

政策建议

中国人工智能治理政策的结构特征与优化路径

刘肖肖^{1,2}, 杨国梁^{2,3*}

摘要 新一轮科技革命和产业变革的背景下,人工智能已经成为核心驱动力,正在对世界经济、社会治理和人民生活产生极其深刻的影响。同时,也引发了对人工智能发展的诸多担忧,如隐私泄露、数据滥用、算法歧视、信息茧房、就业替代等。中国快速响应,正在构建“治理主体-治理议题-治理工具”的三维治理框架。选取 2017—2024 年有关部门发布的规划意见、法律法规、标准规范、原则倡议 4 类政策文本,聚焦数据、算法、算力等人工智能治理关键要素进行政策文本量化分析。分析发现,中国人工智能治理面临治理议题上“前端重、中后端轻”的治理格局,治理主体上“顶层设计优先”原则体现,以及治理工具上“硬法规制主导、软约束不足”的显著特征等挑战。为此,提出人工智能治理议题上的平衡,逐步构建“前端预防—中端管控—后端保障”的全链条治理框架;治理主体间的协同,在继续强化管理、研发、供应主体治理的同时,将使用主体纳入治理核心圈层;治理工具上的协调,优化政策工具组合,推进硬法托底与软法协同等建议。

关键词 人工智能治理;数据保护;算法安全;算力规范

人工智能(artificial intelligence, AI)概念的提出,最早出现于 1950 年计算机科学与人工智能之父——图灵发表的《计算机器与智能》(《Computing machinery and intelligence》)^[1]一文中,该文章提出“机器是否能思考”的问题,并设计了“模仿游戏”来验证机器是否拥有智能,即“图灵测试”。1956 年,在美国达特茅斯会议上,麦卡

锡(John McCarthy)首次提出了人工智能这一概念^[2]。当下,人工智能已经成为新一轮科技革命和产业变革的核心驱动力,正在对世界经济、社会治理和人民生活产生越来越深刻的影响。在经济发展上,全球把人工智能视作引领未来的战略性技术,提升国家竞争力、推动国家经济增长;在社会治理上,人工智能可以模拟政策实施结果,一定程度降低社会治理成本、提升治理效能、减少治理干扰;在工作生活中,深度学习、图像识别、语音识别等人工智能技术已经广泛应用于智能家居、移动支

付、智能医疗、智慧交通等领域,为普通民众提供更加便利的生活服务。然而,人工智能也带来了技术失控风险、社会伦理挑战等问题,具体包括隐私泄露、数据滥用、算法歧视、信息茧房、就业替代等,由此引起了人们对人工智能治理的思考^[3-5]。

基于上述对人工智能技术安全风险、伦理挑战的描绘,当前亟待开展对人工智能风险、挑战的应对防范。为此,本研究在数据、算法、算力等人工智能伦理关键治理要素的框架下,遵循“治理主体、治理议题、治理工具”的三维

1. 中国标准化研究院,北京 100191

2. 中国科学院科技战略咨询研究院,北京 100190

3. 中国科学院大学,北京 100049

收稿日期: 2025-02-17; 修回日期: 2025-05-09

基金项目: 国家市场监督管理总局科技计划资助项目(2023MK190); 中国标准化研究院院长基金项目(572023Y-10377)

作者简介: 刘肖肖,助理研究员,研究方向为科技规划与管理、标准化原理与方法,电子信箱: liuxx@cnis.ac.cn; 杨国梁(通信作者),研究员,研究方向为科技规划与管理、决策理论与方法,电子信箱: glyang@casipm.ac.cn

引用格式: 刘肖肖, 杨国梁. 中国人工智能治理政策的结构特征与优化路径[J]. 科技导报, 2026, 44(5): 112-119; doi:10.3981/j.issn.1000-7857.2025.02.00184

治理框架,运用政策文本量化分析方法揭示出当前中国人工智能伦理治理现状及挑战,并尝试提出符合中国特色的未来人工智能治理路径。

1 中国人工智能治理的现状

1.1 数据保护

随着人工智能技术的广泛应用,数据作为关键要素,其重要性愈发凸显。同时,数据安全及隐私保护问题也逐渐成为全球各国关注的焦点。欧盟采取“自上而下的规则监管”模式,通过设定法律法规来对数据进行监管,《通用数据保护条例》(《General Data Protection Regulation》)生效后得到进一步巩固^[6]。美国秉持“州级自立与分领域监管并存”模式。一方面,以州和地方政府制定符合自身情况的相关政策并成立治理机构,如《加州消费者隐私法》(《California Consumer Privacy Act》);另一方面,针对金融、医疗、儿童隐私等领域主体出台相应法案,如《儿童在线隐私保护法》(《Children's Online Privacy Protection Act》)^[7]。

