全球人工智能治理：进展、困境与前景*

□ 叶淑兰　李孟婷

〔提　　要〕人工智能加剧大国战略竞争，引发新的地缘政治博弈，对国际格局、国际秩序构成重大挑战。如何防范人工智能科技价值异化，引导人工智能朝“善治”方向发展，是事关世界和平与发展的重大议题。全球人工智能治理初步形成以国家为中心、多元主体参与的治理格局，国际社会开始规范人工智能伦理治理，积极探索国家、地区、全球多层次的治理结构，形成风格各异的治理特色。但是，全球人工智能治理仍然面临着主体利益各异、治理对象复杂、排他治理突出、协同治理困难等挑战。全球人工智能治理的前景取决于“善治”、“合治”与“法治”路径能否顺利推进。作为负责任大国，中国积极探索人工智能治理的规则规范，为加强全球人工智能治理提出中国方案、贡献中国经验，成为全球人工智能治理的先行者和助力者。

〔关 键 词〕人工智能、全球治理、大国竞争、人类命运共同体

〔作者简介〕叶淑兰，华东师范大学政治与国际关系学院教授、国际问题研究所所长

李孟婷，华东师范大学政治与国际关系学院博士生

〔中图分类号〕TP18, D815

〔文献标识码〕A

〔文章编号〕0452 8832（2024）4 期 0100-19

* 本文系上海市科学技术委员会重大项目“人工智能新型社会实验与治理方法研究及应用示范”（21511101200）、中央高校基本科研业务费项目华东师范大学哲学社会科学创新团队项目（2024QKT001）的阶段性成果。

人工智能的快速发展驱动全球技术政治的深刻变革，深刻改变国家安全与国际和平的基本范式。[1] 如何防范人工智能科技价值异化，引导人工智能朝“善治”方向发展，是事关世界和平与发展的重大议题，也是全球治理的前沿难题。2023 年 10 月，联合国人工智能高级别咨询机构宣告成立。同月，中国提出《全球人工智能治理倡议》。2024 年 5 月，中国与法国发布关于人工智能和全球治理的联合声明，宣布两国致力于促进安全、可靠和可信的人工智能系统。[2] 在当前大国战略竞争加剧的背景下，习近平总书记指出：“世界各国更加需要加强科技开放合作，通过科技创新共同探索解决重要全球性问题的途径和方法。”[3] 加强对全球人工智能治理既有进展、现实困境与发展前景的研究，总结全球人工智能治理模式与治理特色，不但有利于拓展与丰富面向人工智能的全球治理理论研究，而且有助于凝聚国际社会共识，突破全球人工智能治理的瓶颈，探索有效的人工智能治理框架，推动全球人工智能治理朝着更加公平、包容、负责任、可持续的方向发展。

一、人工智能带来新挑战

人工智能推动科技巨头崛起，改变国家经济、军事力量对比并影响国际格局，冲击国际经济、安全等方面的秩序规范。如果治理不善，人工智能将加剧大国安全困境，导致科技体系的封闭化与分裂化发展，甚至可能危及人类文明生存与发展，颠覆人类主体性。面对人工智能挑战，国际社会亟需加强全球人工智能治理，构建一个公平、透明与负责任的治理体系。

（一）对世界格局的挑战

人工智能的发展将推动科技巨头等非国家行为体崛起，导致国际权力结

[1] 鲁传颖：《人工智能：一项战略性技术的应用及治理》，《人民论坛》2024 年第 1 期，第 72 页。

[2] 《中华人民共和国和法兰西共和国关于人工智能和全球治理的联合声明》，外交部网站，2024 年 5 月 7 日，https://www.mfa.gov.cn/web/wjb_673085/zzjg_673183/xws_674681/xgxw_674683/202405/t20240507_11293821.shtml。

[3] 习近平：《论科技自立自强》，中央文献出版社 2023 年版，第 260 页。

构的重新洗牌，强化美国的科技霸权地位。人工智能还加剧地缘政治竞争，进一步扩大南北差距。

第一，人工智能的发展侵蚀主权国家权力。拥有海量数据的科技巨头等非国家行为体迅速崛起，成为“数字利维坦”，容易导致“数据霸权”。它们拥有巨大的数据建构能力与舆论引导能力，可能在某种程度上侵蚀主权国家权力。脸书（现更名为“Meta”）、谷歌等高科技公司被称为“网络国家”，有些在民众心中甚至拥有比政府高得多的威信。[1] 科技公司在其领域拥有曾经专属于民族国家的权力，[2] 打破了国家的制度化惯例与自我叙事，可能会对国家的本体安全带来持续影响。[3] 科技巨头已经成为国际社会中不可或缺的重要行为体，在人类和平与发展的进程中起到至关重要的作用。面对技术的复杂性与不确定性，如果没有高科技公司的协同治理，全球人工智能治理将陷入失灵困境之中。

第二，人工智能的竞争加速国际权力极化。人工智能作为一项颠覆性技术，将深刻改变国家实力对比，推动非对称性国际权力极化，重塑国际格局。[4] 人工智能技术发展迅速，而且有赖于配套的产业生态发展。当前人工智能技术竞争的核心在于算法、数据与算力，高端芯片则是大国算力竞争的基础。美国企图通过重点产业供应链“脱钩断链”策略遏制中国人工智能技术的发展，维持自身先发优势，在某种程度上强化了人工智能“强者恒强、弱者恒弱”的不平衡发展规律，使“超级大国政治”更为突出。在这场科技竞争中，后发国家非但难以赶超，而且可能被更远地甩在后面，更深地陷入国际分工中心—边缘的格局。

---

［1］ 吴雁飞：《人工智能时代的国际关系研究：挑战与机遇》，《国际论坛》2018 年第 6 期，第 39 页。

［2］ Ian Bremmer and Mustafa Suleyman, “The AI Power Paradox: Can States Learn to Govern Artificial Intelligence - Before It’s Too Late?,” Foreign Affairs, August 16, 2023, https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox.

