大国竞争背景下的人工智能安全治理与战略稳定

沈　逸　　高　瑜

【内容摘要】　在大国战略竞争加剧的背景下，实现对人工智能的有效治理，不仅关乎人类社会的发展，而且影响国际战略稳定。中国和美国作为全球人工智能领域的两大关键行为体，在该领域的治理理念和方式上存在差异。美国依赖技术优势推行单边主导型治理，将人工智能界定为战略竞争的前沿领域，坚持绝对安全理念，试图将监管体系按照人工智能技术和产业生态延展并聚焦于数据资源，尝试构建以美国为中心的非对称单向度的数据流动秩序。中国则坚持真正的多边主义，以总体国家安全观为指引，朝着构建网络空间命运共同体的方向，推动人工智能的全球治理朝着普惠、包容、弹性、灵活的方向发展。中国也非常关切支撑人工智能创新与产业发展所必需的数据资源，努力构建并持续完善以有序流动、分级分类治理为核心特征的数据治理架构，强调均衡考量各方利益，统筹安全与发展，以此作为推进人工智能全球安全治理的核心思路。展望未来，全球人工智能安全治理面临诸多挑战和机遇，各国需要加强沟通与协调，共同应对安全风险，推动人工智能技术健康发展。

【关键词】　大国竞争　人工智能　战略稳定　中美关系

【作者简介】　沈逸，复旦大学国际关系与公共事务学院教授、复旦大学网络空间国际治理研究基地主任（上海　邮编：200433）；高瑜，复旦大学国际关系与公共事务学院博士研究生（上海　邮编：200433）

【中图分类号】 D815.5　D822　　**【文献标识码】** A

【文章编号】 1006-1568-(2024)03-0033-18

【DOI 编号】 10.13851/j.cnki.gjzw.202403003

当前人工智能领域的高速发展，尤其是生成式人工智能治理领域的突破，不仅推动人工智能成为信息技术革命的前沿领域，而且使得围绕人工智能发展带来的影响以及治理模式的讨论成为全球各方普遍关切的新兴议题。2023 年 11 月 1 日，首届全球人工智能安全峰会在英国布莱切利庄园开幕，会议发布了由 28 个国家以及地区组织共同签署的《布莱切利宣言》，就积极推进人工智能安全治理达成共识，承诺用多种方式推进和完善聚焦人工智能安全治理的国际实践。①

这一轮人工智能安全治理与战略稳定的探索与实践，是在国际体系深度调整、国际格局加速演变的特殊背景下进行的。大国战略竞争赋予人工智能安全治理以及战略稳定以特殊路径和方向，人工智能发展依靠的生态（包括技术生态和商业生态）日趋明显地受到大国战略竞争的显著影响。2022—2023 年，美国两次基于“小院高墙”的科技竞争战略，实施针对中国的算力芯片禁令，试图对中国获取发展人工智能至关重要的算力基础设施进行有效限制与制约；② 2024 年 2 月 28 日，美国总统拜登签署行政令，限制向中国、俄罗斯和另外四个国家出售美国的敏感数据，相关数据包括生物识别信息、健康记录以及财务和地理位置等信息，其中重要的考量因素之一就是约束和限制中国在人工智能发展过程中所必需的数据资源。③ 另外，2023 年 11 月 15 日中美元首旧金山会晤所取得的建设性成果之一，就是建立人工智能政府间对话机制。这一对话机制的核心关切是确保人工智能的应用和发展不会对中美战略稳定带来负面影响和挑战。④ 具体地说，美方提出的讨论场

① “Britain Publishes ‘Bletchley Declaration’ on AI Safety,” Reuters, November 2, 2023, https://www.reuters. com/technology/britain-publishes-bletchley-declaration-ai-safety-2023-11-01/.

② Stephen Nellis and Jane Lee, “U.S. Officials Order Nvidia to Halt Sales of Top AI Chips to China,” Reuters, September 1, 2022, https://www.reuters.com/technology/nvidia-says-us-has-imposed-new-license-requirement -future-exports-china-2022-08-31/; Karen Freifeld, “Exclusive: US Tackles Loopholes in Curbs on AI Chip Exports to China,” Reuters, October 17, 2023, https://www.reuters.com/technology/upcoming-us-rules-ai-chip-exports-aim-stop-workarounds-us-official-2023-10-15/.

③ David McCabe, “Biden Acts to Stop Sales of Sensitive Personal Data to China and Russia,” *New York Times*, February 28, 2024, https://www.nytimes.com/2024/02/28/technology/biden-data-sales-china-russia.html.

④ Stephen Collinson, “Analysis: Takeaways from the Biden-Xi Summit, Where Low Expectations Were Met,” CNN, November 16, 2023, https://www.cnn.com/2023/11/15/politics/takeaways-biden-xi-summit/index.html.

