中国审计学自主知识体系

智能审计赋能下新时代中国特色算法审计体系构建研究

刘国城

（南京审计大学 会计学院，江苏 南京 211815）

[摘　要]在审计行业智能化转型与算法技术深度融入社会各领域的双重背景下，构建新时代中国特色算法审计体系，对于防范化解算法风险、提升国家治理现代化水平、夯实数字中国建设基础具有重要意义。智能审计依托自身发展积淀的实时监控、自动化检测、数据穿透分析等技术工具与方法论，为算法审计的落地实施提供重要赋能条件。聚焦智能审计赋能下新时代中国特色算法审计体系的构建，以理论内涵、实践框架、深化路径为主线，首先剖析新时代中国特色算法审计的基本内涵及本质特征，辨析中国算法审计与西方国家算法审计的本质区别；在此基础上，建立以“评估”“监督”“问责”为核心的中国特色算法审计三维实践框架，并提出“本土化创新”与“国际化引领”双轮驱动的发展路径，从而为健全新时代国家数字治理体系和完善全球算法审计机制提供理论支撑与实践指引。

[关键词]算法审计；审计学自主知识体系；智能审计；穿透式审计；数字治理；算法治理；国家治理

[中图分类号]F239.44　　[文献标志码]A　　[文章编号]1004－4833(2026)06－0012－09

一、引言

近年来，数字中国逐步成为新时代中国特色社会主义建设的重要内容。数字技术是数字中国建设的重要引擎，算法模型是数字技术的基础。随着数字中国建设的快速推进，算法模型已深度融入社会治理、公共服务与经济运行等领域，成为驱动国家治理现代化的重要工具。然而，算法技术具有黑箱性且社会影响广泛，可能引发诸如算法歧视、权力滥用等一系列风险，甚至对国家安全与社会公平构成威胁。作为规范算法应用、防范算法风险的重要手段，算法审计已成为新时代全球数字治理的焦点。与此同时，智能审计正在驱动审计行业变革，其自动化检测工具使算法偏见与合规性问题得以快速识别，数据穿透分析技术使算法决策逻辑的可追溯性大幅提升。智能审计不仅在技术工具层面提供持续监控、自动化检测、数据穿透分析等能力，而且将风险画像、全周期动态穿透等成熟审计方法论迁移至算法审计领域，实现对算法审计的多维赋能。

西方国家算法审计实践多以技术中立为前提，侧重在个体权利保障与市场运行效率之间寻求平衡，这一取向难以适配中国超大规模社会的现实条件、复杂治理情境以及“以人民为中心”的算法治理要求。国内算法审计的研究与实践仍处于初步探索阶段，尚未形成可供实务界借鉴应用的系统化体系。如何依托智能审计技术能力构建符合中国国情及自主知识体系需求的算法审计框架，是当前推进国家治理现代化、加快数字中国建设过程中亟待破解的议题。构建新时代中国特色算法审计体系具有特定的理论意义和实践价值。从理论层面来看，构建新时代中国特色算法审计体系的意义在于突破西方算法治理模式，以“以人民为中心”为根本，以“总体国家安全观”为边界，以“新发展理念”为路径，以“技术向善”为约束，以“人类命运共同体”为依托，揭示中国特色算法审计的本质特征与运行逻辑，丰富算法审计的中国话语体系。从实践层面来看，构建新时代中国特色算法审计体系的价值在于从审计评估、审计监督、审计问责等方面提出覆盖算法审计全生命周期作业流程，并在推动本土化创新的基础上为全球算法治理提供可供参考的中国经验。本文探究新时代中国特色算法审计的理论内涵、实践框架与深化路径，旨在丰富我国算法审计体系建设以及为应对算法技术引发的跨国治理挑战提供理论借鉴。

[收稿日期]2025－03－17

[基金项目]国家社会科学基金项目(24BJY066)；江苏省社科应用研究精品工程财经发展专项课题(26SCB－02)

[作者简介]刘国城(1978—)，男，内蒙古赤峰人，南京审计大学会计学院教授，硕士生导师，博士，从事算法审计、审计数智化研究，E-mail：liu8190@126.com。

二、新时代中国特色算法审计的理论内涵

(一)算法审计的理论演进

算法审计的理论发展经历了从技术审查观到影响评估观,再到综合治理观的演进历程。算法审计的早期研究遵循技术审查观,聚焦算法技术活动的合规性和风险性审查,审计内容主要涵盖算法系统输入[1],算法代码[2],算法模型的开发、构建与应用[3-4],算法决策逻辑[5]以及算法风险控制技术[6]。随着算法的社会影响不断泛化,算法审计的理论视角逐步转向影响评估观,学术界不断关注算法与社会规范交互中的歧视和偏见问题,并评估算法对社会群体的差异化影响。影响评估观下,算法审计从伦理视角出发,对算法运行逻辑进行系统性检查,核查算法黑箱的群体性歧视[7],检测模型算法中的偏见性操作流程[8],评估算法系统对个人、组织及社会可能产生的负面影响[9],揭示算法对社会公平的深层影响[10]。近年来,智能审计的实践发展加大了算法全生命周期风险综合治理的现实诉求,倒逼算法审计跳出影响评估观的局限,走向综合治理观。综合治理观主张整合技术工具、法律规制、伦理价值、社会影响与公众参与[11-13],强调构建多主体协同、全流程覆盖的治理框架[14-15],通过建立算法备案、影响评估、过程监督和动态纠偏的闭环机制[16-17],实现算法开发、部署、应用全生命周期的透明化与可问责性[18-19],进而推动算法治理从被动响应转向主动建构,推进治理重心从单纯的风险识别转向制度性纠偏与生态重塑。