中国强调“国家安全为导向”的制度构建策略,持续强化对数据安全和隐私保护的监管力度,从顶层设计到具体实施细则,出台了一系列政策文件,特别是在跨境数据治理中,采取的是“属地原则”的“数据本地化”政策,强调国家数据主权和数据控制权来满足自身安全需要^[8]。2021年11月,《个人信息保护法》^[9]正式施行,这是中国首部专门针对个人信息保护的系统性、综合性法律,为破解个人

信息保护中的热点难点问题提供了强有力的法律保障。自此,《个人信息保护法》与《网络安全法》^[10]《数据安全法》^[11]共同形成了中国数据安全领域法律体系的“三驾马车”,以推动中国数据要素市场的法治化建设。

1.2 算法安全

当前,信息茧房、大数据杀熟、算法歧视等应用乱象层出不穷。美国将算法作为直接监管对象,采取外部问责和政府风险监管并重的治理方式,反对算法歧视、追求决策公平^[12]。2022年,美国参议院形成《算法责任法案(2022)》(《Algorithmic Accountability Act of 2022》),意在进一步要求科技企业在使用自动化决策系统时,对偏见、有效性和相关因素进行系统化影响评估,提升算法应用的普及性、精准性和透明度。欧盟重点规制互联网巨头,解决算法安全问题,防止其滥用市场支配地位。2022年,欧盟委员会发布《数字市场法案》(《Digital Markets Act》),将大型在线平台充当数字市场的“守门人”,规定其在日常运营中必须遵守的要求,例如,承担起数据共享、算法公开等义务^[13]。

近年来,中国逐步将数据安全治理的重心向算法安全治理倾斜,2021年9月,国家互联网信息办公室(以下简称国家网信办)等9个部门联合发布《关于加强互联网信息服务算法综合治理的指导意见》,创新性地形成了算法备案、算法监督检查、算法风险监测、算法安全评估等多维一体的监管举措。2022年3月1日,由国家网信办等4个部门联合发布

的中国首部专门以“算法”为名的规定——《互联网信息服务算法推荐管理规定》正式实施^[14]。该规定为应对算法推荐乱象提供指引和规范,并为国家建立算法安全治理顶层设计奠定制度基础,明确了信息服务规范、用户权益保护、监督管理责任和法律责任。自此,中国也有了针对算法黑箱、算法歧视、信息茧房、舆论操控、不正当竞争等算法侵害的规范。

1.3 算力规范

算力已成为人工智能时代的核心生产力,同时国际上关于算力规范的政策与研究正围绕人工智能的技术演进与社会影响展开,各国形成了差异化的治理框架。欧盟倾向风险导向的算力监管,将算力使用与风险等级挂钩,医疗、交通等高风险领域的人工智能系统需满足算力基础设施的安全性与透明度要求^[15]。美国则支持算力竞争与安全并重,一方面通过《芯片和科学法案》来提升其算力基础设施水平和创新能力,另一方面通过《人工智能风险管理框架》等政策要求算力使用符合公平性、安全性原则^[16]。

在中国,以“东数西算”工程为例,算力网络涉及多区域、多主体的协同,网络架构动态调整频繁,原有安全防护体系难以适配,一旦出现安全防护疏漏,极易引发重大安全事件。为应对挑战,中国出台了2项相关政策。2023年10月,工业和信息化部发布的《算力基础设施高质量发展行动计划》^[17],从提升算力综合供给、强化运力高效承载、增强算力应用赋能等多方面发力,其中也包含对算

力安全规范的相关要求,为算力安全发展提供了总体方向。2023年12月,国家发展和改革委员会发布的《深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》^[18]明确提出要统筹算力发展与安全保障一体化推进,要完善算力安全保障体系,构建促发展保安全机制。

1.4 综合治理

在综合治理方面,美欧均相继出台较为权威的人工智能治理政策。欧盟坚持“以人为本,伦理优先”的治理原则,已生效的《人工智能法案》(《Artificial Intelligence Act》)主张构建协作治理结构,该结构超越了单纯的协商,意在促使公共机构或机构与不同利益相关者(公民、企业、非政府组织、专家等)在政策制定过程中持续进行知识交流^[17]。美国秉持“美国优先,创新优先”策略,拜登政府2023年发布《安全、可靠和可信赖的人工智能开发和使用行政命令》(《The Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence》),要求AI开发者在发布前提交安全测试结果,并对高风险模型实施监管。但执政党不同,治理理念存在差异。2025年1月特朗普政府签署行政令废除该行政命令,采取“轻监管”模式,强调减少对企业创新的干预^[19]。

