

• 管理纵横 •

DOI: 10.3724/BNSFC-2025.04.16.0002

关于人工智能全球治理的分析

葛慧林^{1*} 谢国² 王乐³ 王志衡² 吴国政²

1. 江苏科技大学 科学技术研究院, 镇江 212100

2. 国家自然科学基金委员会 信息科学部, 北京 100085

3. 西安交通大学 人工智能学院, 西安 710049

[摘要] 随着人工智能(Artificial Intelligence, AI)技术的迅猛发展,特别是生成式AI如ChatGPT、Sora等产品的涌现,人工智能已深度融入人们的日常生活和生产工作中,对社会产生了深远影响。然而,这种“狂飙突进”式的发展也带来了安全隐患和风险挑战,如数据安全、算法偏见、伦理失范等系统性风险逐渐显现。在此背景下,亟需探索人工智能全球治理问题,以确保符合全球公共利益。本文基于人工智能全球治理的现状,从技术、协调机制和伦理法律三个方面系统分析了人工智能全球治理的挑战;以Web of Science核心合集数据库920篇文献为研究样本的人工智能全球治理趋势与方向分析表明:当前相关研究主要聚焦伦理框架与道德原则、法律与政策治理以及技术挑战与实践应用三个方面。未来可从构建隐私保护与数据安全的全生命周期管理体系、推动算法透明度标准化与伦理治理框架落地、探索“以AI治理AI”的动态模式和加速全球价值对齐框架建设等方向展开进一步研究;最后结合我国实际,构建立法规范、技术协同与全球合作三位一体的治理路径,为全球规则制定提供实践方案。

[关键词] 人工智能;全球治理;前沿趋势;中国治理实践

人工智能(Artificial Intelligence, AI)作为科技革命的前沿领域,正以其独特的技术能力和创新潜力,深刻地重塑着全球经济、政治、文化及社会发展的格局。在《人工智能安全标准化白皮书(2023版)》中提到,当前以生成式人工智能为代表的新技术、新应用不断打破人们对人工智能的固有认知,但其带来的网络意识形态安全、数据安全、个人信息安全等方面的新风险、新挑战也日益成为国际社会共同关注的焦点。作为负责任的人工智能大国,中国一贯高度重视人工智能治理,在人工智能全球治理中扮演了积极角色。习近平总书记多次作出重要指示,强调“要深入把握新一代人工智能发展的特点,加强人工智能和产业发展融合,为高质量发展提供新动能”。2023年10月18日,国家主席习近平在第三届“一带一路”国际合作高峰论坛开幕式发表主旨演讲,提出“全球人工智能治理倡议”(以下简称“倡议”)。

倡议着眼于数字文明时代的人工智能治理全球合作,为世界提供了基于人类命运共同体理念的人工智能治理新视角。我国在人工智能领域治理的实践生动诠释了人类命运共同体理念的深刻内涵,通过发布《人工智能军事应用规范立场文件》《人工智能伦理治理框架文件》和《人工智能安全治理框架》2.0,既将全球发展倡议、安全倡议与文明倡议转化为具体的治理行动,又构建起涵盖技术规制、伦理准则的多维治理路径。这些举措不仅彰显了中国在人工智能全球治理中的参与深度与引领力度,更为人工智能技术的健康发展注入了规则新动能。

作为一个跨学科、跨领域的综合性议题,人工智能全球治理不仅涉及技术发展的伦理、法律、安全等多个维度,还关乎全球治理体系的有效性和公正性。目前,人工智能市场规模突破5 000亿元,市场主体高度集中于

收稿日期:2025-04-16; 修回日期:2025-06-13

* 通信作者, Email: ghl1989@just.edu.cn

本文受到江苏省社会科学基金省市协作项目(24XZB027)的资助。

引用格式: 葛慧林, 谢国, 王乐, 等. 关于人工智能全球治理的分析. 中国科学基金, 2025, 39(5): 835–843.

Ge HL, Xie G, Wang L, et al. Analysis of the global governance of artificial intelligence. Bulletin of National Natural Science Foundation of China, 2025, 39(5): 835–843. (in Chinese)

服务业。依托政策红利、资本注入与人才储备三驾马车,我国AI产业实现跨越式发展,跻身全球第一梯队。中国信息通信研究院数据显示:自2019年起,产业规模呈现爆发式增长,2023年市场规模攀升至5 784亿元,增速回调至13.9%,显示出由规模扩张向质量提升的发展转型。在此背景下,探讨人工智能全球治理的现实困境与未来路径,对保障AI技术的健康发展、促进国际合作交流以及维护全球和平稳定具有重大的学术价值和现实意义。本研究旨在深入分析人工智能全球治理的现状、挑战与机遇,提出相应的治理策略和建议,以期构建一个开放、包容、可持续的全球AI治理框架提供理论支持和实践指导。

