保护原则,人工智能系统不应侵犯用户和第三方的隐私;(7) 伦理原则,在人工智能系统开发过程中尊重人类的尊严和个人自主权;(8) 可问责性原则,人工智能系统研发应该为用户和利益相关方设置问责机制^①;(9) 合作原则,要求人工智能系统开发者重视不同人工智能系统之间的互联性和互操作性。^② 在应用方面,考虑到人工智能系统在数据使用过程中不断迭代训练的特性,人工智能网络社会推进会议还对人工智能系统的应用过程提出原则性指导。该会议于2018年推出《人工智能应用原则草案》(Draft AI Utilization Principles),提出了十项应用原则建议。其中,“协作原则”“安全保护原则”“安全性原则”“隐私原则”“伦理原则”“透明原则”“问责制原则”与《人工智

^① Ministry of International Affairs and Communications, “Proposal of Discussion toward Formulation of AI R&D Guideline,” https://www.soumu.go.jp/joho_kokusai/g7ict/main_content/ai.pdf, 2024-09-16.

^② The Conference toward AI Network Society, “Draft AI R&D Guidelines for International Discussions,” https://www.soumu.go.jp/main_content/000507517.pdf, 2024-09-16.

能研发指南草案》中的研发原则对应,在应用阶段规定人工智能应该符合的规范。应用原则将日本在2016、2017年强调的“用户协助原则”和“可控性原则”合并为“合理利用原则”,要求人工智能系统用户应在和人工智能的协作中合理分配角色。此外,应用原则还新增了“数据质量原则”和“公平原则”。“数据质量原则”要求数据提供者应注意用于人工智能学习的数据的质量;“公平原则”要求人工智能服务供应商、用户和数据供应商应确保个人不会因为系统的判断受到歧视。^①

日本政府基于国内政策讨论的成果,在国际社会推广相应的原则内容,突出日本对人工智能国际规则的话语贡献。2016年4月29日,“七国集团”信息通信技术部长会议在日本香川县高松市召开。日本身为东道国,在会议上提交了经过其国内讨论的人工智能研发原则草案,希望以此为蓝本,在“七国集团”成员国之间促成关于人工智能开发的国际共识。该草案随即受到了德国、意大利和英国方面的支持,四国在《“七国集团”联合行动》中表示,欢迎各方在人工智能开发方面开展合作研究,讨论人工智能的社会和经济影响,并形成关于人工智能的前瞻性原则。^②同时,从2016年开始,日本就积极参与经合组织关于人工智能的讨论。2016年,在日本总务省的大力支持下,经合组织技术前瞻性会议(The OECD Technology Foresight Forum)选择人工智能作为本次大会主题,帮助政策制定者识别人工智能带来的机会和风险。^③2018年5月,经合组织成立了人工智能专家组(AIGO),汇聚来自政府、业界、社会和学术界的50多名专家为经合组织成员国提供人工智能政策建议。^④其中,来自日本的专家积极向经合组织成员国介绍日本《人工智能指南》草案的内容^⑤,并分享日本国内倡导的“第五代社会”理念和“以人为主”的人工智能社会原则(Social Principle of Human-Centric AI),鼓励经合组织基于日本的伦理价值构建人工智能原则。

2. 通过补充性解释保留对国际共识的差异性理解。这既扩展了人工智能国

① The Conference toward AI Network Society, “Draft AI Utilization Principles,” https://www.soumu.go.jp/main_content/000581310.pdf, 2024-09-16.

② Ministry of International Affairs and Communications, “G7 Opportunities for Collaboration,” https://www.soumu.go.jp/joho_kokusai/g7ict/english/main_content/000416960.pdf, 2024-09-16.

③ OECD Conference Centre, “Summary of the CDEP Technology Foresight Forum Economic and Social Implications of Artificial Intelligence,” [https://one.oecd.org/document/DSTI/CDEP\(2016\)17/en/pdf](https://one.oecd.org/document/DSTI/CDEP(2016)17/en/pdf), 2024-09-16.

④ OECD. AI Policy Observatory, “List of Participants in the OECD Expert Group on AI (AIGO),” <https://oecd.ai/en/list-of-participants-oecd-expert-group-on-ai>, 2024-9-016.

⑤ OECD Conference Centre, “Summary of the CDEP Technology Foresight Forum Economic and Social Implications of Artificial Intelligence,” [https://one.oecd.org/document/DSTI/CDEP\(2016\)17/en/pdf](https://one.oecd.org/document/DSTI/CDEP(2016)17/en/pdf), 2024-9-16.