［3］ 何祎金：《生成式人工智能技术治理的三重困境与应对》，《北京工业大学学报（社会科学版）》2024 年第 2 期，第 128 页。

［4］ 余南平：《新一代通用人工智能对国际关系的影响探究》，《国际问题研究》2023 年第 4 期，第 89 页。

第三，人工智能的博弈加剧地缘政治竞争。人工智能将对国家战略竞争产生重要影响，加剧地缘政治博弈。俄罗斯总统普京曾指出：“谁能成为人工智能领域的领先者，谁就能统治世界。”[1] 美国、俄罗斯、欧洲等大国与地区均积极布局人工智能发展战略。人工智能展现出强大的军事应用潜力，围绕人工智能领域的军备竞赛日趋激烈。美国在传统地缘政治基础上，加紧组建“技术联盟”，主导人工智能国际规则的制定，把地缘政治竞争推向数字空间，意图使中国等发展中国家陷于人工智能产业链下游。高科技公司的权力扩张又将进一步巩固新型“技术极”秩序，[2] 动摇传统地缘政治以国家为中心的主体参与格局，并加剧战略博弈中的话语权竞争。

（二）对国际秩序的挑战

人工智能的快速发展导致数字鸿沟不断扩大，加剧国际经济秩序的不平衡发展，对国际安全秩序产生巨大影响。

一方面，人工智能技术加剧国际经济秩序的不平衡发展。对人工智能的过度依赖可能使人类逐渐丧失创造性思维，并使劳动、工作、创造和谋生变得更加困难。[3] 牛津大学研究团队评估了美国 702 种职业未来被计算机替代的可能性，指出 47% 的岗位面临高度被替代的风险。[4]OpenAI 公司在 2023 年 3 月发布的报告中指出，GPT 模型及相关技术的出现将影响约 80% 的美国劳动力至少 10% 的工作任务。[5] 人工智能技术的运用大大增加了社会的物质

---

［1］ 朱利安·诺塞提：《人工智能：地缘政治的新棋子》，毛志遥译，《国外社会科学文摘》2018 年第 6 期，第 54 页。

［2］ Ian Bremmer and Mustafa Suleyman, “The AI Power Paradox: Can States Learn to Govern Artificial Intelligence - Before It’s Too Late?.”

［3］ 焦建利：《生成式人工智能与人类创造力》，《中国信息技术教育》2023 年第 21 期，第 9 页。

［4］ 中国信息通信研究院、人工智能与经济社会研究中心：《全球人工智能治理体系报告》，2020 年 12 月，第 4 页，http://www.caict.ac.cn/kxyj/qwfb/ztbg/202012/P020201229534156065317.pdf。

［5］ Tyna Eloundou et al., “GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models,” OpenAI, March 17, 2023, https://openai.com/research/gpts-are-gpts.

财富，但是受益的是少数精英，大多数人没有过得更好。[1] 随着人工智能的发展，资本权力将进一步依托技术垄断的优势持续扩张，愈演愈烈的失业潮将加速社会财富的两极分化。从长期来看，受自动化影响最深的可能不是美国和其他发达国家的劳动者，而是以低成本劳动力作为竞争优势的发展中国家。[2] 人工智能提升劳动生产率，降低生产成本，将削弱发展中国家廉价劳动力优势，并减少其参与国际分工的机会。

另一方面，人工智能技术对国际安全秩序造成巨大影响。人工智能技术将更广泛地对军事力量、战略竞争和世界政治变革产生潜在但具有决定性的影响。[3] 军事智能化发展可以实现更科学精准的情报分析与决策部署，以及更准确高效的目标击杀率。人工智能技术的使用大大加快战争的决策与执行速度，使“极速战”（hyperwar）成为可能。[4] 军事领域的人工智能应用引发武器开发及使用方面的伦理道德争议，可能加剧武装冲突与人工智能武器的滥用，并使全球公共安全治理压力上升。低成本、高收益的自主武器系统成为执行危险性较高军事行动的有效手段，因此也降低了发动战争的门槛。此外，致命性自主武器落入恐怖组织的风险正在增大。[5] 当前主要大国纷纷加强人工智能战略规划，力求在大国战略竞争中抢占先机，围绕人工智能领域的军备竞赛也日趋激烈。美国组建“技术联盟”，极力打压中国人工智能发展，维护自身科技霸权地位。马斯克曾对人工智能的军事应用态势表示极大担忧，警告全球人工智能竞赛有可能引发第三次世界大战。[6] 霍金生前更

[1] 杰瑞·卡普兰：《人工智能时代：人机共生下财富、工作与思维的大未来》，李盼译，浙江人民出版社 2016 年版，第 XVIII 页。

[2] 埃里克·布莱恩约弗森、安德鲁·麦卡菲：《第二次机器革命》，蒋永军译，中信出版社 2016 年版，第 252 页。

[3] James Johnson, “Artificial Intelligence and Future Warfare: Implications for International Security,” *Defense & Security Analysis*, Vol.35, No.2, 2019, p.147.