景是避免在与核武器有关的指挥控制系统中不当植入人工智能程序，导致在特定场景下，出现类似冷战时期“未经授权发射”或者“意外发射”等导致的战略级风险。

除了上述大国竞争背景下的国家安全与战略稳定维度外，人工智能的发展已经在具体应用场景中迅速深入政治、经济和社会生活的方方面面。如何预防因人工智能不当应用对政治安全、经济安全、金融安全、社会稳定、隐私保护等带来的全面冲击和挑战，确保人工智能服务于人类社会可持续发展目标，继而对人类社会的发展带来更多正面而非负面效应，正在持续引起各方的关切。此外，一些具有显著示范效应的做法和思路也正在梯次出现并逐渐成熟。

一、人工智能安全治理对高质量全球协作提出了新要求

整体而言，在不涉及具有显著地缘政治属性和大国战略竞争的功能性议题领域，全球各方（无论是国家行为体还是非国家行为体）都普遍意识到人工智能安全治理对全球协作提出了更高的要求，这种新要求客观上成为各方深度强化沟通与协调的外部推动力。具体来说，新要求集中表现在如下三个领域。

第一，避免人工智能威胁主权国家的政治安全是推进人工智能安全治理的普遍共识。尽管中国和西方发达国家在政策制定、执行以及关键术语使用方面存在显著差别，但是自 2016 年美国总统选举之后，西方发达国家对奥巴马政府时期构建和完善的“进攻性互联网自由战略”做出了适当调整，通过提出并推动“选举安全”等概念，微妙但实质性地接受和认可了中国等国家倡导的“政治安全”概念，认识到所谓西方发达民主国家的政治体制并不能抵御来自网络空间错误信息和虚假信息的威胁与攻击。[①] 以 ChatGPT 和 Sora 等为代表的生成式人工智能取得突破性进展之后，美国国土安全部、布

① 沈逸：《美国撕掉“互联网自由”的温柔外衣》，环球网，2016 年 12 月 30 日，https://opinion.huanqiu.com/article/9CaKrnJZsPl。

鲁金斯学会等均发布报告，称密切关注通过使用生成式人工智能工具，在选举过程中发布合成图片、生成虚假音频和虚假视频来进行信息操控等风险，并结合 2024 年美国总统选举提出了相应的警告。[①] 在联合国框架内，在保障“信息一致性”、避免人工智能应用被滥用与虚假信息生成和发布等问题的基础上，各国提出了较为明确的治理目标。[②] 对发展中国家和新兴大国来说，考虑到与发达国家在人工智能创新、发展、能力和应用领域的差距，以及发达国家因滥用互联网导致威胁政治安全的历史记忆，如何在人工智能时代有效保障政治安全、实施有效的人工智能安全治理是它们的主要关切。

第二，避免人工智能在无人平台的不当使用日趋成为军备控制领域最有可能取得有限突破的问题之一。人工智能技术在军事领域应用的关键场景之一是与无人平台结合。从战场的实践来看，这种结合及其近似模拟已经出现在不同的场景中。美国偏好使用无人机实现对恐怖分子的定点清除；[③] 在乌克兰危机中，无人机的大量使用已经在相当程度上改变了地面战场的态势，形成了对传统人员和装甲平台的非对称杀伤优势；[④] 在新一轮巴以冲突中，以及也门胡塞武装介入之后，总体实力处于劣势的哈马斯凭借无人机平台，展现出了令各方高度关注的非对称作战能力。从国际军控谈判的实践来看，各方讨论的是一种可能还没有真正投入使用的组合，即所谓“自主致命武器系统”，它是由人工智能程序而非真人决定实施的致命攻击。[⑤] 虽然存在各种不同的认知，但主流观点倾向于认为应对此类系统的研发、部署和使用保持谨慎与克制，尽可能避免由于此类系统的不当使用造成人道主义灾难，甚至诱发大国直接军事冲突的风险，导致危机事态失控。尽管迄今为止各国在

① Rehan Mirza, “AI, Cybersecurity, Election Integrity, Threats to Election,” *Homeland Security News Wire*, February 22, 2024, https://www.homelandsecuritynewswire.com/dr20240222-ai-and-election-integrity; Norman Eisen, Nicol Turner Lee, and Samara Angel, “8 Best Practices for State Election Officials on AI,” Brookings Institution, March 11, 2024, https://www.brookings.edu/articles/8-best-practices-for-state-election-officials-on-ai/.

② *Interim Report: Governing AI for Humanity*, United Nations, December 2023, https://www.un.org/en/ai-advisory-body.

③ 佘纲正、罗天宇：《军用无人机的使用偏好及安全影响》，《国际政治科学》2022 年第 2 期，第 42—85 页。

④ “More than 3,500 FPV Drone Operates Trained - Russian Defense Ministry,” Tass, January 5, 2023, https://tass.com/defense/1729393.

⑤ “Lethal Autonomous Weapon Systems (LAWS),” UNODA, https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/.

避免人工智能在无人平台的不当使用方面的实质性进展依然有限，但是仍然有理由对这一领域的突破保持审慎乐观。

第三，各国在人工智能治理的指导原则上基本具备达成共识的基础。2023 年联合国成立“人工智能高级别咨询机构”，并于同年 12 月发布了《以人为本的人工智能治理》（Interim Report: Governing AI for Humanity）的中期报告，重点关注人工智能全球治理原则的确立。该报告认可了现有的人工智能全球治理的举措，并提出了初步建议，承诺在 2024 年 8 月之前在最终报告中详细阐述这些建议。具体来说，目前在人工智能安全治理领域达成的国际共识有五点：一是包容性治理，即确保全球范围内的民众（不论其地理位置、经济状况或其他背景）都能平等地访问和使用人工智能工具；二是公共利益原则，即人工智能治理工作必须与公共政策目标保持一致，并扩大利益攸关方的代表性；三是以数据治理为核心，即人工智能治理应与数据治理同步进行，优先考虑数据隐私和安全，同时促进公共数据共享；四是普遍性、网络化与多利益相关方协作，即人工智能的治理需要得到广泛的认可和支持，以推动人工智能治理的全球化和普及化；五是遵守国际法，即人工智能治理需要以国际法为基础，特别是《联合国宪章》、国际人权法和可持续发展目标等。这些观点为人工智能发展提供了一个全面而深入的框架，推动人工智能安全治理朝着更加公正、透明和可持续的方向发展。