中国已在数据保护、算法安全、算力规范层面作出了不少探索,同时也在人工智能综合治理层面也提出了系列倡议、原则、指南及管理办法。2019年初,科技部成立新一代人工智能治理专业委

员会,同年6月发布《新一代人工智能治理原则——发展负责任的人工智能》,提出和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理等8条人工智能治理原则。2020年7月,国家标准化管理委员会等部门联合印发《国家新一代人工智能标准体系建设指南》,在安全及伦理标准部分,提出围绕基础安全,数据、算法和模型安全,技术和系统安全,安全管理和服务,安全测试评估,产品和应用安全等6个部分重点开展标准体系建设。2021年9月,新一代人工智能治理专业委员会发布《新一代人工智能伦理规范》^[20],提出了增进人类福祉、促进公平公正、保护隐私安全、确保可控可信、强化责任担当、提升伦理素养等6项基本伦理要求。2023年7月10日,国家网信办等7个部门联合发布出台了《生成式人工智能服务管理暂行办法》^[21],重点监管利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等内容服务主体。

2 中国人工智能治理的政策文本量化分析

2.1 政策分析框架构建

治理主体指的是治理的“相关者网络”,是参与治理过程的行为主体,包括政府(核心主体)、市场主体(企业)、社会主体(社会组织、公众)、国际组织等。治理议题指的是治理的“问题导向”,是治理活动所针对的具体问题或领域,如数据隐私问题、算法歧视问题等。治理工具指的是治理的“手段工具箱”,是为实现治理目标而采用的政策工具、机制或方式,包括强制性工具、自愿性工具、混合性工具等。通过“治理主体-治理议题-治理工具”的三维分析框架,运用政策文本量化分析法,深入剖析其在单维度、双维度的政策特征,进而提出中国人工智能治理的针对性建议,如图1所示。

2.1.1 治理议题(X维度)

在治理议题的X维度,分别从综合治理和精细治理维度提出了系列政策措施,既包含通识性、系统化人工智能综合治理政策,也包含数据保护、算法安全、算力规范的分环节治理政策^[7],如表1所示。

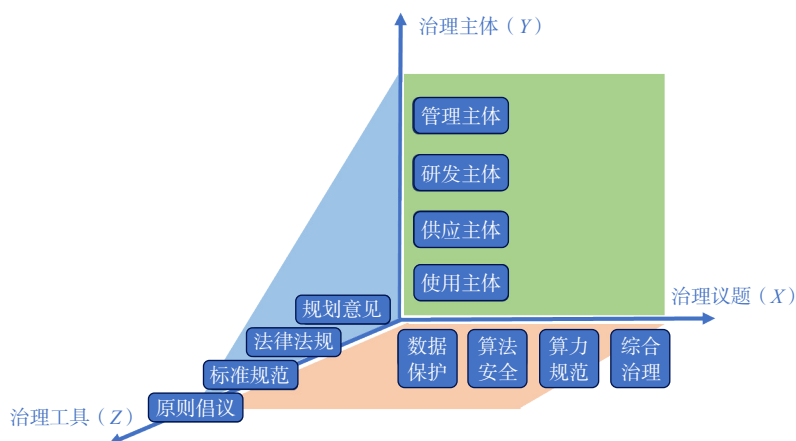


图1 中国人工智能治理的量化分析框架

表 1 人工智能治理过程维度(X 维度)指标及其描述

治理议题	治理议题描述	编码
数据保护	包含数据跨境流动、网络安全、个人信息等相关的法律、条例、审查办法、管理规定、指导意见、安全倡议等	X1
算法安全	包含信息服务相关的算法推荐、弹窗推送、跟帖评论、深度合成、账号管理等	X2
算力规范	包含提升算力相关的算网安全、算力低碳等行动计划、实施意见等	X3
综合治理	包含人工智能相关的发展规划、治理原则、建设指南、伦理规范、管理办法、安全要求、治理倡议等	X4

2.1.2 治理主体(Y 维度)

在治理主体的 Y 维度,中国制定的人工智能治理政策实现了全周期覆盖,包括管理主体、研发主体、供应主体和使用主体 4 类主体^[22],如表 2 所示。

2.1.3 治理工具(Z 维度)

