

引用格式:王恺乐,陈云伟,熊永兰. 人工智能监管政策与措施的国际比较及启示[J]. 世界科技研究与发展, 2024(4): 456–468.

人工智能监管政策与措施的国际比较及启示*

王恺乐¹ 陈云伟^{* 1,2} 熊永兰¹

(1. 中国科学院成都文献情报中心科技创新评价研究中心, 成都 610299;
2. 中国科学院大学经济与管理学院信息资源管理系, 北京 100190)

摘要: 随着技术的飞速发展, 人工智能正在成为产业创新的重要抓手和新质生产力的重要驱动力, 然而技术进步的同时也带来了伦理问题及安全问题等挑战。人工智能技术的安全和可控是科技安全的重要组成部分, 有效治理与监管人工智能已经成为全球共识, 而欧盟、美国和英国在人工智能监管方面已走在国际前列。本文从政策角度出发, 系统梳理欧盟、美国和英国的人工智能监管措施发展脉络, 并比较其在监管机构设置、监管范围及监管路径上的特色做法。研究发现, 欧盟、美国和英国的人工智能监管模式各具特点, 但均存在监管共性与共同趋势。论文基于此提出了我国在人工智能监管方面的对策建议, 以期为制定人工智能监管策略提供决策参考。

关键词: 人工智能; 政策; 科技安全; 监管

DOI: 10.16507/j.issn.1006-6055.2024.07.007

International Comparison and Inspiration of Artificial Intelligence Regulatory Policies and Measures*

WANG Kaile¹ CHEN Yunwei^{* 1,2} XIONG Yonglan¹

(1. Sci-tech innovation Evaluation Research Center, National Science Library (Chengdu), Chinese Academy of Sciences, Chengdu 610299, China; 2. Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: With the rapid development of technology, artificial intelligence is becoming an important gripper for industrial innovation and a driving force for new quality productivity. However, technological progress also brings challenges, such as ethical and security issues. The security and controllability of artificial intelligence technology are important components of technological security, effective governance and regulation of artificial intelligence have become a global consensus, and the European Union, the United States, and the United Kingdom have taken the lead in artificial intelligence regulation internationally. This article systematically reviews the development of artificial intelligence regulatory measures in three

* 四川省科技计划项目“天府实验室运行机制创新策略研究”(2022JDR0360), 中国科学院成都文献情报中心创新基金青年项目“科技人才政策绩效评价模型研究——以我国五大科技创新中心为例”(E3Z0000306)

** E-mail: chenyw@clas.ac.cn

regions from a policy perspective and compares their distinctive practices in regulatory agency settings, scope, and paths. Research has found that the European Union, the United States, and the United Kingdom have their own unique regulatory models, but there are also regulatory commonalities and common trends. Based on this, countermeasures and suggestions for artificial intelligence regulation are proposed for China to provide a decision-making reference for formulating artificial intelligence regulatory strategies.

Keywords: Artificial Intelligence; Policy; Technology Security; Regulation

作为工业社会向智能社会迈进的创新“催化剂”,人工智能(Artificial Intelligence, AI)对人类社产生了深远影响^[1]。然而随着人工智能技术的不断迭代和应用场景的不断泛化,特别是以ChatGPT和Sora等为代表的生成式人工智能的快速发展,数据隐私、算法歧视、安全伦理、知识产权等新挑战加速衍生^[2]。人工智能开始展现出影响个人行为、社会资源分配和政府决策的能力属性,产生权利异化风险^[3],成为科技安全体系中的重要一环。因此,人工智能监管开始成为世界各国重点关注的问题。Mokyr等也从历史的角度强调了对技术管理的重要性,他认为工业革命爆发后经济的持续增长不仅归功于具体的发明创造,还得益于人类对这些发明的理解和管理^[4]。

关于人工智能的监管问题,国内外相关研究主要集中在四个方面。一是针对人工智能应用监管的探讨,主要集中在医学^[5-9]和金融领域^[10-12]。二是针对人工智能监管中的具体问题进行探讨,主要集中在伦理治理领域^[13-16]。三是针对某一国家的人工智能监管举措或监管政策的内容进行分析。如周辉等^[17]对英国人工智能监管实践进行了梳理,董杨等^[18]对美国推动各主体在治理中更负责地使用人工智能系统的做法进行了梳理,雷挺等^[19]分析了欧盟的人工智能监管经验,张广伟^[20]和黄静怡^[21]对欧盟《人工智能法案》进行了解析。四是对生成式人工智能的风险与治理的分析^[22,23]。

已有研究丰富了对人工智能监管的探索,但

少有从国际视角出发对人工智能监管政策与举措的系统性梳理和比较。

人工智能监管的国际比较有助于我国完善科技安全制度,提高对人工智能治理的认识并增强对人工智能监管的能力,避免陷入孤立监管的处境。欧盟、美国和英国对人工智能监管认识较早,已形成可参考可借鉴的经验。因此本文在详细分析欧盟、美国和英国人工智能监管举措演变态势的基础上,比较其在监管机构、监管范围、监管路径等方面的做法,总结其共性特征,为我国制定人工智能监管措施提供参考。

1 人工智能监管政策措施的国际实践

1.1 欧盟:数据监管与法规制定两手抓

1.1.1 前期探索阶段(2017年以前)

欧盟自2016年起就不断探索对人工智能技术应用的监管体系构建^[24],其以欧洲整体层面为视角,致力于构建协调一致、适合欧洲发展的治理体系。如欧盟法律事务委员会出台的《欧盟机器人民事法律规则》就是确保基于人工智能控制的机器人符合法律和伦理标准。在此阶段,尽管欧盟在持续关注人工智能领域发展的潜在风险,但尚未出台专门针对人工智能监管的指导性文件,主要依据2002年制定的反垄断法和2005年制定的反不正当竞争法进行执法。