(二)新时代中国特色算法审计的基本内涵

我国算法审计既应体现新时代特征,又应彰显中国特色,以凸显其在数字中国建设中的独特价值。首先,结合党的二十大精神,与算法审计相关的新时代特征体现在政治、经济、社会、技术、治理等维度,具体表现为政治维度下党的全面领导与国家治理体系现代化,经济维度下的社会创新发展与数字中国建设,社会维度下的公平正义与民生福祉保障,技术维度下的科技创新与算法伦理治理,以及治理维度下的国家安全与风险防控。其次,“中国特色”应体现在政治导向、文化传统、社会制度三个方面。在政治导向方面,算法审计需要符合国家政治导向和政策要求,保障国家安全和社会稳定,谨防算法技术沦为传播有害信息、扰乱公共秩序的工具;在文化传统方面,算法审计实践应以公平、正义、和合的价值取向为标尺,据以检视算法有无偏见、歧视以及有违公平公正的情形;在社会制度方面,算法审计实践须发挥社会主义审计制度在组织与治理上的特有优势,凸显审计在算法监管中的主导地位,并将“以人民为中心”思想贯穿审计实践全过程。综上所述,新时代中国特色算法审计应立足新时代中国发展需求和具体国情,坚持习近平新时代中国特色社会主义思想,将“以人民为中心”作为根本立场,依托总体国家安全观划定安全边界,贯彻落实新发展理念,以技术向善为伦理约束,以人类命运共同体为对外依托,以智能审计技术与方法为主导,对算法系统的设计、部署、运行及其社会影响进行评估、监督与问责,以此实现算法的可信可控、公平公正、安全合规与向善而行。

(三)新时代中国特色算法审计的本质特征

新时代,中国特色算法审计的本质特征可从政治、价值、治理三个维度加以阐释。其一是政治维度。新时代中国特色算法审计在党的领导下展开,服务国家战略与整体利益。从目标上看,算法审计以推进国家治理体系和治理能力现代化为方向,通过规范算法应用强化政府对数字经济的调控、对社会风险的预警以及对算法行政服务的优化;从实践上看,算法审计内容应与党和国家的中心任务相契合,譬如在推进乡村振兴战略实施中核验农业大数据算法的普惠程度,或是在科技创新中严格把控算法伦理的合规性,切实保障算法技术发展成果由人民共享。其二是价值维度。中国特色算法审计应将“算法是否增进人民福祉”作为判断标准,并聚焦算法对个体权益与公共利益的影响这两个方面:针对个体权益,算法审计须重点审查算法是否出现内含歧视、侵害个人隐私或削弱用户选择权的现象,从而推动其价值取向由效率优先转向公平与效率兼顾;针对公共利益,算法审计应关注算法在教育、医疗、养老等民生领域的普惠程度,审视算法在公共安全领域的应用是否合理,算法成效是否惠及全体人民。其三是治理维度。新时代中国特色算法审计须遵循“法治为基、德治为补”理念。一方面,算法审计以法治为根基,依据现行法律法规体系与行业标准展开取证;另一方面,针对算法技术超前于法律供给的领域,算法审计应以德治为柔性补充,通过引入社会主义核心价值观等标准审度算法是否违背公序良俗,并借助软约束弥补法律滞后造成的治理空缺。

(四)新时代中国特色算法审计与西方国家算法审计的本质区别

新时代中国特色算法审计与西方国家算法审计的本质区别具体表现为:(1)理论基础。我国算法审计根植

于“以人民为中心”思想以及总体国家安全观、新发展理念、技术向善、人类命运共同体等理论。西方国家算法审计植根于西方自由主义哲学、新自由主义经济学、技术中立性预设等理论。西方自由主义哲学认为,算法审计的目标是防止算法侵犯个体权利;新自由主义经济学认为,算法审计应依托企业“自律审计”和第三方机构的市场化服务,而非政府干预;技术中立性预设认为,算法风险源于技术漏洞而非社会制度设计,算法审计应聚焦技术层面的公平性修复。(2)领导体制。我国算法审计接受党和国家的统一领导,且与国家战略方向一致,服务于国家安全、社会公平和公共利益。西方国家算法审计多采用分权制衡模式,如美国立法型审计体制,审计机关直接向议会负责并报告工作,在这种体制下,审计机关与政府保持一定距离,以制衡行政权力。(3)价值基础。我国算法审计以“人民利益最大化”为价值基础,既保护个体隐私权利,又注重算法对社会公平、公共利益、国家安全的影响。而部分西方国家算法审计的核心是“保护个体隐私权利”与“维护市场效率”的二元平衡,常常忽视社会整体利益和弱势群体权益。例如,欧盟《通用数据保护条例》虽强化个人数据权利审计,但对因算法导致的“数字鸿沟”缺乏制约。(4)治理目标。我国算法审计目标是保障算法安全,防范系统性风险,注重集体利益和社会整体效益,引导算法维护社会公平,促进共享发展。而部分西方国家算法审计的目标是被动应对技术风险,在资本与权利之间寻求平衡,注重对个体权利保护、市场竞争以及技术创新的支持。