际规则的内涵,也表现出日本面对欧美发达国家科技外交一定的自主性。比如,当欧美国家和“七国集团”成员国在讨论人工智能与所谓“民主”概念时,日本尽管大致赞同这些国家的基本立场,同时也保留了对所谓“‘民主’地使用人工智能”的独特理解。

在人工智能的国际讨论中,“七国集团”和经合组织讨论的所谓“民主价值”更多是从风险防范层面,强调人工智能技术对用户基本权利、个人隐私和网络暴力的影响。比如,在2016年经合组织举行的技术前瞻性会议中,人工智能对民主价值的挑战被定义为:人工智能可能影响治理机制,使决策过程和责任分配变得模糊。在“二十国集团”和经合组织的人工智能原则中,人工智能应该遵循的民主价值是指“自由、尊严、自主性、隐私保护、不受歧视和公平正义”。在“七国集团”的人工智能原则中,人工智能不应该破坏的民主价值观是指“公平、可问责性、透明、安全和防止网络暴力”等内容。为了对接国际社会关于人工智能与民主价值的讨论,日本在《人工智能研发指南》中强调,应该在尊重民主社会价值的同时,降低人工智能系统可能侵犯个人权利的风险。而在日本国内讨论人工智能的民主价值时,还特别强调技术的公平使用和保护本国市场自主性。日本政府认为,应支持公众不受技术和专业知识背景限制,平等地使用人工智能。以日本人工智能学会为首,当下日本人工智能“民主化”的讨论热点在于:作为非英语国家,如何使基于日语语料库训练的人工智能大模型应用于日本社会经济发展。^① 以及讨论如何采取技术评估手段,减少缺乏专业知识背景的用户在利用生成式大语言模型时被虚假信息误导的风险。^②

(三) 实践共享

日本首先在国内社会搭建起人工智能治理和创新体系,并通过积极改革形成了“软法范式”“敏捷治理”的治理模式。在此基础上,日本将相关实践经验投射到国际层面加以推广,试图吸引其他国家采取与之相似的人工智能政策,从而确立日本人工智能治理路径的影响力。

1. 日本积极改革国内人工智能战略架构,形成以“软法范式”和“敏捷治理”为主要特征的人工智能治理体系。日本将人工智能技术发展提升到国家战略高

① 沢田慶、シーン誠、趙天雨「日本語におけるAIの民主化を目指した事前学習モデルの公開」、『人工知能学会研究会資料 言語・音声理解と対話処理研究会』、第96号、2022年、32頁。

② 白松俊、末永彩羽、吉村友希、伊藤孝行「ChatGPTや大規模言語モデルは合意形成や市民共創にどう活用できるか?」、『人工知能学会研究会資料 言語・音声理解と対話処理研究会』、第97号、2023年、30—37頁;太田真人、ファイサル ハディブトラ「文章分類モデルの不確実性に基づく人間によるデバッグ手法の提案」、『人工知能学会全国大会論文集』、2023年。

度是从2016年开始的。日本内阁通过综合科学技术创新会议,在2016年发布《第五期科学技术基本计划(2016—2020)》,并在2017年发布《人工智能技术战略》和《人工智能研究开发目标与产业化路线图》,形成了初步的人工智能政策架构。其中,总务省、文部科学省和经济产业省主要负责人工智能研发相关的政策,总务省分别于2016、2017和2018年出台了《下一代人工智能促进战略》《人工智能研发指南》和《人工智能利用指南》;农林水产省、厚生劳动省和国土交通省则主要负责在技术应用层面为人工智能政策提供建议和指导。^①到2018年,内阁重新调整了综合创新战略推进会议的架构,设立了若干创新政策强化推进小组。其中,人工智能战略小组专门负责日本人工智能政策的设计和完善,使人工智能战略政策制定更加专业化。人工智能战略小组下设人工智能战略专家会议,分别就“以人为中心”的人工智能原则、人工智能教育和人工智能战略审查等议题进行专门政策研究。专家会议向人工智能战略小组提供政策建议,再由人工智能战略小组汇报给综合创新战略推进会议,并最终体现在政策文本中。通过这样的架构调整,日本政府将产业界和学界的不同利益相关方引入人工智能战略框架,形成了多元化的治理和发展格局。通过2018年以来对人工智能战略组织架构和顶层设计的调整,日本国内基本形成了以“软法范式”和“敏捷治理”为主要特征的人工智能治理体系。

日本之所以倾向于“软法范式”,主要还是考虑到支持人工智能技术的发展和应⽤,避免由于过分严苛的监管扼杀创新。尤其在新一轮人工智能技术发展的早期阶段,政府对于技术发展和应⽤的了解也存在局限性,往往很难制定非常明确的监管规定。因此,“软法范式”在新兴技术发展早期较为有效。除此之外,日本国内严重的老龄少⼦化、劳动力不足、财政负担加剧等结构性问题也迫使政府对人工智能的补偿和替代效⽤寄予厚望。鉴于此,日本政府不得不小心翼翼地平衡人工智能技术发展与负责任地使用之间的关系。2024年2月,日本自民党发布了一份关于“促进负责任人工智能基本法”的概念文件^②,并于4月发布了《人工智能白皮书2024》。^③尽管这两份文件展示出石破茂政府建立人工智能监管法律体系的尝试,但从文件内容来看,日本人工智能立法的根本目的还是在于

① 周生升、秦炎铭:《日本人工智能发展战略与全球价值链能力再提升:基于顶层设计与产业发展的竞争力分析》,《国际关系研究》2020年第1期,第67—90页;刘湘丽、肖红军:《软法范式的人工智能伦理监管:日本制度探析》,《现代日本经济》2023年第4期,第28—44页;张鹏飞:《日本人工智能产业发展和政策研究》,《现代日本经济》2021年第5期,第25—40页。

② AIPT, “Overview of the Basic Law for the Promotion of Responsible AI,” February 2024, <https://note.com/api/v2/attachments/download/006badee3e4d847b3a0c92358b2de63a>, 2025-12-15.