1.1.2 逐步推进阶段(2018—2020年)

在这个阶段,为了应对人工智能技术的快速发展及其对社会、经济和伦理的复杂影响,欧盟开

始转向更加具体和系统的监管政策,并制定了一系列涵盖协调计划、伦理指南和白皮书等战略规划,为人工智能的可信任发展和监管奠定了基础,并提出了较为具体的监管框架和建议,以确保人工智能的发展与应用符合欧盟的价值观和法律框架。

2018年4月,欧盟委员会公布了《欧洲人工智能战略》,这份文档标志着欧盟首次对人工智能领域进行系统规划,并被认为是欧盟在人工智能政策监管历史上的重要标志,文件提出要在促进创新与保护公民权利和安全之间取得平衡。2018年5月,《通用数据保护条例》(General Data Protection Regulation, GDPR)正式生效,成为该时期重要的监管举措之一。尽管GDPR不是专门针对人工智能的,但它对人工智能应用有着重要影响,特别是关于个人数据的收集、处理和使用方面^[25],GDPR强调了数据主体的权利,包括对个人数据如何被人工智能系统使用有更清晰的了解。2018年6月,欧盟组建了由52位分别来自学术、产业和民间组织的独立专家构成的人工智能高级顾问团队(High-Level Expert Group on Artificial Intelligence, HLEG),加快建立人工智能监管框架的步伐。2018年12月,为协调各成员国落实《欧洲人工智能战略》,欧盟委员会发布了第一版《人工智能协调计划》,并提出了“设计伦理”和“设计安全”两大原则^[26]。这一计划象征着欧盟在人工智能领域开启了统一和协调的监管进程。到了2019年4月,为了推动人工智能的可靠发展,欧盟委员会发布了由HLEG撰写的《可信人工智能伦理指南》,该指南为欧盟人工智能监管体系建构指明了方向。2020年2月,欧盟委员会公布《欧洲数据战略》和《人工智能白皮书》,提出了针对人工智能的监管架构,建议将不同风险级别的人工

智能应用分类进行监管。

1.1.3 全面规范阶段(2021年以后)

2021年4月,欧盟《人工智能法案(提案)》正式发布,该提案试图建立一个全面的法律框架,以应对人工智能技术带来的各种挑战和机遇,从而促进人工智能的可持续发展和创新。作为对欧盟先前发布的《欧洲数据战略》的响应,欧洲议会在2022年通过了《数据治理法》,旨在加强欧洲范围内的数据共享机制,并促进数据的可用性。可见欧盟一方面从底层逻辑出发,加强对数据使用、流通和共享的监管,另一方面加快立法,开启了对人工智能的全面规范。

在以ChatGPT为代表的人工智能大模型快速发展的背景下,欧盟加快了立法的步伐,在经过多轮讨论与修改后,2024年7月,欧盟发布了全球首个《人工智能法案》(EU AI Act),这也是迄今为止全球范围内最为全面的、针对人工智能监管的法案^[27],进一步巩固了欧盟在全球科技监管上的领导地位。

1.2 美国:力求在监管与非监管中取得平衡

1.2.1 基础奠定阶段(2018年以前)

2018年以前,美国针对人工智能领域出台的政策中,涉及监管方面较少;在奥巴马政府领导下,美国首次开始在国家层面讨论人工智能政策。2016年美国颁布了《国家人工智能研究发展战略规划》^[28]以及《为人工智能的未来做好准备》^[29]两份文件,详细描述了联邦政府需关注的主要政策领域,涵盖了人工智能在公共福祉、研究及劳动力市场、公平性、安全性及治理等多个方面的运用。

1.2.2 扩大布局阶段(2019—2020年)

2019—2020年美国开始加速在人工智能领域的布局。2019年2月,特朗普政府通过的

《关于维持美国人工智能领导地位的行政命令》(EO13859)为美国人工智能发展奠定了总基调;随后发布《美国人工智能倡议》及《国家人工智能研究发展战略计划》2019更新版,以维持其在人工智能技术上的领先地位。

在这一阶段美国开始出台针对人工智能伦理、标准及监管的政策。如应EO13859号行政命令的要求,美国商务部下属的国家标准与技术研究院于2019年8月宣布优先考虑联邦机构参与人工智能标准的制定。2019年11月,美国国防部下属的国防创新委员会发布《人工智能伦理道德标准》最终版本,其目的是确保美军在利用和研发人工智能技术时能够遵循明确的限制和指导原则,以最大程度地减少潜在安全风险和不良影响。2020年1月美国发布首个人工智能监管指南——《人工智能应用监管指南》,旨在提供针对新出现的人工智能问题的监管和非监管措施指引;该指南强调以减少应用障碍和促进技术创新为宗旨,体现了美国政府在监管和非监管措施中的平衡考量,既要确保技术的安全和合规性,又要不阻碍技术的发展和 innovation。2020年2月,美国发布《人工智能倡议首年年度报告》,标志着美国对于人工智能发展的审视和总结,同时明确了未来的发展目标和愿景;该报告提出了在人工智能领域的战略方向和重点任务,包括加强研究投入、推动创新应用、确保技术安全等。2020年12月,特朗普政府签署的《关于在联邦政府中促进使用可信赖人工智能的行政命令》提出了人工智能设计和开发的九项原则,确保公民隐私、权利和自由^[30]。

1.2.3 加速变革阶段(2021年以后)