三、新时代中国特色算法审计的实践框架

在智能审计所提供的自动化检测、持续监控与智能分析支持下,算法审计突破了传统人工审计的效率边界,实现从事后抽样检查向全流程、动态化、智能化、穿透式的范式转变,使得构建面向算法全生命周期的系统化审计实践框架具备了现实的技术可行性。因此,在明晰“以人民为中心”的根本立场、“总体国家安全观”的安全边界、“新发展理念”的实践路径、“技术向善”的伦理约束、“人类命运共同体”的视野依托等理论要义后,亟须构建一套能够将上述理论内核系统化落地且适配智能审计技术生态的新时代中国特色算法审计实践框架。

本部分所构建的以“评估”“监督”“问责”为中心的三维一体算法审计实践框架正是连接理论原则与现实治理的桥梁,而且契合算法治理的全生命周期特性、新时代风险防控需求和中国特色治理逻辑。从算法治理的全生命周期来看,将新时代中国特色算法审计实践框架划分为“评估”“监督”“问责”三个维度,并非任意组合,而是根植于现代系统治理中的“前馈控制、过程控制、反馈控制”闭环逻辑。从新时代风险防控需求来看,算法风险已从技术层面延伸至社会、安全领域,单一维度难以应对复合型挑战,唯有“评估定基准、监督控过程、问责保底线”三维协同,才能实现对算法全链条、多维度的综合治理。从中国特色治理逻辑来看,上述三个维度呼应了“源头治理、过程治理、结果治理”的系统思维,既体现“防患于未然”的治理智慧,又强化“全程可控”的实践要求,更凸显“权责统一”的制度刚性,与西方侧重事后追责或单一合规审查的模式形成显著差异。

(一)新时代中国特色算法审计评估维度

1. 针对算法风险的审计评估

针对算法风险的审计评估是指运用多源异构数据融合、机器学习异常检测等智能审计技术,系统识别算法设计、开发、部署、运行等整个生命周期中可能引发的技术安全、法律合规与经济运行等风险,并通过量化分析、场景模拟、证据验证等手段,评估上述风险发生的可能性、影响范围及严重程度。针对技术安全风险评估,算法审计不应止步于评价算法技术本身是否完善,更须追问算法究竟为谁服务、能否真正利民惠民。上述考量在医疗、教育等民生领域尤为紧要,若算法偏见转化为差别对待,不同地区与群体享受均等可及的公共服务便无从谈起。针对法律合规风险评估,算法审计主体应依据现行法律法规对算法从设计到迭代的全生命周期开展合规性评估,确保算法活动于法有据、有章可循。针对经济运行风险评估,算法审计需要确保算法应用符合国家创新发展战略,促进公平竞争,防止算法垄断破坏市场秩序,保障经济稳定健康发展,如在电商平台的算法推荐中,防止通过算法操纵排名、限制竞争等行为扰乱市场秩序。

2. 针对算法环境适应性的审计评估

针对算法环境适应性的审计评估是指审计主体依托跨模态环境特征匹配校验、数据漂移与环境变化监测等智能审计技术,对算法系统在不同的法规政策环境、技术生态环境、社会经济环境下运行时所表现出的兼容性、可迁移性与动态演化能力进行测试与评价的过程。在法规政策环境适应性评估方面,审计主体需建立“映射 - 验证 - 响应”机制,首先将抽象的法规政策条款转化为可审计的算法参数;其次,验证算法全生命周期活动是否

契合相关政策要求或是否紧跟政策更新,评估算法在政策更新后的快速迁移能力;最后,评价审计客体的响应预案,确保算法活动与各类法规政策同频共振。在技术生态环境适应性评估方面,审计主体需围绕自主可控目标,首先评价算法对国产硬件与软件的兼容性与性能表现;其次,重点考察算法技术架构是否具备松耦合、可重构等特性,以支持其在国内外技术生态变化中的低成本平滑迁移与快速迭代;最后,评估算法系统是否具备应对供应链中断等极端场景的应急预案与动态演化能力。在社会经济环境适应性评估方面,审计主体需关注经济周期波动、产业结构调整及民生诉求变化等方面对算法决策的冲击,重点评估算法在就业、收入分配、区域协调及公共服务等领域是否会产生排斥效应或加剧不平等,并通过压力测试模拟极端社会经济场景,检验算法的公平性与普惠性,使其在动态复杂的社会环境中始终服务于增进人民福祉与国家长期稳定。

(二)新时代中国特色算法审计监督维度

1. 面向算法合规性的审计监督

面向算法合规性的审计监督是指依据《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《生成式人工智能服务管理暂行办法》《生成式人工智能服务安全基本要求》等中国特色社会主义法律体系、行业监管标准及审计客体内部管理制度,运用智能审计方法与工具,对算法设计、开发、部署、运行、维护等活动的合规性开展穿透式审查与监督。针对数据合规层面,首先审查算法所依赖的数据来源是否合法合规,是否遵循相关法律法规,是否体现“数据主权”“数据安全”“数据伦理”等新时代要求;其次,审查数据采集是否获得用户明示同意,是否符合最小必要原则,是否体现“以人民为中心”的思想;最后,审查数据使用是否与原始声明用途一致,是否存在数据滥用或算法歧视风险,是否符合中国特色社会主义法治体系对数据治理的规范要求。针对流程合规层面,首先确认合规性要求是否贯穿算法设计、开发、部署、迭代全流程;其次,审查算法设计阶段是否开展合规先行评估;再次,审查算法开发阶段是否建立跨部门协同机制,法务、风控、技术、伦理等多部门人员是否共同参与审核;最后,审查算法运营阶段是否建立算法可追溯机制,是否落实国家关于“算法备案”“算法日志留存”等监管要求。针对决策合规层面,一是审查算法在作出决策时,是否遵守相关法律法规,是否存在违反行业相关规定的情形,是否体现公平、公正、透明的原则[20];二是对于社保发放等民生领域,审查算法决策是否体现“兜底保障”的社会主义制度优势,坚决杜绝算法“效率至上”,预防民生空间被挤压。

