

人工智能国际法与国内法的双向互动

沈 伟 陈徐安黎

摘 要 人工智能（AI）国际法与国内法的有效互动是规制 AI 风险、保障 AI 发展的法治要求。现有 AI 国际法包含部分直接规制的倡议型软法和区域性硬法以及间接规制的领域国际法。但受限于传统国际法的渊源和主体，既有国际法在适用于 AI 时适配性不足。欧美中分别制定了 AI 国内法，既需要 AI 国际法调和差异分歧，也为 AI 国际法提供了共识基础。国际法的国内法化和国内法的国际化在 AI 国际法与国内法的双向互动中都有所表现。AI 领域的非国家主体有赖于国内法规制，使得国际法向国内法延伸成为可能和必要。中国既需要不断调整和优化国内法，对 AI 发展作出立法回应，又需要提出 AI 国际法方案，推动国际法回应各国国内法需求，促进 AI 国际法软硬法二元协同治理模式的形成。进一步发挥国内法场景式规制的快速响应优势，实现国内法外溢作用，是中国在 AI 国际法方面有所作为的途径。

关键词 人工智能 国际法 国内法 规则外溢 硬法

作者沈伟，上海交通大学凯原法学院特聘教授（上海 200030）；陈徐安黎，上海交通大学凯原法学院博士研究生（上海 200030）。

中图分类号 D9

文献标识码 A

文章编号 0439-8041(2025)10-0110-11

引言

21 世纪以来，AI 技术作为一种底层技术，已经嵌入社会生活的各方面与全过程，深刻改变了人类的生产和生活方式、组织形式、社会结构和国际局势。习近平主席指出，加快发展新一代人工智能是“推动我国科技跨越发展、产业优化升级、生产力整体跃升的重要战略资源”^①，需要通过法治的途径规制 AI 的相应风险，保障中国 AI 技术的健康发展。

目前，AI 法治存在着国内法和国际法两种发展和研究路径。^② 但现有研究或是专注于国内法的制定，或是聚焦于国际法的法理分析，国内法和国际法的研究路径彼此分离，自成体系，缺乏对 AI 国内法与国际法之间互动的分析，未能全面揭示 AI 法治发展的完整图景。一方面，AI 技术具有突破单一国家边界的特性，涉及跨境数据流通、芯片和半导体全球供应链等方面，需要全球协同配合，这为 AI 国际法的发展提供了物理基础。单纯的国内法路径无法满足 AI 跨国界风险扩散的规制需求，也难以形成对全球 AI 安全博弈和技术竞争的调和机制和应对方案。另一方面，现有 AI 的国际法发展尚处于初步阶段，单纯的国际法路径缺少实在法的支持，且各国目前主要倾向于通过国内法的方式对 AI 应用进行监管，AI 国际法的发展无法脱离国内法的语境。

① 习近平：《加强领导做好规划明确任务夯实基础，推动我国新一代人工智能健康发展》，《人民日报》2018 年 11 月 1 日，第 1 版。

② 国内法路径参见林涓民：《论人工智能立法的基本路径》，《中国法学》2024 年第 5 期等。国际法路径参见沈伟、吴柄萱：《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》，《德国研究》2025 年第 2 期等。

国内法治与国际法治虽然分属两个不同的法治体系，但相互关联、影响和作用。国际法和国内法之间存在着两种互动模式，包括以国际法的国内法化为代表的开放型模式和以国内法的国际法化为代表的内顾型模式^①，在当前 AI 国际法与国内法的互动过程中都有所表现。在全球化推进过程中，国际法对国内法的影响力大于国内法对国际法的影响力；在逆全球化思潮下，国内法的影响力更为明显。基于不同的时代背景，二者的相互作用会有所转向，体现了全球化和逆全球化的张力。^② AI 技术发展恰恰发生在逆全球化的周期阶段，国内法的影响力更加明显，国际法的作用明显式微。AI 的应用与发展需要国内法与国际法的双向有效互动。

本文以 AI 国内法与国际法的互动为研究对象，在梳理 AI 现有国际法和主要国家国内法的基础上，分析国际法和国内法具有的共识与呈现的差异，最终提出保障中国 AI 发展与安全的国内法与国际法互动路径。

一、AI 国际法的多重困境

目前，AI 国际法主要包括直接规制型和间接规制型两类：直接规制型国际法是专门针对 AI 应用制定的法律规则，以软法和伦理规范为主，以区域性和部门法硬法为辅，但约束力和适用范围有限，缺乏 AI 多边国际条约。间接规制型国际法则是利用既有的国际法体系对 AI 应用进行规制，主要通过现有国际经贸条约的解释加以实现，但存在碎片化的现象，针对性不足。基于此，目前 AI 国际法存在着局限性，需要与国内法配合和互补。

（一）直接规制路径的约束力问题

直接规制 AI 的国际软法不是有效的国际法渊源，缺乏强制约束力，实际效用有限。国际法渊源通常是指国际法的存在和表现形式。^③ 1945 年《联合国宪章》序言指出各国应“尊重由条约与国际法其他渊源而起之义务”，首次提出国际法渊源的概念，并将其与国家的法律义务绑定。^④《国际法院规约》第 38 条被认为是国际法渊源的权威表述，规定了国际条约、国际习惯和一般法律原则三种国际法渊源。^⑤ 目前国际社会缺乏 AI 国际条约，而国际习惯和一般法律原则的形成具有长期性特征，难以适应 AI 技术的迅速发展。即使部分学者提出“速成国际法理论”，主张不考虑国家实践的持续性或时间因素^⑥，但国际习惯需要各国通过一般实践接受法律规则，至少要求国家对于该规则具有法律确信^⑦，而 AI 国际法难以满足这一条件。