2021年后,美国开始加速对人工智能的监管,多个部门和组织加强了对各自所涉领域的人工智能技术的监管。2021年1月,根据《2020年国

家人工智能倡议法案》和《2021财年国防授权法案》,美国设立了国家人工智能倡议办公室,该机构担当着监管和推行国家人工智能策略的角色。2021年6月,隶属于美国国会的政府问责办公室(Government Accountability Office, GAO)发布了题为《人工智能:联邦机构及其他实体的责任框架》的报告,探讨了人工智能管理、数据、性能及监控方面的核心议题和责任流程^[31]。2021年9月,美国与欧盟共同建立了欧盟-美国贸易与技术理事会(EU-US Trade and Technology Council, TTC),该理事会设立了专门的工作小组,以应对技术滥用对安全和人权带来的威胁,其目的在于围绕人工智能、网络安全等建立“跨大西洋标准”。2022年6月,联邦贸易委员会(Federal Trade Commission, FTC)首先明确表示认为人工智能歧视、欺诈和相关数据滥用问题属于其职权范围^[33]。2022年10月底,美国平等就业机会委员会宣布启动了一项针对招聘和工作场所中人工智能应用的监管和保护的倡议^[34]。2022年10月《人工智能权利法案蓝图》发布,强调公平性和隐私权,并要求相关实体在开发人工智能系统时,必须维护公正与正义,确保使用者不受算法偏见和数据隐私侵犯的影响。

2023年1月,美国国家标准与技术研究院推出《人工智能风险管理框架》的首个版本,旨在识别与人工智能系统部署相关的各类风险。2023年6月,美国总统拜登在旧金山会见人工智能专家和研究人员,讨论管理人工智能技术风险。同时,民主党和共和党议员提交《国家人工智能委员会法案》,并提出配套立法,以加强人工智能监管。2023年10月末,美国白宫公布了由拜登总统签署的《关于安全、可靠和可信的人工智能行政令》,该行政令是针对人工智能技术的全面监管

原则,旨在确保美国在人工智能技术开发方面处于领先地位,并保护公民权利、市场竞争和国家安全不受人工智能技术的负面影响。2024年3月,美国白宫管理和预算办公室(Office of Management and Budget, OMB)发布了首个政府范围内的人工智能政策备忘录,旨在减轻人工智能的风险并利用其优势,包括指定首席人工智能官、提高透明度等措施。

2024年斯坦福大学的人本人工智能研究所(Stanford HAI)公布的《2024年人工智能指数报告》揭示,2023年美国机构研制了61个知名的人工智能模型,这一数字超过了欧盟的21个和中国的15个。此外,美国在人工智能领域的私人资本投入总额达到了672亿美元,几乎是中国相应投资的九倍之多,可见美国在对人工智能的监管上仍以促进发展为核心。

1.3 英国:争取掌握人工智能监管话语权

1.3.1 初步引入阶段(2018年以前)

2018年以前,英国对人工智能领域的监管尚未出台框架性政策或指导意见,但初步引入了加强人工智能监管的立场,监管的方向主要在数据的使用和保护方面。2014年英国信息专员办公室(Information Commissioner's Office, ICO)发布了一份讨论性质文件——《大数据、人工智能、机器学习和数据保护》^[35],探讨了在数字化时代监管新技术所面临的挑战和应对方法;同年,英国下议院科学技术委员会提出了关于社交媒体背景下大数据隐私问题的重要报告《负责任地使用数据》^[13],并强调了在维护个人隐私的同时实现技术创新的重要性。2016年10月,英国下议院科学技术委员会公布了关于人工智能及机器人技术的调研报告,提出英国应当在全球范围内引领人工智能和机器人技术伦理标准的研究,并将领导地位扩展至人

工智能监管领域。2017年,英国颁布了《产业战略:打造面向未来的英国》,强调在数据与人工智能安全领域维持领先地位的重要性。2018年4月,英国政府对外公布了由上议院人工智能特别委员会委托撰写的《英国的人工智能:准备、意愿与能力》报告,这标志着英国首次从国家层面就人工智能伦理治理问题提出见解。报告指出,对人工智能采取一概而论的监管方法并不适宜,而应当依赖于现有的行业特定监管机构来评估并管理人工智能的冲击。2018年8月,英国推出《数据伦理框架》并设立了数据伦理与创新中心(Centre for Data Ethics and Innovation, CDEI),其目的是在伦理道德的指导下探索新兴技术的最大可能性,并向政府提供专业建议。

1.3.2 战略规划阶段(2019—2022年)

在该阶段,英国开始重点加强对人工智能的审计以及具体领域的技术监管。2020年,ICO发布了关于人工智能审计的指南《人工智能审计框架指南:征求意见稿》^[36],该指南是专门为人工智能系统进行量身定制式监管设计的重要开端。2021年,为确保人工智能技术发展与伦理监管并行,英国在数据伦理和多个领域的监管上出台了一系列举措,如卫生和社会保障部(Department of Health and Social Care, DHSC)推出了《数据驱动的健康和护理技术行为准则》,目的是解决国家医疗服务和健康护理系统中由数据驱动技术带来的伦理挑战;英国安全行业协会(British Security Industry Association, BSIA)颁布了《自动面部识别:道德和法律使用指南》,为安全领域在使用自动面部识别技术时提供道德和法律方面的指引^[37];政府通信总部(Government Communications Headquarters, GCHQ)公布了《开创新的国家安全——人工智能伦理》报告,承诺在应用人工智能时坚持