在治理工具的 Z 维度,从“硬性法律”到“软性原则”进行了全范围的覆盖,主要涉及规划意见、法律法规、标准规范、原则倡议等 4 类^[23],如表 3 所示。

2.2 研究样本及编码

2.2.1 研究样本检索与选取

关于政策选取时间,2017 年 7 月,国务院正式印发《新一代人工智能发展规划》,确立了新一代人工智能发展“三步走”战略目标,人工智能的发展至此上升到国家战略层面,推动中国人工智能步入新阶段。因此,本文选取政策时间自 2017 年 1 月—2024 年 4 月。

关于政策检索原则,本研究基于“治理主体—治理议题—治理工具”的三维分析框架遴选政策文本,具体涉及人工智能数据、算

法、算力以及综合治理等关键要素;涉及人工智能治理周期,即“管理、研发、供应、使用”4 个重点环节所涉及的其中一环;涉及人工智能规划意见、法律法规、标准规范、原则倡议等任意文本类型。为此,本研究重点从中国共产党中央委员会、国务院、工业和信息化部、科学技术部、教育部、国家网信办、国家标准化管理委员会等官方网站,以“人工智能”“数据”“网络”“信息”“算法”“算力”“伦理”等关键词搜集相关政策,最终删除重复政策后,遴选出 42 份关联度强的政策文本。通过该图可以看出,自 2021 至今中国人工智能治理政策密集出台,对其重视程度达到高点,如图 2 所示。

2.2.2 研究样本使用与编码

关于政策编码方法,参考 NVivo 软件中文本标注方法对已经选定的 42 份政策进行节点标注(考虑到本研究中政策文本数量并未超过百份,同时为了降低标注误差,采取 2 轮人工标注的方式)。

具体地,对分别符合“治理主体—治理议题—治理工具”三维框架下的任何一个维度,包括治理议题(数据、算法、算力、综合治理)、治理主体(管理、研发、供应、使用)、治理工具(规划意见、法律法规、标准规范、原则倡议)的相关章节有且仅标注 1 次,均标注为 1 个节点。

以《新一代人工智能伦理规范》为例,核心章节分别是人工智能的管理规范、研发规范、供应规范和使用规范,对其文本进行分析和编码后得到,在治理议题维度,节点标注为“数据保护 X1(0 个节点)—算法安全 X2(0 个节点)—算力规范 X3(0 个节点)—综合治理 X4(1 个节点)”;在治理主体维度,节点标注为“管理主体 Y1(1 个节点)—研发主体 Y2(1 个节点)—供应主体 Y3(1 个节点)—使用主体 Y4(1 个节点)”;在治理工具维度,节点标注为“规划意见 Z1(0 个节点)—法律法规 Z2(0 个节点)—标准规范 Z3(1 个节点)—原则倡议 Z4(0 个节点)”。

表 2 人工智能治理过程维度(Y 维度)指标及其描述

治理主体	治理主体描述	编码
管理主体	承担人工智能发展相关的规划、组织、协调等职能的主体,涉及国务院、核心监管部门、行业自律组织等	Y1
研发主体	负责人工智能技术研发与试验相关的主体,涉及科技巨头企业、高校与科研机构、开源社区等	Y2
供应主体	在人工智能产品供应链中承担原材料、零部件等供给职能的主体,包括硬件供应、软件与算法供应、数据服务等	Y3
使用主体	人工智能产品和服务的最终承载者,如政务、金融、医疗、教育、交通、工业、能源、农业等智能化转型	Y4

表 3 人工智能治理过程维度(Z 维度)指标及其描述

治理工具	治理工具描述	编码
规划意见	为人工智能发展与治理绘制了宏观蓝图	Z1
法律法规	是人工智能治理的刚性约束	Z2
标准规范	为人工智能技术和产品的研发、应用提供了统一尺度	Z3
原则倡议	从伦理道德、发展理念等层面为人工智能发展提供软性指引	Z4

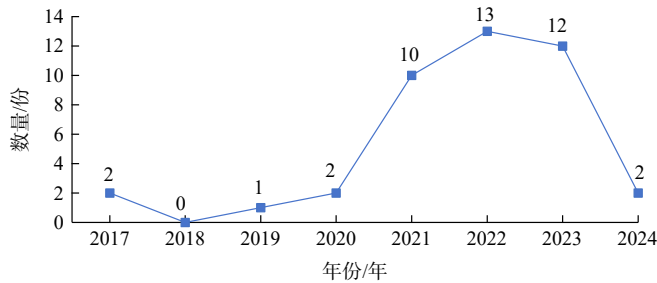


图 2 2017—2024 年中国人工智能治理政策数量的时间变化态势

3 基于“治理主体-治理议题-治理工具”三维框架的中国人工智能治理的政策文本量化分析

3.1 基于三维分析框架的中国人工智能治理政策单维度分析

3.1.1 治理议题维度(X 维度)