2024 年 9 月 5 日，欧洲委员会发布的《人工智能与人权、民主和法治框架公约》（以下简称《框架公约》），是全球第一份 AI 领域具有法律约束力的国际条约，也是目前直接规制 AI 最主要的国际硬法规范，但尚未生效。以《框架公约》为代表的 AI 国际法一定程度上弥补了国际软法的强制力缺陷，通过对缔约方设置一定的实体义务，并建构相应的组织和机制加以协调配合，为直接规制 AI 应用提供了具有一定法律约束力的强制性规范。不过，《框架公约》对于缔约方义务的表述仍然较为模糊和抽象，在多数义务条款中使用了“鼓励”“呼吁”“视情况”“努力”等措辞，将遵守缔约方的国内法作为缔约方履行《框架公约》义务的前提条件，事实上赋予了缔约方在履行条约义务时较为宽泛的自由裁量权，实际约束力有限。从法律效力和责任分配看，《框架公约》只是一个 AI 国际倡议的有限升级版或弱条约，其中的规定如何进一步受到国际社会的认可和接受存在未知数。

（二）间接规制路径的碎片化问题

目前 AI 领域以间接规制型国际法为主，相关规范散见于不同领域的部门国际法，缺乏统一的协调路径。

① 参见徐秀军：《规则内化与规则外溢——中美参与全球治理的内在逻辑》，《世界经济与政治》2017 年第 9 期。

② 参见沈伟、苏可桢：《变局之下国际法与国内法互动的转向》，《探索与争鸣》2024 年第 4 期。

③ 参见李浩培：《国际法的概念与渊源》，贵阳：贵州人民出版社，1994 年，第 52 页。

④ 参见《联合国宪章》序言。

⑤ See Statute of the International Court of Justice, Article 38.

⑥ See Bin Zheng, *Studies in International Space Law*, Oxford: Oxford University Press, 1997, pp. 136–141.

⑦ 参见郑斌：《国际法的渊源：国际习惯法和条约》，李斐南译，《环球法律评论》1986 年第 1 期。

第一，将 AI 相关的商务贸易内容纳入国际经贸协定或数字经济协定。例如，澳大利亚和新加坡签署的《数字经济协定》专条规定缔约双方关于 AI 的共识，包括双方应通过相关的区域和国际论坛，合作并促进框架的制定和采用，以支持可信、安全和负责任地使用 AI 技术，在制定此类 AI 法律框架时，考虑国际公认的原则或指导方针。^① 尽管上述国际条约在规范层面具有法律约束力，但由于其使用的措辞具有较大的自由裁量空间和非强制性，法律强制效果更接近于宣示性立场表述^②或“最佳努力”（best efforts）的承诺，不能等同于缔约方作出了 AI 方面的权利分配或义务承诺。

第二，通过演进解释的方式将 AI 纳入传统的国际法框架，并利用具有强法律约束力的国际法机制解决 AI 技术发展产生的法律争议。尽管从文义解释和目的解释的角度看，WTO 成员方无意将条约缔结时的承诺扩张适用于 AI 技术，WTO 规则不应被解释为涵盖 AI 技术的应用问题。^③ 但从国际法实践来看，国际法院在判例中认为，对条约的概念可以采用演进解释的方法，结合时代背景对国际法的文本含义作出进化性的解释^④，这一理念也被 WTO 争端解决机构广泛应用于争端解决实践中。^⑤ 由于《服务贸易总协定》（GATS）关于服务的定义十分广泛，AI 技术可以在一定程度上被纳入服务的范畴，从而适用 WTO 的相关规定。

第三，基于一般国际法原则对 AI 的技术风险进行规制。例如，国家主权原则不仅可以直接适用于实体空间，也可以适用于网络空间等虚拟空间，对网络空间的国家行为施加约束，违反主权原则也将引发国际责任。^⑥ 部分国家还认为，任何未经一国授权侵入其信息系统或通过数字方式在其领土范围内造成影响的行为，都构成侵犯该国主权的行。⑦ 当一国利用 AI 技术对他国主权施加不利影响或侵入他国 AI 系统时，主权原则是当然的适用原则。再如，不干涉内政原则禁止一国对他国事务施加旨在强加某种行为或后果的强制的或专断的干预^⑧，有学者认为强迫性源于国家间传统交互语境，在数字时代更宜通过持续且低烈度的“利用”（exploitation）标准代替强迫性标准，以规制数字虚拟空间的干涉行为^⑨，这为适用不干涉内政原则规制 AI 国际安全风险提供了理论基础。

二、AI 国内法的差异化路径

欧盟、美国与中国是当前全球 AI 发展中的“主导性数字力量”^⑩，但在国内法层面存在着差异化的模式选择，代表着 AI 国内法的主要发展方向。国内法的差异化路径既为开展 AI 领域的制度竞争提供了条件，又加深了全球 AI 法律体系的碎片化倾向^⑪，仅关注国内法的规制路径将加剧全球 AI 法律体系的复杂性和破碎化。三方国内法模式的共识为国际法提供了规范性基础。

（一）主流 AI 国内法的差异化路径

欧美中 AI 国内法采取了三种路径，关注的焦点有所不同。

作为一个特殊的区域性国际组织，欧盟立法存在多种形态，最常见的包括一级立法和二级立法两种形式。

① See Singapore-Australia Digital Economy Agreement, Article 31.

② See Marta Soprana, “The Digital Economy Partnership Agreement (DEPA): Assessing the Significance of the New Trade Agreement on the Block,” *Trade, Law and Development*, vol. 13, no. 1, 2021, p. 161.

③ See Anupam Chander, “Artificial Intelligence and Trade,” in *Big Data and Global Trade Law*, Mira Burri (ed.), Cambridge: Cambridge University Press, 2021, p. 120.

④ See Aegean Sea Continental Shelf Case, (1978) I. C. J. Report.

⑤ See Appellate Body Report, United States-Import Prohibition of Certain Shrimp and Shrimp Products, WT/DS58/AB/R, November 6, 1998.

⑥ 参见王超：《主权原则在网络空间适用的理论冲突及应对》，《法学》2021 年第 3 期。

⑦ See United Nations Office for Disarmament Affairs, International Law Applied to Operations in Cyberspace, December 7, 2021, <https://documents.unoda.org/wp-content/uploads/2021/12/French-position-on-international-law-applied-to-cyberspace.pdf>.

⑧ 参见詹宁斯、瓦茨修订：《奥本海国际法》，王铁崖等译，北京：中国大百科全书出版社，1995 年，第 313—315 页。

⑨ See Michael P. Fischerkeller, Current International Law Is Not an Adequate Regime for Cyberspace, April 22, 2021, <https://www.lawfaremedia.org/article/current-international-law-not-adequate-regime-cyberspace>.