负责任的原则;英国统计局(Office for National Statistics, ONS)设立应用数据伦理中心,并推出了《统计数据伦理政策》^[38],以促进数据伦理在统计实践中的应用;英国数字、文化、媒体和体育部(Department for Digital, Culture, Media and Sport, DCMS)发布的《国家数据保护法改革咨询方案》涵盖了数据管理、问责、传输、访问规则等重要领域。2021年9月22日,英国政府发布《国家人工智能战略》^[39],并将对人工智能进行有效监管视为达成其宏伟目标的关键支撑之一。2022年1月12日,英国宣告试点成立人工智能标准中心,旨在增强英国在国际技术标准制定中的领导作用。

1.3.3 政策执行阶段(2022年至今)

英国为实现人工智能监管全球领先的目标,近年来出台了多份监管文件。2022年7月,英国数字、文化、媒体和体育部发布《建立一种支持创新的人工智能监管方法》^[40],通过非固定性的方法定义人工智能,为英国政府未来重新界定人工智能留下了空间。2023年3月,英国推出《促进创新的人工智能监管方法》白皮书^[41],提出了基于原则性的人工智能治理策略,目的是促进创新的同时,为公众和企业带来信心。2023年9月,英国反垄断监管机构竞争与市场管理局(Competition and Markets Authority, CMA)提出了管理人工智能模式的新标准^[42]。2023年11月,首届全球人工智能安全峰会在英国召开,会上发布了第一份针对人工智能这一快速新兴技术的国际性声明《布莱切利宣言》,突显出英国在人工智能监管上试图制定全球策略的决心。

2 人工智能监管的国际比较

欧盟、美国和英国作为走在国际人工智能监管前列的国家/组织,其监管的方式、手段有所不

同,本部分对比三个国家/组织在监管机构设置、监管范围以及监管路径上的区分(表1),明晰国际主流人工智能监管策略。

表1 人工智能监管的国际比较

Tab.1 International Comparison of Artificial Intelligence Regulation

比较维度	欧盟	美国	英国
监管机构	已建立人工智能高级专家组和ChatGPT治理特别小组;持续谋划推进综合性监管机构。	已成立系列职能机构,强化统筹协调;提议成立专门监管机构。	由不同领域和层级的部门共同承担监管职责。
监管范围	全生命周期监管,涵盖事前、事中和事后。	针对具体应用场景监管,重点针对事后。	针对具体应用场景监管,重点针对事后。
监管路径	双重监管体系。	立法、行政与司法在各自职权范围内监管。	依托现有机构,去中心化监管。

2.1 监管机构设置:集中和分散

欧盟已建立了人工智能高级专家组和ChatGPT治理特别小组,并一直在谋划推进综合性监管机构的设立。如欧盟委员会于2021年发布的《人工智能条例》预设了一个专门的国家监管机构来确保法律的适用和实施。该条例规定,对于欧盟机构和组织开发的人工智能系统,由欧洲数据保护监督员主管负责。2023年欧盟委员提出将设立新的欧洲人工智能办公室负责成员国间的协调,并监督GPAI(类似ChatGPT的通用AI系统)模型规则的执行情况。

美国尚未确定人工智能监管的主管机构,但已陆续成立系列职能机构以强化对人工智能战略实施的统筹协调,如白宫科技政策办公室牵头成立了国家人工智能行动办公室,商务部人工智能行动办公室成立了人工智能咨询委员会,将人工智能特别委员会改组成人工智能机构间委员会。此外美国也正在积极准备部署全国层面的监管主

管机构。2023年6月,美国众议院提交的《国家人工智能委员会法案》建议成立一个“国家人工智能委员会”,目的是通过监管措施来降低人工智能带来的风险和潜在损害。同时,参议院提交的一项配套法案要求国会和白宫任命来自政府、行业、民间社会和计算机科学领域的20人成立委员会,制定人工智能监管战略。在2023年11月召开的首届全球人工智能安全峰会上,美国商务部长雷蒙多宣布,美国计划建立其专属的人工智能监管机构。

英国并未设立一个专责监管人工智能应用的中央内阁部门,而是由分散在不同领域和级别的多个部门共同肩负此任务。例如,在数据保护领域,成立了信息专员办公室;在市场竞争方面,设立了竞争与市场管理局;在金融服务领域,金融行为管理局等机构负责监管事宜。但在2023年11月,英国向议会提交了《人工智能(监管)法案》,该法案要求设立人工智能管理局。此外,英国政府还计划仿效国际原子能机构(International Atomic Energy Agency, IAEA),在伦敦建立一个对全球人工智能进行监管的权威机构。

2.2 监管范围:全面规制和结果监管

欧盟实行了涵盖人工智能产品从市场准入前到准入后整个生命周期的监管模式,实现了对人工智能的事前、事中及事后全方位管理。在预备阶段,根据人工智能潜在风险的性质,将其分为四个类别,并根据不同的应用环境和风险级别,制定相应的监管措施;在实施阶段,突出了人工智能监管沙盒的功能(监管沙盒被定义为“一个安全的环境,企业可在其中试验创新性的产品、服务、商业模式和交付机制,而不会因为这些活动立刻面临常规的监管后果”)^[43];在产品上市后,要求系统供应商对高风险产品进行持续的售后监控,并

及时向国家监管机构报告任何故障信息。美国和英国则主要针对人工智能技术在具体应用场景下的监管,更倾向于事后及结果监管,强调增强隐私保护、消除滥用人工智能导致的司法和医疗等领域的歧视、维护消费者的合法权益及加强对劳动者的保护等。

2.3 监管路径——协同和去中心化

欧盟的监管体系展现出双重特性。在协调层面,欧盟委员会、欧洲议会以及欧洲理事会联手进行人工智能监管的法律和政策制定;在实施层面,欧盟及其成员国设立了多个级别的执行机构^[44],并成立了人工智能高级专家组、人工智能联盟等,充分发挥组织协会的作用。