1 人工智能全球治理的必要性

随着人工智能技术的迅猛发展,尤其是生成式人工智能如ChatGPT等技术的兴起,人工智能技术已广泛应用于人们的日常生活及各行各业,涵盖医疗、文化、金融等诸多领域,成为推动经济增长与社会进步的关键动力。然而,这种井喷式的发展及其对全球治理结构的潜在影响,迫切要求国际社会构建一个全面、协同的全球治理框架^[1]。鉴于人工智能技术所带来的跨国界风险、伦理挑战和治理赤字,全球治理的必要性日益凸显。人工智能究竟会对人类社会发展产生何种作用,取决于能否确保人工智能技术的发展方向与全球公共利益相一致,促进国际合作,以实现技术的普惠性、安全性和可持续性。因此,全球治理不仅关乎技术发展的全球协调,也关乎维护国际秩序、保障全球安全和推动包容性增长。AI全球治理的构建成为国际关系和全球政策领域中的一个核心议题,其必要性源于对全球共同未来的责任和对人类共同福祉的追求。对此,我国应在迭代完善国内AI治理体系的基础上,通过推动全球共治共享、构建监管合作规则、加快国际标准制定和创新多边协调机制等举措,持续提升国际人工智能治理的影响力。

2 人工智能全球治理的现状

在全球范围内,超过60个国家和地区已发布人工智能相关规划,国际人工智能政策的发展情况呈现出多元化和积极向前的态势。随着人工智能技术的不断发展和普及,各国政府纷纷出台相关政策,以推动人工智能技术的研发、应用和产业发展。欧盟2018年发布的《人工智能协调计划》和2019年发布的《可信任人工智能伦理指南》等政策方案,构建了卓越可信任的人工智能生态系统方案和欧洲统一的人工智能监管框架,并于2021年推出了《人工智能法案》,以全面覆盖人工智能

领域。欧盟更注重个人安全与隐私安全,推行的强监管模式可能不适应发展中国家的技术需求与治理现状,并可能制约技术创新。美国则发布了《维护美国人工智能领导地位的行政命令(2019)》等文件,推动技术创新与国际竞争力,强调技术领先构想,并倡导软性监管,重视技术发展保障,更强调本国发展利益。此外,美国还发布了一系列人工智能治理相关政策,2020年发布的《联邦政府人工智能治理指南》针对联邦机构提出AI技术的应用和监管指南,倡导负责任和透明的AI使用;2021年发布的《人工智能的自动化和公平报告》探讨人工智能技术对公平、隐私和社会的影响,推动AI治理的社会性研究;2022年发布的《美国人工智能治理战略》提出五大原则,旨在保护个人隐私、反对算法歧视、促进透明和公正的AI系统。美国对全球技术合作与治理的合作伙伴的要求较高,难以推动人工智能的协同发展^[2]。相比之下,中国倡导人类命运共同体思想,聚焦人工智能的整体性安全与发展,强调治理的包容、公平、安全与人本问题。中、美、欧三方在人工智能发展与安全问题的治理侧重点与核心关注点存在差异。中国在全球治理中强调提高各方参与感,反对技术垄断与单边强制措施制造发展壁垒行为,重视人工智能全球治理的公平性、动态性,主张建立健全人工智能的法律和规章制度,保障数据安全。针对人工智能发展中的社会问题,中国引导技术的健康发展,为技术发展内置伦理属性,强调全球治理的有益性。中国以负责任的态度,为全球共同应对人工智能发展带来的安全风险与发展风险提供中国方案,展现大国胸怀。

相较于国际人工智能发展轨迹,中国AI产业虽起步较晚,技术萌芽始于1978年,但乘着互联网革命与技术迭代的东风,历经二十载产研攻坚,已突破探索培育期,进入爆发式增长阶段。作为后发先至的国家,中国坚持发展与安全的双重原则,高度重视人工智能的颠覆性影响,并将其提升至国家战略高度。国家对这一新兴产业的支持力度持续增强,如表1所示,近年来密集出台了一系列政策举措,旨在加速人工智能的创新步伐,并推动其与各行业的深度融合。自党的十八大以来,中国陆续出台了《“互联网+”人工智能三年行动实施方案》等政策,推动人工智能发展,并提出了人工智能“三步走”战略和“8条治理原则”,以加快对人工智能有效发展和治理的探索^[3]。政府不仅鼓励人工智能技术的快速发展,还积极促进其与实体经济的紧密结合,从而拓宽了人工智能技术的应用领域。从2016年国务院发布的《“十三五”国家科技创新规划》将人工智能确立为国家战略层面的重大科技项目,到2017年7月印发的《新一代人工

表1 中国人工智能行业国家层面发展政策(部分)
Table 1 National Development Policies for Artificial Intelligence Industry of China (Partial)

