

人工智能治理需要什么样的国际条约？

——以欧洲委员会《人工智能与人权、民主和法治框架公约》为对象^{*}

沈 伟 吴柄萱

摘 要：人工智能(AI)带来的风险亟需法律治理。随着 AI 全球多方治理格局逐步形成，AI 国际法治理取得一定进展，其最新实践体现为欧盟《人工智能法》和欧洲委员会《人工智能与人权、民主和法治框架公约》(简称《AI 公约》)。作为全球首个 AI 国际条约，《AI 公约》试图梳理 AI 治理基本原则、硬法化 AI 治理国际软法，以及强化欧洲 AI 法治的“布鲁塞尔效应”。受限于表现形式、立法技术和缔约参与度，《AI 公约》并不是一份理想的 AI 国际条约。我国应当在发展 AI 治理国际倡议等软法规则的同时，积极参与和推动 AI 国际法制定生成，供给 AI 治理中国方案。

关 键 词：人工智能治理； 国际条约； 国际法； 欧盟《人工智能法》； 欧洲委员会《人工智能与人权、民主和法治框架公约》

作者简介：上海交通大学 凯原法学院 教授 上海 200030

上海交通大学 凯原法学院 博士研究生 上海 200030

中图分类号：D996

文献标识码：A

文 章 编 号：1005-4871(2025)02-0084-27

^{*} 本文系国家社科基金重大项目“美国全球单边经济制裁中涉华制裁案例分析与对策研究”(项目编号:21&ZD208)的阶段性研究成果。

引 言

当今世界人工智能(AI)蓬勃发展,AI技术和各国社会经济发展联系日趋紧密。通过数据整理、信息整合、知识加工和再创造等过程,AI不仅推动了各类科技产品生成,还极大地刺激了数字技术发展。阿尔法围棋(AlphaGo)、自动驾驶汽车、生成式人工智能(Generative AI, Gen AI)等一批新兴技术应用场景涌现,以深度求索(DeepSeek)为代表的新产品得到市场积极反响并产生全球性影响,许多企业开始借助AI重塑市场竞争格局。^①此外,AI还是元宇宙等技术形态的生成逻辑,其深度学习、强化学习能力有强大的创新驱动作用。^②然而,AI技术突破了旧有规范框架,引发了多层次风险。在政治风险层面,AI已具备介入政治安全议题的能力。^③在经济风险层面,AI技术滥用引发信息伪造、数据合规、算法偏见等科技伦理问题。^④在当前多重风险交织的背景下,基于风险的治理成为AI全球治理的基本共识。

全球化视野下AI治理的表现形式多样。各国对AI的治理主要关注生产消费活动,AI技术标准、产业促进政策和安全治理框架都是典型的国内治理对象。国际上的AI治理聚焦技术流通环节,注重产品和技术行业规则、伦理规范,服务于AI风险可控前提下的可持续发展。近年来,各国纷纷制定本国AI产业战略部署。国际利益攸关者也出台多份AI治理原则、全球倡议、国际标准,如电气与电子工程师协会(Institute of Electrical and Electronics Engineers)的《人工智能设计伦理准则》、美国计算机协会(US Association for Computing Machinery)的《算法透明度和责任原则》等,^⑤但这些都不构成严格意义上的国际法。

2024年,继欧盟推出全球首部AI监管法规《人工智能法》(简称《AI法》)后,^⑥欧洲委员会(Council of Europe)公布了全球首个具有法律约束力的AI国际条约

① Sinéad Carew et al., “DeepSeek Sparks AI Stock Selloff; Nvidia Posts Record Market-cap Loss”, *Reuters*, 2025 - 01 - 28, <https://www.reuters.com/technology/chinas-deepseek-sets-off-ai-market-rout-2025-01-27/>, 访问日期:2025 - 02 - 27。

② 沈伟:《元宇宙金融:全新场景与风险监管》,载《上海师范大学学报(哲学社会科学版)》,2025年第1期,第71-84页,这里第71-72页。

③ 封帅、薛世锟:《人工智能与政治安全形态演变:一项系统研究的尝试》,载《国际安全研究》,2025年第1期,第54-72页,这里第60-62页。

④ 赵精武:《人工智能科技伦理审查制度的体系化建构》,载《当代法学》,2025年第1期,第84-96页,这里第84-87页。

⑤ 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》,载《国际经济评论》,2024年第3期,第52-69页,这里第53、62页。

⑥ 《全球首部〈人工智能法案〉正式生效》,新华网,2024 - 08 - 09, <http://www.xinhuanet.com/tech/20240809/61e8802ebbf34d85a1d350fb539317eb/c.html>, 访问日期:2025 - 02 - 01。

《人工智能与人权、民主和法治框架公约》(简称《AI公约》),开辟了AI治理新纪元。^① 本文主要分析《AI公约》出台的背景、内容和推动AI治理的可能效果,辅以对全球需要怎样AI国际条约的辨明,为我国积极推动制定理想的AI国际条约、深度参与AI全球治理提供思路和借鉴。

一、《AI公约》出台的背景

(一)AI全球治理图景

AI带来的风险越发受到关注,AI风险在国际安全层面主要表现为武器、能源等特定领域的风险。例如,AI技术能够应用于杀手机器人、军用无人机等致命性自主武器,对现有联合国《特定常规武器公约》等国际条约的适用构成挑战。^② 又如,AI可能导致能源控制权虚拟化、去人力化,阻扰乃至爬取一国能源关键基础设施的管网数据,直接危及国家能源安全。^③ 在社会民事和经济领域,AI风险体现为算法偏见及数据侵权问题。AI在金融信贷行业错误识别信用不良人士,^④ 自动识别算法、网络爬虫等技术在获取数据时,存在非法收集、买卖数据等情况,这可能会造成个人隐私泄露。^⑤ AI作品也干扰了产权保护体系。AI换脸技术(AI face swap)的使用可能侵害个人信息安全,未经同意的AI伪造手段则可能侵害权利人肖像权、名誉权。^⑥ AI技术在伦理和法律上产生新的争议,^⑦ 是全人类共同面对的风险。因此,AI治理是各国的共同需求。各国还积极探索制定AI国际监管规则,试图通过国际法调整AI治理的权利义务关系。

目前,AI全球治理呈现技术大国主导、国际合作协同的多方治理格局。技术大国层面,美国于2021年颁布《2020年国家人工智能计划法案》(National Artificial Intelligence Initiative Act of 2020, NAIIA),在白宫科技政策办公室(Office of

① 姚旭、张傲:《美欧英签署全球首个人工智能法律公约,AI监管“全球化”进程加快》,复旦发展研究院,2024-09-16, https://fddi.fudan.edu.cn/_t2515/90/e8/c18965a692456/page.htm,访问日期:2025-01-29。

② 杨成铭、魏庆:《人工智能时代致命性自主武器的国际法规制》,载《政法论坛》,2020年第4期,第133-143页,这里第133-135页。

③ 朱荣生、冯紫雯、陈琪、陈劲:《人工智能的国际安全挑战及其治理》,载《中国科技论坛》,2023年第3期,第160-167页,这里第161-165页。

④ 陈兵:《人工智能应用的科技伦理与法治化建设》,载《人民论坛》,2024年第12期,第66-70页,这里第68页。

⑤ 解正山:《数据泄露损害问题研究》,载《清华法学》,2020年第4期,第140-158页,这里第141页。

⑥ 林爱琚、林倩敏:《AI换脸的技术风险与多元规制》,载《未来传播》,2023年第1期,第60-69页,这里第61-64页。

⑦ 郑戈:《人工智能伦理的机制设计》,载《中国法律评论》,2025年第1期,第48-62页,这里第48-49页。

Science and Technology Policy, OSTP)下设立国家 AI 计划办公室,以保障在 AI 领域的国际领导力。随后,OSTP 发布《人工智能权利法案蓝图》政策计划,^①美国国家标准和技术研究院(National Institute of Standards and Technology)发布《人工智能风险管理框架 1.0》,进一步完善 AI 法律监管框架。^②自 2017 年发布《新一代人工智能伦理规范》(以下简称《AI 伦理规范》)以来,我国多次出台相关法律与政策。我国于 2019 年发布《人工智能安全标准化白皮书(2019 版)》,2021 年颁布《中华人民共和国个人信息保护法》(以下简称《个保法》)和《人工智能伦理安全风险防范指引》,2023 年公布《人工智能伦理治理标准化指南》,2024 年出台《国家人工智能产业综合标准化体系建设指南(2024 版)》和《人工智能安全治理框架》。在国际合作层面,2019 年,经济合作与发展组织(Organization for Economic Cooperation and Development, OECD)发布全球首个政府间 AI 行业标准《人工智能发展建议》,提出了“负责任地管理可信 AI”的 AI 全球治理原则。^③此后,世界卫生组织(World Health Organization, WHO)发布《世界卫生组织卫生健康领域人工智能伦理与治理指南》,联合国教科文组织(United Nations Educational, Scientific and Cultural Organization, UNESCO)也发布《人工智能伦理问题建议》(以下简称《AI 伦理建议》)等文件,^④共同推动 AI 全球多方法律治理的进程。

(二)AI 治理模式分歧

主权国家对 AI 实施法律规制虽然保证了 AI 监管规则的法律约束力,但也引发了 AI 治理国别分化现象,造成了 AI 治理模式的差异。^⑤AI 治理国内模式的差异不仅无益于国际社会整体性应对 AI 风险,而且阻碍了国际共识的形成,使得 AI 国际法治陷入困局。AI 治理国别差异主要体现为治理思路、监管标准和规则范式的分化。

一是 AI 治理思路各不相同。欧盟素来重视对私营企业的监管,在 AI 监管上秉持以基本人权为本、重视科技公益属性的治理思路。《AI 法》出台前,欧盟已通

① Katherine B O'Keefe, "White House Publishes Blueprint for AI Bill of Rights", 2022-11-10, <https://www.morganlewis.com/blogs/sourcingatmorganlewis/2022/11/white-house-publishes-blueprint-for-ai-bill-of-rights>, 访问日期:2025-01-29.

② NIST, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile", NIST, 2024-07-26, p. 1, <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>, 访问日期:2025-01-29.

③ OECD, "Recommendation of the Council on Artificial Intelligence", OECD Legal Instruments, 2019-05-22, p. 1, <https://oecd.ai/en/assets/files/OECD-LEGAL-0449-en.pdf>, 访问日期:2025-01-29.

④ 李洋:《联合国教科文组织正式推出〈人工智能伦理问题建议书〉》,中国新闻网,2021-11-26, <https://www.chinanews.com.cn/gj/2021/11-26/9616424.shtml>, 访问日期:2025-01-29.