美国联邦政体及其三权分立的政治体系决定了美国人工智能监管的路径也是三位一体的,立法、行政与司法在各自职权范围内发挥监管能力^[45]。在立法层面,美国没有从联邦层面制定相关法案,但美国州政府已制定监管立法;早在2011年,内华达州便出台了首个针对自动驾驶车辆检测的法律,规定了自动驾驶车辆的监管要求^[46]。在行政层面,美国行政机关具有制定监管政策和执行监管规定的责任。在司法层面,美国法院通过判决的方式使人工智能风险主体承担相应责任。

英国基于希冀保持在全球人工智能技术发展领先地位的愿景,倾向于依靠现有机构进行监管,采取去中心化的监管原则,其监管内容散见于其他法律法规中。

2.4 监管共性

1) 对人工智能监管政策出台趋紧趋“硬”。欧盟《人工智能法案》的出台为全球奠定了人工智能监管的法制基础,而随着人工智能技术飞速发展,公众对人工智能技术可能带来国家安全危

机的担心不断增加,美国 and 英国也在转变对人工智能宽松监管的态度,由过去更强调产业发展与技术创新转向监管与发展并重,监管的措施也由强调企业自我监督和承诺规范转向通过更加频繁出台的政策和更“硬”的措施来实现监管。

2) 加强国际互动与合作磋商同时注重“主权人工智能”。美英等国积极在国际上发声并组建同盟、联盟等,抢占人工智能话语权。如美国在2023年11月联合17个国家发表了一项协议,以确保人工智能在开发和部署过程中的安全^[47],同时联合30个国家签署了对军事人工智能进行合理限制的宣言^[48]。英国举办了首届人工智能安全峰会,以奠定未来国际合作中重要贡献方的身份。在人工智能技术发展上,各国也在积极开展符合本国文化和思维方式的数据训练模型,以形成本土化产品。2024年2月美国总统拜登签署了“关于防止有关国家访问美国敏感数据”的行政令,限制数据被他国用于人工智能训练。

3) 注重监管与产业发展的平衡。欧、美、英在制定人工智能监管举措的同时也强调对产业发展的促进。欧盟的《人工智能法案》强调“规制为主兼顾促进”,努力为产业创新发展和中小企业创造比较友好的监管环境。美国的监管政策相较于欧盟更为宽松,没有像欧盟一样对所有人工智能进行风险分级监管,而是重点关注数据隐私保护、国防安全以及算法歧视,且以人工智能技术发展为侧重。英国对人工智能的监管则更加灵活,鼓励技术创新,在监管方向上更加注重对伦理的监管。

3 启示建议

中国的人工智能技术发展虽然起步较晚,但发展速度迅猛。在促进技术发展的同时,人工智能涉及到的安全、伦理等问题也不可忽视。既需

要利用人工智能技术实现降低生产成本、提高生产效率并刺激新质生产力的形成,同时也需要通过有效的监管与治理手段确保技术的发展不产生危害国家和社会稳定的因素,维护我国科技安全,为此提出以下建议。

1) 设立人工智能管理机构,完善监管政策体系。人工智能的监管能力与水平是支撑规则制定权和话语权的关键因素,因此美国、英国、欧盟提出要设立自己的人工智能监管机构。我国也应设立专门的人工智能监管与治理机构,制定法律法规及监管措施,统筹人工智能监管,不断完善监管政策体系。

2) 积极参与国际合作,把握行业话语权。我国的人工智能发展水平目前处于全球领先地位,其发展基础是基于国际上尚未形成规范统一的监管标准。美国、欧盟正在加紧谋划,一旦形成与我国人工智能发展意识相悖的标准,将极大阻碍和制约我国人工智能技术进一步发展的空间。因此我国应积极参与国际合作与国际联盟的建立,扩大我国人工智能在国际上的影响力,参与甚至主导国际人工智能监管框架与措施的制定,争取掌握行业监管标准的制定权和话语权。

3) 加强研究评估,注重产业创新与监管规制的平衡。在人工智能的技术与场景中应秉持发展与安全并重的理念,加强柔性治理、司法先行先试等路径,避免形成过度威慑。让市场优先演化或发展出规则的形成,随之辅助市场失灵时的强制性规则。可以通过培育发展专门的智库力量,实时跟踪人工智能技术应用的风险和创新需求,周期性评估监管状况和效果,持续优化监管体系和措施。在预测和评估技术相关风险的同时,激发人工智能技术驱动数字经济和实体经济的创新活力,持续推进我国科技产业高质量发展。

4) 加强对人工智能企业的引导, 重视遵守监管规则。欧盟的监管框架延续了长臂管辖规则, 即在欧盟境内提供人工智能相关产品或服务的所有相关主体都需要受到监管, 遵守前述强制性要求, 无论其是否在欧盟境内设有营业场所, 这意味着欧盟以外的科技公司可能受到监管。为避免未来我国人工智能企业技术出口时遭遇封锁, 可由人工智能管理机构建立联系企业制度, 进而引导企业提前布局并重视对国际监管规则的遵守。

5) 加快人工智能人才培养, 以人才建设引领产业与监管并行。没有人才对人工智能技术的理解与领悟, 也就无法出台适应技术发展的监管举措。而根据赛迪顾问的统计数据, 2021年我国高水平人工智能人才留存率不足50%, 中国籍人才已经成为美国人工智能技术研发的中坚力量。人才的流失会遏制我国人工智能技术的发展, 监管也将变得无意义。因此一方面我国应积极承办国际化人工智能顶尖学术会议和高峰论坛, 建立与全球人工智能领域顶尖高校的合作交流机制, 加强对高层次人工智能人才的引进。另一方面要以国内重点高校、科研院所为突破口, 加快推进人工智能一级学科“增量”建设工作。