时间	政策	发布单位	主要内容
2017年3月	《2017年政府工作报告》	国务院	一方面要加快培育新材料、人工智能、集成电路、生物制药、第五代移动通信等新兴产业,另一方面要应用大数据、云计算、物联网等技术加快改造提升传统产业,把发展智能制造作为主攻方向
2017年7月	《2017年政府工作报告》	国务院	推动我国人工智能产业的发展,构建人工智能科技创新体系,培育智能经济,建设智能社会,加强军民融合,构建智能化基础设施和前瞻布局重大科技项目等方面
2018年11月	《新一代人工智能产业创新重点任务揭榜工作方案》	工业和信息化部	通过将人工智能细分主要领域,加快我国人工智能产业与实体经济深度融合
2019年3月	《关于促进人工智能和实体经济深度融合的指导意见》	中央全面深化改革委员会	把握新一代人工智能的发展特点,结合不同行业、不同区域特点,探索创新成果应用转化的路径和方法
2020年7月	《国家新一代人工智能标准体系建设指南》	国家标准化管理委员会等	提出了一个包含八个主要部分的人工智能标准体系结构
2021年3月	《国民经济和社会发展规划第十四个五年规划和二〇三五年远景目标》	中共中央	瞄准人工智能等前沿领域,实施一批具有前瞻性、战略性的国家重大科技项目
2021年9月	《新一代人工智能伦理规范》	国家新一代人工智能治理专业委员会	提出人工智能管理、研发、供应、使用等特定活动的18项具体伦理要求
2023年2月	《质量强国建设纲要》	中共中央、国务院	加快大数据、网络、人工智能等新技术的深度应用,促进现代服务业与先进制造业、现代农业融合发展
2023年7月	《生成式人工智能服务管理暂行办法》	国家网信办等	加快大数据、网络、人工智能等新技术的深度应用,促进现代服务业与先进制造业、现代农业融合发展
2024年6月	《四部门关于印发国家人工智能产业综合标准化体系建设指南(2024版)的通知》	工业和信息化部等	到2026年,标准与产业科技创新的联动水平持续提升,新制定国家标准和行业标准50项以上,引领人工智能产业高质量发展的标准体系加快形成
2025年5月	《数字中国建设2025年行动方案》	国家数据局综合司	持续深挖人工智能等数字技术应用新场景,以数字化驱动生产生活和社会治理方式变革,加快推动数字领域国际合作,全面提升数字中国建设水平
2025年9月	《人工智能安全治理框架》2.0	国家互联网应急中心	持续深挖人工智能等数字技术应用新场景,以数字化驱动生产生活和社会治理方式变革,加快推动数字领域国际合作,全面提升数字中国建设水平

智能发展规划》明确提出“加快人工智能深度应用”,再到2023年发布的《关于推进IPv6技术演进和应用创新发展的实施意见》及《质量强国建设纲要》,这些政策文件均强调了加速人工智能等新技术在各类场景中的应用与融合发展。当前政策导向的迭代升级,表明中国政府对人工智能领域的扶持战略已从突破单一技术维度,转向构建“技术—产业”双向赋能的产研融合范式,旨在通过智能技术的深度应用,为经济社会高质量发展注入新动能,实现全链条价值提升。

3 人工智能全球治理的挑战

当前,全球人工智能领域正经历着快速的技术进步和治理挑战。主要经济体在人工智能的基本概念、价值理念和路径策略上存在显著分歧,这导致了全球治理的区域化困境和治理道路的分散化与碎片化。同时,各国也在激烈争夺人工智能治理的话语权,使得全球治理整体上缺乏积极性与协同性。在这种背景下,人工智能的

快速发展和监管挑战并存,全球迫切需要在AI治理尤其是AI监管方面达成共识,防范可能出现的重大风险。

(1)技术挑战。人工智能安全治理议题正在从传统风险向深层次社会问题演进,包括隐私保护、数据安全、算法黑箱、模型漏洞等传统技术安全议题,以及应用于生物、化学、核能、关键基础设施等领域的底层安全问题。从政策制定方面,由于人工智能技术发展迅猛,相关法律法规的制定存在滞后性,因此人工智能治理要求有前瞻性和可持续性。当前,企业厂商对人工智能技术的宣传主要聚焦于其对生活品质的提升,如通过智能助手提高工作效率、通过智能家居系统改善居住环境等。这些宣传多强调技术的积极面,旨在吸引消费者和投资者,推动技术的广泛应用和市场拓展,甚至出现了过度渲染技术能力的问题,尤其在汽车智能驾驶领域表现显著。以小米SU7事故为例,安徽高速事故与湛江电动二轮车起火事件叠加,引发公众对小米汽车安全性的质疑。尽管后者起火源于外部碰撞,但舆论场仍将矛头指

向技术缺陷,反映出公众对专业术语的认知隔阂。技术宣传需在创新激励与风险警示间寻求平衡,避免将营销凌驾于安全之上。由此可见,技术具有两面性,人工智能在创造便捷效率的同时,也暗含隐患与风险,如数据泄露、算法偏见、就业替代等问题。

(2)协调机制挑战。人工智能全球治理面临多重矛盾。参与方理念差异导致治理碎片化,产生规则、决策和执行赤字。地缘政治加剧使大国协调困难,个别国家以技术封锁、制裁等手段阻碍AI供应链,这些行为既延缓了技术发展,也削弱了全球治理效能。面对这些挑战,全球治理体系需要更加积极地推动共治共享、构建监管合作规则、加快国际标准制定和创新多边协调机制,以提升人工智能治理的国际影响力与话语权。