2. 面向算法安全性的审计监督

面向算法安全性的审计监督是指运用科学的智能审计方法与工具,对算法全生命周期中涉及安全的机制、措施与能力进行系统性审查与验证的一系列活动。针对算法自身安全,首先应审查算法设计是否符合“最小权限原则”和“默认拒绝原则”,是否存在逻辑漏洞、权限越界、决策冲突等安全隐患;其次,应审查算法代码运行是否遵循国家推荐的安全编码规范,是否具备抵御缓冲区溢出、代码注入、越权访问等常见攻击的能力;最后,应关注算法设计是否嵌入“以人民为中心”的安全考量,如政务服务算法是否为残障群体保留“安全兜底”的人工干预通道,金融信贷算法是否存在基于残疾、户籍等因素的歧视性拒绝。针对算法抗攻击能力,一来应通过模拟对抗样本注入、模型窃取、数据投毒等攻击方式,全方位测试算法防御能力,如测试推荐算法是否具备识别与抵御“刷量攻击”的能力,或图像识别算法能否识别“对抗样本攻击”并作出合理响应;二来需秉承新时代“三化六防”网络安全领域指导思想,即以“实战化、体系化、正规化”为新理念,以“动态防御、主动防御、纵深防御、精准防护、整体防护、联防联控”为新举措,通过审计活动有效推进算法安全从被动响应转向主动防控。针对算法安全应急机制,一来应审查审计客体是否建立完善的算法安全事件应急响应机制,是否制定覆盖漏洞发现、评估、优先级排序、修复实施和修复验证等全流程应急预案,是否具备对数据泄露、模型篡改、服务中断等安全事件的快速响应与恢复能力;二来应关注算法安全应急预案是否包含技术自主替代方案,如当使用国外开源算法组件并出现漏洞时,能否及时启用国产化替代模块,以求通过审计监督强化“算法安全自主可控”。

3. 面向算法社会影响的审计监督

面向算法社会影响的审计监督是指运用科学的智能审计方法和工具,对算法应用引发的社会结构变迁、群体权益博弈及社会价值观异化等影响进行识别、检测、分析与预警的监督活动。针对算法群体影响层面,应审查算法是否服务于共同富裕、乡村振兴等国家战略,审查算法对性别、年龄、地域、收入等不同维度下社会群体的差异化影响,识别算法应用是否带有歧视,是否可能侵害弱势群体的合法权益。针对算法社会效应层面,算法审计一来应秉持“算法服务于社会和谐稳定”理念,依托用户调研数据、行业报告数据和舆情数据等信息,全面查验

算法对公共舆论、市场竞争、公共秩序、社会行为等方面的影响，如算法激励是否导致非理性消费，政务算法是否加剧“技术赋权下的基层负担”等；二来审计人员应特别关注算法应用对平台经济健康发展的潜在冲击，通过监督与促进职能的发挥，推动形成与社会主义市场经济体制相适应的公平竞争秩序。针对算法价值冲突层面，一来应以“社会主义意识形态”为根本遵循，审查算法的输出结果和应用方式是否符合社会主义核心价值观、中华优秀传统文化、公序良俗及公共利益导向；二来分析算法效率目标与社会伦理准则之间的冲突程度，推动算法设计在技术目标与社会价值之间实现平衡，如历史题材内容推荐算法是否坚守“历史唯物主义”导向，物流配送算法是否兼顾“服务效率”与“骑手安全”等。

4. 面向算法效能性的审计监督

面向算法效能性的审计监督是指依据明确的目标和标准，运用科学的智能审计方法与工具，对算法在特定应用场景下实现预期目标的程度、资源利用效率以及综合性能表现进行审查与验证的过程。针对算法效能指标测试层面，一是围绕国家战略导向和公共服务目标，设定具有中国特色的算法效能指标体系，如在政务服务领域，重点关注算法审批的准确率、流程压缩时长、群众满意度等体现“放管服”改革成效的指标；二是基于效能指标体系，对算法开展在不同负载条件下的性能测试，查验算法在不同负载下的目标达成度及执行效率。针对算法资源消耗分析层面，一是对算法运行中的资源消耗情况进行评价，包括 CPU、内存、磁盘使用率、能耗比、迭代优化成本等；二是推动算法设计向“低功耗、轻量化、国产化”方向发展。针对算法可扩展性测试层面，模拟数据量增长、业务规模扩大、跨区域协同等典型场景，评估算法在面对大规模数据处理、高并发请求、多主体协同等复杂环境下的性能表现。在新时代“系统治理”理念指导下，对算法进行可扩展性测试，需特别关注其是否具备服务全国统一大市场、服务多民族地区的能力，能否在不同人群、不同网络条件下保持稳定高效输出。