参考文献

- [1] 邓悦, 许弘楷, 王诗菲. 人工智能风险治理: 模式、工具与策略[J]. 改革, 2024(1): 144-158. (DENG Yue, XU Hongkai, WANG Shifei. Artificial Intelligence Risk Governance: Models, Tools, and Strategies [J]. Reform, 2024(1): 144-158.)
- [2] 陈凤仙, 连雨璐, 王娜. 欧美人工智能监管模式及政策启示[J]. 中国行政管理, 2024(1): 77-88. (CHEN Fengxian, LIAN Yulu, WANG Na. European and American Artificial Intelligence Regulatory Models and Policy Implications [J]. Chinese Public Administration, 2024(1): 77-88.)
- [3] 谭九生, 杨建武. 人工智能技术的伦理风险及其协同治理[J]. 中国行政管理, 2019(10): 44-50. (TAN Jiusheng, YANG Jianwu. The Ethical Risk of Artificial Intelligence and Its Cooperative Governance [J]. Chinese Public Administration, 2019(10): 44-50.)
- [4] GOOS M, SAVONA M. The Governance of Artificial Intelligence: Harnessing Opportunities and Mitigating Challenges [J]. Research Policy, 2024, 53(3): 1-10.
- [5] 叶卓俊, 沈艳丽, 江晓, 等. 医学人工智能领域伦理治理重点研究[J]. 中国医学伦理学, 2024, 37(1): 39-44. (YE Zhuojun, SHEN Yanli, JIANG Xiao, et al. Research on the Focus of Ethical Governance in the Field of Artificial Intelligence in Medicine [J]. Chinese Medical Ethics, 2024, 37(1): 39-44.)
- [6] 石佳友, 徐靖仪. 医疗人工智能应用的法律挑战及其治理[J]. 西北大学学报(哲学社会科学版), 2024, 54(2): 91-103. (SHI Jiayou, XU Jingyi. Legal Governance of Medical Artificial Intelligence [J]. Journal of Northwest University (Philosophy and Social Sciences Edition), 2024, 54(2): 91-103.)
- [7] 肖非易, 李雪, 李睿, 等. 医疗人工智能技术评估与监管的国际经验及启示[J]. 中国卫生质量管理, 2023, 30(7): 58-62. (XIAO Feiyi, LI Xue, LI Rui, et al. International Experience and Implications of Medical Artificial Intelligence Technology Evaluation and Regulation [J]. Chinese Health Quality Management, 2023, 30(7): 58-62.)
- [8] REDDY S, ALLAN S, COGHLAN S, et al. A Governance Model for the Application of AI in Health Care [J]. Journal of the American Medical Informatics Association, 2020, 27(3): 491-497.
- [9] WANG Chao, ZHANG Jieyu, LASSI N, et al. Privacy Protection in Using Artificial Intelligence for Healthcare: Chinese Regulation in Comparative Perspective [J]. Healthcare (Basel), 2022, 10(10): 1878-1897.

- [10] 黄琦, 赵大伟, 代会琳, 等. 人工智能“火爆”背景下的金融安全“冷”思考[J]. 金融发展研究, 2023(11): 77-81. (HUANG Qi, ZHAO Dawei, DAI Huilin, et al. Rethinking of Financial Security under the Background of Artificial Intelligence Hype [J]. Journal of Financial Development Research, 2023(11): 77-81.)
- [11] 陈帆, 孙凯. 自然语言大模型在税收监管中的应用——以非居民税收合同审核为例[J]. 税收经济研究, 2024, 29(3): 38-46. (CHEN Fan, SUN Kai. Application of Natural Language Large Model in Tax Supervision: Taking non Resident Tax Contract Review as an Example [J]. Tax and Economic Research, 2024, 29(3): 38-46.)
- [12] 赵炳昊. 金融领域中人工智能应用的监管与法律规制[J]. 郑州大学学报(哲学社会科学版), 2024, 57(3): 51-57, 142. (ZHAO Binghao. Regulation and Legal Control of Artificial Intelligence Applications in the Financial Sector [J]. Journal of Zhengzhou University (Philosophy and Social Sciences Edition), 2024, 57(3): 51-57, 142.)
- [13] 肖红军, 李书苑, 阳镇. 数字科技伦理监管的政策布局与实践模式: 来自英国的考察[J]. 经济体制改革, 2023(5): 156-166. (XIAO Hongjun, LI Shuyuan, YANG Zhen. Policy Framework and Implementation Models for Ethical Regulation of Digital Technology: Insights from the United Kingdom [J]. Reform of Economic System, 2023(5): 156-166.)
- [14] 刘露, 杨晓雷, 高文. 我国人工智能伦理监管需求分析及对策研究[J]. 中国工程科学, 2021, 23(3): 106-112. (LIU Lu, YANG Xiaolei, GAO Wen. Artificial Intelligence Ethics Supervision in China: Demand Analysis and Countermeasures [J]. Strategic Study of CAE, 2021, 23(3): 106-112.)
- [15] 唐晓玲, 郑楚月. 技术伦理视域下美国教育人工智能伦理问题审视: 现实表征与规避策略[J]. 广西师范大学学报(哲学社会科学版), 2024, 60(3): 118-129. (TANG Xiaoling, ZHENG Chuyue. An Examination of the Ethical Issues of Artificial Intelligence in American Education from the Perspective of Technical Ethics: Realistic Representation and Avoidance Strategies [J]. Journal of Guangxi Normal University (Philosophy and Social Sciences Edition), 2024, 60(3): 118-129.)
- [16] 吴逸菲, 樊春良. 中国语境下人工智能伦理治理的路径塑造——基于价值-工具二维理性的融合框架[J/OL]. 科学学研究 (WU Yifei, FAN Chunliang. Shaping the Ethical Governance Path of Artificial Intelligence in the Chinese Context: Based on Value-Instrument Rationality [J/OL]. Studies in Science of Science) [2024-06-28]. <https://doi.org/10.16192/j.cnki.1003-2053.20240306.001>.
- [17] 周辉, 金僖艾. 英国人工智能监管实践、创新与借鉴[J]. 数字法治, 2023(5): 194-206. (ZHOU Hui, JIN Xiai. Practice, Innovation, and Reference of Artificial Intelligence Regulation in the UK [J]. Digital Law, 2023(5): 194-206.)
- [18] 董杨, 张蕴萌, 张雪然. 基于公共责任的人工智能监管: 美国的关键做法及启示[J]. 中国行政管理, 2024(1): 112-120. (DONG Yang, ZHANG Yunmeng, ZHANG Xueran. Artificial Intelligence Accountability Based on Public Responsibility: Key Approaches and Implications in the United States [J]. Chinese Public Administration, 2024(1): 112-120.)
- [19] 雷挺, 吴义熔. 他山之石: 欧盟人工智能监管的经验与启示[J]. 科技管理研究, 2023, 43(24): 43-52. (LEI Ting, WU Yirong. Artificial Intelligence Accountability Based on Public Responsibility: Key Approaches and Implications in the United States [J]. Science and Technology Management Research, 2023, 43(24): 43-52.)
- [20] 张广伟. 欧盟人工智能监管沙盒制度的功能、局限及其启示——基于欧盟《人工智能法》的解析[J]. 德国研究, 2024, 39(2): 116-132, 136. (ZHANG Guangwei. The Functions, Limitations,