⑩ See Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology*, Oxford: Oxford University Press, 2023, p. 13.

⑪ See United Nations Secretary-General's High-level Advisory Body, Interim Report: Governing AI for Humanity, December 2023, https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf.

不同于一级立法以《欧洲联盟条约》等欧盟成立和运行的基础性条约为核心，二级立法主要表现为欧盟立法机构根据基础条约的规定，对社会生活各领域进行管理和规制的条例（Regulation）、指令（Directive）和决定（Decision）等规范。^① 虽然欧盟在许多方面不同于主权国家，但在行使自主立法权制定二级立法时，具有与主权国家非常近似的性质。从制定机关、调整对象以及法律效力来看，欧盟这类立法完全符合国内法的特征。^② 2024年3月，欧洲议会正式通过全球第一部AI专门法——《人工智能法》（以下简称《AI法》），采用基于风险的监管模式，与欧盟《通用数据保护条例》相结合对AI系统的风险进行评估与定级。欧盟的《AI法》在性质上属于二级立法，可以代表欧盟的“国内法”模式。该法采用基于风险分级的全过程规制体系，设置了综合性的AI风险治理框架，覆盖更多的AI技术、相关主体与应用场景。^③ 欧盟《AI法》根据AI应用可能对安全或基本权利造成的不利影响，区分不可接受、高风险、有限风险以及低风险或最小风险四类风险等级。除了低风险或最小风险类的AI服务提供者无需遵循相应义务外，其余三类的AI服务提供者需依照不同的风险等级承担不同程度的义务。欧盟《AI法》采取了综合性的AI立法模式，以统一化的风险分类分级管理作为规制AI应用的主要方式，较为保守。

美国的AI国内法则是采用市场化、分散化的进路。联邦层面的AI国内法缺乏综合性AI法律，也较少设置专门适用于AI的国内法，更多是针对AI国际投资、国际贸易、数据传输、产业研发等国内、国际事项加以规制。例如，2018年美国《出口管制改革法》（ECRA）对新兴和基础技术（emerging and foundational technologies）规定专门的识别程序，实施长期监管和限制。^④ 美国商务部工业安全局（BIS）随后发布的《拟议规则制定事先通知》，将AI纳入上述出口管制范围。^⑤ 2025年1月13日，美国发布《人工智能扩散框架》，对特定封闭式AI模型权重实施出口管控，同时引入了新的许可例外条款。^⑥ 美国的许多AI国内法具有较强的外向战略意图，与全球AI博弈相联系。在州一级，AI国内法主要聚焦于基本人权保障、消费者权益保护、反垄断和反歧视等方面。2024年5月17日，美国科罗拉多州颁布《关于在与人工智能系统的互动中进行消费者保护的法案》，成为美国第一部对AI重点应用场景进行规制的州法。^⑦ 该法从防止算法歧视入手，要求AI的开发者和提供者尽到合理注意义务（reasonably care）以保障相关AI场景下的消费者权益。^⑧ 美国AI国内法对技术发展和科技创新具有较高的包容性^⑨，采用了较为宽松的监管路径，赋予相关市场主体更大的经营自主权。

中国AI国内法没有采用风险导向或者市场导向的AI立法模式，注重平衡AI发展与安全需求。目前，中国暂未制定AI的专门性法律，现有AI监管的国内法体系主要以《民法典》与《网络安全法》《数据安全法》和《个人信息保护法》四部法律中的个人信息与数据处理规范为基础性规则，以部门规章或规范性文件的具体规则为补充，对特定场景的AI应用进行规制。2023年7月国家互联网信息办公室等部门联合发布的《生成式人工智能服务管理暂行办法》（以下简称《暂行办法》），从数据收集、算法训练、服务提供、反歧视等多方面对生成式AI的输入和输出端进行了较为全面的规范，是我国现阶段专门性的AI国内法。整体而言，中国的AI国内法充分平衡AI的创新发展需求和风险防控需求，基于AI的不同应用场景和阶段进行区分，

① 参见蒋小红：《欧盟法的域外适用：价值目标、生成路径和自我限制》，《国际法研究》2022年第6期。

② 参见赵怀普：《欧盟政治与外交》，北京：世界知识出版社，2021年，第178页。

③ 参见沈伟、陈徐安黎：《全球人工智能治理模式比较与中国进路》，《亚太经济》2025年第3期。

④ See ECRA Article 1758.

⑤ See BIS, Review of Controls for Certain Emerging Technologies, November 19, 2018, <https://www.federalregister.gov/documents/2018/11/19/2018-25221/review-of-controls-for-certain-emerging-technologies>.

⑥ See BIS, “Framework for Artificial Intelligence Diffusion,” *Federal Register*, January 15, 2015, <https://www.federalregister.gov/documents/2015/01/15/2015-00636/framework-for-artificial-intelligence-diffusion>.

⑦ See Craig A. Waldman et al., Colorado Enacts AI Consumer Protection Legislation, June 19, 2024, <https://www.mondaq.com/unitedstates/dodd-frank-consumer-protection-act/1480624/colorado-enacts-ai-consumer-protection-legislation#authors>.

⑧ See Colorado General Assembly, Consumer Protections for Artificial Intelligence Concerning Consumer Protections in Interactions with Artificial Intelligence Systems, May 17, 2024, <https://leg.colorado.gov/bills/sb24-205>.

⑨ See Daniel J. Gervais, “The Regulation of Inchoate Technologies,” *Houston Law Review*, vol. 47, no. 3, 2010, pp. 665–706.