(3)伦理法律挑战。人工智能治理涉及多个层面,包括但不限于数据安全、算法透明度、伦理规范、责任界定等。人工智能治理面临伦理规范缺失和法律法规严重滞后的问题,难以覆盖技术快速发展带来的新型社会关系和法律问题。社会大众对人工智能治理的重要性认识不够,缺乏广泛的社会共识。例如,数据安全是人工智能治理的关键领域之一,数据的不当收集、存储和使用可能引发个人隐私泄露等严重后果。然而,由于缺乏足够的宣传和教育,公众对这些问题的了解有限,这不仅影响了公众对人工智能技术的信任度,也阻碍了相关治理政策的有效推行和实施。以数据安全政策为例,政策的制定和实施需要社会各界的广泛支持和参与。当公众对人工智能治理缺乏基本认识时,他们可能无法理解数据安全政策的必要性和紧迫性,从而在政策推行过程中遇到阻力。例如,一些数据安全法规要求企业采取更严格的数据保护措施,这可能会增加企业的运营成本,如果没有公众的支持和理解,企业可能会对这些法规产生抵触情绪,影响政策的执行效果。

因此,为了推动人工智能的健康发展,构建一个安全、可靠、公平的人工智能生态系统,必须加强“人工智能治理”的宣传力度。这不仅需要政府和相关机构的积极引导,也需要企业、学术界和媒体等各方的共同努力。通过多渠道、多形式的宣传,提高公众对人工智能治理重要性的认识,促进社会各界形成共识,为相关政策的顺利实施创造良好的社会环境。

4 人工智能全球治理的趋势与方向

人工智能创新属于技术、资本、数据三要素驱动型领域,对算力支撑、算法研发及数据质量设有较高准入门槛。本文以Artificial intelligence for global governance和Artificial intelligence governance为关键词,以收录在Web of Science核心合集数据库的发表于2016—2024年的920篇文献为研究对象。研究表明:自2016年人工智能元年以来,相关学者持续关注人工智能全球治理研究,2020—2024年Web of Science核心合集数据库年收录文献量呈爆发式增长态势(图1)。Web of Science核心合集数据库中重要文献前十名如表2所示,均涉及人工智

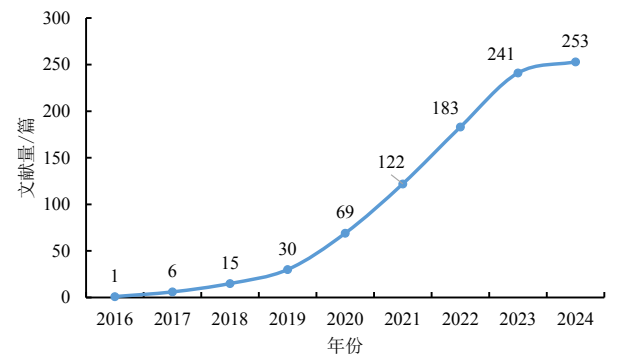


图1 人工智能全球治理领域相关文献收录的时间变化趋势
Fig.1 The Temporal Trend of Relevant References Included in the Field of Artificial Intelligence Global Governance

表2 Web of Science核心合集数据库重要文献前十名
Table 2 The Most Important Top10 References in the Web of Science Database

序号	题名	作者
1	The moral machine experiment	Awad等 ^[4]
2	AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations	Floridi等 ^[5]
3	Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation	Díaz-Rodríguez等 ^[6]
4	Principles alone cannot guarantee ethical AI	Mittelstadt ^[7]
5	Governing AI safety through independent audits	Falco等 ^[8]
6	Preserving the rule of law in the era of artificial intelligence (AI)	Greenstein ^[9]
7	A survey of blockchain applications in sustainable and smart cities	Makani等 ^[10]
8	Boundaries between research ethics and ethical research use in Artificial Intelligence health research	Samuel等 ^[11]
9	Toward responsible AI: An overview of federated learning for user-centered privacy-preserving computing	Yang ^[12]
10	Industry-4.0-enabled digital transformation: Prospects, instruments, challenges, and implications for business strategies	Yaqub等 ^[13]

能安全、道德、风险和挑战等问题,凸显了人工智能全球治理的重要性和紧迫性。

将检索到的920篇文献数据导入Citespace软件绘制共被引网络图(图2),图中每一个节点代表一篇文献,节点之间的连线代表两篇文献之间存在共引用关系,文献引用关系复杂,网络节点聚集,被引量较高,表明“人工智能全球治理”相关研究受到了越来越多的关注。

相关文献主要聚焦以下主题:

主题一:伦理框架与道德原则。相关学者主要探讨AI的伦理框架、道德原则以及公众对于AI道德决策的看法。它涉及AI技术的伦理维度,包括如何确保AI系统的行为符合社会公认的道德标准。例如,Awad等^[4]构建了名为“Moral Machine”的在线平台,用于模拟和评估公众对于自动驾驶车辆面临道德困境时的决策偏好,探讨了不同文化、经济和社会背景下道德偏好的多样性和复杂性,分析了这些道德偏好如何影响自主车辆技术的发展和部署,以及政策制定者应如何考虑这些因素来制定相关政策;Floridi等^[5]采用多学科交叉的研究方法构建了

一个关于AI伦理的框架,提出了五个核心伦理原则(受益性、非伤害性、自主性、公正性和可解释性),并围绕这些原则制定了20条具体建议,构成了一个全面的AI伦理框架,旨在指导AI技术的开发和应用;Díaz-Rodríguez等^[6]通过分析可信AI的四大轴(伦理原则、AI伦理、立法和技术要求),构建了从可信AI原则和要求到可操作协议的映射模型,研究了如何实现负责任的AI系统。Mittelstadt^[7]将人工智能发展与医学发展相对比,指出目前人工智能伦理发展的4重困境:缺乏共同目标和受托责任、缺乏专业历史和规范、缺乏将原则转化为实践的成熟方法和缺乏健全的法律和专业问责机制。