⑤ 张新宝、魏艳伟:《我国人工智能立法基本问题研究》,载《法制与社会发展》,2024 年第 6 期,第 5-21 页,这里第 6-7 页。

过2016年《通用数据保护条例》(General Data Protection Regulation, GDPR)建立了数据保护监管框架,以应对AI技术引发的数据保护问题和个人隐私风险。《AI法》是欧盟为AI监管制定的专门法律规则,对在欧洲从事商业经营的AI技术公司进行针对性监管,要求私营主体履行风险报告等多项合规义务。我国在AI监管方面主要侧重特定敏感领域,采用“小切口”监管策略,凸显“敏捷治理”理念。^①2017年《AI伦理规范》出台后,仅四年时间,有关部门就发布了相关司法解释,法律化人脸识别技术和敏感人脸信息监管规则。^②2021年颁布的《个保法》要求有关部门制定AI个人信息保护规范,为相关规定出台提供了制度窗口。^③除上述围绕敏感个人信息的规范外,我国对AI治理总体呈开放态度。美国、英国则采取积极治理思路,通过AI治理驱动技术创新、创造商业价值。NAIIA在保证美国领导全球AI技术发展、优先在政府部门内推广AI技术利用等方面作出规定,对AI同社会经济活动的进一步结合持鼓励态度。英国《人工智能监管政策》等政策文件专注AI数据和个人信息违法行为,并不抑制商业活动,同时倡导适度监管下的科技创新。AI治理思路的区别决定了监管侧重、政策重点的不同。

二是AI监管标准有所差别。欧盟《AI法》旨在建立欧洲乃至全球最高AI监管标准,从监管AI商业活动。一方面,根据《欧盟运行条约》第288条第2款,欧盟法律对各成员国具有普遍适用性和整体拘束力,^④各成员国原则上都应遵守。^⑤欧盟成员国需要按照《AI法》监管标准在本国境内对AI产品与技术进行管控。《AI法》作为超国家法发挥了在欧盟内促进AI研发应用、防范AI滥用风险的上位法作用,^⑥增进了欧盟AI治理标准和治理水平。另一方面,《AI法》也确立了高标准罚则制度。《AI法》规定对违法企业最高可处3500万欧元或该年全球营业额7%的罚款,^⑦形成制度威慑。相较之下,中美都未采取如此严格的监管措施。根

① 《人工智能全球监管迎来重磅一周》,新华网,2024-05-20, https://www.news.cn/2024-05/20/c_1212364104.htm, 访问日期:2025-02-11。

② 中华人民共和国最高人民法院:《最高人民法院关于审理使用人脸识别技术处理个人信息相关民事案件适用法律若干问题的规定》,中华人民共和国最高人民法院公报,2021-07-27, <http://gongbao.court.gov.cn/Details/118ff4e615bc74154664ceaf3bf39.html>, 访问日期:2025-01-30。

③ 中华人民共和国中央人民政府:《中华人民共和国个人信息保护法》,中国政府网,2021-08-20, https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm, 访问日期:2025-01-30。

④ EU, “Consolidated Version of the Treaty on the Functioning of the European Union”, EUR-Lex, 2012-10-26, pp. 171-172, <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:12012E/TXT;en:PDF>, 访问日期:2025-02-05。

⑤ [德]沃尔夫冈·多伊普勒:《〈欧盟人工智能法案〉的背景、主要内容与评价——兼论该法案对劳动法的影响》,载《环球法律评论》,2024年第3期,第5-27页,这里第9页。

⑥ 张新宝、魏艳伟:《我国人工智能立法基本问题研究》,第7页。

⑦ Matt Kosinski:《什么是〈欧盟AI法案〉?》,IBM,2024-04-08, <https://www.ibm.com/cn-zh/topics/eu-ai-act>, 访问日期:2025-02-11。

据我国现有 AI 监管制度,除非违反数据安全或个人信息保护规则,开发利用 AI 技术的科技企业不会面临高额罚款。此外,《AI 伦理规范》等伦理规则、行业规范本身也没有规定罚则,这为企业探索 AI 技术孵化模式留有空间。美国联邦法律亦未规定类似《AI 法》的高额罚款。在各州法律中,德克萨斯州众议院于 2024 年 12 月 23 日提交审议了《负责任的人工智能治理法案》(Texas Responsible AI Governance Act),将违法引起的处罚上限规定为 10 万美元。^① 由此看来,欧洲 AI 监管标准的严厉程度的确在全球居于首位。

三是 AI 规则范式存在差异。欧盟依赖专门成文法律实施 AI 治理,《AI 法》本身吸收了利益相关各方的考量。在《AI 法》草案内容中,欧洲标准化委员会、欧洲电工标准化委员会及欧盟各成员国技术标准都被引为法案的参考标准。审议《AI 法》议案时,欧盟理事会、欧洲议会就是否引入“基础模型”理念展开辩论。法、德、意等国还在三方对话机制达成一致后提出反对意见,要求不监管 AI 基础模型、保护本国 AI 孵化企业。^② 我国尚未制定 AI 单行法,而是以 AI 相关法律、法规、规章的方式展开 AI 治理,形成了行为治理、法益治理所需的法律框架。《中华人民共和国网络安全法》(以下简称《网络安全法》)、《数据安全法》和《个保法》适用于 AI 引发的互联网、数据和个人信息安全问题。我国还发布了首部 AI 监管法规《生成式人工智能服务管理暂行办法》(以下简称《暂行办法》),并配套出台了《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》等行政规章。^③ 直至近期,我国才将 AI 监管专门法《人工智能法》列入立法工作计划。^④ 美国侧重于通过行政命令等规范类型完善 AI 治理。例如,美国逐步推广 AI 技术的合法开发与利用,先后于 2019 年出台《维护美国 AI 领导地位的行政命令》,2020 年发布《促进政府内可信 AI 技术使用行政命令》,2023 年推出《关于安全、可靠、可信地开发和使用 AI 的行政命令》。2021 年,NAIIA 在形式上表现为法案,但实质上是 2019 年、2020 年两则行政命令的整合。^⑤ 由此可见,各国 AI 治理的法律表现形式不同,归根结底是在权利义务规制、责任归属等问题上的立场不同。

① 《美国德州提出〈人工智能治理法案〉,每次违规最高罚 10 万美元》,安全内参,2025-01-02, <https://www.secrss.com/articles/74225>, 访问日期:2025-02-11。

② [德]沃尔夫冈·多伊普勒:《〈欧盟人工智能法案〉的背景、主要内容与评价——兼论该法案对劳动法的影响》,第 8 页。

③ 谭寅亮:《全球 AI 监管新纪元——欧盟 AI 法案的影响与挑战》,CEIBS,2024-11-18, <https://cn.ceibs.edu/media/press-clippings/faculty/25963>, 访问日期:2025-01-06。

④ 张凌寒:《人工智能法律治理的路径拓展》,载《中国社会科学》,2025 年第 1 期,第 91-110 页,这里第 99 页。

⑤ 刘晓春、夏杰:《美国人工智能立法态势介评》,载《中国对外贸易》,2023 年第 10 期,第 38-41 页,这里第 38-39 页。

(三) AI 国际法治理的进展

在经济全球化和技术数字化、信息化进程推动下, AI 国际法治理逐渐受到各国关注。^① AI 治理模式的分歧凸显了国际法治理的必要性, 推动跨国监管规则不断涌现。AI 国际法的治理对象是 AI 技术引发的全球风险, 治理目标在于识别、应对技术风险, 使 AI 总体处于安全可控状态。行业驱动型道德规范(industry-driven ethics)曾被视为 AI 风险治理的主要或唯一监管模式。近年来, AI 技术流通改变了国际产业格局, 推动国际经贸规则、数字经济协定作出相应调整, 由此形成了一些直接和间接的规制规则。^② 这些国际规则具有一定普遍性和灵活性, 能够保障国际法随 AI 技术的发展得到“进化解释”(evolutionary interpretation), 丰富和推进 AI 全球治理的内容和进程。^③ 当前 AI 国际法治理有如下特征。

一是非国家主体参与驱动。对 AI 挑战的国际法回应主要来自国际组织。面对 AI 技术在一国领域内的传播与应用, 主权国家可以通过法律强制力保证 AI 治理规则的有效性。然而, 由于不存在全球政府, 以《AI 法》为代表的超国家机构的区域国际法或“超国家法”(supranational law)^④填补了 AI 国际法治理的制度空白。欧盟是政府间区域性国际组织, 具有独立法律人格及超国家性。欧盟法是“调整欧盟各国对内和对外关系的国际法和国内法规范的总称”,^⑤既是区域国际法, 也对各成员国具有直接和最高效力。欧盟《AI 法》对各成员国起到约束和指导作用。^⑥

二是以软法治理为主。目前 AI 相关国际法律尚不完善。根据《国际法院规约》(Statute of the International Court of Justice)第 38 条第 1 款的规定, 国际条约和国际习惯是国际法的主要渊源, 是国际硬法的构成部分。国际条约是明确成文的国家间条约、公约、协定, 国际习惯的成立要满足“国家实践+法律确信”要求。按此标准, AI 领域国际硬法明显缺位, 软法在 AI 治理中扮演着重要角色。在 AI 领域, 截至 2021 年, 全球共产生 634 个软法文件, 仅 2017—2019 年就有超过 500

① 沈伟、赵尔雅:《数字经济背景下的人工智能国际法规制》,载《上海财经大学学报》,2022 年第 5 期,第 123—137 页,这里第 125—126 页。

② Mark Findlay/Jolyon Ford, “Regulatory Insights on Artificial Intelligence: Research for Policy”, in Mark Findlay/Jolyon Ford/Josephine Seah/Dilan Thampapillai (eds.), *Regulatory Insights on Artificial Intelligence*, Cheltenham: Edward Elgar Publishing Limited, 2022, pp. 1—304, here p. 1.

③ Shin-yi Peng/Ching-Fu Lin/Thomas Streinz (eds.), *Artificial Intelligence and International Economic Law*, Cambridge, UK: Cambridge University Press, 2021, p. 2.