5. 面向算法鲁棒性的审计监督

面向算法鲁棒性的审计监督是指依据既定的标准、规范以及特定业务场景需求，通过敏感性分析、压力测试、异常场景模拟等智能审计方法，对算法在面对输入数据异常、环境条件变化、恶意扰动等干扰因素时的稳定性、适应性、抗干扰性进行审查与验证的监督活动。针对算法内部数据扰动测试层面，应通过向算法输入经过扰动的数据，如添加噪声、修改数据分布、引入异常值等，观察算法的输出结果和性能变化，查看算法对输入数据变化的敏感程度以及能否正确处理不完整、不准确或恶意篡改的数据。在测试中，需要特别关注算法在公共服务、民生保障、社会治理等关键领域是否具备抵御“数据陷阱”“信息误导”的能力，防止因输入数据异常导致对弱势群体的误判、歧视或服务中断，确保算法真正服务于人民利益。针对算法外部环境变化测试层面，应模拟算法在实际应用中可能遇到的外部环境变化，如网络延迟、硬件故障、系统资源不足等，测试算法在上述情况下的稳定性和容错能力，确保算法能够在不利的外部条件下持续提供可靠服务。特别是在政务、医疗、教育、能源等涉及国家安全与民生福祉的关键领域，还应深度关注算法系统是否具备“高可用性”和“高适应性”，以保障算法系统在极端环境下仍然能够保持可持续服务能力。

（三）新时代中国特色算法审计问责维度

算法审计问责是指在出现算法引发歧视、偏见、侵权、安全风险或社会危害时，以保障算法应用的合法性、安全性、公正性、效能性、鲁棒性为目标，通过责任溯源与追责问责，依法依规追溯算法开发者、算法运营者、算法使用者等相关主体的责任，并推进责任整改与风险闭环管控的一系列过程。

1. 责任溯源

针对算法责任溯源，审计主体一是依托算法全生命周期，全面拆解责任节点，在设计开发阶段追溯“技术缺陷”与“伦理合规”责任，在部署运行阶段追溯“场景适配”与“过程管控”责任，在迭代升级阶段追溯“变更透明”与“影响评估”责任，在风险处置阶段追溯“应急响应”与“损失补救”责任，进而实现从技术链到责任链的穿透；二是基于技术、制度、价值层面，厘清每一具体层面所面临的问题。从技术层面看，需解决三个问题，即谁开发的代码、谁做出的决策以及谁应该担责。从制度层面看，需解决四个问题，即算法责任主体的法定边界、全生命周期的算法责任分配、算法责任认定的法定标准以及算法责任溯源证据的法定效力。从价值层面看，需解决五个问题，即算法责任溯源是否符合公平正义原则、是否体现技术向善与社会价值的统一、是否尊重和保护弱势群体的权益、是否体现公众参与和民主监督的价值以及是否服务于广大人民群众的根本利益。

2. 追责问责

当算法开发者、运营者、使用者等主体违反法律法规、技术标准和伦理规范时,需从党政联合问责、技术追责、民生补责三个层面开展追责问责。针对党政联合问责,以审计为主导的党政联合问责流程分为四步:首先,当审计机关发现算法应用存在严重违规行为,特别是涉及违反党的政策、国家重大战略部署或国家安全等问题时,应及时启动联合问责程序,将初步情况提交至党委和政府;其次,党委和政府根据审计汇报内容,迅速组织审计机关、纪检监察机关、司法机关、行业主管部门等成立联合问责小组;再次,问责小组进一步核实审计证据,准确界定算法责任主体和类型;最后,党委和政府依据调查结果,对党政机关、国有企业及其负责人等追责问责对象,综合运用通报批评、党纪处分、行政处分等手段实施政治问责。针对技术追责,若因算法技术问题而引发责任事故,审计机关应同其他问责主体一起,对直接参与算法设计、开发、测试的技术团队成员及技术管理者等进行批评教育或暂停相关业务权限。针对民生补责,审计机关应敦促相关责任主体对算法的实际使用者或受影响的用户群体进行民事补偿,补偿方式包括但不限于经济损失补偿、精神损害赔偿、公共服务改进等。

四、新时代中国特色算法审计的深化路径

在智能审计技术深度赋能算法审计作业的情形下,算法审计的深化路径既须立足本土实践,筑牢自主根基,又须放眼全球治理,贡献中国方案。智能审计所提供的自动化检测、持续监控与智能分析能力,为算法审计的技术创新、制度优化与文化建构注入新动能,也为中国算法审计走向国际舞台提供了技术支撑与范式参照。

(一)构筑本土化根基,建立基于技术、制度、文化层面的创新机制

算法审计本土化根基的构筑根植于制度逻辑、实践逻辑和安全逻辑三个层面。我国算法审计工作应以适配中国制度基因、实践生态与安全需求为前提,深度复用智能审计领域的成熟建设成果,通过建立基于技术、制度、文化等层面的协同创新机制,形成扎根中国土壤、服务国家治理、彰显中国价值的算法审计体系。算法审计本土化既拒绝简单套用西方审计框架,也反对无视国外有益经验的闭门造车,而是以“解决中国问题、保障人民利益、维护国家主权”为根本导向,实现算法审计与中国特色社会主义发展的深度契合。