二、大国战略竞争使人工智能具有特殊意义

目前人工智能技术的突破，发生在中美战略竞争的大背景下。从特朗普政府时期开始，人工智能就被美国确立为赢得中美战略竞争、保障美国战略优势的前沿颠覆性技术领域之一。[①] 人工智能在当前中美战略竞争中具有双重属性。一方面，人工智能的不当使用可能引发核战争等灾难性后果，对全球安全构成严重威胁，被认为是有可能导致大国战略竞争失控的直接因素；

① “Maintaining American Leadership in Artificial Intelligence,” Federal Register, February 11, 2019, https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence.

另一方面，人工智能也为大国战略稳定提供了新的契机。通过加强在人工智能领域的合作，可以有效缓解战略竞争的压力，推动全球战略稳定。

第一，人工智能的安全化。在大国战略竞争的背景下，人工智能被安全化，相关技术和产业的发展面临很大的不确定性。美国将人工智能技术和产业的发展提升至国家安全的高度，不仅增加了相关领域的不确定性，还导致原本正常的技术竞争和产业竞争被扭曲。出于对大国博弈的警惕，与人工智能相关的技术和产业，无论其是否具有军事意图，是否直接涉及主权安全，都被视为国家安全和战略竞争的重要组成部分。[①] 美国人工智能国家安全委员会发布的相关报告明确表示，当前的大国竞争是以二战以来美国的技术优势首次受到“威胁”为背景。在美国的认知中，中国作为一个“战略竞争对手”，可能会通过人工智能技术“渗透美国社会、窃取美国数据并通过网络攻击和虚假信息宣传干扰美国民主体制的运行”。[②] 因此，美国采取了一系列措施来限制中国在人工智能领域的发展。其中，在高端芯片上对中国进行出口管制是美国政府的重要手段之一。2022 年 10 月 7 日，拜登政府发布了向中国出口人工智能和半导体技术的管制新规。具体来说，拜登政府试图从芯片、软件、设备和零部件四个方面，全方位限制中国人工智能的发展。[③] 这些新的出口管制措施的出台，成为中美科技竞争中的一个转折点，在加剧双方竞争的同时，还可能引发更多的贸易和技术摩擦。

第二，人工智能的应用成为维持大国战略稳定的新要素。尽管当前大国之间的竞争和战略博弈正在如火如荼地进行，但目前大国之间鲜有通过发动战争侵吞对方领土的战略意图。一方面，全球化使各国在政治、经济、文化等多个领域形成相互依赖，这种相互依赖使得战争的成本和风险大大增加。另一方面，核武器和其他大规模杀伤性武器的存在，也使得战争的成本和风

① 刘国柱、尹楠楠：《美国国家安全认知的新视阈：人工智能与国家安全》，《国际安全研究》2020 年第 2 期，第 135—155 页。

② *National Security Commission on Artificial Intelligence Final Report*, NSCAI, October 5, 2021, https://assets.foleon.com/eu-central-1/de-uploads-7e3kk3/48187/nscai_full_report_digital.04d6b124173c.pdf.

③ Gregory C. Allen, “Choking off China’s Access to the Future of AI,” Center for Strategic and International Studies, October 11, 2022, https://www.csis.org/analysis/choking-chinas-access-future-ai.

险变得极高。因此，在二战结束之后，世界进入现代国际体系，与以往频繁爆发战争的时代不同，国际社会进入了以利益互换为特征的“大国无战争时代”。[①] 当前，主要大国的核指挥和控制系统对人工智能程序的依赖逐步加深，依靠机器学习和算法系统来强化数据流动，以此增强对安全态势的感知能力。[②] 在人工智能时代，核打击的定位能力、突防能力和运载能力等提升至新的层级，进一步强化了核武器的威慑能力。[③] 在此背景下，一旦爆发大规模冲突，人工智能将推动核武器攻击“自主升级”，无需人工操作就会激活升级系统，传统战争模式下的“门槛”机制很难发挥其应有的效力，最终导致战争的规模、持续时间、打击范围和破坏力度不受人类所控制。[④] 因此，各国在面临冲突升级的可能性时会采取更加审慎的态度，严格避免因为误判或误操作带来灾难性的后果。从这个角度来看，人工智能的应用为各方提供了一种新型的、更为理性的竞争方式，从而为维护战略稳定发挥积极作用。

三、数据治理成为人工智能安全治理的抓手与切入点

数据是人工智能的核心驱动力，其数量和质量直接影响着人工智能的性能和应用效果。随着全球化的发展和信息技术的进步，跨境数据流动日益频繁，为人工智能的发展提供了丰富的数据资源。然而，跨境数据流动也带来了一系列挑战，数据隐私泄露、数据安全风险以及不同国家之间的数据保护政策差异等随之而来。因此，如何在保障数据安全的前提下实现跨境数据的有效流动和利用，成为人工智能治理中的一个重要课题。在大国竞争的背景下，人工智能治理博弈的焦点问题在于跨境数据流动的规则博弈。

第一，数据和人工智能技术的发展在政治、经济和社会领域带来了前所

① 杨原：《大国无战争时代的大国权力竞争：行为原理与互动机制》，中国社会科学出版社 2017 年版，第 286 页。

② Mark Fitzpatrick, “Artificial Intelligence and Nuclear Command and Control,” *Survival*, Vol. 61, No. 3, 2019, pp. 81-92.