④ 李斌主编:《现代国际法学》,北京:科学出版社,2004 年版,第 2 页。

⑤ 张晓东:《论欧盟法的性质及其对现代国际法的贡献》,载《欧洲研究》,2010 年第 1 期,第 67—80 页,这里第 68 页。

⑥ 凌慧明:《欧洲联盟法与一般国际法之比较》,载《法学杂志》,1997 年第 2 期,第 32—33 页,这里第 32 页。

个文件。2021年，UNESCO发布的《人工智能伦理问题建议书》(Recommendation on Ethics of AI)获193个会员国协商一致通过，成为AI全球治理准则之一。^①也有学者将上述软法规规范称为“准监管”(para-regulation)，这类规范虽不具有强制执行力，但具备主导AI议题、设定监管路径依赖、塑造AI治理可追溯性的能力。在AI领域，真正意义上的监管恰始于准监管。^②

三是监管目标趋同化。在AI国际法治理中，监管目标和规制原则有趋同倾向。在全球现存的84个AI伦理规范中，大部分以公平、正义、隐私、无害、自由等11个价值原则为共识基础。^③OECD、联合国、WHO也在AI伦理方面提出了人类社会普遍关切的共识性原则，折射出AI软法治理的共同偏好。AI可能在技术层面引发数据训练失真、算法技术非中立、算力资源配置不均等系统性风险，在非技术层面引发数据流动中的泄露、算法黑箱和算法歧视、治理规则缺位等非系统性风险。^④AI国际法的监管目标也较为一致，均聚焦安全、人权等风险治理面向。

四是治理需求差异化。尽管目标趋同，但各国AI技术发展水平的差异导致各自风险治理的侧重点不同，进而引发了AI国际法治理需求的差异。美国在AI国际倡议中强调市场的自我调节和技术创新作用，主张依托倡议性文件、分散监管模式实现AI全球治理，实则是希望借助宽松的AI国际治理体系，利用市场机制扩大资源集聚效应，掌握AI技术主动权。欧盟AI法沿袭了在数据跨境流动、碳排放和碳交易等新兴领域已经展现的“以市场换监管”思路，力图制定有全球影响力的区域国际法，避免在市场环境碎片化背景下各国自行其是。我国则在监管中关注发达国家和发展中国家的立场差异，倾向于推动AI国际法向服务发展中国家利益的方向转变，为AI技术后发国家争取制度保护，实现全球监管的协同发展。^⑤

五是以间接规制为主。国际法通常不直接供给AI监管规则，而是围绕AI产业技术进行间接规制。《全面与进步跨太平洋伙伴关系协定》(Comprehensive and Progressive Agreement for Trans-Pacific Partnership, CPTPP)、《区域全面经济伙伴关系协定》(Regional Comprehensive Economic Partnership, RCEP)、《数字经济

① 曾雄、梁正、张辉：《人工智能软法治理的优化进路：由软法先行到软法与硬法协同》，载《电子政务》，2024年第6期，第96-107页，这里第96页。

② José-Miguel Bello y Villarino, "Global Standard-Setting for Artificial Intelligence: Para-regulating International Law for AI?", *The Australian Year Book of International Law*, Vol. 41, 2023, pp. 157-181, here p. 161.

③ Anna Jobin et al., "The Global Landscape of AI Ethics Guidelines", *Nature Machine Intelligence*, Vol. 1, 2019, pp. 389-399, here p. 395.

④ 邓悦、许弘楷、王诗菲：《人工智能风险治理：模式、工具与策略》，载《改革》，2024年第1期，第144-158页，这里第145-146页。

⑤ 谭寅亮：《全球AI监管新纪元——欧盟AI法案的影响与挑战》。

伙伴关系协定》(Digital Economy Partnership Agreement, DEPA)等多边协定,以及《新加坡-澳大利亚数字经济协定》(Singapore-Australia Digital Economy Agreement, SADEA)等双边协定文本均有与 AI 有关的条款,如 CPTPP 以第 14 章、RCEP 以第 12 章为电子商务专章,DEPA 第 8 条和 SADEA 第 31 条是 AI 共识性专条。这些国际协定中存在跨境数据流通、数据保护、数据访问和政府公开、知识产权与算法、金融科技、密钥和网络安全等常见 AI 应用场景的规定,形塑了数字经济时代的国际经济法,也是 AI 全球治理议程的构成部分。^① 国际法作为国内法和全球共识的“摆渡人”居中发挥“结构性耦合”的作用,保障 AI 全球“善治”。^②

二、《AI 公约》的主要内容和目标

欧洲在 AI 治理方面具有先进经验,是全球 AI 立法最活跃的地区之一。自 2018 年欧盟提出基于风险的 AI 管理方法^③以来,风险评估理念逐步深入欧盟 AI 法律与政策。《AI 公约》是全球 AI 法律治理的重要成果,既呼应了具有多层架构特征的欧盟运行体系,又深化了 AI 国际法治理。^④ 欧洲委员会认为《AI 公约》有望解决将人权制度应用于 AI 领域的解释性难题,在法律上引入 AI 治理基本原则,确定适用于经贸领域的 AI 国际人权法规范。^⑤ 分析《AI 公约》的内容和目标,有助于更加直观、清晰地认识其制度贡献。

(一)《AI 公约》的主要内容

就其内容而言,《AI 公约》在汇总欧洲内部共识、初步统一 AI 治理规则、实现欧洲 AI 治理模式国际化等环节取得了一定进展。

首先,《AI 公约》汇集了欧洲各国在 AI 领域的治理共识。随着全球 AI 监管竞争提速,各国将国内法协调提上日程,推动国际法治理框架加速成型,促使 AI 国际法文本陆续涌现。^⑥ 多方知识汇总、多边协商谈判、协调一致的规则制定模式是 AI

① 沈伟、赵尔雅:《数字经济背景下的人工智能国际法规制》,第 128 页;沈伟、赵尔雅:《数字经济和数字博弈双重背景下人工智能的国际法治理》,载《武大国际法评论》,2023 年第 4 期,第 81-105 页,这里第 90 页。

② 曾雄、梁正、张辉:《人工智能软法治理的优化进路:由软法先行到软法与硬法协同》,第 96、106 页。

③ [德]沃尔夫冈·多伊普勒:《〈欧盟人工智能法案〉的背景、主要内容与评价——兼论该法案对劳动法的影响》,第 7-9 页。

④ 李益斌、李浩洋:《欧美中人工智能监管规范比较研究》,载《当代世界与社会主义》,2024 年第 5 期,第 161-169 页,这里第 164-165 页。

⑤ Hannah van Kolschooten/Carmel Shachar, “The Council of Europe’s AI Convention (2023-2024): Promises and Pitfalls for Health Protection”, *Health Policy*, Vol. 138, No. 104935, 2023, pp. 1-5, here p. 2.

⑥ Nathalie Alisa Smuha, “Biden, Bletchley, and the Emerging International Law of AI”, *Verfassungsblog*, 2023-11-15, <https://verfassungsblog.de/biden-bletchley-and-the-emerging-international-law-of-ai/>, 访问日期:2025-02-27。

国际法形成的主要路径。英国曾于 2023 年召开全球首届 AI 安全峰会，为规则制定寻求多方磋商。OECD、七国集团(G7)、UNESCO 的 AI 治理框架在确立过程中均体现了这一模式。^① 如今的《AI 公约》也反映了该模式的合理性。欧洲委员会自 2019 年开始研究出台 AI 国际条约的可能性，于 2022 年 1 月主持成立了特设政府间 AI 委员会(Ad hoc Committee on Artificial Intelligence, CAHAI)，主导了为期超过两年的缔约谈判进程，期间欧洲委员会全体成员国、观察员国都参与了公约起草和谈判制定。欧盟委员会、欧洲数据机构以及私营部门的代表也参加了《AI 公约》的制定工作。^② 在序言部分，《AI 公约》至少确立了人权、民主和法治三个宏观价值和加强进步创新、促进个体平等、保障个人隐私权、开展数字扫盲等具体共识。在正文条款中，《AI 公约》第一章“总则”、第二章“一般义务”是条约内容的概括总结；第三章“与 AI 系统生命周期内活动有关的原则”列举了 AI 系统应当保障的各项法益；第四章“补救措施”中提出了利益受侵害主体的救济路径；第五章“评估与减轻风险及不利影响”中要求建立风险预测和监控框架；第六章“公约的实施”阐述了 AI 系统开发利用中的特别注意事项。上述章节基本整合了欧洲各国对 AI 治理的观点和态度，形成了一种底线共识，发挥了增进 AI 国际法治的正面作用。

其次，《AI 公约》通过国际条约尝试确立 AI 全球治理统一规则。欧洲数据保护监管机构表示，《AI 公约》具有指导未来 AI 法律出台的重要作用，是 AI 全球治理的一环。^③ CAHAI 指出，《AI 公约》旨在创建“共同的法律框架，其中包含 AI 在人权、民主和法治发展、设计和实施方面的某些最低标准”。作为铺垫，《AI 法》已给出所谓“某些最低标准”的具体内涵。其第 6—9 条将 AI 系统引发的风险分为不可接受风险、高风险、有限风险和最低风险四类，^④ 高风险 AI 系统的开发者和用户应当履行合规性评估、数据治理和透明性报告等义务。^⑤ 第 21—26 条还规定了高风险

① IAPP, “Global AI Law and Policy Tracker”, 2024 - 11, p. 2, https://iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf, 访问日期:2025-02-07。

② 安峥:《美英欧盟等签署首个具法律约束力的国际 AI 公约,为何有欧洲专家“失望”?》,上观新闻,2024-09-06, <https://www.shobserver.com/wx/detail.do?id=793237>, 访问日期:2025-01-30。

③ 吴沈括、许林逸:《欧洲委员会〈人工智能公约〉要旨及其启示》,载《中国审判》,2023 年 13 期,第 68-71 页,这里第 69-71 页。

④ EU, “Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)”, EUR-Lex, 2024-07-12, pp. 46-57, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689, 访问日期:2025-02-05。

⑤ 江海洋、魏书敏:《基于风险的通用人工智能监管——从欧盟〈人工智能法案〉视角展开》,载《科技与法律》,2024 年第 2 期,第 88-97 页,这里第 90-91 页。

AI相关主体义务群。^①《AI公约》在此基础上进一步引入了通用最低保护标准条款。例如,《AI公约》第21条载明公约不能限制或减损缔约方国内法或其他国际条约保障的权利,第22条倡导缔约方采取超出公约的保护措施。^②通过这些处理,《AI公约》事实上将《AI法》的监督标准固定为底线标准,鼓励各缔约方采取更高标准。《AI公约》第6条强调,一切AI系统的使用都须遵守各国国内法律、履行公约项下义务,以建立一套能够产生外部性的法律治理体系。对于公约的欧洲缔约国而言,此处的“各国国内法律”即《AI法》,欧洲的AI区域法和国际法得以经法律解释形成有机整体。《AI公约》还对AI全生命周期风险评估提出明确要求。其第16条提出各缔约方须采取措施识别、评估、预防、缓解AI系统的风险和潜在影响,通过听证会等形式吸纳公众意见,以增强AI治理的透明性与社会认可度。^③该条所涉AI系统风险指向《AI法》规定的风险类型。此外,在条款表述上明确采取“权利”“义务”规定是《AI公约》在形式上的突破之处,各国参与治理也体现了《AI公约》统一AI全球治理口径的基本态度。这些特征都使《AI公约》具备制度积极性。

最后,《AI公约》弥补了各国AI治理的不足。欧洲委员会是由46个成员国、5个正式观察员国和1个准观察员国组成的国际组织。其46个成员国包括全体27个欧盟成员国和英国、乌克兰等19个非欧盟成员国,5个观察员国分别是加拿大、梵蒂冈、日本、墨西哥、美国,1个准观察员国指以色列。^④欧洲委员会在欧盟框架

① 具体而言,主管机关须履行信息保密义务,授权代表须尽到监管配合义务,进口商、分销商、提供者和部署者按照各自商业行为履行对应义务。See EU, “Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)”, pp. 64 – 69.