通过统计发现,总节点数 64,数据保护维度(X1)35 个节点,占比 54.69%,占据主导地位;其次是算法安全维度(X2)15 个节点,占比 23.44%;再次是综合治理维度(X4)12 个节点,占比 18.75%;最后是算力规范维度(X3)2 个节点,占比 3.12%。这一结构反映出当前中国治理框架的阶段性与潜在优化空间。

这种“前端重、中后端轻”的治理格局,与人工智能技术发展的现实逻辑高度相关。数据作为智能算法的“生产资料”,其安全与否直接关系到治理体系的基础稳定性。《数据安全法》《个人信息保护法》等顶层设计的落地,推动治理资源向数据采集、存储、使用

全流程倾斜,形成以数据合规为核心的“防护墙”。算法安全维度的 23.44% 占比,则体现了对算法透明性、可解释性及公平性的制度探索,例如,《互联网信息服务算法推荐管理规定》等政策的出台,初步构建了算法治理的规则框架。综合治理维度的 18.75% 占比,说明治理层已开始关注技术应用的社会影响,如《新一代人工智能治理原则》对人类福祉、公平正义等价值的强调。

然而,算力规范维度的显著薄弱(仅 2 个节点)暴露出治理体系的结构短板。算力作为人工智能的“基础设施”,其能源消耗、资源分配及安全防护等问题尚未得到足够重视。在算力全球化布局与碳中和目标双重背景下,缺乏系统性规范可能导致 2 方面风险:一是算力资源的无序竞争加剧能源浪费,二是算力节点的安全防护漏洞可能引发供应链危机。这种“重数据轻算力”的治理偏向,与人工智能“数据-算法-算力”三位一体的技术逻辑存在内在张力,亟

待从全生命周期治理视角进行优化(图 3)。

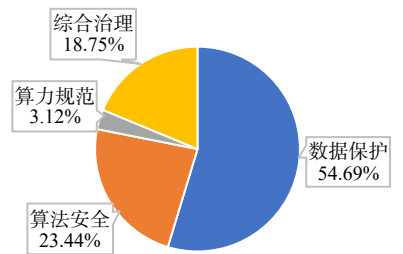


图 3 基于治理议题维度的政策量化分析结果

3.1.2 治理主体维度(Y 维度)

通过统计发现,总节点数 115,管理主体维度(Y1)37 个节点,占比 32.17%;研发主体维度(Y2)34 个节点,占比 29.57%;供应主体维度(Y3)34 个节点,占比 29.57%;使用主体维度(Y4)10 个节点,占比 8.69%。这一量化统计反映出中国治理体系呈现出显著的“主体结构偏向性”特征。

管理主体的高占比本质上是治理体系“顶层设计优先”原则的体现。《新一代人工智能发展规划》《国家新一代人工智能标准体系建设指南》等政策文件通过明确监管权责划分(如国家网信办、工业和信息化部、科学技术部等多部门协同机制),构建起“管理先行”治理模式,这有助于快速建立行业准入门槛,例如对人工智能企业的备案制度、算法安全评估机制等,为产业发展划定基本安全边界。同时,研发主体(Y2)与供应主体(Y3)的高占比(合计 59.14%),反映出研发与供应主体的协同治理,即政策对技术创新链与产业链的双重关注。

使用主体维度(Y4)的低占比(8.69%),暴露出治理体系对“人

工智能技术应用场景”的关注不足,造成终端风险防控的“最后一公里”缺失。一是权益保障的缺口。当智能算法侵害用户权益(如算法歧视、隐私泄露)时,现有政策侧重对企业的追责(如《个人信息保护法》),但用户端的救济渠道(如集体诉讼机制、算法解释请求权)尚未充分建立。二是公众参与缺位。使用主体治理的薄弱,也导致普通用户难以参与技术伦理决策。例如,欧盟《人工智能法案》明确要求高风险系统需进行“社会影响评估”,而中国相关政策中用户反馈机制仍待完善(图4)。

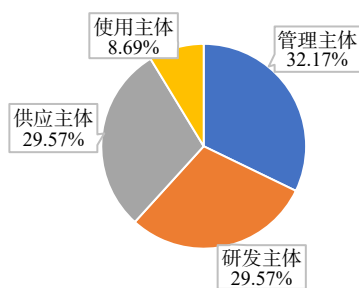


图4 基于治理主体维度的政策量化分析结果

3.1.3 治理工具维度(Z维度)