1. 算法审计技术本土化创新

算法审计技术本土化创新是指以自主研发、跨界攻关、适配改造为关键路径,构建自主可控、场景适配、前瞻引领且契合本土算法应用场景与治理目标的技术体系。智能审计领域多年沉淀的成熟工具与分析方法应当有针对性地迁移至算法审计的评估与监督环节。可供自主研发与跨界攻关的算法审计技术有很多,如算法输出一致性检测技术、算法公平性校验技术、鲁棒性测试技术等。在技术研发与攻关中,审计人员首先应锚定本土场景的技术痛点,明晰风险识别、合规验证、伦理评估等维度的现实需求;其次,应确定技术开发方向,以跨学科融合驱动技术迭代创新;再次,可选取政务、民生等典型领域开展试点,采集应用反馈数据,进一步优化技术特性;最后,待技术成熟后,制定相应技术标准,在全国范围内推广,并保持持续监控与迭代更新。至于外来技术的适配改造,需要搭建本土化“检测工具库”,工具库至少涵盖四类:(1)算法模型透明性检测工具,用以剖析“黑箱”模型的决策逻辑,支撑责任认定与合规性审查,如 LIME、SHAP 等;(2)公平性与偏见检测工具,用以甄别算法的公平性缺陷,支撑伦理审查与公平性校验,代表工具有 Fairlearn、Themis-ML 等;(3)鲁棒性与安全性检测工具,用以评估模型在对抗性攻击、数据扰动下的运行稳定性,支撑技术安全审查,如 Foolbox、ART 等;(4)数据溯源与合规检测工具,用以验证训练数据的合法性、隐私保护与数据处理合规性,如 TensorFlow Privacy 等。检测工具库的构建以及外来技术适配改造,不能仅停留在技术能力层面,更应锚定我国创新驱动发展、国家安全等战略目标,推动算法审计由技术验证迈向治理赋能。

2. 算法审计制度本土化创新

算法审计制度本土化创新是指审计主体立足我国现行法律法规体系、社会治理模式与技术产业生态,对算法审计配套的政策法规、组织结构与操作流程进行系统性重构、调整与优化,使之深度适配我国现实国情和社会主义核心价值观。国内的智能审计经过长期实践,已经积累了大量可借鉴的制度经验,这些经验成果可为算法审计制度建构提供参考,但算法审计面向的审计客体为算法模型,算法审计制度需要结合算法规制的特有规律进行再造,不能照搬智能审计既有的制度框架。在政策法规层面,现阶段国内专门面向算法审计的法规、标准和规范极为稀缺。这就要求审计主体持续跟进国家关于算法治理的新政策、新规定,及时修订不适应当前算法审

计实践需求的规则，不断完善算法审计政策、法规和制度，有针对性地制定特定行业、领域内的算法审计指南和规范。在组织结构层面，算法审计宜采用国家审计、内部审计、社会审计多元主体协同共治的组织体系。其中，国家审计机关需在中央审计委员会的统一领导下，依法履行职责，着力强化对涉及公共利益和国家安全等关键领域算法的强制性审计，依法调取算法参数、训练日志等关键资料，并就重大违规情形提出问责建议；内部审计部门是推行算法审计的中坚力量，其优势是便于深入算法内部测试模型算法的训练过程，审查算法参数选择，从而精准排查算法系统的风险隐患；社会审计机构可承接政府或企业自主委托的审计业务，能够在一定程度上缓和算法开发者与监管者之间的利益冲突，其结果可为政府监管与企业合规提供重要参考。在操作流程层面，算法审计须坚持分级分类原则，根据算法的风险等级、社会影响范围及技术复杂度，对审计对象实施差异化监管[21]。对高风险算法实行全流程穿透式审计，审查数据来源，验证模型训练过程，分析输出结果；对影响局部群体权益但风险可控的中风险算法执行标准审计程序，将审计力量集中于关键环节；对仅涉及个体微小权益或技术透明度较高的低风险算法，可采取自动化检测或抽样测试等简化流程，辅以定期监控即可实现有效监督。

3. 算法审计文化本土化创新

算法审计文化的本土化创新根植于中华优秀传统文化，是以社会主义核心价值观为灵魂、以“中国式现代化”目标为方向，在审计实务中重塑算法审计的价值取向、执业规范与社会共识，从而形成适配我国算法治理场景、能够化解中国式算法治理难题的特色审计文化体系。在价值理念层面，算法审计文化建设须以社会主义核心价值观筑牢文化根基，深度挖掘中华传统治理哲学与人文智慧，适配现代算法治理下的伦理准则与实践范式。在行为准则层面，算法审计制度设计须将抽象的本土价值理念，如以人为本、公平正义、民主文明、诚信爱国等，转化为可落地、可约束、可评估的具体行为规范，力求从本土价值体系中提取算法审计行为准则的“标尺”，促使算法审计主体在算法审计实践中有章可循。在社会共识层面，行业与监管部门可选取国内典型的算法争议案例作为切入点，依托媒体宣传、基层宣讲、公众座谈等传播载体，向社会公众普及算法审计的内涵，同时阐释算法审计的重要性及其影响，以此增进社会公众对算法审计工作的理解、认同与支持。

（二）推动国际化发展，贡献中国智慧与中国方案

算法审计在全球化算法治理中扮演着至关重要的角色。放眼全球，智能审计的国际标准、跨国审计协作机制不断发展，为算法审计开展国际对话奠定了很好的沟通基础。但算法审计兼具算法规制、数字伦理的特殊属性，不能简单照搬全球智能审计的规则体系，需要立足我国本土化算法审计建设成果，走出一条适配数字文明新形态的国际化发展道路。中国算法审计应在本土化创新的基础上，通过规则输出、标准引领、案例示范和国际协作，积极参与并引领全球算法审计规则制定，推动构建“技术自主可控、规则包容共治、文化多元共生”的国际算法综合治理新秩序，为全球数字社会可持续发展贡献中国智慧与方案。