[4] 傅莹：《人工智能对国际关系的影响初析》，《国际政治科学》2019 年第 1 期，第 11-12 页。

[5] 龙坤、徐能武：《人工智能军事应用的国际安全风险与治理路径》，《国际展望》2022 年第 5 期，第 126 页。

[6] “AI Could Spark World War III, Warns Elon Musk,” OECD.AI, https://oecd.ai/en/incidents/1899.

是不无担忧地指出，人工智能的全面发展可能意味着人类的终结。[1]

（三）对全球技术治理的挑战

全球人工智能治理是全球技术治理的前沿议题和棘手问题，关系到全球技术治理的创新发展。人工智能以惊人的速度完成迭代升级，但技术治理却远远滞后于技术本身的发展。人工智能带来的隐私安全、算法“黑箱”、伦理风险与科技价值异化对全球技术治理构成重大挑战，全球人工智能治理亟需加强。

首先，人工智能存在隐私侵犯与泄漏的风险。英国数据监管机构指出，Snapchat 人工智能聊天机器人或危及儿童隐私。[2]2023 年，微软的人工智能研究团队在 GitHub 上发布开源数据时意外泄漏了 38TB 的隐私数据。[3] 其次，人工智能算法“黑箱”可能导致算法歧视问题。2022 年的一项研究表明，机器人大规模地表现出性别、种族等方面的刻板印象。[4] 再次，人工智能存在技术滥用与道德伦理风险。2023 年，一名年轻的比利时男子在与智能聊天机器人聊天数周后自杀身亡，[5] 引发人们对人工智能安全性的关注，加强人工智能的技术治理与伦理治理势在必行。最后，人工智能的使用存在人机责任不清的问题。2022 年 5 月，一辆特斯拉汽车在启用全自动驾驶功能后冲出道路撞上树，随后爆炸起火最终导致车内人员身亡，[6] 人机责任如何界定与区

---

［1］ Rory Cellan-Jones, “Stephen Hawking Warns Artificial Intelligence Could End Mankind,” BBC, December 2, 2014, https://www.bbc.com/news/technology-30290540.

［2］ Shiona McCallum, “Snapchat: Snap AI Chatbot ‘May Risk Children’s Privacy’,” BBC, October 6, 2023, https://www.bbc.com/news/technology-67027282.

［3］ Hillai Ben-Sasson and Ronny Greenberg, “38TB of Data Accidentally Exposed by Microsoft AI Researchers,” September 18, 2023, https://www.wiz.io/blog/38-terabytes-of-private-data-accidentally-exposed-by-microsoft-ai-researchers.

［4］ Andrew Hundt et al., “Robots Enact Malignant Stereotypes,” *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, New York: Association for Computing Machinery, 2022, p.743.

［5］ Lauren Walker, “Belgian Man Dies by Suicide Following Exchanges with Chatbot,” The Brussels Times, March 28, 2023, https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt.

［6］ Trisha Thadani et al., “Tesla Worker Killed in Fiery Crash May Be First ‘Full Self-Driving’ Fatality,” The Washington Post, February 13, 2024, https://www.washingtonpost.com/technology/interactive/2024/tesla-full-self-driving-fatal-crash/.

分成为棘手的问题。

二、全球人工智能治理的进展

人工智能给人类社会带来的巨大挑战引发国际社会的普遍焦虑与不安全感，全球人工智能治理亟需加强。经合组织相关数据显示，当前已有 69 个国家及地区出台了 1000 多项关于人工智能的政策措施。[1] 在主要大国和地区、联合国等国际组织、行业协会、科技企业等多元主体的努力下，全球人工智能治理已经开始起步并取得显著进展，呈现从多方进程主导向多边进程主导的过渡与转移特征。[2] 国际社会开始规范人工智能伦理治理，初步探索多元主体参与的多层次治理结构，形成风格各异的治理特色。

（一）初步形成多元主体治理格局

2016 年以来，已有 40 余个国家和地区先后将人工智能发展上升到国家战略高度。[3] 政府与民间、公共部门与私人部门合作等治理方式 [4] 已经成为人工智能治理不可或缺的一种方式。当前国际社会已初步形成以主权国家为主，国际组织、行业组织、学术团体、高科技公司等多元主体共同参与的人工智能治理网络。

第一，主权国家加强以硬法治理为特征的国家中心治理模式。以国家为中心的治理模式是当前全球人工智能治理最为显著的模式，中美等大国均积极通过制定法律法规等方式加强对国内人工智能发展的布局、规划与监管。国家中心治理模式具有有效使用公权力以及运用硬法治理的优势，有助于取得更为显著的治理效果，但是主权国家各自为政的单一行动已完全无法应对

[1] “National AI Policies and Strategies,” OECD.AI, https://oecd.ai/en/dashboards/overview.

[2] 贾开、俞晗之、薛澜：《人工智能全球治理新阶段的特征、赤字与改革方向》，《国际论坛》2024 年第 3 期，第 68 页。

[3] 中国信息通信研究院：《人工智能白皮书》，2022 年 4 月，第 1 页，http://www.caict.ac.cn/kxyj/qwfb/bps/202204/P020220412613255124271.pdf。

[4] Roderick Rhodes, “The New Governance: Governing without Government,” *Political Studies*, Vol.44, No.4, 1996, pp.652-667.