③ 自由民主党デジタル社会推進本部「AI ホワイトペーパー 2024」,2024年4月,https://www.soumu.go.jp/main_content/000944148.pdf, 2025-12-15.

促进创新,通过指导方针,鼓励企业自主进行风险管理。^①正如人工智能战略会议在2025年2月发布的《中期报告》中指出:“根据技术中立原则,日本倾向于依赖现有的行业特定法律,而非制定全面的人工智能专项法规。”^②因此,有日本法律界人士认为,2025年5月生效的《促进人工智能技术研发和应用法案》本质上是一部“基本法”,不包含任何新的法律规定或处罚措施,只是概述了日本政府发展人工智能的战略路线。^③

“敏捷治理”是日本近年来在国内监管和国际治理中展现出的另一特点。所谓“敏捷治理”是指针对新兴技术不断变化的特点,能够不断适应、修改和完善治理方案,从而避免传统治理模式的碎片化和滞后性等问题。在日本国内的多份指导原则和“七国集团”的声明文件中都提到,由于人工智能技术的复杂性和不确定性,治理目标会随着技术和应用的变化而不断变化。单一固定的治理模式往往难以奏效。因此,相关治理原则强调可持续性,针对人工智能技术从研发前到部署后的不同阶段动态调整监管措施。鼓励人工智能企业主动透明并及时披露安全信息,从而避免由于信息不对称和政府监管滞后导致的安全风险。推动政府和企业合作制定相关标准,并开展持续评估和修订,从而灵活应对人工智能技术的快速发展。“软法范式”本身也和“敏捷治理”相辅相成,能够最大限度体现治理主体的包容性,激发自主性,实现以推动创新为目的的监管模式。

2. 日本将国内实践共享到国际层面,通过国际人工智能合作,吸引其他国家对其治理实践的认可,进而确立其国际影响力。在多边机制中,日本积极向国际社会分享国内人工智能治理的实践经验。2024年1月,日本政府发布了《人工智能运营商指南(草案)》,尝试整合更新现有的《人工智能开发指南》《人工智能应用指南》和《实施人工智能原则的治理指南》,形成一体化的人工智能指导性文件。该文件继续以指导原则的形式,对人工智能开发人员、供应商和业务用户提出安全要求,一以贯之体现出日本的“软法范式”特征。2024年4月,《人工智能运营商指南》正式出台,时任日本首相岸田文雄在5月举行的首尔人工智能峰会

① Hiroki Habuka, “New Government Policy Shows Japan Favors a Light Touch for AI Regulation,” CSIS, February 25, 2025, <https://www.csis.org/analysis/new-government-policy-shows-japan-favors-light-touch-ai-regulation>, 2025-12-15.

② AI 戦略会議・AI 制度研究会「中間とりまとめ」, 2025年2月4日, https://www8.cao.go.jp/cstp/ai/interim_report.pdf, 2025-12-15.

③ 内閣府「人工知能関連技術の研究開発及び活用の推進に関する法律(AI法)」, 2025年6月, https://www8.cao.go.jp/cstp/ai/ai_act/ai_act.html, 2025-12-15.

上将该指南作为日本国内呼应广岛进程的重要成果向国际社会分享。^① 2024年5月,人工智能战略会议发布《关于人工智能制度的思考》白皮书,文件提出,日本政府应该思考如何建设成为一个“人工智能友好”的国家,建设“理解和应用人工智能”的社会。^② 从文件内容来看,促进人工智能创新和应用,增强日本人工智能国际影响力仍然是当前日本政府发展人工智能的首要目标,对于建立法律监管体系,文件则采取了一贯慎重的态度。这与“软法范式”的治理思路高度一致。白皮书还强调,对于生成式人工智能带来的虚假信息风险,日本政府应该考虑采取适当措施对知识产权进行法律保护。为此,日本政府正在尝试建立“发起人档案”(Originator Profile, OP)框架作为知识产权保护的重要手段。发起人档案技术是一个可以识别信息发送者身份的第三方系统,旨在帮助读者区分内容真实性并防止虚假和错误信息以及广告欺诈。^③ 目前,发起人档案技术体系以企业自愿而非政府强加为主要的参与方式,进一步体现了人工智能知识产权保护的“软法范式”。在2024年5月召开的经合组织部长级会议上,日本首相岸田文雄重点介绍了日本的这一做法,并呼吁经合组织成员国加入发起人档案技术体系,以期建立知识产权保护的“软法”国际合作框架。^④