1. 规则输出

中国算法审计的国际化需从“被动适应国际规则”转向“主动参与规则制定”，将自身在算法审计领域的制度规范、监管逻辑、价值导向等转化为具有国际影响力的规则性成果，并通过多元渠道推动相关规则在国际社会得到认可、采纳或参考。对此，一来我国应对接国际算法治理平台，扩大中国算法审计规则的影响力。一方面，可积极参加 ISO、IEEE 等相关国际组织有关算法审计规则的制定工作，提出具有建设性的算法审计规则建议；另一方面，在联合国教科文组织人工智能伦理框架、G20 数字经济发展与合作倡议中，将中国算法审计规则纳入国际规则文本，以为其他国家提供有益参考。二来我国可与其他国家签订双边或多边协议，在特定领域内稳妥推行符合多方共同利益的中国算法审计规则，为全球算法治理提供多元视角，贡献中国智慧，着力构建具有中国特色的国际规则生态。2023 年 10 月，我国同古巴等 13 国共同发布《“一带一路”数字经济国际合作北京倡议》，同阿富汗等 34 国签署《数字经济和绿色发展国际经贸合作框架倡议》，为中国算法审计规则走向国际提供了重要契机。如果能够将自主算法审计规则体系嵌入“一带一路”数字经济合作框架，或与沿线国家另行签订双边、多边协议，则既可共同防范“数字丝绸之路”中的人工智能算法风险，也有助于缩小沿线国家数字鸿沟，推动数字资源共建共享。

2. 标准引领

中国算法审计标准之国际化，须扎根国内本土化创新实践，通过主导或深度参与国际标准的制定、推广与落地，将流程规范、合规要求、风险评估框架等具有中国特色的算法审计准则转化为具有国际共识或广泛影响力的

行业基准,使算法审计的实施路径、研究方法与治理智慧融入全球数字治理体系。至于如何推进算法审计标准引领,首先,应构建一套既契合国内要求、又能回应国际需求的算法审计标准体系;其次,积极投身国际标准化组织,广泛参与相关标准的讨论与制定,提交具有科学性与可操作性的提案,助力更多中国算法审计标准获得国际认可并融入国际标准体系;最后,编撰并对外发布有关算法审计的技术报告与白皮书,或定期举办国际研讨会与论坛,清晰界定中国标准同国际主流标准的异同及相互转换路径,凸显中国算法审计标准的优势与实际价值,进而吸引全球组织自发采用中国算法审计标准开展本国算法治理。

3. 案例示范

中国算法审计的案例示范是以本土或跨境场景中算法审计的实践案例成果为载体,通过对其进行系统性梳理、标准化封装、场景化传播和国际化应用,将中国在算法审计中的技术方案、伦理原则、制度设计转化为可复制、可验证、可对话的全球性参考模板,进而打破国际社会对中国技术治理模式的认知壁垒,推动中国算法审计规则、标准与协作机制被国际社会主动采纳或借鉴。关于推进案例示范的做法,首先,应选择具有代表性的行业或应用场景,实施高标准的算法审计项目,研发可供国际借鉴的优秀案例成果;其次,编制案例白皮书,将技术细节、操作规程、实际效果等转化为国际通用的方法论语言,并开发“算法审计案例工具箱”,配置审计模板、技术工具包和线上教程等,让案例经验可供其他国家借鉴;再次,在联合国人工智能治理论坛、金砖国家数字经济会议等各种国际平台,或依托 IEEE 算法治理委员会、全球数字经济联盟等各类国际行业组织,强化中国算法审计案例在世界各国中的示范指导作用;最后,跟踪相关案例在其他国家的实施情况,收集国外用户的反馈及建议,进一步优化案例的普适性。

4. 国际协作

我国算法审计主体有必要与国际组织、跨国企业及全球科研机构等主体建立广泛而深入的合作关系,在政策制定、技术研发、模式探索、人才培养及产业合作等层面实现资源互补、经验互通、技术共创、能力共建与成果共享。推进算法审计国际协作需要从两方面入手:一是建立广泛的国际协作渠道,提升中国算法审计在全球的影响力和认可度。例如,积极加入国际算法审计相关组织,与国际同行建立合作关系,定期举办国际会议、论坛或研讨会,促进国际学术交流与合作。二是构建深入的国际合作框架,通过签署合作协议或备忘录,促使算法审计在实践、研究及培训等领域开展实质性合作。例如,共同开展跨国算法审计实践活动,协同解决跨国算法应用以及跨境数据流动中的跨域问题;再如,与国外科研机构、高校、企业共同建立联合研发团队,攻克算法审计关键技术;还如,举办跨国培训活动,提升各相关方从业人员的技术水平。中国可通过与其他国家开展多方面的协同合作,在国际算法治理中更高效地发挥建设性作用,从而实现互利共赢。

五、结束语

自 2022 年 8 月我国实施算法备案制度以来,国家网信办已陆续公布 19 批共计 1003 个境内互联网信息服务算法,以及 18 批共计 7718 个境内深度合成服务算法。可以预见,随着智能技术与经济社会各领域加速融合,未来纳入备案管理的算法数量仍将快速增长。作为以防范和规制算法风险为核心的新兴治理领域,有关算法审计的理论研究与实践探索目前仍处于起步阶段。西方国家多以技术中立和个体权利为价值取向,并据此形成自身的算法审计范式,上述做法难以充分适应我国超大规模社会治理情境下统筹发展与安全、兼顾效率与公平的现实需求。因此,立足中国国情,构建融政治性、人民性与专业性于一体的中国算法审计自主知识体系和实践框架,不仅是算法治理领域亟待深入研究的重要学术议题,也是推进数字中国建设、提升数字治理现代化水平所面临的重要时代命题。基于此,本文系统阐释智能审计赋能下新时代中国特色算法审计的理论内核、实践框架与深化路径,力求突破既有西方算法治理范式的路径依赖,始终坚持“以人民为中心”的根本立场,通过“本土化创新筑牢根基、国际化发展贡献方案”双轮驱动,共同为我国数字治理领域算法制度建设提供具有可操作性和实践支撑力的审计方案。

参考文献:

[1]张永忠,张宝山. 算法规制的路径创新:论我国算法审计制度的构建[J]. 电子政务,2022(10):48-61.