对于其中存在较为突出风险的部分先行制定相应的国内法加以回应, AI 国内法处于快速发展和完善的阶段。

(二) AI 国内法的分歧与共识

欧美中差异化的国内法路径体现在三者不同的立法模式、理念上, 导致了规范 AI 应用的立法碎片化(见表 1)。欧盟《AI 法》提供了以风险为基础的国内法模板, 但其静态化的风险评估模式, 难以应对国际风险社会背景下 AI 风险快速发展变化的现状。在风险社会中, AI 技术发展的复杂性和不确定性要求国内法对 AI 安全风险的规制具有灵活性与适应性, 等级化的控制难以有效应对动态环境实施针对性管控。^① 与此相对, 中国的场景式立法具有快速响应的优势,《暂行办法》即使与理想的生成式 AI 系统规则可能尚有距离, 但从征求意见到落地实施仅历经三个月左右的时间, 可以有效应对 AI 实践变化。中国 AI 国内法通过重点场景规制的模式, 在保持法律稳定性的同时, 能以更为灵活的方式应对 AI 跨领域和多场景的发展, 对规制 AI 的具体应用发挥了一定的作用^②, 但在法律的体系性和监管的整体性方面还有提升空间。美国的 AI 国内法模式包含国内和国际的双重视角, 在促进国内 AI 快速发展并保持美国技术和经济优势的同时, 对主要竞争对手国家采取遏制、限制和制裁措施, 阻碍 AI 技术流动, 加剧国际 AI 博弈与竞争。美国对于风险规制的关注不足也为日后 AI 风险的集中爆发留下隐患, 可能导致相应安全风险外溢, 对全球安全造成威胁。

表 1 欧美中 AI 国内法模式比较

	欧盟	美国	中国
立法模式	综合性专门立法	分散化立法	不同层级立法结合
立法理念	风险导向	市场导向	兼顾发展与安全
优势	体系性强	包容性强	快速响应
不足	滞后性	风险规制不足	体系性较弱

欧美中在 AI 国内法路径上的差异和具体适用上的偏向给跨国 AI 企业或 AI 跨国服务的发展和规制造成障碍, 导致法律适用冲突或规制“长臂管辖”, 也使相关企业存在监管套利的空间, 影响 AI 风险规制的成本和成效。部分国家甚至可能利用 AI 国内法中的人权、法治等具有较为抽象和较高自由裁量权的概念对竞争对手国家实施变相的限制、约束和制裁。例如, 欧美等已在军事、关键矿产或其他领域以可持续发展、人权保护、劳工保护等因素为由对中国企业跨境投资、贸易、数据流动等活动施加限制, 进行经济制裁^③, 直接影响 AI 的发展。同时, 一旦 AI 成为大国博弈和竞争的底层技术, AI 国内法也会陷入安全困境, 个别国家的国内法会存在非理性因素, 加剧技术博弈和恶性竞争, 导致供应链断裂, 扩大技术鸿沟, 不利于全球技术合作。AI 国内法需要国际法的协调和约束。

三、AI 国际法的国内法化形态

(一) AI 国际法的调整对象向国内法延伸

当前 AI 国际法与国内法的调整内容和对象存在着大量重叠。欧洲《框架公约》作为 AI 直接规制型国际法的典型代表, 以确保 AI 活动符合人权、民主和法治要求作为立法目的, 涉及定义与范围、缔约方需要遵守的一般义务、AI 活动的原则、风险评估与缓解等实体性内容, 并对公约的救济措施、实施机制和后续的合作机制等程序性内容加以规定。^④

① 参见范如国:《“全球风险社会”治理: 复杂性范式与中国参与》,《中国社会科学》2017 年第 2 期。

② 参见史凤林、张志远:《论人工智能的公法规制: 美欧模式与中国路径》,《理论月刊》2023 年第 8 期。

③ 参见沈伟、陈徐安黎:《供应链尽职调查国内立法的国际法分析——以 WTO 规则为切入点》,《国际商务——对外经济贸易大学学报》2024 年第 6 期; 沈伟、陈徐安黎:《可持续供应链合规规则的国际法边界与中国路径》,《财贸研究》2025 年第 1 期。

④ See Council of Europe, Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, September 5, 2024, <https://rm.coe.int/1680afae3c>.

在一般义务层面，欧洲《框架公约》规定了缔约方需要承担的两项基本义务，即保护人权和尊重法治。^①在活动原则层面，《框架公约》多维度设置了包括尊重人格尊严和自由原则、透明度原则、问责制原则、平等和非歧视原则、隐私和数据保护原则、可靠性原则以及促进安全创新原则在内的7项AI活动原则。^②在救济措施层面，欧洲《框架公约》从国内救济、信息共享和保障知情权、申诉权三方面对缔约方的救济义务作出规定。首先，缔约方应通过符合国际法和国内法的措施确保AI活动导致的侵犯人权问题存在可救济的途径，包括建立相应的法律制度、为受害人提供公平公正的赔偿等方面。其次，缔约方有义务记录对人权存在重大影响的AI活动信息，并保障公众对相关信息的获取权和对AI活动的监督权。最后，缔约方应保障受影响主体对相关信息的知情权以及对于AI活动的申诉权。^③

由于AI国际法的调整对象向着国内法延伸，将约束各国国内规则的形成和国内法的制定。整体上看，《框架公约》虽然以国家义务与责任作为媒介，但其主要目的不在于实现国家之间的利益和责任分配，而是保护国家内部个人的基本权利，最终调整的法律关系主要是国内的AI企业与消费者、AI服务提供者与使用者之间的关系，本质上是国内法调整的范围。对于美国等缔约方而言，欧洲《框架公约》规定缔约方需要酌情建立分类分级的风险管理体系，会产生溢出效应，影响缔约方的国内法，可能会改变美国现有市场化、分散化的国内法路径选择。

（二）AI国际法促使国内法趋同

目前，AI领域形成的一些国际共识，正在通过国际软法的形式得以呈现。2023年11月，欧美中等26国签署了全球第一份针对AI的国际性声明《布莱切利宣言》，承诺以安全可靠、以人为本、可信赖及负责的方式设计、开发、部署并使用AI，主张建立有效的问责机制，提高AI决策的透明度以解决“算法黑箱”问题，确保AI的应用符合公众利益。^④该宣言体现了世界主要国家和地区在AI监管方面形成的国际共识，其中强调AI系统的安全性和可靠性，呼吁采取必要措施确保其安全可靠运行，防止AI系统的误用或滥用。2024年9月23日，联合国未来峰会通过了《全球数字契约》，与会各方在弥合数字鸿沟、促进数字经济、营造数字空间、推进数据治理和加强AI治理五个方面，对全球数字治理的目标、原则和行动达成基本一致。^⑤《全球数字契约》虽不具有法律约束力，但代表着多数国家在数字治理中的共识，奠定了AI国际法的基本框架，为相关国际硬法的制定提供了方向和指引，也为各国国内法与国际法的协调提供了兼容性方案。