② Council of Europe, “Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law”, 2024 – 09 – 05, p. 7, <https://rm.coe.int/1680afae3c>, 访问日期:2025 – 01 – 29。

③ CAHAI, “Feasibility Study: A Legal Framework on AI Design, Development and Application Based on Council of Europe Standards on Human Rights, Democracy and the Rule of Law”, Council of Europe, 2020 – 12 – 17, p. 2, <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>, 访问日期:2025 – 01 – 29; Council of Europe, “Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law”, p. 6。

④ 关于观察员国和准观察员国的区别,根据欧洲委员会官网的介绍,除正式成员国外,委员会名下观察员国分为正式观察员国(Observer State)和议会大会观察员国(Observer to the Parliamentary Assembly)。前者的权限来自欧洲委员会部长委员会(Committee of Ministers)的直接授权,是全职观察员国;后者的权限基于欧洲委员会《议会大会议事规则》第62条有关“特别访客”的规定获得,其观察权仅限于议会大会程序。See Council of Europe, “Israel // Observer to the Parliamentary Assembly”, Council of Europe, 2025, <https://www.coe.int/en/web/portal/israel>, 访问日期:2025 – 03 – 27; Parliamentary Assembly of Council of Europe, “Rules of Procedure of the Assembly”, Council of Europe, 2023 – 01, <https://assembly.coe.int/nw/xml/RoP/RoP-XML2HTML-en.asp>, 访问日期:2025 – 03 – 27。

外发挥议事、磋商、谈判和国际协调功能。相较《AI 法》展现的区域国际法模式，《AI 公约》在酝酿过程中充分吸收了欧洲国家及非欧洲成员国家的意见建议，扩展了造法思路的来源。在制定谈判过程中，澳大利亚、阿根廷、哥斯达黎加、秘鲁、乌拉圭等 5 个域外国家也加入了这一进程，使《AI 公约》的条款更具有广泛性。《AI 公约》出台至今已获得美国、英国、欧盟、以色列、挪威、冰岛、格鲁吉亚、黑山、摩尔多瓦、安道尔和圣马力诺等 11 个缔约方签署。该公约的生效须获得至少 5 个签署国（其中包括至少 3 个欧洲委员会成员国）批准，且将于 3 个月期满后的下个月首日开始施行。^① 目前批准程序还在进行中。由于《AI 公约》仍对其他国家开放缔约，其有望持续吸纳新缔约方，从而增强其包容性、拓展其适用性。

（二）《AI 公约》的造法目标

一是体系化 AI 国际法律治理。《AI 公约》在界定基础概念、分配风险规制义务、制定监管原则等方面有所创新，延伸和体系化了欧盟《AI 法》。《AI 公约》聚焦 AI 相关基本权利。其序言部分将规制对象限缩于《AI 法》第 9 条涉及的“AI 系统全生命周期内活动”，将 AI 风险具体表述为造成数字歧视、侵害隐私和个人自由的 AI 系统滥用，将基于产业活动的监管调整为旨在保障人权的监管。^② 为兼顾 AI 技术发展和公民权利保护两大立法目标，《AI 公约》采取了概括性的风险规制义务表述。^③ 《AI 公约》第 4 条阐释了基于现行有效国内法和国际法、贯穿 AI 系统全生命周期的人权保护义务。第 5 条确立了保障民主程序、遵守法治原则的义务。^④ 基本原则方面，《AI 公约》整合了监管原则专章，构建基于原则的（principle-based）人权监管模式。^⑤ 《AI 法》的监管原则条款散见于各个章节。^⑥ 不同于《AI 法》基于规则

^① Council of Europe, “Council of Europe Opens First Ever Global Treaty on AI for Signature”, 2024-09-05, <https://www.coe.int/en/web/portal/-/council-of-europe-opens-first-ever-global-treaty-on-ai-for-signature>, 访问日期:2025-03-27。

^② Hannah van Kolfschooten/Carmel Shachar, “The Council of Europe’s AI Convention (2023-2024): Promises and Pitfalls for Health Protection”, p. 1.

^③ Tareq Al-Billeh/Ruba Hmaidan/Ali Al-Hammouri/Mohammed Al Makhmari, “The Risks of Using Artificial Intelligence on Privacy and Human Rights: Unifying Global Standards”, *Jurnal Media Hukum*, Vol. 31, No. 2, 2024, pp. 333-350, here p. 344.

^④ Council of Europe, “Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law”, p. 4.

^⑤ Hannah van Kolfschooten, “The Council of Europe’s Artificial Intelligence Convention: Implications for Health and Patients”, Law Centre for Health and Life of University of Amsterdam, 2023-06-27, p. 1, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4492186, 访问日期:2025-02-26。

^⑥ EU, “Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)”, pp. 55-67.

的(rule-based)物项监管,《AI公约》第7—13条分别树立了人类尊严和个体自治、透明度和有效监管等7项基本原则。^①通过体系解释,《AI公约》要求监管主体遵循人类尊严和个体自治原则,并在监管中体现有效问责、可责性及平等和非歧视原则。透明度和有效监管、创新安全原则是该公约的监管目标,隐私和个人数据保护原则则为监管手段提供了指导。^②根据《AI公约》第11条“隐私和个人数据保护”原则专条,国内法和国际法规定的隐私和数据保护义务以及技术标准和监管框架中的隐私和数据保护规范对《AI公约》缔约国具有约束力。这种造法模式使《AI公约》项下的监管更为有法可依。^③

二是硬法化 AI 国际监管规则。《AI公约》体现了 AI 国际法治理在法律解释、立法推动、软硬法协同中的积极作用,推动了 AI 治理硬法化。法律解释方面,欧盟委员会已签署批准《AI公约》,该公约成为解释《AI法》等法律条款的重要依据。例如,对具有不可接受风险的 AI 系统,《AI公约》第16条要求缔约方评估是否需要禁用该 AI 系统,并进一步完善执行机制。《AI公约》第23条将“缔约方会议”确立为监督与合作核心机制,并授予其解释条款、发布建议、修订公约内容等职能。^④这也是《AI法》希望实现但未能实现的制度设计。就国际法而言,《AI公约》将散见于各种宣言、倡议中的 AI 治理理念转化为国际条约,使 AI 软法治理规范硬法化。哈佛大学研究机构通过分析 30 余份关于 AI 伦理的主要国际宣言,归纳出隐私保护、问责机制、公平与非歧视性等原则要求。^⑤《AI公约》第三章将上述碎片化软法原则纳入国际公约,形成了 7 项基本原则及其概括性义务和监督保障框架。《AI公约》第四章第14条要求各缔约方在“符合国内法规定的情况下”对 AI 侵犯人权的行为实行补救,第15条进一步约定各缔约方应“根据适用的国内法和国际法”保障知情权。这些配套制度增强了国际通用 AI 治理规则的权威性。《AI公约》的部分内容已超出欧盟法律范畴,实现了软硬法协同。例如,《AI公约》第13条创新安全原则条款建立监管沙盒(regulatory sandbox)、发布特别指导或“无行

① Council of Europe, “Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law”, pp. 4–5.

② CAHAI, “Feasibility Study: A Legal Framework on AI Design, Development and Application Based on Council of Europe Standards on Human Rights, Democracy and the Rule of Law”, pp. 7–8.

③ 同注①。

④ 同注①, pp. 6–7。

⑤ Jessica Fjeld et al., “Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI”, Berkman Klein Center for Internet & Society Research at Harvard University, 2020–02–14, pp. 4–6, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482, 访问日期:2025–01–29。

动函”等，都是具体的执行方式。^①《AI 公约》第 27 条规定了缔约方之间既有相关协议或条约的优先适用性。^②这一条款协调了《AI 公约》与《AI 法》的关系，^③或可成为沟通国内法和国际法的桥梁型条款。

三是强化欧洲 AI 治理的“布鲁塞尔效应”。《AI 公约》反映了欧洲委员会扩大 AI 治理规则接受度、推进 AI 国际条约普适化的战略意图。其根本目的是适应 AI 全球治理的区域竞争格局，在竞争中获取先发优势和比较优势，强化欧洲 AI 治理的“布鲁塞尔效应”。欧盟立法产生“布鲁塞尔效应”需要满足五个条件：庞大的市场规模，有效的监管能力，严格的监管标准，监管对象与监管制度的深度绑定，以及技术、经济、法律上的不可分割性。^④例如，GDPR 对全球数据保护立法起到示范作用，推动了美国加州《消费者隐私法案》等一系列法律的制定。^⑤《AI 公约》旨在进一步强化这种“布鲁塞尔效应”。《AI 公约》能够提升各缔约方的 AI 监管能力，即“颁布规则并确保规则得到遵守的能力”。这一能力的发挥取决于监管专业知识、政策制定一致性和制裁权限。^⑥《AI 法》第 64—70 条在欧盟层面建立了欧盟 AI 办公室、AI 委员会、独立专家科学小组和成员国专家库利用机制，^⑦提高了 AI 监管知识共享程度。关于政策制定一致性，欧盟委员会和欧洲委员会正在推动更新“AI 协调计划”，以提高各国行业规则和产业标准协调程度。欧盟还和美国通过欧美贸易与技术委员会、数字伙伴关系网络等机制增进合作，初步确立了美欧 AI 治理协同合作政策。^⑧制裁权限上，《AI 公约》第 27 条第 2 款允许作为欧盟成员国的缔约方就公约相关事项适用欧盟法律，等同于将《AI 法》的执行程序和制裁权限引入《AI 公约》。《AI 公约》扩张了欧洲 AI 监管标准。欧洲是最早发布具有规范意义

① CAHAI, “Feasibility Study: A Legal Framework on AI Design, Development and Application Based on Council of Europe Standards on Human Rights, Democracy and the Rule of Law”, pp. 54–55.

② Council of Europe, “Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law”, p. 9.