主题二:法律与政策治理。该主题下,相关学者关注AI的法律与政策治理,包括如何确保AI系统在法律框架内运行,以及如何通过政策来规范和引导AI的发展,涉及AI技术的法律维度,包括隐私保护、数据安全、责任归属等问题。例如,Falco等^[8]构建了独立审计框架(Independent Audit of AI Systems, IAAIS),并提出“AAA”治理原则(前瞻性评估、审计追踪、系统遵守管辖

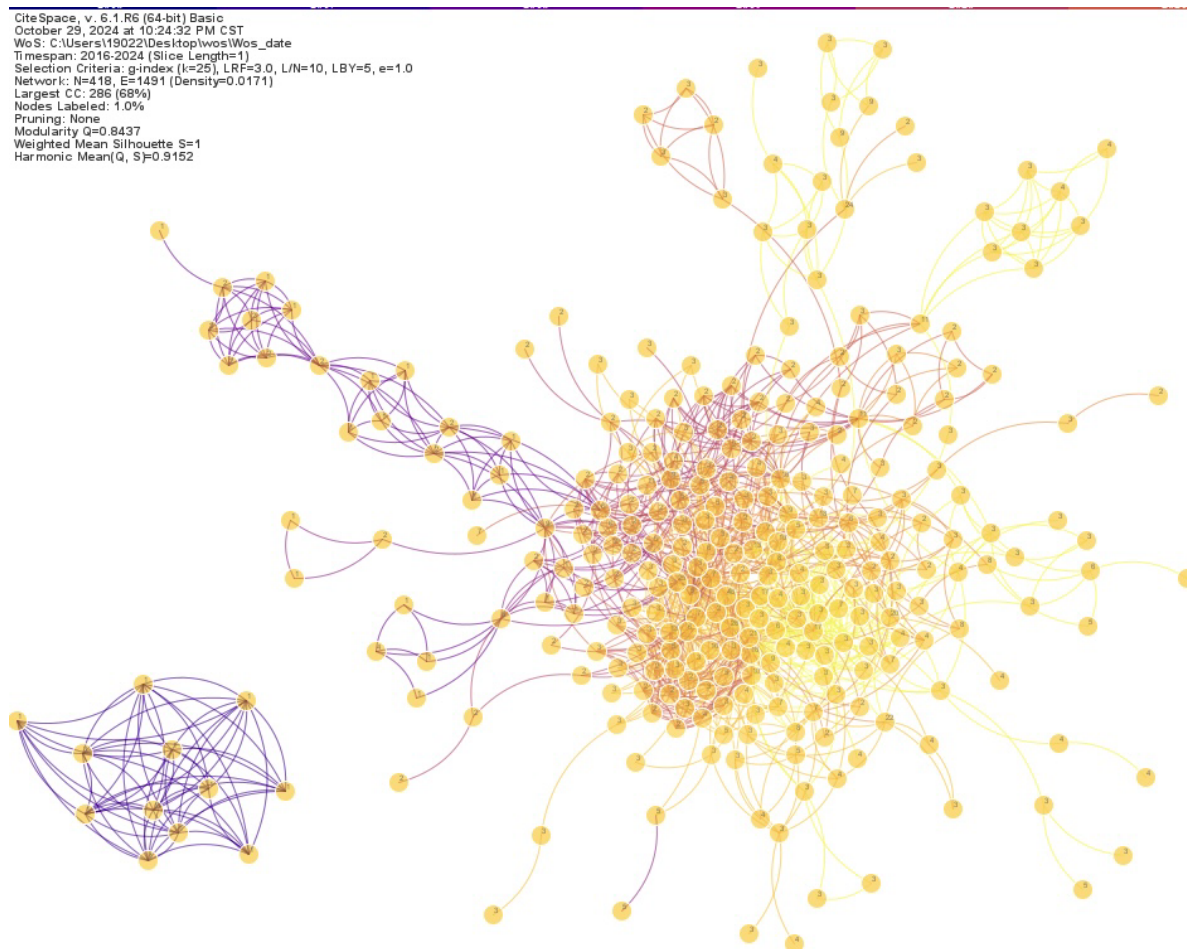


图2 人工智能全球治理相关文献共被引图(Web of Science核心合集数据库)
Fig.2 Related References on Jointly Cited Global Governance of Artificial Intelligence

要求),为AI治理提供可执行和可强制执行的基线;Greenstein^[9]探讨了如何在人工智能时代维护法治,包括法律的可访问性、歧视与偏见、权力滥用及决策质疑等方面,探讨了AI如何因其技术复杂性和不透明性等特点影响到法律的可访问性、可预测性和可理解性。