③ Zachary Kallenborn and Philipp C. Bleek, “Swarming Destruction: Drone Swarms and Chemical, Biological, Radiological, and Nuclear Weapons,” *Nonproliferation Review*, Vol. 25, No.5-6, 2018, pp. 523-543.

④ 张煌、杜雁芸：《人工智能军事化发展态势及其安全影响》，《外交评论》2022 年第 3 期，第 99—130 页。

未有的巨大机遇。数据要素与人工智能的发展催生了数字经济、智能经济等新经济形态，为全球经济增长提供了新的增长点。[①] 数据作为新型生产要素，通过人工智能技术广泛应用于经济领域，有助于提升生产效率，优化资源配置，实现产业升级和创新。同时，政府机构利用人工智能技术处理海量数据，能够在很大程度上提高政府决策的科学性、精准性和时效性，推动政府治理向现代化、智能化方向发展。例如，大数据分析有助于政府更准确地把握社情民意，优化政策的制定和实施。人工智能技术深度挖掘医疗、教育、交通等领域的海量数据，能够制定个性化方案并在生活中提供最优路径，为提升人们的生活质量带来革命性改变。

第二，人工智能存在数据依赖性和数据敏感性的特征，[②] 数据自身的安全性能会直接决定人工智能的安全程度。人工智能作为新兴技术，其发展尚未达到成熟、完备的程度，还有许多不足和缺陷需要加以克服和改进。因而在当前阶段，人工智能技术存在着天然的逻辑风险。此外，处理和训练数据的算法模型也是人工智能安全风险的一大来源，若输入的数据对原本的算法模式产生干扰，导致机器学习的侧重点和部分参数发生小幅改变，输出的结果就会发生颠覆性的变化。[③] 无论从技术优化治理的角度看，还是深度学习的数据对象看，都不能放任人工智能技术野蛮生长，否则就可能会出现背离人类设计初衷、违背伦理道德、冲击社会规范的风险。

第三，不同的数据治理偏好是区分不同国家人工智能治理模式的核心。这些政策偏好往往受到一个国家的政治、经济、文化和技术背景的影响，反映了各国在平衡创新与风险管理、个人隐私保护与社会安全需求、商业发展与伦理道德约束之间的不同取向。[④] 目前，中美两国在人工智能安全治理方

① 黄丽华、杜万里、吴蔽余：《基于数据要素流通价值链的数据产权结构性分置》，《大数据》2023 年第 2 期，第 3—15 页。

② 皮勇、张明诚：《总体国家安全观视域下人工智能安全风险治理研究》，《中国科技论坛》2023 年第 6 期，第 86—96 页。

③ Matthew Jagielski, et al., “Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning,” *2018 IEEE Symposium on Security and Privacy (SP)*, 2018, pp. 19-35.

④ 郎平：《数字时代国家安全困境与网络空间命运共同体构建》，《当代世界》2023 年第 10 期，第 34—39 页。

面有不同的模式。美国通过数据单向度流入境内的方式构建单边主导型人工智能安全治理模式，这种模式体现了美国在技术创新和经济发展上对于全球主导地位的追求。相比之下，中国通过均衡有序的数据流动方式构建多边普惠型人工智能安全治理模式，更加注重数据的共享和普惠性。总的来说，不同国家的数据治理偏好和人工智能安全治理模式反映了它们在促进技术发展与保障社会利益之间寻求平衡的不同路径。一方面，这种多样性增加了国际社会在人工智能安全治理上达成一致的难度；另一方面，多种治理模式也为其他国家和地区提供了可供参考和借鉴的多种可能。

四、美国：单边主导型人工智能安全治理

在人工智能治理领域，美国的基本考量侧重于发展利益优先，特别是以人工智能企业的数据红利作为政策出发点，保障业务遍布世界各地的美国人工智能企业能够持续不断地从自由流动的全球数据中获益。在治理主体方面，美国政府与大型人工智能企业开展了各种形式的合作，政府要求企业通过或明或暗的方式提供其服务器上来自世界各国的用户数据。在此基础上，美国充分利用企业的数据抓取能力和政府的数据分析能力，对世界各国的数据进行直接或间接的监管，旨在形成一套以美国为中心的单向数据流动体系，从而为美国人工智能模型的开发与训练提供充足的数据优势资源。