从政策工具维度观察中国人工智能治理体系,呈现出“硬法规制主导、软约束不足”的显著特征。在总计42个治理节点中,法律法规维度(Z2)以22个节点、52.38%的绝对优势位居首位,战略规划维度(Z1)占比21.43%(9个节点),标准指南维度(Z3)占14.29%(6个节点),而原则倡议维度(Z4)仅占11.90%(5个节点)。这一结构既体现了治理体系的制度建构逻辑,也揭示出从“刚性管控”向“刚柔并济”转型的必要性。

52.38%的法律法规占比,本质上是对人工智能技术“强渗透

性风险”的制度回应。《数据安全法》《个人信息保护法》《算法推荐管理规定》等刚性规范,通过设定禁止性条款(如数据跨境传输限制)、义务性规则(如算法备案要求)和法律责任(如行政处罚标准),为技术应用划定不可逾越的安全红线。

但是,标准指南维度(Z3)和原则倡议维度(Z4)分别仅占14.29%和11.90%,反映出治理体系对技术落地的“规则衔接”不足,也反映出治理体系对“技术伦理柔性治理”的重视不足。这将造成如下风险:一是创新抑制风险,过度依赖法律强制和行政规划,可能导致“合规成本挤占创新空间”;二是“软约束”不足,价值引领的“弹性空间”缺失。总体上,当前政策工具呈现“法律—规划—标准—倡议”的单向传导链条,但缺乏“硬法授权软法、软法补充硬法”的互动机制(图5)。

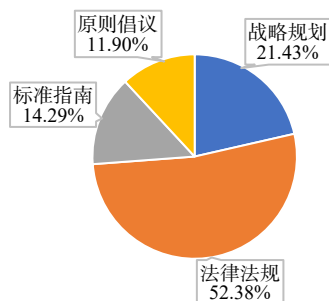


图5 基于治理工具维度的政策量化分析结果

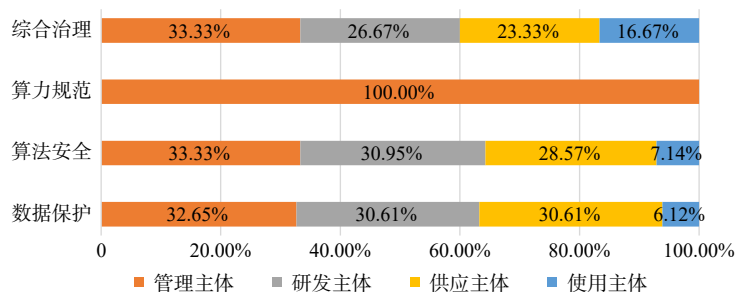


图6 基于治理议题(X)—治理主体(Y)维度的量化分析结果

3.2 三维分析框架下的中国人工智能治理政策双维度分析

3.2.1 治理议题(X)—治理主体(Y)维度分析

通过治理议题(X)—治理主体(Y)二维量化分析发现,在数据保护、算法安全、综合治理3大核心议题中,管理主体(政府)、研发主体(企业/科研机构)、供应主体(产业链上下游)的政策参与度均维持在20%~30%的高位区间,形成“治理铁三角”结构。然而,使用主体(终端用户)在3大议题中的参与度均低于10%,形成治理结构的“断裂带”。此外,算力规范维度相关政策发布数量较少,且均集中在管理规范上。

人工智能技术的“规划—开发—生产—使用”本应形成完整治理链条,但当前人工智能治理体系呈现出“重上游、轻下游”的结构,可能导致以下问题:一是造成风险传导效应,研发与供应环节的合规性,可能因使用环节的失控而失效,例如数据合规采集后,被用户滥用;二是产业生态的“供需错配”,供给侧治理政策的过度聚焦,导致人工智能技术停滞,而需求侧培育机制(如用户数字素养教育、公众参与式伦理审查)严重缺位(图6)。

3.2.2 治理议题(X)—治理工具(Z)维度分析

通过量化分析发现,治理议题中数据保护、算法安全聚焦法律法规类文本,占比分别是 57.14% 和 60.00%,形成“硬法优先”的治理模式。数据保护、算法安全的硬法化,实质是运用法律工具对“负外部性”进行定价(如违法收集个人信息的高额罚款),迫使企业将社会成本纳入技术创新决策。但过度刚性的规制可能引发“寒蝉效应”,降低企业创新活力。综合治理则聚焦于规划意见类文本,占比为 33.33%,呈现“软法主导”特征。软法的灵活性虽能快速响应技术变化,但缺乏法律强制力,难以形成稳定的预期制度环境(图 7)。