在双边和区域合作中,日本尝试以本国的治理特色吸引其他国家的响应。日本在2024年2月建立了人工智能安全研究所(AI Safety Institute),与国内外政府部门合作研究人工智能的安全评估方法。该研究所的设立既体现了日本建立人工智能监管体系的审慎态度,符合“软法范式”的原则,也反映出日本政府根据人工智能技术发展及时调整治理目标的“敏捷治理”特色。更重要的是,通过设立人工智能安全研究所,日本可以与治理路径不同的其他国家展开充分讨论,将其作为向国际社会推广其实践经验的窗口。目前,日本人工智能安全研究所已经与美国国家技术与标准研究院(NIST)展开了深度合作,讨论了日本《人工智能运营商指南》与美国《人工智能风险管理框架》之间的互操作性。两国政府在

① Prime Minister's Office of Japan, "Prime Minister Kishida's Participation in the AI Seoul Summit Leaders' Session," https://japan.kantei.go.jp/101_kishida/diplomatic/202405/21aisummit.html, 2024-09-16.

② AI 戦略チーム「「AI 制度に関する考え方」について」, https://www8.cao.go.jp/cstp/ai/ai_senryaku/9kai/shiryu2-1.pdf, 2024年5月。

③ Originator Profile 技術、「Originator Profile 技術について」, <https://originator-profile.org/ja-JP/overview/>, 2024-09-16.

④ Daisuke Akimoto, "Japan's AI Diplomacy," *The Diplomat*, April 26, 2024, https://my.nda.news/artificial-intelligence-st_38009/news/japans-ai-diplomacy-the-diplomat-sn_5788591/, 2024-11-01.

领导人联合声明中,将该项目作为重点合作案例予以肯定。^①同时,日本人工智能安全研究所还参与了包括首尔人工智能峰会、新加坡科技峰会在内的一系列多边国际会议,以研究机构的身份与采取“硬法范式”的欧盟交换意见。

数据跨境流通和人才培养是全球人工智能治理的基础。在数据跨境流通方面,日本已经借助“可信数据自由流动”框架,与美国、欧盟在双边层面发展高水平数字贸易和数据跨境治理合作,将日本国内的数据治理实践与欧美国家对接,构成双方人工智能治理的基础。在人才培养方面,日本通过双边和区域范围内的国际合作,搭建国际人工智能合作网络,将日本的人才培养计划向外部世界开放。日本政府早在2016年就已经意识到,缺乏专业的人工智能人才是阻碍本国人工智能技术水平发展的重要原因之一,因此,日本政府分别在《综合创新战略》《人工智能战略》《人工智能技术战略实施计划》等文件中,多次强调建立人工智能数字人才培养体系。2021年3月,日本内阁发布《数理、数据科学、人工智能教育课程认证制度》进一步强调建立国内人才培养体系的重要性。2023年12月,在纪念日本与东南亚国家建交50周年峰会上,日本与东盟以联合声明形式确定将在人工智能和网络安全领域展开合作。^②2024年5月,时任日本首相岸田文雄宣布,将在未来五年内与东盟合作培养10万名数字专业人才。^③日本与东盟就数字专业人才培养开展合作,正是利用了其国内的治理经验,将国内的人工智能人才培养体系投射到国际合作中,通过吸引东盟成员国深度融入人工智能市场,进而提升日本对发展中国家的影响力。

三、日本参与人工智能国际规则制定的策略效果

通过梳理日本的政策实践发现,全球人工智能治理的话语权不仅取决于一国本身的技术水平发展和治理能力,也需要在国际合作中采取灵活的外交手段。在方法论上,日本形成了体系化的科技外交策略,使制度参与、话语构建、实践共享三条路径之间形成了有机互动;在实践效果层面,三条策略各自发挥作用,帮助日本树立起“积极行为者”的形象。

① White House, “United States-Japan Joint Leaders’ Statement,” <https://www.whitehouse.gov/briefing-room/statements-releases/2024/04/10/united-states-japan-joint-leaders-statement/>, 2024-09-16.

② NIKKEI Asia, “Japan and ASEAN to Team on AI and Cybersecurity,” <https://asia.nikkei.com/Politics/International-relations/Japan-and-ASEAN-to-team-on-AI-and-cybersecurity>, 2024-09-16.

③ Kyodo News, “Japan to Work with ASEAN to Train 100,000 Digital Professionals,” https://english.kyodonews.net/news/2024/05/eacdc302251c-japan-to-work-with-asean-to-train-100000-digital-professionals.html#google_vignette, 2024-09-16.