（一）严格监管境内人工智能数据

美国对于人工智能数据的监管主要由联邦通信委员会（Federal Communications Commission）和联邦贸易委员会（Federal Trade Commission）两大机构负责。一方面，联邦通信委员会旨在全方位监管国内外通信情况，确保美国的通信能力位于世界前列。联邦通信委员会建立于 1934 年，是受国会监督的独立的美国政府机构，负责实施和执行美国通信法律法规。具体职责包括：促进宽带服务设施的竞争、创新与投资，建立适当的竞争框架来推动通信技术革新以支持美国经济发展，鼓励国内外充分利用频谱，修订通

信法规以推动新技术蓬勃发展，加强国家通信基础设施的防御能力。[①] 另一方面，联邦贸易委员会旨在管控商业数据，确保美国的商业市场处于高效、良性的竞争态势。该委员会成立于 1914 年，最初的目的在于防止商业领域出现不公平竞争；从 1938 年起开始执行各种消费者保护法律；在 1975 年被国会授权负责制定和执行整个行业的贸易法规，具有保护消费者和促进良性竞争的双重使命。因此，在跨境数据流动规制中，联邦贸易委员会主要负责商业数据流动管理，确保企业履行用户隐私保护承诺，并对侵害消费者隐私权的行为提起诉讼。[②] 这两大机构相配合，不仅能够确保美国政府有效管控国内的数据和人工智能发展，而且能够帮助美国政府掌握其他国家的个人数据和商业数据，为美国“棱镜计划”“梯队系统”等全球监听系统的落实和运作奠定了组织架构基础。

此外，美国针对关键部门和重要行业的数据进行严格监管。例如，2018 年《外国投资风险审查现代化法案》规定，由美国外国投资委员会（CFIUS）对公民敏感个人信息的投资和交易行为进行安全审查；[③] 美国《出口管理条例》（EAR）赋予商务部对特定领域数据传输的许可权，如军用或军民两用技术等；[④] 《金融服务现代化法案》《健康保险流通与责任法案》以及《电子通信隐私法》等法案分别针对金融、医疗、通信等领域的数据出境作出相应的监管规定。

事实上，从 2015 年起，美国对大国竞争的焦虑持续上升，尤其担忧以人工智能为代表的新兴技术带来的数据安全威胁。在 2015 年的《国家安全战略》中，美国将大国竞争视为国家安全的首要挑战，认为大国的行为和态度是国际社会能否有效应对全球性安全风险的决定性因素。由于发展中大国的崛起，经济力量对比的变化决定了国家间权力分配的动态变迁，大国之间的关系将对未来的国际格局和全球安全产生重大影响。美国认为在当时的国

① “About the FCC,” FCC Commission, https://www.fcc.gov/about/overview.
② “About the FTC,” FTC Commission, https://www.ftc.gov/about-ftc.
③ “Foreign Investment Risk Review Modernization Act of 2018,” Congress Gov, https://www.congress.gov/bill/115th-congress/house-bill/5841/text.
④ “Export Administration Regulations (EAR),” Bureau of Industry and Security, https://www.bis.doc.gov/index. php/regulations/export-administration-regulations-ear.

际环境中，网络空间中的大数据、云计算和人工智能等新兴技术发展势头迅猛，导致美国本土和关键基础设施的安全面临更加复杂的外部威胁，甚至有可能受到“灾难性打击”。因此，网络空间安全被置于太空、海洋安全之前，成为集体行动最应该保护的领域。美国政府认为上述共享空间是人员、商品、服务和思想自由流动的基础，是连接全球经济和公民社会的动脉。但是，网络基础设施脆弱性明显，容易遭到非法势力的恶意攻击，政府需要与私营部门、民间社会和其他利益攸关方加强合作，构建国家网络安全保护框架。[①]

（二）广泛调取境外人工智能数据

在数据获取方面，美国拥有全球最大的互联网企业，在世界各地有广泛的用户，在数据挖掘、收集和分析上的技术较为成熟。在此基础上，政府机构通过行政权力介入其中，与互联网巨头企业进行或明或暗的数据合作，建立千丝万缕的联系。美国作为互联网的诞生地，其科技实力也处于世界领先地位，谷歌、微软、脸书、X（原推特）、苹果、亚马逊等大型跨国公司基本已经覆盖了全球的数据流动范围。对于这些美国企业而言，在全球范围内获取尽可能多的数据是现代跨国企业特别是互联网企业获利的重要手段。简言之，在数字时代，数据就意味着利润，数据广泛流动就意味着利润最大化。因此，消除数据跨境流动的壁垒会给美国带来显著的商业利益。早在 1997 年，时任美国总统克林顿颁布《全球电子商务纲要》（A Framework for Global Electronic Commerce），初步形成了跨境数据自由流动的治理战略，主张在亚太经合组织、美洲国家首脑会议、《北美自由贸易协定》机制等多边平台展开对话，尽量减少各国为保护公民隐私而设置的非关税贸易壁垒。[②] 尽管在“9・11”事件后，美国政府在一段时间内将网络空间治理的重点放在保护国家安全方面，但自希拉里担任国务卿后，“互联网自由”重新被确定为美国网络空间治理的主要战略，主张将网络空间看作是“全球公域”，提倡数据在全球范围内自由流动，并指责他国设置“数字屏障”的行为违反了《世

① “Fact Sheet: The 2015 National Security Strategy,” White House, February 6, 2015, https://obamawhitehouse.archives.gov/the-press-office/2015/02/06/fact-sheet-2015-national-security-strategy.