③ 沈伟、苏可桢：《变局之下国际法与国内法互动的转向》，载《探索与争鸣》，2024 年第 4 期，第 96–106 页，这里第 99 页。

④ Anu Bradford, “The Brussels Effect”, *Northwestern University Law Review*, Vol. 107, No. 1, 2012, pp. 1–68, here pp. 10–18; 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》，载《德国研究》，2024 年第 2 期，第 66–88 页，这里第 75–79 页。

⑤ 姚旭、张傲：《美欧英签署全球首个人工智能法律公约，AI 监管“全球化”进程加快》。

⑥ Anu Bradford, *The Brussels Effect*, Oxford, UK: Oxford University Press, 2020, p. 31.

⑦ EU, “Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)”, pp. 95–100.

⑧ 郭梦迎、张薇薇：《欧盟人工智能发展与治理面临的机遇和挑战》，载《和平与发展》，2024 年第 5 期，第 50–76 页，这里第 58–60 页。

的专门性 AI 伦理规则的法域之一。^① 欧盟法律持续投入行政司法资源,遏制 AI 技术同市场、媒体资源绑定带来的算法垄断权力,维护欧洲公民的言论自由和政治平等权。^② 《AI 公约》践行了这种技术民主理念,初步建立了公共领域信息共享机制,复制和扩大了《AI 法》的“最大公约数”,^③ 其不仅满足了欧盟公民的风险厌恶偏好,还将监管标准推广至格鲁吉亚、以色列等非欧洲国家。《AI 公约》还增强了各缔约方在 AI 治理上的不可切割性,其原则性、包容性和开放性可助力更广泛的 AI 国际合作。例如,《AI 公约》第 25 条第 1 款专条规定 AI 治理国际合作,鼓励各缔约方协助非缔约方按该公约条款行事,并推动其加入公约。^④ 《AI 公约》借鉴了 G7《开发先进人工智能系统组织的国际行为准则》中关于 AI 监管的表述,^⑤ 也承继了美、英、欧盟关于 AI 竞争联合声明的精神。^⑥ 上述效应也将重塑全球 AI 市场格局,推动各国加强对国际法律的重视,扩大欧洲在 AI 国际法治上的影响力。

三、《AI 公约》作为 AI 治理国际条约的不足

(一) AI 国际条约的理想状态

AI 国际法治离不开国际硬法治理,国际条约是国际硬法的代表。AI 的全产业链影响为 AI 全球治理提供了必要性和正当性基础,AI 技术的迅猛发展也构成对国家主权、国家有效监管等国际法基本原则的冲突,造成对国际人权法、国际知识产权法等部门国际法的挑战,因此,构建和完善 AI 治理的国际条约体系尤为必要。^⑦ 为应对上述挑战,国际社会提出多种应对策略,形成了对 AI 国际条约的理想预期。

首先,理想的 AI 国际条约应当具备国际条约的核心构成要件。国际条约有形式、内容和规范效果三个基本构成要件。一份成熟的 AI 国际条约至少需要在形式

① 李学尧:《人工智能伦理的法律性质》,载《中外法学》,2024年第4期,第884—898页,这里第886页。

② Ugur Aytac, “Big Tech, Algorithmic Power, and Democratic Control”, *Journal of Politics*, Vol. 86, No. 4, 2024, pp. 1431—1445, here pp. 1437—1443.

③ 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》,第83—84页。

④ Council of Europe, “Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law”, pp. 8—9.

⑤ 方晓:《AI 监管加速:白宫发行政令,G7 提自愿行为准则,英国开全球峰会》,澎湃新闻,2023—10—31, https://www.thepaper.cn/newsDetail_forward_25126009, 访问日期:2025—01—31。

⑥ 《美英欧监管机构签署关于人工智能竞争的联合声明》,新华网,2024—07—24, <http://www.news.cn/world/20240724/d1124218f501455d8dd63140db54f31c/c.html>, 访问日期:2025—01—31。

⑦ 赵骏、李婉贞:《人工智能对国际法的挑战及其应对》,载《浙江大学学报(人文社会科学版)》,2020年第2期,第10—25页,这里第10—19页。

上满足多边要求,在内容上为 AI 治理确定各国权利义务和条约执行机制,在规范效果上为后续国际合作设置通用 AI 监管框架。^① 如此方能形成有效的国际法规制,完善以国际条约为核心的 AI 国际法治。形式上的多边要求建立在 AI 风险治理的时代背景下,传统的主权边界已不再能限制 AI 产品与技术的国际流通, AI 风险治理依赖国内法的域外效果。国际条约提炼了各国 AI 治理共识,推动了 AI 国内法律规范国际化,能够增强 AI 全球治理的生命力。内容上,不同于传统国际条约,在法律制定层面, AI 国际条约对造法技术提出更高要求。在参与主体层面, AI 国际条约要求各国政府、国际组织、企业和个人的多元主体参与,对各种主体课以新的义务类型,利用有限监管资源实现 AI 系统全生命周期监管。为确保 AI 国际条约的有效执行,可以采取在全球范围内将 AI 软法治理硬法化、修改与制定现行双多边经贸条约等造法路线。 AI 国际条约建立了通用 AI 监管框架,其效果在于鼓励主权国家履行义务与承担责任、设立专门 AI 监管国际机构协调各国 AI 治理进程,最终深化 AI 国际法治。以治理手段多元化、规制思路多样化 as 特征,围绕国际条约构建的 AI 国际法治体系将为 AI 全球治理带来积极变量,有望改变当前各国疲于以国内法手段应对 AI 新产品、新技术、新路线的被动态势。

其次,理想的 AI 国际条约应当兼容国际“超级软法”(supersoft law)和国际硬法的优势,做到监管细腻、执行有力。所谓“超级软法”,是指 AI 治理中不具备正式法律约束力,但在合规层面上认可度、接受度足够高的指引性文件,如前文提及的行业规则、产业标准、伦理规范等。这些“超级软法”填充了国际法的定义释明、范畴澄清、适用指南等细节,基于这些规范制定的 AI 国际条约在规则上更有普适效力。国际硬法或是具体调整各缔约方 AI 治理法律关系、改变各国权利义务布局的专门性 AI 国际公约,或是已被纳入国际经贸关系的双边和多边贸易、投资、金融协定。两者在制度优势上有相通之处。国际硬法的优势在于更高效的执行力、更稳定的实用性以及更可靠的制度支持能力,能够为各缔约方提供法律确信、指导各国开展 AI 治理国际法实践,从而形成有法律效力的国际习惯。^② 无论是“超级软法”还是国际硬法,都能巩固 AI 国际法治有效性、提升 AI 国际法治的权威。成功的“超级软法”或国际硬法文本还能指引一国提高 AI 国内立法技术,帮助该国对接国际高标准 AI 法律规范,为该国乃至该国所在区域经济政治安全护航。因此,好的 AI 国际条约需要同时以国际“超级软法”和国际硬法的造法技术为支撑,方能经受

^① Michael Bowman/Dino Kritsiotis (eds.), *Conceptual and Contextual Perspectives on the Modern Law of Treaties*, Cambridge, UK: Cambridge University Press, 2018, pp. 85-98.

^② 沈伟、赵尔雅:《数字经济背景下的人工智能国际法规制》,第 133 页;李猛:《人类命运共同体视角下人工智能风险全球治理的国际法规制路径探究》,载《宁夏社会科学》,2024 年第 2 期,第 112-124 页,这里第 116 页。

住实践检验,在技术迅猛迭代、风险暗流涌动的当今世界行稳致远。

最后,理想的AI国际条约应当以全球AI相关领域为覆盖范围,效力与影响不拘于一隅。AI国际条约应超出某一个或几个特定国家的利益需要,符合世界各国共同利益。党的二十届三中全会《决定》明确提出要“推进人工智能治理法治化”。^①推进AI治理法治化,就是发挥AI全球治理的善治属性,推动AI国际条约等制度成果生成与发展。AI国际条约应发挥国际公共产品的作用,符合人类社会对AI健康发展的期待。各国对AI国际条约也抱有类似的期许,通过国内国际治理表达共享AI国际法治成果的愿望。除欧盟《AI法》、欧洲委员会《AI公约》外,英国科技部门提出《人工智能监管的创新方法》草案,美国国会通过针对AI深度伪造的监管法律,法国数据监管部门出具《AI法》和GDPR的制度衔接建议,各国分别采取行动,从整体上推动AI国际法治议程进展。在跨国合作方面,2024年5月,我国与美国举办首届AI政府间对话,同法国发布AI全球治理联合声明。我国还于2024年7月在上海举办2024世界人工智能大会暨人工智能全球治理高级别会议,会议通过《人工智能全球治理上海宣言》,倡议构建AI全球治理机制。^②这些国内和国际努力都有助于取各方AI治理规则交集、衔接监管制度,促进通用权利义务和执行条款的塑造,最终促成一份或数份AI国际条约。整合各国制度贡献的AI国际条约也自然具备全球视野与国际适用性,从而有望实现真正的“放之四海而皆准”。

(二)《AI公约》还不是理想的AI国际条约

《AI公约》相较前文所述理想的AI国际条约仍有差距,在构成要件、制度能力和效力与影响上都有所不足,难以满足AI治理的根本需要。

首先,《AI公约》存在要素缺陷,在国际条约的形式、内容和规范效果等方面均有局限性。《AI公约》仅以国际公约的形式呈现,全文仅有36条,过度追求条文表达简洁性,而忽视了形式协调性。例如,《AI公约》最重要的创新点在于第二章“一般义务”专章,然而整章仅有第4条、第5条两个条款。同样,第四章“补救措施”专章在重要性上仅次于一般义务,但全章也只有第14条、第15条两个条款。相较之下,第三章规定的基本原则、第六章公约实施中的特别注意事项共包含13个条款,占到全文篇幅比重的三分之一。《AI公约》的内容也欠丰富,并未实质性地为各缔约方课以AI治理权利义务和责任分配。该公约在权利救济方面仅保留宽泛而难

^① 张吉豫:《保障人工智能健康发展 推进人工智能治理法治化(有的放矢)》,中华人民共和国司法部,2024-09-06, https://www.moj.gov.cn/pub/sfbgw/zwgkztzl/xxxxcgxjpfzxs/fzsxllqy/202409/t20240906_505500.html, 访问日期:2025-02-02。

^② 王卫:《2024年全球加快推进人工智能监管》,《法治日报》,2024-12-16, <http://epaper.legaldaily.com.cn/fzrb/content/20241216/Article06002GN.htm>, 访问日期:2025-02-03。

以落实的申诉程序,且未约定违约罚则和争端解决机制。^① 在规范效果上,弗朗西斯卡·法努奇(Francesca Fanucci)认为该公约已被“稀释”,难以在后续国内国际立法中充当监管框架基础。各国 AI 执法能力差异也使公约的同质化规定难以落实。即便公约得到适用,服务国家安全的 AI 系统得到监管豁免,各缔约方在执行公约时仍有相当自由裁量权限,^② 私营企业等 AI 治理主体尚未受到公约拘束。这些缺陷都使《AI 公约》欠缺监管强度和执法力度,在关键监管规范上浮于表面、流于形式。