人工智能的全球风险，无法遏制人工智能安全化、武器化趋势，还容易陷入为保持技术领先地位“重研发而轻治理”的怪圈。为控制国家中心治理模式的负外部性，欧盟在地区层面加强伦理治理，中国则在加强负责任人工智能国家治理基础上积极推进全球治理。

第二，国际组织推进以伦理治理与机制治理为主要特征的治理模式。经合组织、联合国教科文组织等国际组织在推动人工智能伦理治理方面起到引领性作用。2019 年《经合组织人工智能原则》提出以人为本，打造值得信赖的人工智能。[1]2021 年联合国教科文组织发布的《人工智能伦理问题建议书》提出人工智能系统生命周期中应遵守的价值观。[2]2023 年 7 月，在联合国安理会人工智能与安全问题高级别公开会上，秘书长古特雷斯强烈呼吁通过成立新的联合国机构强化对人工智能的全球治理。同年 10 月，联合国人工智能高级别咨询机构宣布成立，这是全球人工智能治理机制化建设向前迈出的重要一步。

第三，行业协会与高科技企业加强以治理联盟与技术治理为主要特征的治理模式。加强行业间治理联盟合作是人工智能治理的重要途径。亚马逊、谷歌等高科技公司早在 2016 年就联合发起人工智能合作伙伴关系（Partnership on AI），至今在全球已有 126 个合作伙伴。[3] 世界经济论坛在 2023 年也发起人工智能治理联盟（AI Governance Alliance）。[4] 此外，行业协会与高科技企业在技术治理上扮演重要角色，美国电气和电子工程师协会（IEEE）提供人工智能伦理与治理方面的行业技术标准，[5] 中国计算机协会、美国计算机协会、国际标准化组织等行业组织以及微软、谷歌等高科

[1] “OECD AI Principles Overview,” OECD.AI, https://oecd.ai/en/ai-principles.

[2] 《人工智能伦理问题建议书》，UNESCO 数字图书馆网站，https://unesdoc.unesco.org/ark:/48223/pf0000380455_chi。

[3] “Our Partners,” Partnership on AI, https://partnershiponai.org/partners/.

[4] 李响：《多方合作才能实现负责任的人工智能治理》，世界经济论坛网站，2023 年 11 月 18 日，https://cn.weforum.org/agenda/2023/11/ai-development-multistakeholder-governance-cn/。

[5] “IEEE GET Program for AI Ethics and Governance Standards,” IEEE Xplore, https://ieeexplore.ieee.org/browse/standards/get-program/page/series?id=93.

技公司，也纷纷制定人工智能技术标准与伦理规范。

（二）加强人工智能伦理治理规范

人工智能存在数据泄露、隐私侵犯、技术滥用等道德风险，并对人类生存与发展构成重大挑战。为引导人工智能向善发展，规范人工智能伦理治理尤为重要。2017 年，美国“有益的人工智能”（Beneficial AI）会议讨论并制定了阿西洛马人工智能原则（Asilomar AI Principles），明确规定人工智能应该是有益的智能，不能不受（人类）控制；其中 23 条原则涉及研究、道德标准与价值观念等方面，是最早且最具影响力的人工智能治理原则之一。[1] 近年来，美国、中国、欧盟等主要国家与地区组织均将伦理治理纳入自身的人工智能发展战略中。此外，联合国、二十国集团、经合组织等也不断提出人工智能伦理原则与倡议。联合国教科文组织 2021 年通过《人工智能伦理问题建议书》，从价值观、伦理原则与政策指导三方面确立了全球首个人工智能伦理框架，高达 193 个成员国一致通过该建议书，促使人工智能伦理规范成为更大范围的国际共识。

欧盟在规范人工智能伦理治理上颇具代表性。欧盟人工智能技术和产业发展均落后于美国，其将人工智能治理的重心放在伦理道德与法律法规上。作为人工智能监管在全球范围内的“第一推动者”，[2] 欧盟强调人工智能应合法、合伦理、稳健，要构建可信赖的人工智能生态系统。[3]2024 年通过的《人工智能法案》进一步强化了欧盟对高风险人工智能应用的监管力度。

中国较早开始加强人工智能伦理治理，早在 2017 年发布的《新一代人工智能发展规划》就提出 2025 年要初步建立人工智能伦理规范。[4] 随后，中

---

［1］ “Asilomar AI Principles,” Future of Life Institute, August 11, 2017, https://futureoflife.org/open-letter/ai-principles/.

［2］ Nathan Benaich and Ian Hogarth, “State of AI Report 2021,” October 12, 2021, p.166, https://www.stateof.ai/2021.

［3］ European Commission, “Ethics Guidelines for Trustworthy AI,” April 8, 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

［4］ 《新一代人工智能发展规划》，中国政府网，2017 年 7 月 8 日，https://www.gov.cn/gongbao/content/2017/content_5216427.htm。

国对人工智能伦理治理的相关原则和具体要求作出更为明确的规定。2022 年发布的《中国关于加强人工智能伦理治理的立场文件》则从人工智能监管、研发、使用及国际合作等方面提出主张，推动国际人工智能伦理治理。[1] 加强对人工智能伦理治理的规范，有助于确立技术开发与应用过程中需遵循的基本价值理念，将伦理规范内化到技术发展的全生命周期中，避免颠覆性风险的发生，确保人工智能技术向善发展。