② The Framework for Global Electronic Commerce, White House, 1997, https://clintonwhitehouse4.archives.gov/WH/New/Commerce/.

界人权宣言》，以人道主义保护和现代化发展为由要求各国开放数据流动。美国政府希望尽可能减少数据流动的国界壁垒，鼓励各国在跨境数据流动规制中采取宽松措施。[①] 究其根本，这一政策主张背后的原因是数据在全球范围内自由流动会给美国带来丰厚的经济利润和政治利益，美国在网络空间中占据了先发性优势，其数据治理模式必然以大型企业的商业利益为先。

在数据使用方面，美国依靠其先进的互联网技术在全球网络空间推行数据霸权，企图使各国产生的数据都能流入美国境内，并为美国政府所使用。根据 2001 年颁布的《美国爱国者法案》（USA PATRIOT Act），美国的远程电子服务和通信服务提供商可以直接获取国内外的通信记录（包括语言和文字），并将其披露给美国政府；而对于外国的信息搜集机构，该法案则以“防范恐怖主义”为由对其设置了种种限制。[②]《澄清海外合法使用数据法案》（Clarifying Lawful Overseas Use of Data Act）明确规定，美国政府可以出于国家安全需要对境外数据进行调取，无论该数据主体是否为美国公民或美国企业，[③] 即变相赋予了美国政府通过网络运营商获取境外数据的权力。这些规定暴露了美国对本国政府和外国政府明显的双重标准。

简言之，在对外宣传体系中，美国政府一贯将“自由”奉为核心价值规范，在网络空间更是着力推动数据在全球范围内的自由流动。然而，所谓的“数据全球化”不过是借助“自由”这一价值大旗占领道德高地，让全球的数据都流动到美国境内，在全球范围内尽可能攫取利益，为美国政府所用，推动其人工智能产业发展，然后再凭借其强大的跨国企业进一步获取全球数据，如此循环往复。在大型企业和政府机构的共同推动下，美国在人工智能治理领域形成了以文件表述、官方言论为“虚”，以治理机构、法律法规为“实”的一套有效配合模式，实质上是以“全球化”为包装，在数据空间继续推行美国霸权，监管世界各国的数据流动，从而为美国人工智能技术和相

① Hillary Rodham Clinton, “Remarks on Internet Freedom,” U.S. Department of State, January 21, 2010, https://2009-2017.state.gov/secretary/20092013clinton/rm/2010/01/135519.htm.

② “USA PATRIOT Act,” *FinCEN*, https://www.fincen.gov/resources/statutes-regulations/usa-patriot-act.

③ “CLOUD Act Resources,” *Criminal Division*, https://www.justice.gov/criminal/cloud-act-resources.

关产业的发展提供优渥的数据条件，确保美国人工智能发展占据领先优势。

（三）美国在人工智能安全治理上的政策举措

美国在人工智能安全治理方面体现了其在全球范围内维持领导地位的决心。2018 年，美国成立人工智能国家安全委员会（NSCAI），其使命是保持美国在人工智能技术和产业方面的竞争力，审查人工智能、机器学习开发和相关技术的进展情况，由此展开有关人工智能安全监管各方面的行动。该委员会也是近年来美国国会与政府围绕人工智能开展的一系列体制建设的重要组成部分，它的设立标志着美国将人工智能上升到国家安全层级，从顶层统筹指引全美人工智能领域的发展，进一步完善人工智能监管链条，为聚合各方之力加速推进人工智能发展奠定重要基础。

2023 年 10 月 30 日，美国总统拜登签署《关于安全、可靠、可信地开发和使用人工智能的行政命令》，强调要确保美国在人工智能领域的发展和风险管控上处于世界领先地位。行政命令的一个关键方面是要求人工智能系统的开发者与美国政府共享安全测试结果，实质上赋予了白宫更大的权力，以密切掌控私营部门人工智能技术的发展。此外，该行政命令重视与国际合作伙伴和国际标准化组织加速制定标准，还涉及生物合成筛选的新标准，用于防范使用人工智能设计危险生物材料的风险。[①] 为了更好实现这一政策目标，2024 年 1 月 29 日，拜登政府宣布要在 90 天内采取一系列广泛行动，以应对人工智能对安全造成的一些重大威胁。这些行动涉及对人工智能技术的进一步监管、推广安全标准、促进研发创新、加强国际合作等多个方面，旨在确保美国在人工智能领域的领先地位，并推动人工智能技术的安全、可靠和可持续发展。具体包括：加强人工智能安全和隐私保护标准的制定和实施，推动人工智能技术的研发和创新，加强与其他国家在人工智能领域的合作与交流，以及开展相关教育和培训项目来提升公众对人工智能技术的认识和理解。

这些政策文件和行动表明，美国在人工智能治理上采取了一系列全面且

① “Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence,” White House, October 30, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

深入的措施，旨在保持其在该领域的领导地位，并应对与人工智能发展相关的各种挑战。同时，这些政策文件和行动也体现了美国在人工智能安全治理中的单边主导倾向。另外，美国政府将人工智能视为其维护霸权地位的重要领域，着力确保在与中国的技术竞争中占据优势地位，以增强美国的外部安全防护能力。在此基础上，美国试图在人工智能安全治理规则领域占据主动权，让美国倡导的相关原则、规则和规范融入并内化为国际规范。[①]