其次,《AI 公约》也未达到对其制度能力的立法预期。其既未能充分整合 AI “超级软法”的制度优势,又缺乏 AI 国际硬法应有的约束力。在促进经济发展、实现国际社会发展目标等国际软法的侧重点上,《AI 公约》的实际作用有限。《AI 公约》出台以前,联合国机构已通过数份 AI 治理规范,作为国际软法确立了基本的 AI 治理原则。例如,2021 年,UNESCO 全体 193 个会员国一致通过联合国首个 AI 全球治理性文件《AI 伦理建议》;2023 年,联合国成立 AI 咨询机构;2024 年 9 月,联合国未来峰会通过《未来契约》及其附件《全球数字契约》,为 AI 全球治理提供权威指南。^③ 《全球数字契约》将创造包容、开放、可持续、公平、安全、可靠的数字未来作为宗旨,提出 13 项技术治理基本原则,^④ 呼吁各国弥合数字鸿沟、扩大数字经济覆盖面、加强 AI 全球治理以造福人类。^⑤ 相较而言,《AI 公约》全文多次出现“人权、民主和法治”字样,并强调国际经济治理体系中的“民主”内涵也是公约正当性的来源。^⑥ 然而,公约中仅在第 4—5 条简要介绍了相关内容,对定性和解释一

① 安峥:《美英欧盟等签署首个具法律约束力的国际 AI 公约,为何有欧洲专家“失望”?》。

② Mauricio Figueroa, “The Drawbacks of International Law in Governing Artificial Intelligence”, *Society for Computers & Law*, 2025 - 01 - 14, <https://www.scl.org/the-drawbacks-of-international-law-in-governing-artificial-intelligence/>, 访问日期:2025 - 02 - 27; Karolina Babická/Cristina Giacomini, “Understanding the Scope of the Council of Europe Framework Convention on AI”, *Opinio Juris*, 2024 - 11 - 05, <https://opiniojuris.org/2024/11/05/understanding-the-scope-of-the-council-of-europe-framework-convention-on-ai/>, 访问日期:2025 - 02 - 27。

③ 《聚焦未来:科技创新与数字合作——应对 AI 威胁,弥合数字鸿沟的竞赛》,联合国,2024 - 09 - 13, <https://news.un.org/zh/story/2024/09/1131351>, 访问日期:2025 - 02 - 06。

④ 《全球数字契约》第 8 条涵盖的 13 项技术治理基本原则分别是:1. 包容性;2. 以发展为导向;3. 立足于国际法,促进、保障人权;4. 性别平等、尊重弱势群体;5. 可持续性;6. 公平和非歧视性;7. 可及和可负担性;8. 互操作性和治理兼容;9. 全生命周期的创新安全、责任承担和可责性;10. 创新和竞争;11. 多元主体治理;12. 为发展中国家提供执行资源;13. 前瞻性和预测机制,以应对风险和挑战。

⑤ 《未来契约》,联合国大会,2024 - 09 - 20, <https://docs.un.org/zh/A/79/L.2>, 访问日期:2025 - 02 - 06。

⑥ 沈伟:《国际经济治理体系的正当性:内涵、危机及反思》,载《郑州大学学报(哲学社会科学版)》,2024 年第 5 期,第 25 - 26 页。

笔带过,^①并未充分评价何为人权、民主与法治。^②该公约略过AI带动经济发展的重要内容,未能触及国际AI争议协调、私主体争端解决机制等人权保护机制,也未提供明确成文的民主或法治纪律。《AI公约》仅包含7项基本原则,与《全球数字契约》的13项基本原则相比,在联合国AI全球治理发展导向、可持续性、互操作性和支持发展中国家等方面存在明显缺失。在对国际硬法文本殊为关键的风险控制层面,《AI公约》对新兴技术的适应性也有待提高。作为国际公约,《AI公约》内容未涉及生成式AI、大语言模型等近年来备受关注的技术应用场景。《AI公约》中的风险管理条款未对风险分类和管理具体流程作出说明,对数据治理、算法透明性和跨境数据流动等关键领域的处理也相对笼统,难以应对AI技术发展的复杂多样性。《AI公约》也无法解决其条款滞后性问题。当公约仍执着于以人为本的理想主义监管时,AI已渗入各国的司法制度,通过算法改变该国的执法决策偏好并反过来影响国际争端解决程序。^③《AI公约》仅以建议性语言要求缔约方采取暂停或禁止措施,而这种松散监管不足以应对高风险AI应用的潜在威胁。^④这些结构性缺陷都使《AI公约》难以承担AI全球治理的核心任务。

最后,《AI公约》仍有区域主义色彩,尚不具备全球影响力和约束力。欧洲委员会秘书长玛利亚·佩伊契诺维奇·布里奇(Marija Pejčinović Burić)表示,《AI公约》是“对各大洲有相同价值观国家支持的国际法律标准需求的回应”。欧洲委员会官方新闻指出,《AI公约》是一部世界各国都有资格加入、允许非欧洲委员会成员国签署和遵守的国际公约。^⑤然而,目前《AI公约》仅有11个正式缔约方,且绝大部分在AI产业发展上处于后发地位的发展中国家并未签约,^⑥我国也没有参与签署。截至2024年底,在所有缔约方中,仅挪威一国接受了《AI公约》约束,并将

① Council of Europe, “Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law”, p. 4.

② 在《AI公约》的序言、全部36个正文条文及最后条款中,“人权、民主和法治”(Human Rights, Democracy and the Rule of Law)表述出现在序言、第1条“目标和宗旨”、第3条“范围”、第9条“可责性和责任承担”、第13条“安全创新”、第16条“风险和影响管理框架”和第25条“国际合作”等部分。

③ Ashley Deeks, “Introduction to the Symposium: How Will Artificial Intelligence Affect International Law?”, *American Journal of International Law*, Vol. 114, 2020, pp. 138–140, here p. 138.

④ Madhumita Murgia/Javier Espinoza, “US, Britain and Brussels to Sign Agreement on AI Standards”, *Financial Times*, 2024–09–05, <https://www.ft.com/content/4052e7fe-7b8a-4c42-baa2-b608ba858df5>, 访问日期:2025–02–27。

⑤ Council of Europe, “Council of Europe Opens First Ever Global Treaty on AI for Signature”, 2024–09–05, <https://www.coe.int/en/web/portal/-/council-of-europe-opens-first-ever-global-treaty-on-ai-for-signature>, 访问日期:2025–02–27。

⑥ 并非所有发展中国家都未签署《AI公约》。例如,黑山就是发展中国家,但也作为欧洲委员会成员国参与缔约。

该公约第 2—6 章载明的原则和义务适用于私主体 AI 活动监管。^① 并且,迄今为止没有一个非欧洲委员会成员国签署该公约。客观上,《AI 公约》仍是美国及其盟友的小多边俱乐部式规则,旨在保持美欧制度优势,使发展中国家实际上或感到被排除在外(are or feel excluded)。这可能促使出现“中国领导、协调或培育”下全球南方国家主导的另一版 AI 国际条约。^② 除缔约国家不多以外,各大科技企业也对《AI 公约》持观望态度,对签约不甚积极。苹果、英伟达、Meta、Mistral AI 等知名科技公司都未签署《AI 公约》,并表达了对该公约过于依赖自愿承诺机制、约束力不强的担忧。^③ 这种客观情况令《AI 公约》难以成为受全体联合国会员国广泛欢迎的国际条约,使美欧 AI 治理陷入自行其是的境地。

《AI 公约》首倡国际硬法层面的 AI 治理,的确发挥了法律和社会层面的积极作用。但囿于上述各方面的缺陷,《AI 公约》难以满足全球各国的期待。在 AI 越发展现其风险性、破坏性的当下,全球法律体系面临 AI 可能修补、替代或破坏既有规范的三种可能。即便以《AI 公约》为代表的国际条约能够应付 AI 带来的风险,由于缺乏实用能力和制度响应,这些国际法也很难积极引导 AI 改善现行法并带来善治。^④

四、我国如何推动制定理想的 AI 国际条约

(一)《AI 公约》对我国的影响

我国 AI 技术正处于快速发展阶段,尤其在计算机视觉领域已取得一定优势,监控和追踪技术也走在世界前沿。DeepSeek 等先进 Gen AI 应用在研发成本、算力支持和自主学习程度等方面甚至已取得全球领先优势,冲击了美国、欧洲 AI 技术先发地位。据 BBC 报道,DeepSeek 面世导致英伟达(Nvidia)股价骤跌 17%,对全球 AI 产业格局产生很大影响。也应看到,受计算能力局限等因素影响,商汤科

① Karolina Babická/Cristina Giacomini, “Understanding the Scope of the Council of Europe Framework Convention on AI”, *Opinio Juris*, 2024 - 11 - 05, <https://opiniojuris.org/2024/11/05/understanding-the-scope-of-the-council-of-europe-framework-convention-on-ai/>, 访问日期:2025 - 02 - 27。

② 沈伟、胡耀辉:《美式小多边主义与美国重塑全球经贸体系》,载《国际论坛》,2022 年第 1 期,第 3 - 24 页,这里第 7 - 8 页;José Jaime Villalobos/Matthijs Michiel Maas, “Beyond A Piecemeal Approach: Prospects for A Framework Convention on AI”, *Institute for Law & AI*, 2024 - 09 - 25, pp. 11 - 12, <https://law-ai.org/framework-conventions/>, 访问日期:2025 - 02 - 06。

③ 《苹果、Meta 等缺席欧盟〈人工智能公约〉》,世界互联网大会,2024 - 09 - 28, https://cn.wicinternet.org/2024-09/28/content_37589134.htm, 访问日期:2025 - 02 - 03。

④ Matthijs Michiel Maas, “International Law Does Not Compute: Artificial Intelligence and the Development, Displacement or Destruction of the Global Legal Order”, *Melbourne Journal of International Law*, Vol. 20, No. 1, 2019, pp. 1 - 29, here p. 29.