主题三:技术挑战与实践应用。相关学者主要关注AI技术在特定领域的应用中遇到的挑战和伦理问题,以及如何通过技术创新来解决这些问题。例如,Makani等^[10]搜集和分析了已有的区块链应用案例,采用定性和定量相结合的方式评估了区块链技术的潜在影响和价值,提出了可以通过提高透明度、降低交易成本、增强数据安全和隐私保护等方式为城市的可持续发展和智慧化转型提供有力支持;Samuel等^[11]通过深入访谈,分析了研究者和相关决策者在AI健康研究领域对研究伦理的理解与实践,研究发现受访者在研究伦理与研究成果的社会影响伦理之间存在明确区分,强调了修改伦理生态系统以加强对研究使用伦理监督的必要性;Yang^[12]采用了理论分析与实际应用案例相结合的方式,构建了联邦学习作为隐私保护计算的核心模型,该模型允许不同方在训练模型和进行推理时,传递加密的模型参数而非原始数据,应用案例涵盖了自然语言处理、医疗影像分析、电子商务等多个领域,展示了联邦学习在解决数据安全和隐私保护问题方面的广泛应用前景。Yaqub等^[13]构建了一个涵盖数字转型前因、发展及后果的模型,研究了Industry 4.0(I4.0)驱动的数字转型对现代企业和经济的影响,并描述了关键前景、工具和挑战。

未来人工智能治理将围绕四大方向深化。(一)构建隐私保护与数据安全的全生命周期管理体系,通过联邦学习等技术实现“数据可用不可见”;(二)推动算法透明度标准化与伦理治理框架落地,如欧盟对高风险系统的可追溯性强制要求;(三)探索“以AI治理AI”的动态模式,开发自动化合规工具与分布式协作机制;(四)加速全球价值对齐框架建设,通过《全球人工智能治理倡议》平衡创新监管与跨国协同。此外,人工智能安全治理议题正在从传统的技术层面风险扩展到更深的社会层面,通过技术迭代、制度设计与全球协作实现经济效益与社会福祉的兼容,未来需持续探索超级智能的失控防范与主权安全平衡,推动治理范式从“人工干预”向“智能协同”跃迁^[14]。

5 中国的治理实践与贡献

在全球人工智能技术快速发展的背景下,中国凭借政策创新、技术开放与国际合作,逐步成为全球人工智

能治理的重要引领者。中国人工智能伦理治理处于技术快速发展、经济社会高质量发展、国际秩序深度调整“三期交叠”的特殊时期^[15]。从推动技术民主化到构建多边治理框架,中国的实践不仅塑造了本土产业的繁荣,更通过“中国方案”为全球治理注入新动能。中国在人工智能治理领域的核心贡献之一是率先建立了一套兼顾创新与安全的政策体系。2023年发布的《全球人工智能治理倡议》系统提出了“以人为本”“智能向善”等原则,强调技术发展需服务于人类福祉,而非加剧霸权垄断。在此基础上,中国颁布了全球首部针对生成式人工智能的法规《生成式人工智能服务管理暂行办法》,明确数据安全与隐私保护要求,为国际社会提供了可借鉴的监管模式。2025年发布的《人工智能生成合成内容标识办法》进一步细化技术标准,通过“标识+合规”机制破解虚假信息难题,推动生成式AI的可持续发展。这些举措不仅体现了中国对技术风险的敏锐预判,更展现了通过法治化路径实现“善治”的治理智慧。中国通过开源技术推动人工智能领域的“技术公平”。以初创企业DeepSeek为例,其开源模型迅速成为全球开发者共享的创新平台,助力发展中国家以低成本开发本土AI应用,弥补技术差距。这种模式挑战了传统封闭生态,构建了包容性创新网络——全球开发者、学者均可贡献代码,形成“共研共享”的协作生态。中国的人工智能治理实践表明,技术创新与全球责任并非对立,应当始终坚持发展与安全原则,积极参与多边框架下的治理议程,争取更为有利的国际制度环境^[16]。通过开源共享、多边合作与伦理先行,中国为全球治理提供了“发展与安全平衡”“技术民主化”“包容性增长”三大范式。

工业和信息化部发布的数据显示,截至2023年第二季度,我国人工智能核心产业规模已突破5 000亿元,企业数量超过4 400家,仅次于美国,全球排名第二,广泛应用于包括智慧城市、智能制造、智慧农业和AI for science在内的20个细分领域。由互联网周刊携手eNet研究院、德本咨询联合发布的“2024年人工智能企业百强排行榜”(表3),位居榜单前列的企业,如百度、阿里巴巴和

表3 2024年中国人工智能企业百强榜前十名企业
Table 3 Top 10 Chinese Artificial Intelligence Enterprises in 2024

排名	品牌	突破性	排名	品牌	突破性
1	华为	92.58	6	寒武纪	90.21
2	百度	92.36	7	商汤科技	89.65
3	阿里巴巴	91.87	8	旷视科技	89.12
4	腾讯	91.54	9	云从科技	88.76
5	科大讯飞	90.78	10	依图科技	88.34

为等,在人工智能产业的基础层、技术层和应用层均有全面布局。核心技术取得了重大突破,为人工智能的多领域应用创造可能,如与物联网、大数据、云计算等技术的结合,推动了智能物联网、智能金融等领域的快速发展。随着人工智能技术的不断进步,其应用场景也在不断拓展,从最初的智能制造、智慧医疗等领域,逐渐扩展到金融科技、智慧城市、智慧教育等多个领域。这些应用场景的拓展,不仅推动了人工智能技术的普及和应用,也促进了相关产业的升级和国内人工智能企业积极参与国际竞争与合作,与国际知名企业、科研机构等开展合作与交流,推动人工智能技术创新和应用,提升了企业国际竞争力,也有助于推动全球人工智能产业的共同发展。可见,国内的人工智能产业呈现出综合性企业竞争优势显著、技术创新持续推动产业发展、应用场景不断拓展、政策支持持续加强以及国际合作与交流日益频繁等趋势。这些趋势将共同推动国内人工智能产业的快速发展和升级。