五、中国：多边普惠型人工智能安全治理

2023 年 10 月 18 日，习近平主席在第三届“一带一路”国际合作高峰论坛上提出《全球人工智能治理倡议》，强调人工智能安全、发展和治理三个方面协调并重。[②] 这一理念为全球人工智能治理提供了全局性思路，不仅强调数据的有序流动和平衡开发，而且注重在保障安全的基础上推动经济社会发展，实现安全与发展、治理与创新的良性互动。核心原则是通过分级分类的数据治理方式激发人工智能技术赋能安全与发展的协同效应，释放数据的潜在价值以抵御人工智能安全风险；同时让机器学习助推经济社会发展，从而为我国高质量发展提供不竭动力。

（一）分级分类治理人工智能数据

分级分类治理是人工智能数据治理的核心。通过对数据进行科学分类和合理定级，可以实现对不同级别、不同类型数据的差异化管理和保护，从而确保数据的安全和有效利用。这种治理方式既能够防止数据滥用和泄露，又能够促进数据的合理流动和共享，为人工智能技术的发展提供有力支撑。中国近些年开发出一套数据分级分类保护标准，有效兼顾了个人隐私、经济发展和安全保护的需要。2021 年 11 月 14 日，网信办发布《网络数据安全管理条例（征求意见稿）》，其中包括对国家数据进行分类分级保护，将数据分

① 刘国柱、尹楠楠：《美国国家安全认知的新视阈：人工智能与国家安全》，第 135—155 页。

② 《全球人工智能治理倡议》，第三届“一带一路”国际合作高峰论坛官方网站，2023 年 10 月 19 日，http://www.beltandroadforum.org/n101/2023/1019/c134-1232.html。

为一般数据、重要数据、核心数据三种级别，分级实施不同的保护措施。对个人的一般数据以及重要数据实行重点保护，对核心数据实行严格保护，不同级别数据的处理者需遵循相应级别的数据安全保护义务。

从数据类别上看，对于公共数据，国家在监督的基础上推动数据开放共享，以加强对数据潜力的开发利用。对于个人信息的处理，具体规则应在醒目位置公开展示，在征得主体同意的基础上进行处理。对于跨境数据，需要首先通过网信部门的数据出境安全评估，传输方和接收方均应通过个人信息保护认证，并遵照标准规定签订合同。此外，涉及跨境个人信息和重要数据处理的运营者，需要每年编制数据出境安全报告，内容涵盖数据接收方的身份信息、数据的类别和数量以及出境目的、境外的数据存放方式和时间、地点等。[①]《网络数据安全管理条例（征求意见稿）》对相关内容的规定表明中国的数据治理规则正逐步走向成熟，同时兼顾安全保护和经济发展两方面利益。该条例反映了中国数据流动蓬勃发展的现状，也保障了中国数字经济发展潜力可以得到充分挖掘，在保障数据安全的基础上提升数据处理的公平性，并给予数据流动充分的灵活性，为中国的数字经济和人工智能发展保驾护航。

（二）有序平衡地开发人工智能数据

中国强调的有序数据流动是实现人工智能安全治理的关键。在保障数据安全的前提下，推动数据的跨境流动和共享，有助于释放数据的潜在价值，促进全球范围内的技术创新和合作。这不仅有利于提升中国人工智能产业的国际竞争力，而且有助于推动全球人工智能治理体系的完善和发展。2023 年 9 月 4 日，习近平主席在致 2023 中国国际智能产业博览会的贺信中明确指出，在智能产业和数字经济蓬勃发展、全球要素资源配置方式发生极大改变的背景下，我国高度重视数字经济发展，协同推进数字产业化和产业数字化，加快建设网络强国和数字中国，并构建网络空间命运共同体。[②]

① 《国家互联网信息办公室关于〈网络数据安全管理条例（征求意见稿）〉公开征求意见的通知》，国家网信办，2021 年 11 月 14 日，http://www.cac.gov.cn/2021-11/14/c_1638501991577898.htm。

② 《习近平向 2023 中国国际智能产业博览会致贺信》，新华网，2023 年 9 月 4 日，http://www.news.cn/politics/leaders/2023-09/04/c_1129844700.htm。

2023 年 9 月 28 日，国家网信办发布了《规范和促进数据跨境流动规定（征求意见稿）》，向社会公开征求意见。[①] 这份征求意见稿是继《数据出境安全评估办法》《网络安全标准实践指南——个人信息跨境处理活动安全认证规范》和《个人信息出境标准合同办法》三项规定之后，对中国跨境数据流动的治理制度进行的补充和完善，也是在国内外跨境数据流动发展的新形势下，中国对数据安全和数据流动的再平衡，极大地完善了中国的数据生态系统。从条文的具体内容看，这项新的数据跨境流动规定主要为中国数据流出到境外提供便利，同时强调流出之后要保障数据的安全。