现有AI国际法基本上反映了主要国家国内法的共识，并促使这些国家国内法逐步趋同（convergence）。2023年5月联合国发布的政策简报强调“透明度、公平和负责任是AI治理的核心”。^⑥《全球数字契约》进一步规定了“平衡、包容和基于风险的方法治理AI”的共识。^⑦

欧美中在AI监管思路有所区别的同时，在监管手段和监管内容上存在一定共识，与AI国际法的核心路径基本一致。例如，美国对于高风险AI活动的特别关注和严格监督^⑧，以及中国《互联网信息服务算法推荐管理规定》基于年龄、地域等特征界定算法推荐服务的级别的规范和对算法推荐服务的监管主体与供应者义务进行分级分类的规定^⑨，都体现了与欧盟类似的风险分级管理思路。以欧美中为代表的AI国内法不仅立法

①②③ See Council of Europe, Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Article 4–5, 7–13, 14–15, September 5, 2024, <https://rm.coe.int/1680afae3c>.

④ See “The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023,” Gov. UK, November 1, 2023, <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>.

⑤⑦ See United Nations, The Pact for the Future, September 20, 2024, <https://documents.un.org/doc/undoc/ltd/n24/252/89/pdf/n2425289.pdf>.

⑥ See United Nations Office of the Secretary-Generals Envoy on Technology, Our Common Agenda: Policy Brief 5: A Global Digital Compact — an Open, Free and Secure Digital Future for All, May 2023, <https://indonesia.un.org/sites/default/files/2023-07/our-common-agenda-policy-brief-global-digi-compact-en.pdf>.

⑧ See Colorado General Assembly, Consumer Protections for Artificial Intelligence Concerning Consumer Protections in Interactions with Artificial Intelligence Systems, May 17, 2024, <https://leg.colorado.gov/bills/sb24-205>.

⑨ 参见《互联网信息服务算法推荐管理规定》第3条、第18—19条。

理念存在一定的一致性，也与作为国际硬法的欧洲《框架公约》中的风险评估与缓解的规范相契合，同时符合国际软法的相关原则性要求和国际共识。这说明各国 AI 国内法存在着“最大公约数”，可能成为未来国际法的基础性要素。

（三）AI 国际法空白阻碍国内法运行

目前国际法直接或间接规制 AI 应用存在着一定的局限性和滞后性，导致 AI 国际法与 AI 监管需要的不匹配，形成了法律缝隙。在一般国际法原则层面，国家主权原则、不干涉内政原则虽然可以提供原则上的指引，但违反国家主权原则的门槛难以划定，缺乏权威的判断主体，存在较大争议^①，不干涉内政原则在现有国际法框架中可适用的空间有限。

在国际条约层面，国际法诸多领域均受到 AI 技术发展的挑战，在适用上存在困难。以贸易领域为例，虽然 WTO 相关条款可以适用于 AI 的相关应用，但由于 WTO 文本的滞后性以及分类的僵化，WTO 规则体系采用《关税和贸易总协定》规制货物贸易与 GATS 规制服务贸易的二分法以及服务贸易具体承诺表服务类别的限定对于运用 WTO 规则规制 AI 造成了困难。^② 同时，生成式 AI 已初步具备人类能力特征，新生问题超出法律规制范围。^③ 而这些都源于国际法本身针对传统安全的有限性和时间生成上的滞后性特点与 AI 风险具有的易变性、动态性特点之间的不匹配。

AI 国际法的供给不足对各国监管 AI 的效能和国家安全造成不利影响。例如，Open AI 公司提供的 ChatGPT 生成式 AI 服务运用了全球互联网三分之二的数据进行算法训练从而提升技术的覆盖性^④，其中涉及大量的数据跨境流通。随着 AI 技术的迅猛发展和全球化应用，其所带来的技术复杂性和综合性已经超出传统国际社会安全体系的安全供给能力，AI 技术发展存在的国际安全风险和对国际社会治理体系提出的挑战难以得到既有国际法的有效覆盖和解决，加重全球安全赤字^⑤，威胁各国的国家安全，影响 AI 国内法的发展。

四、AI 国内法的国际化形态

（一）大国 AI 博弈影响 AI 国际法的形成

大国 AI 博弈和 AI 国际法制定权和话语权竞争是 AI 国内法的国际化化的重要原因和表现。新一代 AI 技术的可持续发展不仅是技术层面的竞争，也是法治软实力和规则先发优势的竞争。^⑥ 欧盟《AI 法》是以 AI 国内法影响国际法的欧洲方案，也是欧盟争夺全球 AI 监管模式和规则话语权的法律工具，促进了欧盟《AI 法》向欧洲《框架公约》的转化。欧盟试图借此实现“布鲁塞尔效应”，推动 AI “欧洲标准”全球化。^⑦ 但是，AI 国内法的竞争在促使各国提升本国 AI 立法水平的同时，也可能基于竞争因素的影响，在风险规制中掺杂非理性因素的考量。例如，对于“不可接受”级别的 AI 系统，欧盟强调对于核心权利和民主价值观的维护，但在风险分级时未经风险评估，更多体现了欧盟在 AI 规制上的政治判断，以显示欧盟与美中 AI 国内法范式的区别，体现出更具象征意义的监管姿态和价值取向。^⑧

各国通过扩张 AI 国内法域外管辖范围或者实施 AI 国际经贸往来的国内法限制等方式进行国际竞争，可能违反各国承担的国际法义务，对 AI 国际法的发展和完善造成不利影响。由于 AI 技术具有国家整体战略和

① 参见黄志雄、陈徽：《供应链安全国际法保护的困境与出路——以“太阳风”事件为切入点》，《厦门大学学报（哲学社会科学版）》2022 年第 1 期。

② 参见戴艺晗：《国际贸易法视域下的人工智能规制——以 WTO 规则为视角》，《上海财经大学学报》2023 年第 2 期。

③ 参见袁曾：《生成式人工智能治理的法律回应》，《上海大学学报（社会科学版）》2024 年第 1 期。

④ 参见何哲：《ChatGPT：开启通用人工智能时代及其社会影响》，《电子政务》2023 年第 4 期。

⑤ 参见沈伟、陈徐安黎：《“安全赤字”概念考：政策、学理和国际法》，《国家安全研究》2024 年第 5 期。

⑥ 参见张凌寒：《中国需要一部怎样的〈人工智能法〉？——中国人工智能立法的基本逻辑与制度架构》，《法律科学（西北政法大學學報）》2024 年第 3 期。

⑦ See Anu Bradford, *The Brussels Effect: How the European Union Rules the World*, New York: Oxford University Press, 2020, pp. xiii-xix.