人工智能作为信息科学中研究最活跃、发展最迅速、影响最广泛的领域之一,国家自然科学基金委员会(以下简称“自然科学基金委”)在把握国家国际重大需求前沿问题方面具有第一手的资料。自然科学基金委立足科学前沿,优化学科布局,面向国家社会经济发展中的重大应用需求,于2018年设立了人工智能一级学科代码F06,将人工智能学科作为自然科学基金委的重要资助领域。这一举措不仅体现了自然科学基金委对人工智能领域重要性的认识,也展示了其在推动国家科技进步和应对国际竞争中的前瞻性布局。

2023年度,人工智能领域(申请代码F06)下的“人才”项目系列和“研究”项目申请数逐年增加。例如,国家自然科学基金青年科学基金项目(C类)的申请数从2019年的960项增加到2023年的1 484项,资助率在23%左右。2020—2023年人工智能领域的面上项目申请量相对稳定,资助率起伏在1.4%以内。2024年放开面上申请限制后,面上项目申请量较2023年增加了49.85%。此外,近五年面上项目各二级代码的申请占比情况中,F0610人工智能与其他学科交叉的申请占比均保持第一。地区科学基金项目,2024年全国15个省(自治区、直辖市)共申报项目482项,经两轮专家评审最终立项63项,立项率13.07%。其中,广西壮族自治区、江西省、甘肃省的申请规模与申报单位数量均位居前三,展现出显著的区域集中度。

2024年,自然科学基金委发布了“2024年度国家自然科学基金委员会宏观调控专项项目指南——数据市场制度设计与关键技术”。该专项立足于我国一体化数据

市场的培育及其创新实践,各课题紧密结合医疗健康数据、产业数据、金融数据、交通数据、碳交易数据等典型数据流通行业的实践。同年,信息科学部公布3大重点项目群及118个立项领域,其中人工智能专项聚焦“数据要素流通基础理论与关键技术”研究方向,同步推出8个前沿指南方向。该领域项目覆盖F0601、F0603、F0604、F0606、F0607、F0609和F0610共7个二级代码,经评审从35项申请中择优资助10项。获批项目横跨知识库问答、城市治理、数据治理等多元场景,彰显国家自然科学基金在人工智能全球治理领域的研究深度与战略前瞻性。

目前,我国人工智能开源生态完整,基础大模型不断赋能工业领域,AI+行业应用蓬勃发展,AI标准化工作逐步开展,但在计算框架、数据集建设、基础大模型方面存在差距,开源服务平台建设亟需探索新路径。当前挑战和机遇并存的现状,进一步凸显了自然科学基金委在人工智能领域的资助和支持的重要性,不仅有助于推动技术创新,还能在国际竞争中提升我国的科技竞争力。

6 意见与建议

在人工智能领域,中国正积极塑造其全球角色,通过多维度的策略来确保技术的安全、创新和伦理发展。

(1)法律筑基,立法先行。构建覆盖法律框架、伦理审查、人才培养的治理体系。推进《人工智能促进法》专项立法,将数据加密、算法备案纳入强制规范,由科技部伦理委员会牵头制定医疗、金融等领域伦理指南;在高校开设“人工智能治理”交叉学科,培养兼具技术认知与伦理判断的复合型人才,同步通过工业和信息化部AI安全演练平台强化政企协同治理能力,实现“规则制定—人才培养—实践验证”闭环。

(2)创新驱动,安全护航。建立技术发展与风险防控的动态平衡机制。依据应用场景风险分级实施差异化管理,对自动驾驶等高危领域推行算法透明化与实时监测;整合自然科学基金委资助的216项网络安全研究成果,构建AI安全防御技术库,设立专项基金支持可解释算法研发;在长三角设立治理沙盒试验区,允许企业在监管下测试伦理决策系统,形成“技术突破—安全验证—场景落地”创新链。

(3)全球协作,规则引领。与国际社会合作制定统一标准,加大基础和应用研究投入,完善数据安全与隐私保护法规,重视伦理和社会责任,推动人工智能教育体系建设,重点关注应用场景的治理,确保技术安全可靠发展。通过标准输出与治理援助提升国际话语权,依托亚洲基础设施投资银行设立国际发展基金,推动“一带一路”国家采用中国数据跨境标准;在联合国发起

《AI军控倡议》,联合国禁止生成式AI武器化;将浙江“城市大脑”、深圳算法备案制等本土经验转化为治理工具包,通过世界互联网大会向全球推广;启动“南方国家AI治理援助计划”,为东南亚、非洲提供监管系统开发支持,构建技术援助与规则塑造双轮驱动的外交新范式。