技等我国 AI 企业还有较大上升空间。^① 我国应当客观、准确评估《AI 公约》造成的国内国际法律影响,并做好对于未来技术和制度竞争的应对规划。

《AI 公约》将进一步塑造 AI 治理的区域竞争格局。近年来,中美欧在 AI 等高科技领域出现了竞合博弈的态势。^② 2019 年,欧洲的科技投资额度是五年前的 4 倍,拥有近 600 万专业开发人员。欧洲在 AI 等信息科技领域的投资额占区域国内生产总值(GDP)的 1.7%,同中国的 2.16%、美国的 3.33%相比,差距并不显著。^③ 根据 2020—2024 年企业表现,欧盟培育出瑞士阿西布朗勃法瑞集团(ABB)和德国库卡公司(KUKA)等全球工业机器人领先企业,在 AI 制造和芯片生产方面有蔡司、阿斯麦等知名大公司和 Aleph Alpha、Mistral AI、Helsing 等新锐企业,获得了一定市场支配力。但近年来,欧洲相较中美的 AI 投入和产出差距被不断拉大。在投入方面,据统计,在 2018—2023 年这五年中,美国较欧盟多获得近 900 亿欧元投资;2023 年,中国一国的 AI 私人投资额就超过德国、法国、瑞典、西班牙四个欧盟领先 AI 投资国家的私人投资总额。^④ 在产出方面,截至 2017 年底,全球十大互联网企业没有一家落地欧洲,全球 185 家 AI 独角兽企业中仅有 10%来自欧洲,而美国和中国同类型企业数目已分别占到总数的 54%和 23%。在当时的欧洲公司中,仅有半数左右取得了正回报率,且 89%的全欧营收来自美国。^⑤ 为此,美欧正基于“价值观同盟”^⑥形成并巩固 AI 治理的“小多边主义”(Minilateralism)机制,^⑦试图遏制我国 AI 产业发展。《AI 公约》的出台,标志着欧洲同我国的 AI 制度竞赛进入新阶段,其烈度和波及范围亦有所增强。

《AI 公约》也会同我国 AI 治理规则发生摩擦。结合前文所述《AI 伦理规范》等行业指引,我国初步构建了软硬法兼具的 AI 治理规则体系。《AI 公约》作为国

① Kelly Ng/Brandon Drenon/Tom Gerken/Marc Cieslak, “DeepSeek: The Chinese AI App that Has the World Talking”, *BBC*, 2025-01-28, <https://www.bbc.com/news/articles/c5yv5976z9po>, 访问日期: 2025-02-27; Liza Lin, “China Puts Power of State Behind AI and Risks Strangling It”, *The Wall Street Journal*, 2024-07-16, <https://www.wsj.com/tech/china-puts-power-of-state-behind-ai-and-risks-strangling-it-f045e11d>, 访问日期: 2025-02-27。

② 方兴东、钟祥铭、谢永琪:《“布鲁塞尔效应”与中国数字治理的制度创新——中美欧竞合博弈的建构主义解读》,载《传媒观察》,2024年第3期,第33-44页,这里第40-41页。

③ McKinsey Global Institute, “Notes from the AI Frontier Tackling Europe’s Gap in Digital and AI”, McKinsey, 2019-02-07, p. 6, <https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20europes%20gap%20in%20digital%20and%20ai/mgi-tackling-europes-gap-in-digital-and-ai-feb-2019-vf.pdf>, 访问日期: 2025-02-05。

④ 郭梦迎、张薇薇:《欧盟人工智能发展与治理面临的机遇和挑战》,第61、62、69页。

⑤ 同注③, p. 7。

⑥ 苏可桢、沈伟:《美国“价值观同盟”视角下的经贸规则全球异化与中国因应》,载《亚太经济》,2023年第3期,第13-21页,这里第13-15页。

⑦ 沈伟、胡耀辉:《美式小多边主义与美国重塑全球经贸体系》,第4-8页。

际条约有国际影响力,其 AI 治理思路同我国有明显区别,部分条款在适用范围及规范限度上同我国国内法存在差异。

首先,《AI 公约》和我国法律在 AI 治理上侧重不同。欧洲各国多为经济发达、技术先发国家,能够形成建设数字经济治理体系、建立权利导向型监管模式的共识。《AI 公约》反映了这种追逐 AI 治理最高标准的心态,有“不合规即违法”的意蕴。我国是发展中国家、技术新兴国家,立法及政策导向强调技术突破下的国家主权与数据安全。《网络安全法》《数据安全法》划定了信息维护和数据保护的底线, AI 技术得以在不触及法律红线的前提下开发利用,总体呈“不违法即合规”态势。监管侧重上的不同使《AI 公约》难以适配我国的 AI 治理需求。

其次,《AI 公约》和我国法律对 AI 技术的态度不同。这一点体现为技术发展阶段不同导致的监管口径区别。《AI 公约》要求 AI 技术在开发、部署和使用中符合保护人权、民主和法治的国际义务,欧洲 AI 技术处于细化风险管控、强化合规要求的阶段。我国则尚处在 AI 技术安全和发展并重的阶段。例如,《AI 伦理规范》第 3 条将增进人类福祉置于 6 项 AI 伦理基本原则的首位,表达了对合理利用 AI 技术的信心。^①《暂行办法》也要求鼓励 Gen AI 创新应用、推进基础设施建设,希望实现技术促进和风险治理的平衡。^② 这些规范表明,我国目前仍以发展 AI 技术为法律或政策的优先事项,不倾向于对 AI 技术进行过细过实的制度管控。

最后,《AI 公约》和我国法律在 AI 监管强度上也有差异。《AI 公约》执行的风险分类分级监管标准较为严格,涉及 AI 全产业链诸多门类。如此全面的监管显然不适用于正处于产业发展期的我国 AI 行业。150 余家欧洲科技公司曾联名抵制欧洲 AI 监管模式,认为这种强势监管模式将抑制 AI 技术在当地的发展、打击行业发展的积极性和能动性,最终导致政府独断,与《AI 公约》的技术民主精神南辕北辙。并且,过于细化的 AI 风险划分缺乏实践案例支撑,可能造成法律治理“空中楼阁”现象。^③ 对上述《AI 公约》监管“副作用”的批评恰能引发我国的共鸣。我国的 AI 白皮书、治理指南、治理框架等规范均紧密结合行业发展趋势,虽未能按照风险层次实现文本标准化、规则体系化,但也形成了动态多元的“生态治理”方案,^④ 相较《AI 公约》更为科学合理、注重实际。

① 中华人民共和国科技部:《〈新一代人工智能伦理规范〉发布》,2021-09-26, https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html, 访问日期:2025-04-06。

② Jen Zhu Scott, “Understanding China’s Pragmatic AI Plan”, *Financial Times*, 2024-07-31, <https://www.ft.com/content/e76d54e6-607f-4511-a2dc-db030feea0b9>, 访问日期:2025-02-27。

③ 赵泽睿:《人工智能立法的沟通认知模型与程序治理思路》,上海交通大学中国法与社会研究院, 2025-01-13, <http://www.socio-legal.sjtu.edu.cn/wxzy/info.aspx?itemid=4780&lcid=13>, 访问日期:2025-02-06。

④ 张凌寒:《人工智能法律治理的路径拓展》,第 100 页。

《AI公约》同我国AI治理思路的差异使两者难以无缝衔接。《AI公约》形成的全球话语权和影响力可能反作用于我国的AI治理生态,造成我国在AI全球治理规则上同国际主流趋势相分离的风险,并对我国AI产业发展造成负面影响。我国应当认识到《AI公约》对国际社会和我国法治的双重影响,充分识别其与我国AI治理目标、我国预期下理想AI国际条约的差异。

(二)在《AI公约》启示下推动AI治理中国方案

《AI公约》对我国深度参与国际条约的起草与制定提供了启示。通过制定国际规则并构建符合国家利益的话语体系,一国能够宣介其在AI全球治理上的基本态度和法律观点,输出该国的法律制度与国际实践。我国应当适时总结推动、实施布局AI治理中国方案,加速国际广泛认可的AI国际法生成,扩大我国在AI国际法治中的话语权。^①

我国现已就AI监管参与制定数项国际合作决议、倡议。2023年,我国提出《全球人工智能治理倡议》(以下简称《AI治理倡议》),就提升AI技术的安全性、可靠性、可控性、公平性呼吁国际主体广泛参与。在内容上,该倡议在第1—2条呼吁促进国际合作与国际法建设,第3—13条提倡发展AI技术、实现国际协调普惠,第14—15条提出对联合国下一步AI促进行动的建议,具有跨地域性、无差别性和广泛适用性特征。^②《AI治理倡议》和当前《AI公约》确立的AI治理基本原则存在部分重合。同时,我国在该倡议中强调增强发展中国家的代表性和发言权,确保AI全球治理权利、机会和规则平等,支持国际组织发挥核心协调作用,而这恰好是《AI公约》疏漏的。^③2024年7月1日,联合国大会以143票赞成通过我国提出的《加强人工智能能力建设方面的国际合作决议》(第A/RES/78/311号决议,以下简称《AI能力决议》)。^④该份决议是联合国首份AI能力建设决议,坚持了我国提出的以人为本、智能向善、造福人类等AI治理原则,保留了开放、公平、非歧视等亦体现于《AI公约》的基本原则。^⑤《AI能力决议》同时回应了南北方国家的关切,相较《AI公约》更具国际代表性。

① 张磊:《人工智能国际规则博弈态势、共识分歧与中国“话语势能”建构》,载《新疆社会科学》,2024年第4期,第119—133页,这里第119、133页。

② 《2024年7月1日大会决议78/311. 加强人工智能能力建设方面的国际合作》,联合国大会,2024-07-01, <https://docs.un.org/zh/A/RES/78/311>, 访问日期:2025-02-05。

③ 《全球人工智能治理倡议》,中国政府网,2023-10-20, https://www.mfa.gov.cn/web/ziliao_674904/1179_674909/202310/t20231020_11164831.shtml, 访问日期:2025-02-02。

④ 《联大通过中国提出的加强人工智能能力建设国际合作决议》,中国政府网,2024-07-02, https://www.gov.cn/yaowen/liebiao/202407/content_6960524.htm, 访问日期:2025-02-02。

⑤ 中华人民共和国外交部:《推动普惠发展,弥合智能鸿沟——王毅在人工智能能力建设国际合作高级别会议上的致辞》,2024-09-27, http://la.china-embassy.gov.cn/xwdt/202409/t20240927_11498445.htm, 访问日期:2025-02-05。

同年9月25日,王毅外长提出我国《人工智能能力建设普惠计划》(以下简称《AI普惠计划》),并于10月函送联合国秘书长发布。^①该计划为加强AI全球治理提出10项中国行动,在10项中国行动中有多达9项直接涉及中国对发展中国家AI能力建设的支持和援助。^②这些承诺为执行《AI能力决议》提供现实路径,体现了大国担当,^③同《AI公约》一味确立AI治理高标准、忽视发展中国家AI治理能力需要的造法基调有本质区别。以上述国际行动为基础,我国进一步深化AI全球治理合作,为AI本国立法、区域双多边协定和国际条约的制定确定了基准。