[2]张超. 资讯类推荐算法的算法审计路径、伦理与可审计机制[J]. 中国出版,2021(7):31-35.

[3]Raji I D,Buolamwini J. Actionable auditing:Investigating the impact of publicly naming biased performance results of commercial AI products[C]. Proceedings of the 2019 AAAI/ACM Conference on AI,Ethics,and Society. New York:ACM,2019.
[4]张欣,宋雨鑫. 算法审计的制度逻辑和本土化构建[J]. 郑州大学学报(哲学社会科学版),2022(6):33-42.
[5]Goodman E P,Trehu J. Algorithmic auditing:Chasing AI accountability[J]. Santa Clara High Tech Law Journal,2023,39(3):289-338.
[6]Schwalbe U. Algorithms,machine learning,and collusion[J]. Journal of Competition Law and Economics,2018,14(4):568-607.
[7]师文,陈昌凤. 平台算法的"主流化"偏向与"个性化"特质研究——基于计算实验的算法审计[J]. 新闻记者,2023(11):3-14.
[8]胡耘通,刘婧娴. 算法审计治理:价值逻辑、现实挑战与体系构建[J]. 财会月刊,2025(24):69-74.
[9]Le Merrer E,Pons R,Trédan G. Algorithmic audits of algorithms,and the law[J]. AI and Ethics,2024,4(4):1365-1375.
[10]Birhane A,Steed R,Ojewale V,et al. AI auditing:The broken bus on the road to AI accountability[C]. 2024 IEEE Conference on Secure and Trustworthy Machine Learning(SaTML),IEEE,2024.
[11]史普润,刘晓红,陈杰. 数字平台治理的算法审计融合机制研究——基于新质生产力的视角[J]. 中国注册会计师,2026(1):83-92.
[12]梁力军,陈国渔,张婷莉菲. AI 技术融合下审计伦理的"变"与"不变":内核坚守与框架适配[J]. 财会月刊,2026(13):78-84.
[13]张宸瑜. 基于公共性算法制度建构的新质生产力治理——以新一代生成式人工智能参与数字政府建设为例[J]. 南京审计大学学报,2024(4):97-111.
[14]贺勇,刘筱祎,张珺. 数字平台算法审计:理论框架与实务建议[J]. 审计与经济研究,2026(2):20-31.
[15]崔永梅,杨婷羽,应文池,等. 人工智能审计的理论内涵、问责边界与框架构建——基于负责任创新视角[J]. 审计与经济研究,2025(4):11-22.
[16]刘国城,徐天,缪倩云. 公共治理领域中算法应用的风险样态与审计治理[J]. 财会通讯,2025(9):114-121.
[17]贺勇,李佳蔚,刘筱祎. 数字平台算法审计:现实理据、客观挑战与关键进路[J]. 南京审计大学学报,2025(2):10-20.
[18]师文,程柯银. 算法审计的国际传播视角:框架与研究路径[J]. 传媒观察,2026(2):40-51.
[19]张宝山,张永忠. 从概念到制度:算法审计问责体系研究[J]. 南京审计大学学报,2026(2):34-44.
[20]吴勇,张超,武子豪,等. 深度学习模型的审计应用:工作机理、功能拓展与流程框架[J]. 会计与经济研究,2025(5):15-37.
[21]徐京平,朱沛雯. 数据合规审计的实践向度与风险因应——一个概念建构的分析框架及其考证[J]. 审计与经济研究,2026(4):30-41.

[责任编辑:刘　茜]

Research on the Construction of a New Era Chinese Characteristics Algorithmic Audit System under the Empowerment of Intelligent Auditing

LIU Guocheng
(School of Accounting, Nanjing Audit University, Nanjing 211815, China)

Abstract: Against the background of the intelligent transformation of the audit industry and the deep integration of algorithmic technology into all fields of society, building an algorithmic auditing system with Chinese characteristics in the new era is of great significance for preventing and resolving algorithmic risks, improving the level of national governance modernization, and laying a solid foundation for the construction of Digital China. Intelligent auditing relies on its own development and accumulation of technical tools and methodologies such as real-time monitoring, automated detection, and data penetration analysis to provide important enabling conditions for the implementation of algorithmic auditing. This article focuses on the construction of an algorithmic audit system with Chinese characteristics in the new era under the empowerment of intelligent auditing, taking theoretical connotation, practical framework and deepening path as the main lines, it first analyzes the basic connotation and essential characteristics of algorithmic auditing with Chinese characteristics in the new era, and identifies the essential differences between China's algorithmic auditing and Western countries. Secondly, it establishes a three-dimensional practice framework for algorithmic auditing with Chinese characteristics with "evaluation", "supervision" and "accountability" as the core. Thirdly, it proposes a development path driven by the two wheels of "local innovation" and "international leadership". This research aims to provide theoretical support and practical guidance for improving the national digital governance system in the new era and improving the global algorithm audit mechanism.

Key Words: algorithmic audit; independent knowledge system of auditing; intelligent audit; penetrative audit; digital governance; algorithmic governance; national governance