⑧ See Regine Paul, European Artificial Intelligence “Trusted Throughout the World”: Risk-Based Regulation and The Fashioning of a Competitive Common AI market, December 11, 2023, https://bora.uib.no/bora-xmlui/bitstream/handle/11250/3117993/2023_Paul_RegGov_EU%2brisk-based%2bAI%2band%2bcompetitiveness.pdf?sequence=1&isAllowed=y.

安全的意义, 各国全方位竞争的动机, 使国家重新选择进攻现实主义的对外战略^①, 甚至演变成具有侵略性的积极技术遏制和博弈。例如, 欧盟通过制定外商投资国家安全审查框架, 对于 AI 等新兴技术的外商投资进行严格审查, 并且在缺乏明确理据的情况下经常性地禁止或严格限制特别关注国家的相关对外投资, 显然违反了欧盟各成员国对外缔结的双边投资协定的相关规定。^② 各国通过国内法的对外扩张、“长臂管辖”和歧视性适用等方式影响国际法的适用, 已成为 AI 国际法与国内法互动的一大特点。

(二) AI 国内法的涉外法治建设促进国际法完善

各国 AI 国内法的具体规范正在不同程度地为国际法完善提供“原料”, 影响着 AI 时代国际法的发展方向。一个典型的例子是, 欧盟《AI 法》作为全球最受关注的首部 AI 综合性国内法, 已对各国国内法和国际法产生不同程度的示范效应^③, 尤其体现在欧洲《框架公约》的出台。欧盟《AI 法》基于风险预防的理念, 制定了覆盖全过程的 AI 风险规制体系。与此相类似, 欧洲《框架公约》规定了相应的风险管理框架, 要求缔约方采取措施评估、预防和缓解 AI 活动的风险, 根据 AI 对于人权、法治和民主造成影响的可能性和影响程度, 对 AI 风险进行分类分级管理, 适当考虑潜在影响的严重性和可能性, 充分考虑利益相关方的意见, 并采取风险管理方法以确保 AI 活动的不利影响得到缓解和解决。^④ 尽管欧洲《框架公约》与欧盟《AI 法》在规则的完善程度上有所差异, 但在整体的法律体系建构上具有一致性, 均秉持以风险为基础的立法理念, 均突出人权、民主等价值取向。这深刻反映了欧盟《AI 法》对欧洲《框架公约》的影响, 也体现出 AI 国内法具有引领和带动国际法发展的能力。

加强 AI 国内法的涉外法治建设是增强国内法国际影响力的重要方式, 对于 AI 全球法治发展具有重要意义。2023 年 10 月中国提出的《全球人工智能治理倡议》(以下简称《倡议》)从 AI 的新理念和新业态出发, 着眼于 AI 的发展、安全和治理三重维度, 全面阐发了以人为本、智能向善等 11 项中国主张。《倡议》强调了 AI 发展应当遵守国际法, 尊重各国主权和管辖权, 秉持相互尊重、平等互利的原则, 符合和平、发展、公平、正义、民主、自由的全人类共同价值, 并确保各国 AI 发展与治理的权利平等、机会平等、规则平等。^⑤ 《倡议》反映了中国 AI 国内法积极走向国际法化的趋势, 通过以国内法的形式声明中国在 AI 全球发展中所坚持的理念、宗旨、原则和要求, 为提升全球 AI 治理能力、塑造 AI 国际法新秩序作出了重要贡献。^⑥ 2024 年 7 月, 中国在《倡议》的基础上提出加强 AI 能力建设国际合作决议, 强调以人为本、智能向善的发展原则, 在第 78 届联合国大会以协商一致的方式通过, 全球 140 多个国家参加决议联署, 充分体现出全球 AI 治理的中国影响力。^⑦ 联合国《全球数字契约》强调发展优先、普惠包容、创新驱动、保护数字空间文化多样性, 营造开放、公平、包容和非歧视数字发展环境等原则与中国提出《倡议》以及全球发展倡议、全球文明倡议的发展路径相一致^⑧, 中国 AI 治理理念通过涉外法治建设对于全球 AI 国际法的发展已经产生影响, 发挥了国际规则的塑造功能。

① 参见王悠、陈定定:《迈向进攻性现实主义世界?——人工智能时代的国际关系》,《当代世界》2018 年第 10 期。

② 参见沈伟、陈徐安黎:《欧盟供应链安全规则的合法性认定——以中国与欧盟各国双边投资协定为切入点》,《上海财经大学学报》2024 年第 3 期。

③ 参见张凌寒:《人工智能法律治理的路径拓展》,《中国社会科学》2025 年第 1 期。

④ See Council of Europe, Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Article 16, September 5, 2024, <https://rm.coe.int/1680afae3c>.