（一）日本在规则制定中形成了体系化的参与策略

首先,制度参与为话语建构策略和实践共享策略提供了活动平台。无论是为国际人工智能原则提供内容草案,还是将本国的治理实践共享给国际社会,日本都依赖经合组织、“七国集团”和“二十国集团”等国际制度作为科技外交的操作工具。^①在2016—2023年间,日本分别担任了“二十国集团”、经合组织和“七国集团”峰会的东道国,这赋予其特殊的结构性权力,使日本可以在国际制度之间“牵线搭桥”,通过促进国际制度间合作来彰显自身的影响力。东道国的身份优势和制度参与的外交策略也拓宽了日本的活动空间,使日本对人工智能国际规则的话语建构具有自主性。日本不用一味追随美国或欧盟的立场和做法,而是可以根据自身的考量,对国际规则做出符合日本国家利益的解释,或用自身实践定义国际人工智能治理的模式。

其次,话语构建和实践共享强化了日本制度参与的具体内容。日本主动塑造国际规则话语并分享本土治理经验,使其在参与多边制度时能够有效凸显其治理主张的特点。这不仅塑造了日本作为“积极行为者”的国际形象,更使其避免沦为国际制度中的“搭便车者”(free-rider)。

最后,话语构建策略与实践共享策略相互支撑,提高了日本参与国际规则制定的说服力。一方面,实践共享是对话语构建策略的落实。日本依靠实践共享策略,与国际社会分享自己的治理实践经验,从而实现话语构建过程中日本提出的主张;另一方面,话语建构对日本的实践共享形成指导。日本在国际层面推动话语构建,进而利用国际话语引导国内的治理实践行为,使日本与国际社会形成双向互动。

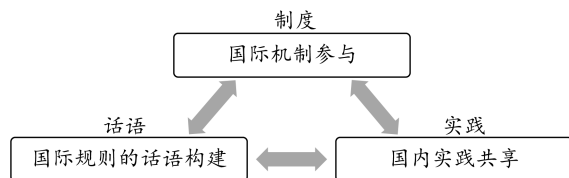


图 日本参与人工智能国际规则制定的体系化策略

图表来源:笔者自制。

^① 张翼燕、许文:《科技外交的工具分类研究》,《全球科技经济瞭望》2017年第6期,第72—76页。

(二) 日本通过体系化的参与策略提升了话语权

日本通过制度参与、话语建构和实践共享三种路径,在人工智能国际规则制定进程中有效提升了国际影响力。制度参与策略成功促进了不同国际制度间的合作,提升了人工智能国际规则体系的整合性;话语构建策略则有效帮助日本将自身的伦理价值观念融入国际人工智能原则,使日本具备了国际话语影响力。

1. 日本的制度参与策略促进不同国际制度间协调合作,推动人工智能国际规则落实。首先,在日本的推动下,经合组织深度参与了“七国集团”和“二十国集团”人工智能原则制定的流程,为“七国集团”和“二十国集团”提供了一系列指导和支持,合作程度不断加深。^① 2019年,“二十国集团”部长会议指出:为了塑造公众对人工智能技术的信任和信心,并充分发挥人工智能技术的潜力,“二十国集团”将坚持“以人民为中心”的人工智能战略,遵循经合组织《人工智能建议书》,制定“二十国集团人工智能原则”(G20 AI Principles),并将遵循这一原则开展人工智能政策研究。^② 2023年4月,“七国集团”数字与技术部长会议在日本群馬县召开。部长会议联合宣言中提到:“经合组织人工智能原则为研发可信赖的人工智能提供了指导,并为部署和开发以人为本的人工智能系统提供了开放和有利的环境。‘七国集团’将继续采纳这些原则,并支持经合组织的相关行动。”^③ 2023年5—10月,“七国集团”峰会启动“广岛人工智能进程”作为协调制定人工智能监管国际规则的平台,并发布《广岛进程组织开发先进人工智能系统的国际指导原则》和《广岛进程组织开发先进人工智能系统的国际行为准则》。^④ 在“七国集团”峰会上,经合组织发表的名为《七国集团关于生成式人工智能的广岛进程》的报告被“七国集团”部长级会议提交给领导人峰会,作为重要政策参考。同时,《七国集团广岛领导人公报》基于经合组织《人工智能建议书》梳理了关于人工智能政策的优先事项和风险。2024年5月,日本政府成立了广岛人工智能进程友

^① 经合组织指出:“多年来,应‘七国集团’主席国邀请,经合组织为‘七国集团’提供了实质性支持,帮助‘七国集团’制定应对共同性全球挑战的多边解决方案。在数字化转型方面,经合组织正在帮助‘七国集团’制定加强人工智能治理的监测和评估机制,包括支持‘广岛人工智能进程’和相关行为准则的实施。”参见 <https://www.oecd.org/en/about/oecd-and-g7.html>, 2026-04-10。

^② Ministry of International Affairs and Communications, “G20 Ministerial Statement on Trade and Digital Economy,” https://www.soumu.go.jp/main_content/000625715.pdf, 2024-09-16.

^③ Ministry of International Affairs and Communications, “Results of G7 Digital and Tech Ministers’ Meeting in Takasaki, Gunma,” https://www.soumu.go.jp/joho_kokusai/g7digital-tech-2023/en/topics/topics_20230430.html, 2024-09-16.

^④ Ministry of International Affairs and Communications, “G7 Hiroshima Leaders’ Communiqué,” https://www.mofa.go.jp/mofaj/gaiko/summit/hiroshima23/documents/pdf/Leaders_Communique_01_en.pdf?v20231006, 2024-09-16.