参 考 文 献

- [1] 让互联网更好造福世界各国人民——世界互联网大会推动构建网络空间命运共同体迈向新阶段理论研讨会发言摘编. 人民日报, 2024-09-11(014).
- [2] 高奇琦. GPT技术的全球治理方案:一个技术多边主义的框架. 当代世界, 2023(5):10—17.
Gao QQ. Global governance solutions for GPT technology: A framework of technological multilateralism. Contemporary World, 2023(5): 10—17. (in Chinese)
- [3] 国务院. 国务院关于印发新一代人工智能发展规划的通知. (2017-07-20)/[2025-06-13]. https://www.gov.cn/zhengce/zhengceku/2017-07/20/content_5211996.htm.
- [4] Awad E, Dsouza S, Kim R, et al. The moral machine experiment. Nature, 2018, 563(7729): 59—64.
- [5] Floridi L, Cowls J, Beltrametti M, et al. AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 2018, 28(4): 689—707.
- [6] Díaz-Rodríguez N, Del Ser J, Coeckelbergh M, et al. Connecting the dots in trustworthy artificial intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 2023, 99: 101896.
- [7] Mittelstadt B. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 2019, 1(11): 501—507.
- [8] Falco G, Shneiderman B, Badger J, et al. Governing AI safety through independent audits. Nature Machine Intelligence, 2021, 3(7): 566—571.
- [9] Greenstein S. Preserving the rule of law in the era of artificial intelligence (AI). Artificial Intelligence and Law, 2022, 30(3): 291—323.
- [10] Makani S, Pittala R, Alsayed E, et al. A survey of blockchain applications in sustainable and smart cities. Cluster Computing, 2022, 25(6): 3915—3936.
- [11] Samuel G, Chubb J, Derrick G. Boundaries between research ethics and ethical research use in artificial intelligence health research. Journal of Empirical Research on Human Research Ethics, 2021, 16(3): 325—337.
- [12] Yang Q. Toward responsible AI: An overview of federated learning for user-centered privacy-preserving computing. ACM Transactions on Interactive Intelligent Systems, 2021, 11(3/4): 1—22.
- [13] Yaqub MZ, Alsabbab A. Industry-4.0-enabled digital transformation: Prospects, instruments, challenges, and implications for business strategies. Sustainability, 2023, 15(11): 8553.
- [14] Roberts H, Hine E, Taddeo M, et al. Global AI governance: Barriers and pathways forward Open Access. International Affairs, 2024, 100(3): 1275—1286.
- [15] 朱依娜, 卢阳旭. 中国人工智能伦理治理实践与挑战——基于三期交叠视角. 中国科技论坛, 2025(1): 148—153, 178.
Zhu YN, Lu YX. Practice and challenges of ethical governance of AI in China: A new perspective. Forum on Science and Technology in China, 2025(1): 148—153, 178. (in Chinese)
- [16] 俎文天. 人工智能全球治理合作: 问题、进路与中国参与. 国际经济评论, 2025(4): 153—176, 8.
Zu WT. Global governance cooperation on artificial intelligence: Issues, approaches and China's participation. International Economic Review, 2025(4): 153—176, 8. (in Chinese)

Analysis of the Global Governance of Artificial Intelligence

Huilin Ge^{1*} Guo Xie² Le Wang³ Zhiheng Wang² Guozheng Wu²

1. Academy of Science and Technology, Jiangsu University of Science and Technology, Zhenjiang 212100, China

2. Department of Information Sciences, National Natural Science Foundation of China, Beijing 100085, China

3. School of Artificial Intelligence, Xi'an Jiaotong University, Xi'an 710049, China

Abstract With the rapid development of Artificial Intelligence (AI) technology, especially the emergence of generative AI products such as ChatGPT and Sora, AI has been deeply integrated into people's daily life and production work, with far-reaching impact on society. However, this kind of "rapid" development has also brought about security risks and challenges, such as data security, algorithmic bias, ethical misconduct and other systemic risks are gradually emerging. In this context, there is an urgent need to explore the issue of global governance of AI to ensure that technological development is consistent with global public interests. This paper systematically analyses the challenges of global governance of artificial intelligence from three perspectives: technology, coordination mechanisms, and ethics and law. These perspectives are based on the current status of global governance of artificial intelligence. The paper employs bibliometric methods to explore the trends and directions of global governance of artificial intelligence. A thorough analysis of the prevailing trends and directions in global governance of artificial intelligence, drawing upon a comprehensive sample of 920 documents from the Web of Science database, reveals that contemporary research is predominantly concentrated on three primary domains: ethical frameworks and moral principles, legal and policy governance, and technical challenges and practical applications. Future research could explore directions such as the establishment of a comprehensive lifecycle management system for privacy protection and data security, the promotion of the standardisation of algorithm transparency and the implementation of ethical governance frameworks, the exploration of dynamic models of "AI governing AI", and the acceleration of the construction of a global value alignment framework. Finally, an analysis of China's current circumstances has been undertaken, resulting in the formulation of a three-pronged governance approach encompassing legislative norms, technological coordination, and global cooperation. This approach is designed to deliver pragmatic solutions for the establishment of global rules.

Keywords artificial intelligence; global governance; frontier trends; China governance practices

葛慧林 博士,江苏科技大学副研究员,主要研究方向为人工智能、科技政策与科研管理。主持国家自然科学基金项目2项。

(责任编辑 王磊 张强)

* Corresponding Author, Email: ghl1989@just.edu.cn