⑤ 参见《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。

⑥ 参见费艳颖、赵凯:《全球人工智能治理倡议的世界意义》,《中国社会科学报》2024 年 6 月 19 日, 第 A06 版。

⑦ 参见《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》, 2024 年 7 月 2 日, https://www.gov.cn/yaowen/liebiao/202407/content_6960524.htm。

⑧ 参见邢翀、郭超凯:《联合国未来峰会通过“全球数字契约”中方:将与国际社会一道推动契约落地》, 2024 年 9 月 23 日, <https://www.chinanews.com.cn/gn/2024/09-23/10290761.shtml>。

五、互动视角下 AI 国际法与国内法发展的中国路径

（一）推动 AI 国际法发展的中国方案

在国内法与国际法互动的视角下，AI 国际法需要回应 AI 国内法的碎片化特征，突出普遍性特征和协调性方案，寻求各国 AI 国内法共识，促进各国国内法趋同，以防止各国国内法竞争造成的规则碎片化甚至是竞争性现象。在国际经贸领域，WTO 为 AI 国际法铺设了框架性指引和协调监管渠道，可以作为促进直接规制型 AI 国际法统一，并提供强制力支撑的重要平台。WTO《贸易技术壁垒协议》（以下简称“TBT 协议”）从货物贸易角度提供了各国 AI 监管协调的具体平台与工具。该协定通过鼓励成员国以国际标准为基础制定措施，从透明度、公开性、公正性、有效性、一致性和发展等六大维度驱动 AI 监管的国际协调进程，以确保技术法规、标准不具有歧视性，平衡 AI 技术所需的国内监管需求和国际流通需求。TBT 协议以促进国内法与国际标准的一致性为目标，其提供的协调性方案以国际标准为桥梁和媒介，提供了国际软法向国内硬法的转化通道。TBT 协议的转化路径为实现各国 AI 的协同监管提供了新的思路，基于国际标准这一关键工具，使各国可以通过与 AI 国际标准对齐的方式达成监管一致性。^① 中国需要积极引领 AI 国际法趋同化和硬法化，使国际法从提供强制约束力的一致性监管方案转向提供平衡各国监管需求的协调性方案。直接规制型国际法可以通过 TBT 协议的协调方案转化为具有强制力的国内法，促进全球 AI 国内法趋向一致，间接赋予 AI 国际软法以事实约束力。2025 年 7 月，中国政府正式提出建立世界 AI 合作组织的倡议，以专门性国际组织为平台汇聚 AI 领域的国际共识，可以缓解各国国内法碎片化和竞争化带来的风险。

AI 国际法主体在规范上和事实上的局限性制约了国际法发挥作用，需要加以调整，回应 AI 的技术规制和国内法的发展需求，纳入 AI 发展中的大量非政府主体，实现多元主体的有效参与。在西方资本主义国家，面对全球市场力量的剧烈冲击，国内的工业、商业、金融和市场经济等一系列经济命脉不再完全受国家主权控制^②，跨国集团、高科技公司、大型金融企业等拥有大量资本的巨头企业或财团对于国内法的制定发挥重要甚至是绝对影响，同时也左右了该国的外交决策和国际法参与路径。^③ 尤其是在 AI 等新兴技术领域，大型科技公司不仅参与和影响西方资本主义国家的国内法制定、实施的全过程，还拥有着更强的技术能力，可以对 AI 的技术治理发挥作用。2024 年 7 月，世界 AI 大会暨 AI 全球治理高级别会议发表《人工智能全球治理上海宣言》（以下简称《上海宣言》），鼓励企业、研究机构、社会组织和个人等多元主体积极参与 AI 治理体系的构建和实施^④，代表着中国推动 AI 治理主体多元化的重要探索。因此，中国可以继续推动 AI 国际法适当突破传统的国际法主体概念，打通主体层面的制度壁垒，使更加多元化的主体可以成为国际法形成与运行过程中的行为主体。

整体上看，AI 国际法发展需要采取二元模式，软法与硬法呈现出结构性耦合的关系，中国可以提出软硬法协同的 AI 国际法方案，提升中国在 AI 领域的话语权和国际影响力。滞后性是数字时代法律的固有特征，软法的出现有助于改善滞后性问题。AI 的可解释性、系统风险、AI 偏差和算法共谋等主要风险带来了技术俘获、算法黑箱和信任失败等诸多新挑战，难以完全通过硬法和传统监管的方式加以化解，需要引入技术治理主义（Technical Governance）的思路^⑤，充分结合技术特征开展智能化监管与算法监管。^⑥ 国际软法兼具灵活性、实验性以及多元主体合作等优势，与 AI“善治”需求相适配，顺应了技术治理主义的时代潮流，也为多元主体的参与提供了空间。国际硬法则为国际软法的实施提供了制度性保障，对 AI 时代的国际法治具有长远

① See IEC Editorial Team, International Standards Key Tools for Regulation and Trade in AI, July 7, 2023, <https://www.iec.ch/blog/international-standards-key-tools-regulation-and-trade-ai>.

② 参见谢长安：《论资本主义全球化的当代困境及其影响》，《贵州师范大学学报（社会科学版）》2021 年第 2 期。

③ 参见袁征：《西方利益集团何以陷入恶性政治博弈》，《人民论坛》2017 年第 22 期。

④ 参见《人工智能全球治理上海宣言（全文）》，2024 年 7 月 4 日，https://www.gov.cn/yaowen/liebiao/202407/content_6961358.htm。

⑤ 参见沈伟：《数字经济时代的金融监管难题和技治主义方案》，《上海经济研究》2024 年第 5 期。

⑥ 参见沈伟：《数字规制的场景、困境与进路：作为法治场域的元宇宙》，《政法论丛》2024 年第 3 期。

意义。中国需要进一步加快 AI 国际法的国际合作机制建设,发挥软法作为“摆渡人”的作用,使软法及时向硬法转化,以弥补其约束力不足的缺陷,进而实现软法和硬法的转化与沟通。

(二) 促进 AI 国内法完善的中國路径

AI 国内法与国际法的契合度将影响一国开展对外贸易、投资等国际经济活动的融入度,中国 AI 国内法需要尽可能与国际法保持同频共振,纳入和融入全球共识。在内容上,中国提出的《倡议》涵盖了 AI 国际法和国内法的“最大公约数”,包括安全性和可靠性、伦理和人权保护、国际合作等,均为《全球数字契约》和欧洲《框架公约》等国际法强调的重点。但上述规则暂未纳入中国具体的国内法中,减弱了中国国内法与国际法的契合度,可能限制中国 AI 行业的发展与国际接轨的程度。未来,中国的 AI 国内法制定中,可以将各国关于 AI 的共识性内容在 AI 国内法的总则中予以规定,作为指导场景规制的基本原则。