（三）积极探索多层次治理结构

人工智能治理已经初步形成国家、地区与全球多层次治理结构。从国家层面看，主要大国均通过制定政策、开展立法等方式加强人工智能国内治理。从地区层面看，欧盟、东盟均通过了各具特色的人工智能治理文件。欧盟《人工智能法案》以严格著称，对人工智能的稳健性、准确性、安全性等方面制定了严格标准。东盟《人工智能治理和伦理指南》则凸显灵活性，强调文化差异，倾向于让成员国自行确定最有效的措施。[2] 从全球层面看，目前人工智能治理主要依赖联合国、七国集团、二十国集团在内的国际组织与国际机制。联合国宣布成立的人工智能高级别咨询机构为联合国框架下人工智能管理机构的形式和功能奠定基础。

人工智能国家治理进展相对较快，全球治理进展较为缓慢。受技术民族主义影响，不同主体与层次之间的协同治理仍较为匮乏。全球人工智能治理大多停留在概念与理念的提出与阐述上，真正落地的实践行动还相当有限。全球人工智能治理的进一步发展需要多元、多层次治理主体之间的协同治理与网络化治理模式的创新。

（四）形成风格各异的治理特色

伴随人工智能技术的发展，美国、欧盟、中国等纷纷加强对人工智能治理方法、模式的探索，结合各自政治制度与人工智能发展特点，逐渐形成风

---

［1］《中国关于加强人工智能伦理治理的立场文件》，外交部网站，2022 年 11 月 17 日，https://www.mfa.gov.cn/web/ziliao_674904/zcwj_674915/202211/t20221117_10976728.shtml。

［2］ ASEAN, “ASEAN Guide on AI Governance and Ethics,” https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf.

格各异的人工智能治理特色。

美国为保持技术领先地位形成相对宽松的治理特色。美国加强人工智能治理的首要目标在于提升美国人工智能的全球领导力。[1] 由于美国将人工智能研发作为优先事项，在治理措施的制定上较为审慎，监管力度也较为宽松，更倾向于行业自律，以免过度治理阻碍科技创新。2023 年美国国会共提出 181 项有关人工智能的议案，但仅有 1 项议案获得通过。[2] 美国还设立了国家人工智能倡议办公室（NAIIO）、国家人工智能咨询委员会（NAIAC）等机构，推动各部门、各行业的协同治理。

欧盟形成重视伦理治理与软硬法相结合的治理特色。欧盟人工智能治理较为严格，尤其重视对伦理规范、数据保护、网络安全等方面的监管。[3] 欧盟 2019 年发布的《可信赖的人工智能伦理准则》从人类自主性、技术安全性、隐私治理、透明度、防止偏见、社会福祉、问责制度七大方面提出人工智能伦理要求。[4] 欧盟不断将以伦理准则为代表的软法向立法推进，2024 年通过全球首部《人工智能法案》，[5] 实现区域性硬法的落地。

中国形成既立足国内又面向全球的负责任人工智能治理特色。中国高度重视人工智能治理，提出公平公正、包容共享、尊重隐私、共担责任等人工智能治理原则。2023 年中国《生成式人工智能服务管理暂行办法》成为全球首部针对生成式人工智能的专门立法。中国还积极向国际社会提出《全球数据安全倡议》（2020 年）、《关于加强人工智能伦理治理的立场文件》（2022 年）以及《全球人工智能治理倡议》（2023 年）等文件。中国支持在联合国

---

［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.

［2］ “The AI Index 2024 Annual Report,” Stanford University, April 2024, p.382, https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf.

［3］ 宋黎磊、戴淑婷：《科技安全化与泛安全化：欧盟人工智能战略研究》，《德国研究》2022 年第 4 期，第 53 页。

［4］ European Commission, “Ethics Guidelines for Trustworthy AI.”

［5］ European Commission, “AI Act,” 2024, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

框架下讨论成立国际人工智能治理机构，协调国际人工智能发展、安全与治理重大问题，[1] 为包括发展中国家在内的国际社会提供中国方案，这对于加强负责任的全球人工智能治理起到积极推动作用。

三、全球人工智能治理的困境

大国战略竞争加剧了颠覆性技术的安全化与武器化倾向，超越地区合作层面的全球性数字治理几乎处于全面停顿的状态。[2] 当前全球人工智能治理面临着主体利益各异、治理对象复杂、治理共识不足、治理方式各异等困境。

（一）主体利益各异，全球治理乏力

全球人工智能治理虽然形成了多元主体参与的治理格局，但多元主体之间的互动与协同还远远不足，治理力量呈现分散化与碎片化的状态。主要大国利益诉求多样化，不同国家间的治理法规政策存在很大差异，而国际组织、行业协会、学术团体、高科技企业又立场各异，各主体难以达成相对统一的全球治理目标、标准与原则。

国家作为重要的国际行为体，是推动全球人工智能治理的中坚力量。然而，在现实主义思维支配下，国家极为重视人工智能国内治理，对全球治理却更多处于观望与呼吁状态，缺乏切实行动。加强全球人工智能治理无疑需要美国等西方发达国家发挥更为重要的作用，但美国为了维护其科技霸权地位，通过技术联盟这一排他性框架 [3] 打压其他国家的科技发展，企图扩大技术代差。美国通过其主导的技术治理多边体系抢占人工智能国际规则制定的主导权，在客观上加大了全球人工智能治理的现实困难。

科技巨头之间治理价值观的不一致令人工智能治理前景变得更为模糊不

[1] 《全球人工智能治理倡议》，外交部网站，2023 年 10 月 20 日，https://www.mfa.gov.cn/web/zyxw/202310/t20231020_11164831.shtml。