受《AI公约》的启发,我国亟需进一步拓展AI治理场景、优化AI治理思路、积累AI治理经验,提出更高水平、更具包容性和普遍性的AI全球治理方案和规则,形成AI治理中国方案。这要求我国做到如下两个方面。

第一,我国应当比照既有国际法和国际规则,以法律法规等高位阶立法形式加强AI风险治理,全面应对AI技术带来的挑战。^④我国首部AI产业促进立法《深圳经济特区人工智能产业促进条例》已于2022年8月30日通过,其中提出了AI创新产品准入制度。然而,该条例的法律阶位较低且针对性弱,主要聚焦于AI相关产业准入问题。^⑤我国应加快推进AI专门立法,推动AI法律监管体系完善,优化AI国内法治理。针对因侧重治理而导致的风险规制局限,我国可以借鉴《AI公约》倡导的AI治理机制,引入利益相关方商谈机制、公共领域信息共享机制并建立合理、有效、可执行的数据保护或技术合规罚则,为底线思维下的AI治理提供利益平衡性和执法尖锐度,逐步提高AI全生命周期法治水平。针对监管口径对技术发展的过度宽容问题,我国可以持续关注法律人格、技术控制和标准化等国际公认的

① 傅聪:《2024年10月15日中国常驻联合国代表给秘书长的信》,联合国大会,2024-10-15, https://www.google.com.hk/url?sa=t&source=web&rct=j&opi=89978449&url=https://digitallibrary.un.org/record/4065096/files/A_79_545-ZH.pdf&ved=2ahUKEwiUzYzAu6uLAXWHN2IAHYNO_MbwQFnoECCwQAQ&usg=AOvVaw3E2A_ZmOYM1vW7LnFIRGsu,访问日期:2025-02-05。

② 该计划的10项中国行动分别是:1.开展人工智能领域南北合作、南南合作和三方合作;2.开展人工智能模型研发和赋能合作等产供应链国际合作;3.同各国尤其是发展中国家共同助力全球气候治理和可持续发展;4.搭建人工智能能力建设国际合作平台;5.同发展中国家共享人工智能教育资源;6.同发展中国家共同加强人力资源援助合作;7.同各国特别是发展中国家共同培育公众人工智能素养;8.同各国特别是发展中国家共同开展人工智能语料库建设;9.同各国特别是发展中国家共同促进和完善数据基础设施;10.同各国特别是发展中国家共同应对人工智能风险。

③ 中华人民共和国外交部:《人工智能能力建设普惠计划》,2024-09-27, https://www.mfa.gov.cn/web/wjzbhd/202409/t20240927_11498463.shtml,访问日期:2025-02-02。

④ Stephanie Laulhe Shaelou et al., “Challenges to Fundamental Human Rights in the Age of Artificial Intelligence Systems: Shaping the Digital Legal Order While Upholding Rule of Law Principles and European Values”, *ERA-Forum*, Vol. 24, No. 4, 2024, pp. 567-587, here pp. 573-579.

⑤ 《深圳经济特区人工智能产业促进条例》,深圳政府在线,2022-11-01, http://www.sz.gov.cn/cn/xxgk/zfxxgj/zwdt/content/post_10085826.html,访问日期:2025-01-31。

AI 重大法律问题,^①跟进国际前沿立法动态。例如,《AI 公约》在 AI 系统和基本权利保护关系上有体系性规定,美国 NAIIA 整合了应对 AI 不同方面影响的多个 AI 监管部门。^② 这些制度都为我国监控 AI 技术发展、实时跟进和法律追踪提供了思路。面对各国 AI 监管强度差异,我国可以参照《AI 公约》的国际条约造法经验,通过区域或多边国际法文本尝试协调和统一多方 AI 技术标准和监管政策,在国际法造法过程中吸收他国 AI 治理经验的可取之处。

第二,鉴于各国 AI 技术发展和治理水平差异,我国应积极适应 AI 技术跨国发展,承担大国责任,持续深化 AI 治理国际合作。《AI 公约》就区域人权保护提供了监管思路,显示了国际法律原则一体化、多国共享的态势。^③ 我国可以顺应这一趋势,在共建“一带一路”倡议、亚洲基础设施投资银行(亚投行)等多边经济框架下继续着力发展 AI 国际法文本,参与 AI 治理标准推广。当前,一系列国际倡议、决议历经积累,形成了 AI 国际软法治理矩阵。除我国《AI 治理倡议》《AI 能力决议》《AI 普惠计划》外,OECD《人工智能发展建议》、UNESCO《AI 伦理建议》、联合国《全球数字契约》、英国《布莱切利宣言》、法国巴黎 AI 行动峰会《关于发展包容、可持续的人工智能造福人类与地球的声明》等文本都在一定程度上整合了 AI 国际共识,^④提高了区域和多边国际法出台的可能性。《AI 公约》实质上也源自 AI 国际倡议,具有鲜明的欧洲色彩,其高标准、高人权关切、高技术要求的立法特征同其他国际软法形成分歧。我国可以在不同类型、立场的国际倡议中实现多方参与、兼容并蓄,借国际倡议的制定传播“中国声音”,形塑 AI 治理的全球多数立场。修订双边和多边协定,将 AI 监管规范引入既有国际条约也是可行路径。例如,我国可以基于已加入的 RCEP 第 12 章电子商务专章规定,在未来同 RCEP 其他成员国的经贸

① Thomas Burri, “International Law and Artificial Intelligence”, *German Yearbook of International Law*, Vol. 60, 2017, pp. 91–108, here p. 91.

② 包括 2020 年美国国防部设立的联合人工智能中心、美国总务管理局设立的人工智能卓越中心、美国能源部设立的人工智能和技术办公室、2021 年 OSTP 设立的国家人工智能研究资源工作组等。参见董汀、黄智尧:《人工智能国际治理与“不确定性”》,清华大学战略与安全研究中心,2023-12-14, <https://ciss.tsinghua.edu.cn/info/zlyaq/6788>, 访问日期:2025-02-06。

③ Christian Stracke et al., “Ethical AI and Education: The Need for International Regulation to Foster Human Rights, Democracy and the Rule of Law”, Paper Presented to the Conference on “Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky”, Communications in Computer and Information Science, Switzerland, 2024-07-02, https://link.springer.com/chapter/10.1007/978-3-031-64312-5_55, 访问日期:2025-02-27; David Leslie et al., “Artificial Intelligence, Human Rights, Democracy, and the Rule of Law: a Primer”, The Council of Europe, 2021-04-06, pp. 30–33, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3817999, 访问日期:2025-01-29。

④ 佐伊·克莱曼、丽芙·麦克马洪:《巴黎 AI 峰会:英国和美国拒绝签署全球 AI 声明》,BBC 中文网,2025-02-12, <https://www.bbc.com/zhongwen/articles/cge74rydqz4o/simp>, 访问日期:2025-03-06。

协定中引入 AI 相关电子商务条款,以体现我国 AI 治理思路。我国还可以借鉴 DEPA 第 8 条和 SADEA 第 31 条的 AI 治理共识性条款,在同其他发展中国家的经贸协作中复现同类型条款,修订或重新议定上述条款。《AI 公约》推出了多种协同监管机制,如创新透明度监管方式、创建监管沙盒、引入特殊监管指南或使用不采取行动函等。CAHAI 的解释性报告也鼓励各缔约方在 AI 系统全生命周期内纳入可靠和有效监督机制。^① 我国也可以在行业规则或产业标准中引入 AI 技术研发和产品生产的人工监督机制,以地方试点形式探索落实高水平治理规则,逐步与 AI 全球治理规则接轨。

结 语

数字时代 AI 治理的重要性与日俱增,国内法在跨国规制上存在功能欠缺,国际法在全球监管上存在制度缺位。《AI 公约》作为以确定统一监管标准、整合风险治理原则为目的的国际条约,在一定程度上推动了 AI 国际法治的进步,但也存在明显不足。《AI 公约》条款缺乏约束力和执行力,其立法模式未能发挥国际“超级软法”和国际硬法的优势,缔约参与度和国际接受度较低。该公约在很大程度上只是欧盟《AI 法》的翻版,没有提供更高层次的全球公共产品,不能真正满足 AI 全球治理的需要,并非国际社会迫切需要的国际法规范。

世界需要的 AI 国际条约应具备一些基本要素。一是具有法律约束力和强制执行力。AI 全球治理长期依赖行业自律和产业监管规范,而国际软法治理的缺乏直接影响了各国权利义务的功能。未来的 AI 国际条约应能直接确定各缔约方在 AI 全球治理层面的权利、分配各方的义务并对各方施以责任,配以惩罚性条款和争端解决条款,并提供较为完备的制度设计。二是具有风险治理导向、发展价值取向和公共产品属性。世界各国需要的 AI 国际条约须能回应各国共同关切,合理处置、评估乃至克服风险,在公共领域发展风险治理路径,为 AI 全球治理营造基于全球优化价值链(globally optimized value chain)的多方治理生态。三是具备普遍性、包容性和开放性。鉴于 AI 的无国界性,一份理想的 AI 国际条约也应当强化全球意识,突破区域保护主义桎梏,推动以指引性、合作性、协商性为特征的规范的形成。AI 国际条约需要为 AI 的自主学习能力留有制度空间,不局限于个别场景或领域,在监管中把握 AI 产业—技术—社会互动的全球互联性。^② 满足上述要求的

^① Council of Europe, “Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law”, 2024-09-05, pp. 22-24, <https://rm.coe.int/1680afae67>, 访问日期:2025-01-29。

^② András Hárs, “AI and International Law-Legal Personality and Avenues for Regulation”, *Hungarian Journal of Legal Studies*, Vol. 62, No. 4, 2021, pp. 320-344, here p. 340.

AI 国际条约,才会是真正具有监管力度、公共视野和全球影响的,世界需要的 AI 国际条约。

《AI 公约》在对我国造成客观影响的同时提供了 AI 全球治理的制度性启示。《AI 公约》意在体系化 AI 国际治理基本原则和规制工具,硬法化 AI 国际风险法治规则,强化欧洲 AI 法治的“布鲁塞尔效应”。《AI 公约》也将加剧全球各国 AI 治理的制度竞争,造成不同治理路径和模式的摩擦。在 AI 治理思路方面,我国可以参考《AI 公约》的规则核心和立法技巧,寻求权利保障与技术发展之间的平衡,以制度试点等方式确定最有利于我国 AI 产业发展和风险控制的监管路径。通过深化发展 AI 国际软法治理、积极参与 AI 国际硬法制定,我国可以进一步提升 AI 全球治理的话语权、影响力和竞争力,在成为全球科技强国的道路上获得国内法和国际法的双重制度助力,为人类社会提供更高水平、符合各国预期的 AI 治理公共产品。

责任编辑:郑春荣