好小组,优先考虑面向全球发展中国家的人工智能外交工作。截至2025年10月,友好小组已吸引57个国家和地区参与,另有包括联合国开发计划署、世界经济论坛等在内的26个国际组织及企业加入合作伙伴社区。^①

其次,日本推出的“可信数据自由流动”框架为人工智能国际合作提供了数据跨境流动的底层保障,帮助相关国际规则有效落实。“可信数据自由流动”框架包括四大关键合作领域:(1)限制数据本地化;(2)监管合作;(3)政府访问数据;(4)行业数据共享。^②首先,鼓励数据跨境流动并展开数据监管合作,可以推动人工智能发展满足“包容性增长、可持续发展和福祉”原则,推动国际数据技术标准制定,使人工智能应对全球性挑战。其次,在“可信数据自由流动”框架下开展限制数据本地化的相关工作,并通过安全例外条款提高政府对用户数据的可信赖访问,将有助于解决保护用户隐私和知识产权与安全监管之间的矛盾,从而维护“以人为中心”的人工智能原则。再次,保证政府对用户数据的访问权限并促进行业内部数据共享,可以保证人工智能系统的透明性和可解释性。在此基础上,加强政府与行业的合作有助于建立人工智能风险识别和预防机制,完善人工智能治理体系,从而保证人工智能系统的稳健性和安全性。最后,实现政府与企业的信息共享,并在国际层面展开监管合作,有利于为人工智能系统建立问责机制,实现“可问责性”原则。

2. 日本通过话语构建策略,将自身的伦理价值观念融入国际人工智能原则。“以人为中心”(human-centric)的人工智能治理理念是日本参与国际规则制定的引领性概念,它强调在人工智能应用中保护个人权利和人的自主性。2019年,日本总务省成立了“以人为中心的人工智能社会原则会议”,并推出《以人为中心的人工智能社会原则》(Social Principles of Human-centric AI),将“以人为中心”的价值理念映射到人工智能领域的发展中。在东京大学的倡议下,日本人工智能学会设立了伦理委员会,并在2016至2017年间召开若干次学术研讨会,制定了《日本人工智能学会伦理准则》。^③日本总务省于2016年成立了人工智能网络化研究会议,制定伦理规范,预防人工智能的开发利用侵犯人的尊严和自由意志,

① 关于广岛 AI 进程友好小组的信息,参见 <https://www.soumu.go.jp/hiroshimaaiprocess/en/supporters.html#:~:text=The%20Friends%20Group%20was%20launched%20in%20May%202024,toward%20globally%20promoting%20safe%2C%20secure%2C%20and%20trustworthy%20AI.>

② G7, “G7 Roadmap for Cooperation on DFET,” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/986160/Annex_2__Roadmap_for_cooperation_on_Data_Free_Flow_with_Trust.pdf, 2024-09-16.

③ 人工智能学会伦理委员会「人工智能学会 伦理指针」,2017年2月, https://www.ai-gakkai.or.jp/ai-elsi/report/ethical_guidelines, 2025-06-22。

并且保证人的主动地位、对决策结果的可解释性以及消费者的安全。^① 这些讨论结果均被引入日本内阁 2019 年举行的人工智能战略实施会议,并汇总在《以人为中心的人工智能社会原则》文本中。日本对“以人为中心”理念的具体阐释是:“人工智能的使用不应侵犯《联合国宪章》和国际准则保障的基本人权。人工智能应该在社会中开发、利用和实施,以扩大人的能力,追求不同人的不同幸福观念。……在使用人工智能时,人们必须自己判断和决定如何使用人工智能。参与人工智能开发、提供和利用的适当利益相关者应根据问题的性质,对人工智能利用的结果负责。为了避免产生数字鸿沟,让所有人都能从人工智能中获益,无论他们的数字专业知识如何,每个利益相关者都应该在人工智能部署过程中考虑系统的用户友好性。”^②

目前,这一概念已经出现于各国和国际组织的人工智能原则中。第一,日本提倡人工智能必须满足基本人权和人类福祉的要求。而经合组织也在人工智能原则中强调“尊重法律规则、人权和民主价值,包括自由、尊严和自主;隐私和数据保护;非歧视性和公平……”。同样,“七国集团”的人工智能指导原则提倡通过国际合作,“优先发展先进人工智能系统以应对全人类共同的挑战”“实施数据输入措施,保护个人资料和知识产权”。第二,日本主张人工智能使用过程中必须确保人类决策的自主性,“避免过度依赖人工智能导致人类决策被人工智能不良操作”,并主张“人类必须自己判断和决定如何使用人工智能”。而经合组织也在《人工智能建议书》中提倡“确保人类的自主性和决定权”,“七国集团”同样强调通过“建立识别机制,使用户意识到人工智能在决策过程中的存在”。第三,日本主张人工智能的推广应该满足公平性原则,要开展“人工智能扫盲和教育机制,促进人工智能的正确使用”并努力消除“数字鸿沟”,经合组织和“七国集团”同样响应日本的号召,主张“培养人类对人工智能系统的总体理解”。

3. 日本的人工智能治理实践在一定程度上得到国际社会的响应和效仿。在个人隐私保护方面,日本采用“软法范式”对科技公司数据采集做出规定。日本在 2024 年 4 月发布的《人工智能运营商指南》中,要求企业自主制定隐私保护规则、鼓励企业通过安全认证和第三方评估提升可信度。^③ 在 2025 年 5 月通过的