[2] 戚凯、周祉含：《全球数字治理：发展、困境与中国角色》，《国际问题研究》2022 年第 6 期，第 59 页。

[3] 唐新华：《西方“技术联盟”：构建新科技霸权的战略路径》，《现代国际关系》2021 年第 1 期，第 38 页。

清。2023 年 3 月，马斯克等上千名科技人士发布公开信，呼吁所有人工智能实验室暂停训练比 GPT-4 更强大的人工智能系统至少 6 个月，反对过快发展人工智能。[1] 马斯克批评 OpenAI 追求利润而忽视人工智能安全，但他很快于 2023 年 7 月成立新人工智能公司 X.AI，声称目标是“了解宇宙的真实本质”，并于当年 11 月推出聊天机器人 Grok，与谷歌、OpenAI 等行业领导者竞争。[2] 高科技公司既无法回避对人工智能快速发展潜在风险的担忧，又希望获取技术领先地位并实现自身的科技价值观，客观上加剧了人工智能的研发竞争，使全球人工智能治理陷入迟滞甚至停顿的状态。

（二）治理对象复杂，治理时机难定

人工智能治理是一个全球性、系统性的工程。在 2021 年世界人工智能大会（WAIC）治理论坛上，清华大学教授薛澜提出“敏捷治理”概念，认为数据、算法、应用场景、平台、企业作为治理对象应自下而上分层治理。[3] 显然，人工智能治理对象的复杂性及其产生的不确定性风险使治理难度进一步加大。算法“黑箱”的透明化困境及其带来的算法歧视问题、人工智能实际应用场景的变化，都成为人工智能治理不可回避的难题。[4]

人工智能技术处于快速迭代发展的过程之中，对其进行治理与监管极易陷入“科林格里奇困境”（Collingridge’s Dilemma）[5]：在技术相对可控的发展早期难以对其未来发展的风险作出准确预测，因此无法过早开始治理；在技术发展成熟后，再对已经显现并深嵌于社会的风险进行治理则异常困难。例如，在人工智能深度合成技术发展早期，人们对其后续可能被不法分子滥

[1] “Pause Giant AI Experiments: An Open Letter,” Future of Life Institute, March 22, 2023, https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

[2] Robert Hart, “Elon Musk’s Artificial Intelligence Startup xAI Reportedly Nears $18 Billion Valuation With Fresh Funding As AI Race Heats Up,” May 9, 2024, https://www.forbes.com/sites/roberthart/2024/05/09/elon-musks-artificial-intelligence-startup-xai-reportedly-nears-18-billion-valuation-with-fresh-funding-as-ai-race-heats-up/.

[3] 《薛澜教授：人工智能治理很关键，创新和治理须协调推动》，清华大学中国科技政策研究中心网站，2021 年 8 月 5 日，http://cistp.sppm.tsinghua.edu.cn/info/1024/1227.htm。

[4] 曹建峰：《人工智能：机器歧视及应对之策》，《信息安全与通信保密》2016 年第 12 期，第 18 页。

[5] David Collingridge, *The Social Control of Technology*, London: Frances Pinter, 1980, p.11.

用进行诈骗等问题认识不足，未能及时介入治理，而当深度合成技术已被滥用于电信诈骗、色情传播、肖像权侵犯、舆论操控之后，又面临更为复杂的治理难题。总之，“前瞻治理”无法准确预测技术的发展方向，“事后治理”又无法弥补已经造成的技术负面影响，这成为人工智能治理的一大困境。

（三）治理共识尚浅，排他治理突出

当前国际社会虽然有加强人工智能发展与治理的基本共识，但具体要如何推进治理，尤其是如何推进全球层面的治理，则难以达成一致。受不同行为体利益取向、意识形态、文化价值观差异等影响，各方治理理念存在较大差异，甚至相互冲突，要达成深度治理共识显得尤为困难。

各国在追求人工智能快速发展的同时，没有进行充分的技术分享。世界各国或是在本国行政管辖权范围内构建相应的人工智能技术发展规范，或是在有限的排他性“联盟集团”内部展开治理问题协商。[1]2019 年 5 月，经合组织数字经济政策委员会（CDEP）发布了全球首份政府间人工智能政策指南，其中关于人工智能发展的五项原则和五点建议基本上与特朗普签署的“维持美国在人工智能领域领先地位”行政令的基调完全一致；而且，恰恰由于美国的参与和支持，该政策指南成为一份被发达国家认可的“国际准则”。[2]但是美国在背后积极推动的此次讨论不仅未邀请中国代表参加，还拒绝中国学者参与。[3] 美国试图在全球人工智能治理中边缘化中国，这种带有排他性的做法极大影响了全球人工智能治理的进程。此外，大国领导力缺失将制约全球治理能力的提升，[4] 使全球人工智能治理的国际合作空间难以拓展。

（四）治理松紧不一，协同治理困难

由于过严或过松的监管都不利于人工智能技术的健康发展和科学治理，

[1] 余南平：《新一代通用人工智能对国际关系的影响探究》，第 94 页。

[2] 李括：《美国科技霸权中的人工智能优势及对全球价值链的重塑》，《国际关系研究》2020 年第 1 期，第 41 页。

[3] 朱荣生、陈琪：《美国对华人工智能政策：权力博弈还是安全驱动》，《和平与发展》2022 年第 6 期，第 60 页。

[4] 罗会钧、查云龙：《人工智能时代的全球治理转型与中国应对》，《上海交通大学学报（哲学社会科学版）》2023 年第 12 期，第 15 页。

因此在鼓励人工智能技术创新、保持发展势头与设置监管边界、制定治理政策之间，各国往往根据其国家利益与现实需求采取松紧不一的治理方式。总体上，中国与欧盟的治理强度高于美国。美国始终强调在人工智能领域的全球领先地位，因此其治理侧重于行业自治、应用与流程管理，通过政策指南、行业共识、行业标准等方式，以相对较弱的监管来保证人工智能产业发展的活力。美国 2020 年发布的《人工智能应用规范指南》强调：“为了促进美国的创新，各机构要谨记人工智能的国际应用，要确保美国公司不受美国监管制度的不利影响。”[1] 欧盟以强监管态度，通过硬法与软法并存的治理方式，进一步将维护人工智能伦理价值观上升至欧洲整体战略层面。[2]2018 年 5 月生效的欧盟《通用数据保护条例》是目前世界上最严格的数据隐私保护法规。