相较于国际法和其他国家的国内法而言,中国 AI 国内法的优势在于场景式规制的动态性、灵活性、敏捷性,未来的国内法及其实施需要强化动态性监管,突出中国的制度优势,为实现国内法的规则外溢奠定基础。一方面,对 AI 技术不宜采用前提与结果的静态考察思维,而应结合 AI 技术发展的动态过程,即时性地评估风险,实时调整 AI 国内法的路径、模式和范式,避免 AI 技术可预测的有害结果发生。中国在 AI 国内法路径选择具有快速响应 AI 规制需求的能力,顺应技术发展和风险治理的时代要求,构成中国特色的制度优势,可以继续予以保持和深化。

另一方面,AI 国内法的制定不是单纯的自然科学问题,也不是单纯的社会科学问题,需要运用跨学科的知识加以解决^①,通过多元社会主体普遍参与风险识别、风险评估和风险应对的全过程,提升防范和化解风险和“科林布里奇困境”的效能。^②《上海宣言》强调建立多元化的参与机制,包括公开征求意见、社会调查等方式,让公众参与到 AI 决策过程中,代表了与会各国的基本共识。中国积极发展全过程人民民主和完善共建共治共享的社会治理制度,公众参与在立法、政府决策以及基层治理已有长足发展,并体现在《宪法》《立法法》所规定的立法、决策和社会事务中的公众参与制度^③,这为多元主体参与 AI 治理和决策提供了制度渠道和中国经验。在 AI 国际法形成较为完备的体系前,中国可以通过技术治理的方式,先行回应 AI 技术应用过程中的突出问题,发挥国内法在纳入多元主体上的优势,通过国内法的多元主体优势引领国际法治发展和完善。

AI 国内法具有主导或者辐射国际法发展的能力,中国需要通过涉外法治建设,积极引领国际法发展,实现国内法的国际化。如果一个国家在特定议题中建立了完善的国内法体系并且取得实效,将降低该国在国际法制定中所面临的国内政治成本,从而获得相对规则优势。^④中国可以在制定国际法的规则博弈中,将运转良好的国内法通过多边渠道“输出”为国际法,提高制定国际法的竞争力。《倡议》的提出体现了我国制定国际 AI 规范的“扩散优先”路径,即在国际层面形成具体法规之前,将中国 AI 技术引入发展中经济体。该倡议有助于减少国际社会在 AI 领域可能出现的“数字鸿沟”,得到了发展中国家的认可^⑤,也部分转化为联合国的国际倡议。中国可以基于国内法的实践,寻求维护国际社会整体安全和利益的最佳伦理、规范和制度,突破个别国家国内法竞争的藩篱,提出中国的 AI 国际法方案。

结语

AI 的技术发展具有联动性、跨国性和多样性的特征,法律是规制 AI 应用、促进 AI 发展的核心手段。AI

① See David H. Guston and Daniel Sarewitz, “Real-time Technology Assessment,” *Technology in Society*, vol. 24, no. 1, 2002, p. 98.

② See Johannes Glückler, Gary Herrigel and Michael Handke, *Knowledge for Governance*, Cham: Springer, 2021, p. 98.

③ 参见肖林:《公众参与社会治理的中国特色与深层挑战》,《世界社会科学》2024 年第 6 期。

④ See Walter Mattli and Tim Büthe, “Setting International Standards: Technological Rationality or Primacy of Power?” *World Politics*, vol. 56, no. 1, 2003, p. 41.

⑤ 参见《巴西专家:中国提出的〈全球人工智能治理倡议〉有助防范“技术贫民窟”》,2023 年 11 月 1 日, http://www.news.cn/world/2023-11/01/c_1129952180.htm。

的发展既需要国际社会通过国际法引导和规范 AI 技术治理,推动 AI 全球应用法治化,也需要各国通过国内法落实 AI 国际法的要求并快速响应 AI 技术变革性发展的相应风险和挑战,推动 AI 技术发展。这要求全面审视 AI 国际法和国内法,妥善把握两者之间的互动形态,并以此为基础制定中国的 AI 发展和治理方案。

现有 AI 国际法已取得了一定的进展,《全球数字契约》和欧洲《框架公约》等专门性国际法的出现为 AI 的应用与发展提供了法律依据,形成了一定的国际共识。中国的 AI 国内法体系可以在此基础上加以借鉴和调整,更好地融入既有国际法体系,促进 AI 国际合作。但是,现有国际法的法律约束力和参与主体的局限性无法适配 AI 国内法的快速发展需求和规则竞争困境,需要引导国际法共识,拓宽国际法在规范和事实层面的参与主体,软硬法结合的二元模式提供了 AI 国际法改革的基本方案。国内法很大程度上影响甚至是主导了 AI 国际法的发展,中国需要通过加强涉外法治建设,突出场景式立法的快速响应优势,增强国内法的外溢效应,弥补 AI 国际法的不足,增强中国在 AI 国际法体系建构上的话语权和影响力。

[本文系国家社会科学基金重大项目“美国全球单边经济制裁中涉华制裁案例分析与对策研究”(21&ZD208)的阶段性成果]

(责任编辑:邱小航)

The Interaction between International Law and Domestic Law on Artificial Intelligence

SHEN Wei, CHEN-XU Anli

Abstract: The effective interaction between international law and domestic law on artificial intelligence (AI) is a requirement of the rule of law for regulating AI risks and ensuring the sound development of AI. Existing international law on AI comprises three components: initiative-based soft law and regional hard law for direct regulation, as well as area-specific international law for indirect regulation. However, constrained by the sources and subjects of traditional international law, existing international law lacks adequate adaptability when applied to AI. Europe, the United States, and China have each formulated national AI laws; these domestic laws not only require international law on AI to reconcile differences and disputes, but also provide a foundation of consensus for the development of such international law. The internalization of international law and the externalization of domestic law are both manifested in the two-way interaction between international law and domestic law on AI. Non-state actors in the AI field rely on regulation by domestic law, which makes the extension of international law to domestic law possible and necessary. China needs to not only continuously adjust and optimize its domestic law to provide a legislative response to the development of AI, but also put forward proposals for international law on AI. It should promote international law to respond to the needs of domestic laws of various countries and facilitate the formation of a dual collaborative governance model of hard and soft law for international law on AI. Further leveraging the advantage of domestic law in rapid response through scenario-based regulation and realizing the spillover effect of domestic law constitutes a way for China to make contributions in the field of international law on AI.

Key words: artificial intelligence, international law, domestic law, rule externalization, hard law