① AIネットワーク化検討会議「AIネットワーク化の影響とリスク」、2016年6月20日、https://www.soumu.go.jp/main_content/000425289.pdf, 2025-06-22。

② Cabinet Office, “Social Principles of Human-centric AI,” <https://www8.cao.go.jp/cstp/stmain/aisocialprinciples.pdf>, 2025-06-22。

③ Prime Minister’s Office of Japan, “Prime Minister Kishida’s Participation in the AI Seoul Summit Leaders’ Session,” https://japan.kantei.go.jp/101_kishida/diplomatic/202405/21aisummit.html, 2024-09-16。

《人工智能相关技术研究开发及应用推进法》中,日本政府也只是要求企业配合进行隐私数据泄露调查,而不对企业设置刑事处罚,避免监管过度遏制创新。^①这一做法已经得到不少经合组织成员国的响应和效仿。目前,法国已经出台了《负责任的面部识别限制政策框架》,德国基于隐私保护的理念推出了《人工智能质量水印》,韩国也出台了《人工智能个人信息保护自查清单》^②,皆是通过政府对企业的柔性引导,协调技术创新与隐私保护之间的潜在冲突。

同时,日本“软法范式”的监管策略为其国内市场营造了良好的科技创新和商业环境,进一步提升了国际市场对日本人工智能产业的关注度。2024 年,美国开放人工智能研究中心(OpenAI)、微软(Microsoft)、甲骨文(Oracle)等科技巨头以及以安德森·霍洛维茨资金(Andreessen Horowitz, a16z)为代表的硅谷资本对在日本布局人工智能产业展现出了极高的热情。OpenAI 首个亚洲办公室落地日本东京,为日企提供日语版 GPT-4^③;微软宣布就云计算和人工智能领域,未来两年内在日本投资 29 亿美元^④;甲骨文宣布就云计算和人工智能基础设施,未来十年内在日本投资 80 亿美元^⑤;安德森·霍洛维茨资金计划在日本开设第二家国际办事处,用于东京的融资活动。^⑥ 硅谷资本的热情也带动了本国资本的积极性,2024 年 4 月,孙正义宣布软银将以人工智能半导体为突破口,预计投资 10 万亿日元。^⑦ 10 月 31 日,日本丰田公司与日本电信电话株式会社(NTT)宣布,

① 内閣府「人工知能関連技術の研究開発及び活用の推進に関する法律」、2025 年 5 月、https://www.cao.go.jp/houan/pdf/217/217anbun_2.pdf, 2025-06-22。

② World Economic Forum, “A Policy Framework for Responsible Limits on Facial Recognition Use Case: Law Enforcement Investigations,” https://www3.weforum.org/docs/WEF_Facial_Recognition_for_Law_Enforcement_Investigations_2022.pdf, 2024-09-16; Lucia Russo and Noah Oder, “How Countries Are Implementing the OECD Principles for Trustworthy AI,” <https://oecd.ai/en/wonk/national-policies-2>, 2024-09-16。

③ OpenAI, “Introducing OpenAI Japan,” <https://openai.com/index/introducing-openai-japan/>, 2026-04-10。

④ Microsoft Source, “Microsoft to Invest US\$ 2.9 Billion in AI and Cloud Infrastructure in Japan while Boosting the Nation’s Skills, Research and Cybersecurity,” <https://news.microsoft.com/apac/2024/04/10/microsoft-to-invest-us2-9-billion-in-ai-and-cloud-infrastructure-in-japan-while-boosting-the-nations-skills-research-and-cybersecurity/>, 2026-04-10。

⑤ Oracle News, “Oracle to Invest More Than \$ 8 Billion in Cloud Computing and AI in Japan,” <https://www.oracle.com/news/announcement/oracle-to-invest-more-than-eight-billion-in-cloud-computing-and-ai-in-japan-2024-04-17/>, 2026-04-10。

⑥ Chain Catcher, “a16z Announced the Establishment of its First Mega Fund in Asia in Japan,” <https://www.chaincatcher.com/en/article/2138596>, 2026-04-10。

⑦ 日经中文网:《软银集团将向“AI 革命”投资 10 万亿日元》,2024 年 5 月,<https://cn.nikkei.com/industry/itelectric-appliance/55579-2024-05-13-09-38-50.html>, 2025-04-23。

将共同投资 5000 亿日元用于研发人工智能基础设施。^① 在蓬勃发展的人工智能市场中,日本涌现出沙卡纳人工智能(Sakana AI)、斯帕蒂克尔((Sparticle)、软银机器人(SoftBank Robotics)等一批高质量的本土人工智能企业,显示出日本在人工智能领域的活力和潜力。