首先，征求意见稿的第一条至第七条规定了可以不申报数据出境安全评估的各种情况，如学术交流、商业往来、跨境消费、国际贸易等。在不涉及重要数据且不危及个人信息安全的前提下，该规定尽可能为企业跨境数据流动“松绑”，减少数据出境安全评估的场景，简化企业在数据出境之前的合规审查流程，给予企业更多前置自主判断的空间。其次，征求意见稿第八条到第十条强调中国的数据安全保障红线不可动摇。尽管数据跨境流动规定的主要目的是赋予企业更多自主空间，但并非等同于跨境数据完全进行自由流动，豁免企业的数据安全保护义务。例如，涉及党政军和涉密单位的敏感信息仍需要由监管部门进行严密的数据出境安全评估。此外，企业在数据出境之后仍要履行数据安全保护义务，面临风险时及时采取补救措施并向政府部门汇报，接受网信部门的指导与监督。

（三）中国的人工智能安全治理模式

自党的十八大以来，习近平主席多次强调人工智能的重要性，并以统筹发展和安全的科学思路为人工智能治理指明方向。2018 年 7 月 25 日，习近平在金砖国家工商论坛上的讲话中就指出，“未来 10 年，将是世界经济新旧动能转换的关键 10 年”，同时，随着以人工智能为代表的新一轮科技革命和产业变革，这也“将是全球治理体系深刻重塑的 10 年”。[②] 人工智能、

① 《国家互联网信息办公室关于〈规范和促进数据跨境流动规定（征求意见稿）〉公开征求意见的通知》，国家网信办，2023 年 9 月 28 日，http://www.cac.gov.cn/2023-09/28/c_1697558914242877.htm。

②《习近平在金砖国家工商论坛上的讲话》，中国政府网，2018 年 7 月 26 日，https://www.gov.cn/xinwen/2018-07/26/content_5309266.htm。

大数据、量子信息、生物技术等创新技术正在积聚力量，催生大量新产业、新业态、新模式，给全球发展和人类生产生活带来翻天覆地的变化。与世界上其他国家相比，中国人工智能治理体系的建设与完善具有更加特殊的意义。从国家安全的角度来看，中国必须保障跨境数据流动始终以总体国家安全观为指引，实现安全与发展的统筹兼顾。中国改革开放以来的发展经验告诉我们，不发展就是最大的不安全。信息技术革命和数字经济发展是推动新时期中国经济持续增长的关键所在，人工智能技术更是中国实现经济转型升级、提高生产力水平、实现高质量发展的重要工具。

从中国发展的经验来看，人工智能安全治理需要考虑到中国在全球产业链、供应链、价值链中的地位与作用，也需要考虑到不同行业的中国企业在全球范围对接各国人工智能治理体系的合规需求，还需要考虑到不同行业的外国企业在中国开展业务的合规认知与需求；在治理体系完备性、系统性、灵活性之间实现均衡，探索一条符合中国实际需求的人工智能创新治理之路。因此，中国人工智能治理体系的建设与完善必须统筹发展与安全的双向要求，为国家发展注入持久的动力。总体而言，中国的人工智能治理模式强调各国应坚持发展和安全并重的原则，即在人工智能治理中均衡把握经济价值创造、国家安全维护与个人隐私数据保护三个主要维度的综合需求，并且将治理的政策成本、社会成本与合规成本控制在可以接受的范围之内。

结语：走向未来的人工智能治理

从战略层面看，中美有责任在人工智能领域管控分歧并寻求合作，避免技术无序扩散带来的风险，并防止因缺乏互信而滑入人工智能的“零和博弈”。从技术层面看，中美在人工智能研发和治理方面具有高度的互补性。例如，中国在数据方面更具优势，而美国则在算力和算法方面更有优势；在应用领域，两国的侧重点各有不同，双方的合作将有利于实现资源互补，推动技术进步。中美两国在开发复杂的人工智能模型方面都具备丰富的人才资源、强大的经济实力和先进的计算能力，两国在人工智能和全球治理方面的

互动将对人类未来产生深远影响。中美可以通过加强交流与合作来共同推动人工智能技术的健康发展，为全球治理贡献智慧和力量。例如，在数据治理、隐私保护、伦理规范等方面，两国可以加强沟通与协作，共同制定国际标准和规则，以应对人工智能带来的挑战。如果两国能够秉持合作共赢的理念，建设性而非否定性地建构中美战略竞争的新框架，则人工智能安全治理的务实合作不仅可以达成，而且有理由谨慎乐观地预期相关合作可以外溢，为中美战略稳定注入新的积极因素，促进、强化而非消解、弱化中美战略稳定。中国对美总体战略具有延续性、一致性和稳定性；中国面临的考验是如何推动美国采取更加负责任的方式来制定和推行相关战略与政策，这也是中国在未来提高对外战略能力和能力体系建设的主要任务之一。

当前全球人工智能安全治理正处在重要的十字路口，既有大国间竞争的趋势，又有突破竞争思维、寻求合作共赢以造福全人类的可能性。通过利用人工智能的技术优势，各国可以在避免直接冲突的同时，展示自身的科技实力和发展潜力。在构建全球合作框架方面，联合国和其他多边机构已经为推进全球人工智能安全治理做出了积极的努力，并取得了一定的共识。这些机构通过召开国际会议、制定指导原则、推广最佳实践等方式，促进了各国在人工智能领域的交流与合作。要实现更加广泛的合作与进步，各国仍需加强政策对话与协调，建立更加紧密的合作关系，共同应对人工智能带来的挑战。

[责任编辑：樊文光]