4 中国人工智能治理建议

中国在人工智能治理上有别于美国、欧盟的“创新优先”和“伦理优先”治理模式,而是将“坚持发展和安全并重、促进创新和依法治理相结合”作为人工智能治理的基本原则,推进人工智能健康发展。当前,中国人工智能治理体系正在逐步完善,结合本文的政策量化分析结果,主要提出治理议题上的平衡、治理主体间的协同、治理

工具上的协调等建议。

一是治理议题上的平衡。从治理现代化的视角审视,这种维度分布折射出中国人工智能治理的“前端治理重心”特征,即通过强化数据安全与隐私保护筑牢治理根基,符合风险预防原则与公众利益优先的治理逻辑。但面向技术产业化的深入推进,需逐步构建“前端预防—中端管控—后端保障”的全链条治理框架。为此,建议建立算力资源分类分级管理制度,将绿色算力、安全算力纳入政策支持范畴。

二是治理主体间的协同。当前中国人工智能治理政策存在“主体结构偏向”,但随着技术应用向纵深发展,必须从“主体割裂”转向“全链条协同”,在继续强化管理、研发、供应主体治理的同时,将使用主体纳入治理核心圈层。为此,建议强化使用主体的“场景化治理”,针对高风险场景(如金融科技、公共安全),建立强制性用户操作标准;探索建立“算法用户委员会”,赋予用户对自动化决策的异议权;推动企业承担使用端治理责任,通过“生产者责任延伸制度”,要求研发与供应主体向用户提供安全培训、风险预警等服务。

三是治理工具上的协调。当

前中国人工智能治理更多侧重“硬约束”(即法律法规)等方式进行规制。关于“软约束”,目前仅达成了较少的共识性“软约束”,且企业在政策制定上参与度相对较弱。为此,建议优化政策工具组合,推进硬法托底与软法协同。在高风险领域强化硬法刚性,针对生物识别、公共监控等涉及基本权利的场景,加快出台《人工智能法》等专门立法,明确数据滥用、算法歧视的刑事责任;在创新活跃领域扩大软法空间,例如生成式人工智能、量子计算等前沿领域,推行“监管沙盒+行业自律”模式。

参考文献(References)

[1] Turing A M. Computing machinery and intelligence[J]. Mind, 1950, 59(236): 433-460.

[2] John M, Marvin L M, Nathaniel R, et al. A proposal for the Dartmouth summer research project on acritical intelligence[J]. AI Magazine, 2006, 27(4): 12-14.

[3] 胡建华, 叶丽霞. 人工智能治理: 内涵、挑战与治理[J]. 江西理工大学学报, 2022, 43(4): 99-104.

[4] 中国社会科学网. 人工智能伦理反思: 风险与应对[EB/OL]. (2022-05-31) [2024-12-03]. https://www.cssn.cn/skgz/bwyc/202208/t20220803_5468336.shtml.

[5] 张兆翔, 张吉豫, 谭铁牛. 人工智能伦理问题的现状分析与对策[J]. 中国科学院院刊, 2021, 36(11): 1270-1277.

[6] 王天禅. 美欧人工智能治理的分化: 基于治理结构、能力势差和战略选择的考察[J]. 国际关系研究, 2025(2): 107-135.

[7] 马海群, 蔡庆平, 崔文波, 等. 美国数据与算法安全治理: 进路、特征与启示[J]. 信息资源管理学报, 2023, 13(1): 52-64.

[8] 梁宇, 郑易平. 全球数据治理的现实困境与中国方案的实践进路[J]. 昆明理工大学学报(社会科学版), 2023, 23(6): 105-114.