由于利益差异以及治理松紧程度不一，国家之间往往难以进行有效的协同治理。同为技术联盟成员，美国与欧盟紧密合作，意图领导全球人工智能治理规则的制定，但在涉及各自人工智能治理领域的具体措施时，双方时有分歧与摩擦。欧盟谋求技术主权，并不想成为美国的战略附庸，因而试图在中美之间寻求平衡。欧盟在数据隐私保护与监管方面的政策要比美国严格得多，两者难以达成相对一致的治理方式与规则，在全球层面上的协同治理则更是难上加难。尽管其他多边国际组织、行业自治网络和技术社团等在人工智能的伦理准则和监管机制方面展开有益探索，但由于不同组织与机制间的主张各异，彼此之间的沟通协调与协同治理也颇为困难。

四、全球人工智能治理的出路

人工智能所涉及利益与技术的复杂程度决定了其全球治理的难度要远高于其他领域的治理。面对人工智能治理困境，存在推动全球人工智能走向“善

[1] “Guidance for Regulation of Artificial Intelligence Applications,” The White House, November 17, 2020, p.5, https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf.

[2] 中国信息通信研究院、人工智能与经济社会研究中心：《全球人工智能治理体系报告》，第 10 页。

治”与全球治理难以推进两种声音。[1] 人工智能治理具有复杂性、长期性与艰巨性的特征，其治理前景取决于“善治”、“合治”与“法治”路径能否顺利推进。

（一）“善治”：共识下的差异

人工智能是把双刃剑，它朝着何种方向发展取决于国际社会的治理能力。要确保人工智能发展是为人类谋福利而非颠覆人类社会，其底线在于“善治”，即发展公平、透明、负责任、向善的人工智能，通过良善的治理来实现全球的正义目标。[2]

“善治”已经成为人工智能治理伦理话语的高度概括。阿西洛马人工智能原则强调人工智能发展要符合人类利益，要创造安全、透明、负责任的人工智能系统。[3] 中国倡导“以人为本”理念与“智能向善”宗旨，确保人工智能始终处于人类控制之下，打造可审核、可监督、可追溯、可信赖的人工智能技术。[4] 无论是欧盟的人工智能发展愿景，[5] 还是 2023 年联合国人工智能高级别咨询机构发布的《以人为本的人工智能治理》报告，[6] 无不体现出“善治”的共性。“善治”中的“以人为本”既源于中国古代的民本思想，又契合西方人权理念，是中西方文化的共通价值，能够成为促进各国人工智能治理共识的价值基础。

尽管各方对“善治”能够产生观念上的共鸣，促成初步共识，但要真正达成“善治”目标并非易事。多元主体对于“善治”概念以及达成“善治”路径的理解仍有差异。对于人工智能“善治”更为强调的应该是治理的公平与正义，还是治理的伦理与规范，抑或是治理的敏捷与效能等，目前尚无定论。

---

［1］ 高奇琦：《全球善智与全球合智：人工智能全球治理的未来》，《世界经济与政治》2019 年第 7 期，第 47-48 页；余南平：《新一代通用人工智能对国际关系的影响探究》，第 93-94 页。

［2］ 高奇琦：《全球善智与全球合智：人工智能全球治理的未来》，第 38 页。

［3］ “Asilomar AI Principles.”

［4］ 《全球人工智能治理倡议》。

［5］ European Commission, “Ethics Guidelines for Trustworthy AI.”

［6］ “Governing AI for Humanity,” UN, December 31, 2023, https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf.

因此，人工智能“善治”仍处于较为理想化的呼吁阶段。如要真正实现“善治”目标，国际社会需要在已有基本共识的基础上，坚持客观审慎的态度，强化人的主体性，明确人工智能的研发禁区，将可解释、可信赖、安全、负责任的人工智能作为未来发展方向。

（二）“合治”：协同治理创新

人工智能的治理目标靠单一行为体难以实现，迫切需要加强不同国际行为体的“合治”。《布莱切利宣言》[1] 指出，人工智能的风险从本质上看是全球性的，最好通过国际合作来解决。[2] 人工智能“合治”需要推动跨国界、跨地区、跨领域、跨学科、跨部门的合作与协同治理。

加强人工智能治理不仅需要人的自觉、反思与克制，还需要在国家主义和全球主义间寻求某种平衡，[3] 实现技术治理、社会治理、政府治理与国际治理之间的合作，从政企协同、政民协同、人机协同等多层面推动多元主体协同治理。人工智能“合治”呈现多主体合治、跨界合治、跨国合治等多种形式，需要推动多边合作机制，充分发挥行业协会、研究机构、高科技公司的积极作用。

当前，不同行为体人工智能治理模式的相互影响与借鉴为推动“合治”创造了一定可能性。分散的人工智能治理机制在相互链接的过程中出现了成员范围、合作内容的重叠，[4] 因此需要加强对人工智能治理机制的有机整合。推动全球人工智能“合治”也是发展中国家共享人工智能知识成果、获得开源人工智能技术的宝贵机会。[5] 但是，由于多元主体的利益诉求与合作基础

---

［1］ 2023 年 11 月，首届全球人工智能安全峰会在英国布莱切利庄园举行。在开幕式上，包括中国在内的 28 个国家和欧盟共同达成的《布莱切利宣言》正式发布，这是全球第一份针对人工智能的国际性声明。

［2］ “The Bletchley Declaration by Countries Attending the AI Safety Summit,” 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.