使人工智能服务于“可持续发展目标”也已经成为各国制定科技政策的共同理念。日本在《人工智能战略 2019》中,论述了建设“第五代社会”和可持续发展目标之间的关系。日本在追求“第五代社会”的同时,企图为解决全球问题做出贡献,创造“循环生态经济”,作为可持续发展目标领域实践的典范。面对人口老龄化、少子化、劳动力短缺和基础设施恶化等社会问题,需要以变革的方式协调人工智能技术的社会经济影响,以实现“融合多样性的可持续社会”。^② 日本的这一阐释受到了国际社会的赞同,2019 年,“二十国集团”部长级会议指出:日本提倡的“以人为中心”的“第五代社会”概念,体现了现实与虚拟世界的日益融合,有助于实现“可持续发展目标”。^③ 呼应日本的“可持续发展”理念,经合组织的《人工智能建议书》指出,利益攸关方“应积极负责地管理可信赖的人工智能,为人类和地球带来有益的结果,如增强人类能力和创造力……保护自然环境,从而促进包容性增长、福祉、可持续发展和环境可持续性。”^④“二十国集团”国家也意识到,人工智能技术可以促进世界上大多数人口和能源消耗集中的城市地区的可持续和包容性增长。^⑤ 目前,世界各国都在将“可持续发展目标”作为人工智能发展应当遵循的原则之一。比如,欧盟推出了“数字孪生地球”(Destination Earth, DestinE)计划,作为欧盟委员会绿色协议和数字战略的一部分,在全球范围内开发高精度的地球数字模型,监测、模拟和预测自然现象与人类活动之间的相互作用。^⑥ 德国也相应推出了《人工智能灯塔:环境、气候、自然和资源倡议》。

① “Toyota, NTT to Invest \$ 3.3 Billion on Development of AI Platform,” *The Asahi Shimbun*, November 1, 2024.

② Cabinet Office, “AI Strategy 2019,” <https://www8.cao.go.jp/cstp/ai/aistrategy2019en.pdf>, 2024-09-16.

③ Ministry of International Affairs and Communications, “G20 Ministerial Statement on Trade and Digital Economy,” https://www.soumu.go.jp/main_content/000625715.pdf, 2024-09-16.

④ OECD, “Recommendation of the Council on Artificial Intelligence,” <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#backgroundInformation>, 2024-09-16.

⑤ Ministry of International Affairs and Communications, “G20 Ministerial Statement on Trade and Digital Economy,” https://www.soumu.go.jp/main_content/000625715.pdf, 2024-09-16.

⑥ “Destination Earth,” <https://digital-strategy.ec.europa.eu/en/policies/destination-earth>, 2024-09-16.

结 语

本文梳理了日本借助结构化的科技外交路径参与并塑造国际规则的过程,发现日本并未因相对滞后的技术水平而被边缘化,反而通过制度参与、话语构建与实践共享三条路径,提升了其在全球人工智能治理中的影响力与话语权。日本通过制度参与策略,推动不同国际制度交流合作,并为人工智能国际合作提供配套制度;通过话语构建策略,为国际人工智能规则制定提供草案文本,将本国治理理念融入国际标准;通过实践共享策略,为其“软法范式”和“敏捷治理”的治理特点提升国际影响。日本在人工智能领域科技外交所取得的进展,不仅是因为这三种策略分别发挥了作用,更是因为三种路径之间存在相互促进、相互支撑的有机互动。话语构建策略与实践共享策略相互支撑,加深了日本参与国际规则制定的程度,而制度参与策略则为前两种策略提供了外交活动的空间和平台。

全球人工智能治理是一个多元主体共同参与的互动进程。在这一过程中,技术水平位于“第二梯队”的国家所扮演的角色不容忽视。不同于中国、美国这样的技术“领先者”,以日本为代表的人工智能技术“追赶者”受限于其国内发展条件,通过特殊的策略模式参与国际规则制定,以实现本国利益诉求。本文对日本策略的分析,有助于对全球人工智能治理中不同行为体之间的互动关系形成更全面的理解。与此同时,日本的规则制定参与路径或许对其他人工智能发展条件同样受限的国家具有一定的借鉴意义。

effectiveness of the above pathway design. Looking ahead, China should incorporate social organizations into the overall planning of its overseas security mechanisms and, through diversified collaborative governance, establish a systematic and efficient overseas security provision system supported by the discipline of national security studies.

Systemic Strategy: Japan's Path toward Participating in International AI Governance Rule-Making

Jiang Tianjiao and Zhang Shuyan (169)

Japan's overall AI capability ranks in the world's second tier. Nevertheless, through a structured science and technology diplomacy approach, it has effectively enhanced its international influence in this field by employing three strategies: institutional participation, discourse construction, and practice sharing. At the institutional level, Japan has actively promoted the "alignment" of AI governance principles among the OECD, the G7, and the G20 and has introduced the "Data Free Flow with Trust" (DFFT) framework as a supporting mechanism for AI governance. At the discourse level, Japan prepares draft texts for international AI discussions to highlight its national interests in international rule-making. At the practical level, Japan shares with the international community its governance practices characterized by a "soft law paradigm" and "agile governance," thereby attracting international recognition. Through the interaction among these three approaches, Japan has shaped its international image as an "active actor" in global AI governance, helping to strengthen its international discursive power.