- and Implications of the EU Artificial Intelligence Regulatory Sandbox System: An Analysis Based on the EU Artificial Intelligence Law [J]. *Deutschland-Studien*, 2024, 39(2): 116-132, 136.)
- [21] 黄静怡. “分级分类”与“契约”风险治理并行的人工智能监管制度构建——以欧盟《人工智能法案》为分析对象 [J]. *海南金融*, 2024(2): 76-85. (HUANG Jingyi. The Construction of an Artificial Intelligence Regulatory System that Combines “Hierarchical Classification” and “Contractual” Risk Governance: Taking the EU’s “Artificial Intelligence Act” as the Analysis Object [J]. *Hainan Finance*, 2024, (2): 76-85.)
- [22] 雷昊楠. 生成式人工智能的竞争风险、监管逻辑与反垄断应对 [J]. *中国科技论坛*, 2024(7): 64-74. (LEI Haonan. Competitive Risk, Regulatory Logic and Anti-monopoly Response of Artificial Intelligence Generated Content [J]. *Forum on Science and Technology in China*, 2024(7): 64-74.)
- [23] 沈芳君. 生成式人工智能的风险与治理——兼论如何打破“科林格里奇困境” [J]. *浙江大学学报(人文社会科学版)*, 2024, 54(6): 73-91. (SHEN Fangjun. The Risks and Governance of Generative Artificial Intelligence: Addressing the “Colling Ridge Dilemma” [J]. *Journal of Zhejiang University (Humanities and Social Sciences)*, 2024, 54(6): 73-91.)
- [24] 宁宣凤, 吴涵, 陈琳珺, 等. 全球人工智能治理大变局之欧盟人工智能治理监管框架评述及启示 (NING Fengxuan, WU Han, CHEN Linjun, et al. Review and Inspiration of the EU Artificial Intelligence Governance Regulatory Framework in the Global Artificial Intelligence Governance Transformation) [EB/OL]. (2024-03-07). <https://www.kwm.com/cn/zh/insights/latest-thinking/eu-artificial-intelligence-governance-and-regulatory-framework.html>.
- [25] HACKER P, PASSOTH J-H. Varieties of AI Explanations under the Law. From the GDPR to the AIA, and Beyond [C]. *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*. Austria: Springer, 2022: 343-373.
- [26] 遂宁市科技局. 欧盟发布人工智能协调计划 (Science and Technology Department of Suining. EU Releases AI Coordination Plan.) [EB/OL]. [2024-03-07]. <https://skjj.suining.gov.cn/xinwen/show/bc24752f8eba4a9390e9fe7eda948ba8.html>.
- [27] 第一财经. 扎紧人工智能护栏: 最强AI法案将生效, “布鲁塞尔效应”波及全球 (First Financial. Tighten the Artificial Intelligence Fence: the Strongest AI Bill will Come into Effect, and the ‘Brussels Effect’ will Affect the World) [EB/OL]. [2024-07-17]. <https://baijiahao.baidu.com/s?id=1804836669127661410&wfr=spider&for=pc>.
- [28] Networking and Information Technology Research and Development Subcommittee. *The National Artificial Intelligence Research and Development Strategic Plan* [R]. Washington, DC: National Science and Technology Council, 2016.
- [29] National Science and Technology Council Committee on Technology. *Preparing for the Future of Artificial Intelligence* [R]. Washington, DC: Executive Office of the President, 2016.
- [30] LAWRENCE C, CUI I, HO D. The Bureaucratic Challenge to AI Governance: An Empirical Assessment of Implementation at US Federal Agencies [C]. *Proceedings of the 2023 A AAI/ACM Conference on AI*. Canada: AIES, 2023: 606-652.
- [31] 澎湃新闻. 美国规制人工智能, 发布问责框架 (Pengpai News. US Regulation of Artificial Intelligence, Release of Accountability Framework.) [EB/OL]. [2024-03-04]. <https://baijiahao.baidu.com/s?id=1705440040983011567&wfr=spider&for=pc>.