［3］ 张东冬：《人类命运共同体理念下的全球人工智能治理：现实困局与中国方案》，《社会主义研究》2021 年第 6 期，第 172 页。

［4］ 陈佳慧：《AI 治理，从分散走向协同？》，网易，2024 年 1 月 9 日，https://www.163.com/dy/article/IO1HPULI0514BTKQ.html。

［5］ 《全球人工智能治理倡议》。

不同，人工智能治理很容易陷入“分而治之”的困境。这需要处理好“分”与“合”的内在张力，平衡好国家主义和全球主义的关系，将碎片化与分散化的治理转化为治理合力，增强全球治理的包容性。

（三）“法治”：软硬法并济

人工智能“法治”意味着通过制定规则、规范、机制避免技术价值的异化，通过制度性的约束与监督推动人工智能走向“善治”。技术本身具有价值中立性，但技术的滥用或误用可能使其价值向度偏离正常轨道。阿西洛马人工智能23条原则暗含了发展通用人工智能甚至是超（高）级人工智能的目标，[1]这导致人工智能在自主性方面过于超前。如果没有外部监管，放任技术自由迭代发展，不仅将加剧技术风险，还可能动摇人类基本价值。加强人工智能“法治”建设需要在有法可治的基础上进行依法治理，加强国内硬法与国际软法建设，尤其重视国际软法治理体系的建设，强调国际规范、国际机制的引导与约束作用，明确人工智能研发、使用与治理的价值理念。“法治”既是对“善治”与“合治”的制度保障，也是实现全球人工智能治理落地的重要推力。

硬法常见于国内法。中国的《个人信息保护法》《网络安全法》《数据安全法》为人工智能治理奠定了法律基础。欧盟《人工智能法案》明确了人工智能在欧盟市场的准入规则，并要求在高风险领域进行人工监督；该法案为人工智能治理从原则性指导走向可审核、可执行的法律程序提供了参考路径，是区域性“法治”的一个例证。但是，国际社会并不存在类似国内社会的硬法治理，只能在有约束性承诺的国际公约的基础上，制定软性治理框架，发挥国际法与国际机制在推动全球人工智能治理方面的作用。面对全球人工智能治理的复杂局面，单纯依靠国内硬法或国际软法并不足以实现全面有效的人工智能治理。因此，要实现“法治”的理想状态，还需要探索软硬并济的混合治理机制，在克服两者弊端的基础上实现优势互补，让软法强化硬法的灵活性，让硬法保障软法的效力性，[2]实现软硬法的协同治理。

---

[1] 高奇琦：《全球善智与全球合智：人工智能全球治理的未来》，第31-32页。

[2] 郝家杰：《人工智能国际协同监管“软法”机制研究》，《全球科技经济瞭望》2023年第9期，第42页。

五、结语

人工智能的发展对地缘政治、国际体系以及人类社会均构成巨大挑战。人工智能大国竞争极易陷入个体理性导致集体非理性的悖论中。科技命运共同体的本质需要国家超越技术民族主义意识，加强全球技术创新与协同发展。人工智能治理失灵问题的解决无法依靠单一国家，需要世界各国的共同努力。推动全球人工智能治理体系的建设，需要进一步加强机制化建设，构建技术监控、安全评估、风险防范“三位一体”的治理体系。为防止人工智能治理领域寡头垄断地位的形成，需要充分发挥主权国家作用，推动多边主义治理，[1] 重视发达国家与发展中国家的协同合作治理，积极应对当前日益扩大的数字鸿沟，推动国际政治经济的平衡发展，实现全球人工智能治理的“善治”目标。

作为负责任的大国，中国始终“深度参与全球科技治理，贡献中国智慧，着力推动构建人类命运共同体”。[2] 中国在国内层面需要积极加强人工智能社会治理实践，建立先行先试实验点与示范点，探索人工智能治理的规则规范，以更好地为全球人工智能治理提供中国经验。在国际层面，中国积极支持联合国框架下的人工智能治理议程，提出《全球数据安全倡议》《全球人工智能治理倡议》《国际科技合作倡议》等中国方案，分享技术治理的中国智慧，成为全球人工智能治理的先行者和助力者。作为世界上最大的发展中国家，中国在人类命运共同体理念引领下，积极提出人工智能治理标准、规则、规范，努力缩小发展中国家与发达国家之间的数字鸿沟，不断推动构建公正、包容、可信、向善的全球人工智能治理体系。

【责任编辑：肖莹莹】

[1] 陈伟光、袁静：《人工智能全球治理：基于治理主体、结构和机制的分析》，《国际观察》2018 年第 4 期，第 31 页。

[2] 《习近平：在中国科学院第十九次院士大会、中国工程院第十四次院士大会上的讲话》，中国政府网，2018 年 5 月 28 日，https://www.gov.cn/xinwen/2018-05/28/content_5294322.htm。