[9] 中华人民共和国个人信息保护法

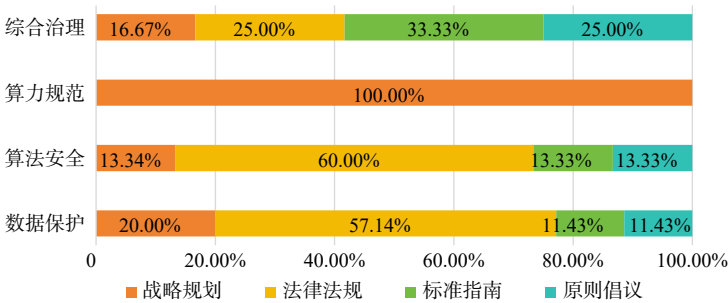


图 7 治理议题(X)—治理工具(Z)维度的量化分析结果

- [EB/OL]. (2021-08-20) [2024-12-25]. <https://flk.npc.gov.cn/detail2.html?ZmY4MDgxODE3YjY0NzJh-MzAxN2I2NTZjYzIwNDAwNDQ>.
- [10] 中华人民共和国网络安全法[EB/OL]. (2016-11-07) [2024-12-25]. http://www.npc.gov.cn/npc/c2/c30834/201905/t20190521_274248.html.
- [11] 中华人民共和国数据安全法[EB/OL]. (2021-06-10) [2024-12-25]. http://www.npc.gov.cn/npc/c2/c30834/202106/t20210610_311888.html.
- [12] 张凌寒. 权力之治: 人工智能时代的算法规制[M]. 上海: 上海人民出版社, 2021.
- [13] 崔文波, 张涛, 马海群, 等. 欧盟数据与算法安全治理: 特征与启示[J]. 信息资源管理学报, 2023, 13(2): 30-41.
- [14] 互联网信息服务算法推荐管理规定[EB/OL]. (2021-12-31) [2024-12-28]. https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm.
- [15] European Union. Artificial intelligence act[EB/OL]. (2021-4-21) [2025-4-22]. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>.
- [16] National Institute of Standards and Technology. Artificial intelligence risk management framework: generative artificial intelligence profile [EB/OL]. (2024-7-26) [2025-4-22]. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>.
- [17] 算力基础设施高质量发展行动计划的通知[EB/OL]. (2021-12-31) [2024-12-28] https://www.gov.cn/zhengce/zhengceku/202310/content_6907900.htm.
- [18] 国家发展改革委员会. 关于深入实施“东数西算”工程加快构建全国一体化算力网的实施意见[EB/OL]. (2023-12-25) [2024-12-28]. https://www.gov.cn/zhengce/zhengceku/202401/content_6924596.htm.
- [19] Cancela-Outeda C. The EU's AI act: A framework for collaborative governance[J]. Internet of Things, 2024, 27: 101291.
- [20] 新一代人工智能伦理规范[EB/OL]. (2021-9-26) [2025-01-02]. https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html.
- [21] 生成式人工智能服务管理暂行办法[EB/OL]. (2023-07-10) [2024-12-29]. https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm.
- [22] 樊春良, 吴逸菲. 人工智能伦理治理的三维分析框架: 核心要素、创新过程与主体责任[J]. 自然辩证法研究, 2023, 39(10): 18-24.
- [23] 郝凯冰. 基于三维框架的中国科技伦理治理政策分析[J]. 科学学研究, 2024, 42(11): 1-16.

Structural characteristics and optimization paths of China's artificial intelligence governance policies

LIU Xiaoxiao^{1,2}, YANG Guoliang^{2,3*}

1. China National Institute of Standardization, Beijing 100191, China

2. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China

3. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Against the backdrop of a new round of scientific and technological revolution and industrial transformation, artificial intelligence (AI) has emerged as a core driving force, exerting an extremely profound impact on the global economy, social governance, and people's lives. Meanwhile, AI has also given rise to numerous concerns regarding its development, such as privacy breaches, data abuse, algorithmic discrimination, information cocoons, and job displacement. China is making a rapid response and constructing a three-dimensional governance framework of "governance subjects, governance issues, and governance tools". This paper selects four types of policy texts issued by relevant Chinese government departments from 2017 to 2024, including plans and opinions, laws and regulations, standards and specifications, and principles and initiatives. It focuses on quantitative analysis of policy texts regarding key elements of AI governance such as data, algorithms, and computing power. The analysis reveals that China's AI governance faces challenges including a governance pattern of "heavy focus on the front-end and light on the mid-and rear-end" in governance issues, the embodiment of the "top-down design priority" principle in governance subjects, and the prominent feature of "dominance of hard regulation and insufficiency of soft constraints" in governance tools. To address these, this paper proposes balancing governance issues in AI governance, gradually constructing a full-chain governance framework of "front-end prevention—mid-end control—rear-end guarantee"; synergizing among governance subjects, while continuing to strengthen the governance of management, R&D, and supply subjects, incorporating user subjects into the core governance circle; and coordinating governance tools, optimizing the combination of policy tools, and promoting the coordination of hard law as the foundation and soft law.

Keywords artificial intelligence governance; data protection; algorithmic security; computing power standardization ●



(责任编辑 王微)