- [32] 中华人民共和国驻欧盟使团经济商务处. 欧盟与美国建立贸易和技术理事会 (Economic and Commercial Office of the Mission of the People's Republic of China to the European Union. Establishment of a Trade and Technology Council Between the European Union and the United States.) [EB/OL]. [2024-03-04]. <http://eu.mofcom.gov.cn/article/jmxw/202106/20210603109684.shtml>.
- [33] 安全内参. 布鲁金斯学会: 欧盟和美国在人工智能监管领域开始趋同对标 (Safety reference. Brookings Institution: EU and US Begin to Converge and Benchmark in the Field of Artificial Intelligence Regulation.) [EB/OL]. [2024-03-04]. <https://www.secrss.com/articles/40752>.
- [34] ENGLER A. The EU and U.S. are Starting to Align on AI Regulation [EB/OL]. [2024-03-04]. <https://www.brookings.edu/articles/the-eu-and-u-s-are-starting-to-align-on-ai-regulation/>.
- [35] Information Commissioner's Office. Big Data, Artificial Intelligence, Machine Learning and Data Protection [EB/OL]. [2024-03-07]. <https://ico.org.uk/media/for-organisation/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.
- [36] Information Commissioner's Office. Guidance on the AI Auditing Framework Draft Guidance for Consultation [EB/OL]. [2024-03-07]. <https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf>.
- [37] ALMEIDA D, SHMARKO K, LOMAS E. The Ethics of Facial Recognition Technologies, Surveillance, and Accountability in an Age of Artificial Intelligence: A Comparative Analysis of US, EU, and UK Regulatory Frameworks [J]. *AI and Ethics*, 2021, 2(3): 11-11.
- [38] Office for National Statistics. Data Ethics Policy [EB/OL]. [2024-03-07]. <https://www.ons.gov.uk/about-us/transparencyandgovernance/datastrategy/datapolicies/dataethicspolicy>.
- [39] GOV. UK. New Ten-year Plan to Make the UK a Global AI Superpower [EB/OL]. [2024-03-07]. <https://www.gov.uk/government/news/new-ten-year-plan-to-make-britain-a-global-ai-superpower>.
- [40] GOV.UK. Establishing a Pro-innovation Approach to Regulating AI [EB/OL]. [2024-03-07]. <https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai>.
- [41] GOV.UK. AI Regulation: a Pro-innovation Approach [EB/OL]. [2024-03-07]. <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach>.
- [42] GOV.UK. Proposed Principles to Guide Competitive AI Markets and Protect Consumers [EB/OL]. [2024-03-07]. <https://www.gov.uk/government/news/proposed-principles-to-guide-competitive-ai-markets-and-protect-consumers>.
- [43] 宁宣凤, 吴涵. 路未央, 花已遍芳——欧盟《人工智能法案》主要监管及激励措施评述 (NING Fengxuan, WU Han. The Road is not Finished, the Flowers are Fragrant: A Review of the Main Regulatory and Incentive Measures of the EU's Artificial Intelligence Act.) [EB/OL]. [2024-03-14]. https://mp.weixin.qq.com/s/VMSAOuWYlg4U8_LcJocIGw.
- [44] 和军, 杨慧. ChatGPT类生成式人工智能监管的国际比较与借鉴 [J]. *湖南科技大学学报 (社会科学版)*, 2023, 26(6): 119-128. (HE Jun, YANG Hui. International Comparison and Reference on ChatGPT Like Generative Artificial Intelligence Regulation [J]. *Journal of Hunan University of Science and Technology (Social Science Edition)*, 2023, 26(6): 119-128.)
- [45] 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 [J]. *理论月刊*, 2023(8): 127-139. (SHI Fenglin, ZHANG Zhiyuan. On Public Law Regulation of Artificial Intelligence: the US Europe Model and China's Path [J]. *Theory Monthly*,

2023(8):127-139.)

- [46] 刘轩, 陈海彬. 人工智能监管: 理论、模式与趋势[J]. 情报理论与实践, 2023, 46(6): 17-23, 37. (LIU Xuan, CHEN Haibin. Artificial Intelligence Regulation: Theory, Model and Trend [J]. Information Studies: Theory & Application, 2023, 46(6): 17-23, 37.)
- [47] NGUYEN B. 美国与其他 17 个国家达成协议, 将致力于保护人工智能安全 (NGUYEN B. The United States has Reached an Agreement with 17 other Countries to Work towards Protecting the Security of Artificial Intelligence.) [EB/OL]. [2024-03-15]. <https://baijiahao.baidu.com/s?id=1783777782595928916&wfr=spider&for=pc>.
- [48] AI智能矩阵. 美国和 30 个国家同意对军事人工智能进行合理限制, 不包括中国和俄罗斯 (AI Intelligence Matrix. The United States and 30 Countries Agree to Impose Reasonable Restrictions on Military Artificial Intelligence, Excluding China and Russia.) [EB/OL]. [2024-03-15]. <https://zhuanlan.zhihu.com/p/666110419>.

作者贡献说明

王恺乐: 论文初稿撰写;

陈云伟: 选题指导, 论文修改;

熊永兰: 论文修改。