

# 数字平台算法审计:理论框架与实务建议

贺 勇<sup>1</sup>,刘筱祎<sup>1</sup>,张 珺<sup>2</sup>

(1. 湖南工商大学 会计学院,湖南 长沙 410205;2. 湖南理工大学 经济与管理学院,湖南 岳阳 414006)

**[摘要]**数字平台是数字经济最成功的组织模式,占据着绝对重要的位置。作为数字经济发展的核心驱动力和底层架构,算法如同一把“双刃剑”,在推动数字平台飞速发展的同时也伴生诸多算法风险。算法审计是数字治理的重要工具,对促进数字平台健康发展具有重要作用。以数字经济健康发展为导向,构建数字平台算法审计理论框架和实务路径,将算法作为审计对象,系统分析了数字平台算法审计的预防、监督、评价和抵御机制,从数字平台治理、数据、模型和应用维度阐释了数字平台算法审计的实务挑战和关注要点、关键审计问题和程序。数字平台算法审计拓宽了审计的领域,为审计学科的多元融合发展提供了新的方向,丰富了我国数字经济治理和审计学自主知识体系。

**[关键词]**数字经济;数字平台;算法风险;算法审计;数字治理;算法治理;审计监督;中国审计学自主知识体系

**[中图分类号]**F239.43 **[文献标志码]**A **[文章编号]**1004-4833(2026)02-0020-12

## 一、引言

算法作为数字平台的核心技术支撑,如同一把“双刃剑”,在促进数字经济繁荣的同时也衍生出一系列算法失当和社会责任缺失的风险,如,算法操控、算法歧视、算法黑箱、算法共谋等,已经深度冲击数字经济发展进程中的市场良性竞争与社会管理秩序,甚至给国家安全与社会稳定带来了诸多隐患和挑战。习近平总书记强调“推动数字经济健康发展,要坚持促进发展和监管规范两手抓、两手都要硬,在发展中规范、在规范中发展”<sup>[1]</sup>。数字平台是数字经济的主要组织形态,具有公共物品属性,其高质量发展代表着数字经济的高质量发展<sup>[2]</sup>。

作为数字时代的重要治理工具,算法审计在揭示数字经济发展进程中深层次矛盾、化解数字平台技术异化的现实问题、契合消费者对健康数字生活的需求等方面发挥不可替代的作用。算法审计对于算法全生命周期的治理进路契合数字经济健康发展理念。我国《“十四五”数字经济发展规划》指出要健全并完善数字经济治理体系,强化协同治理和监管机制,构建多元共治的新格局。本文以数字经济健康发展为导向,构建数字平台算法审计的理论框架,进一步提出可操作的审计关键问题与程序,为数字平台算法审计从理论探讨迈向实务操作提供参考,丰富了中国审计学自主知识体系在数字时代所具有的理论内涵和方法论工具。

## 二、文献回顾

### (一)数字平台算法风险治理路径与困境

#### 1. 数字平台算法风险的治理路径

作为一种重要的基础设施,数字平台已然成为数字经济发展的关键主体<sup>[3]</sup>,其提供的各类产品以及服务已经渗透到公众生活的诸多不同领域,从而呈现出公共物品属性<sup>[4]</sup>。算法是数字平台的核心技术,使数字平台具有“双刃剑效应”<sup>[5]</sup>。因此,数字平台算法风险成为数字治理的核心议题之一。如何治理数字平台的算法风险?答案不在于是否要监管,而在于如何监管<sup>[6]</sup>。从实践看,数字平台算法治理有三种模式,即自我治理、外部治理和共同治理。其一,在自我治理模式下,透明度通常是自愿的,外部监督最小化。自我治理的途径主要包括技术变革、透明度努力或两者的某种组合<sup>[7]</sup>。但自愿依赖善意,而且追索权受到限制<sup>[6]</sup>。其二,外部治理主要是政府干预,包括三类政策措施:实施全面的隐私和数据保护法规、否定中介责任保护以及反垄断促进竞争<sup>[8]</sup>,但有

**[收稿日期]**2025-04-22

**[基金项目]**国家社会科学基金项目(22BJY080);湖南省自然科学基金项目(2024JJ5116);湖南省研究生科研创新项目(CX20240961)

**[作者简介]**贺勇(1976—),男,湖南安化人,湖南工商大学会计学院副教授,硕士生导师,从事数字经济与公司治理研究,E-mail: sternman@126.com;刘筱祎(2000—),女,湖南衡东人,湖南工商大学会计学院硕士研究生,从事数字资产会计研究;张珺(1996—),男,湖南岳阳人,湖南理工大学经济与管理学院讲师,博士,从事公司金融研究。

学者指出,监管政策是碎片化的<sup>[9]</sup>。由于监管严重不足,数字平台公司在全球范围内拥有几乎不受约束的权力<sup>[10]</sup>。其三,共同治理包括成立某种可履行多种职能的组织,从调查用户投诉到为平台公司创建道德框架,通过提供透明度报告、第三方审计和其他问责机制对平台公司的行为进行审查<sup>[11]</sup>。审计规制算法权力不仅具有显著的制度优势,同时还具有突破资源、时空、传统的技术优势<sup>[12]</sup>。

## 2. 数字平台算法风险的治理困境

第一,算法的“黑箱”特性导致认知困境。一方面,外部监管和用户难以理解算法内部的运作逻辑,算法公平性与合理性的有效审查难以展开<sup>[8,13]</sup>;另一方面,算法在持续不断地自我学习和动态优化,带来所谓的“规制滞后”情况<sup>[8]</sup>,监管指标和合规要求失去了时效性,算法治理只能流于形式。第二,算法归责困境。从算法开发到应用的全生命周期,算法责任究竟应当归于算法自身,还是归于开发者,抑或平台公司,学术界尚无共识<sup>[14]</sup>。特别是数字平台公司常以算法自动决策和技术中立为盾牌,规避承担算法责任<sup>[15]</sup>。法律层面的归责原则要求有主观过错,这导致证明平台公司存在主观意图极为困难。第三,算法风险的治理工具存在局限。算法应用场景千差万别,相关法律若过于宽松则无效,若过于严格则扼杀创新<sup>[16]</sup>。在数字平台逐渐形成寡头垄断格局的情况下,算法信息所产生的认知桎梏阻碍了个体理性的思考,用户往往缺乏真实的选择权<sup>[17]</sup>,即使算法伦理原则大量涌现,但由于缺乏落地实施机制,最终还是沦为数字平台进行“伦理洗涤”的工具<sup>[18]</sup>。因此,对数字平台算法风险的治理不仅是一种能力,更是一个跨越不同参与者和行为的特定而复杂的互动网络<sup>[6]</sup>。

### (二) 算法审计的内涵与作用

#### 1. 算法审计的内涵

算法审计的概念最早由 Sandvig 等在研究谷歌平台广告推送算法可能存在的偏见时提出<sup>[11]</sup>。Gaddis 表示算法审计是测试人工智能是否存在盲点或偏见的一系列技术的集合<sup>[12]</sup>。Brown 等从更宽泛的角度进行了研究,并认为算法审计是收集算法在特定环境中的行为数据,据此评定这种行为是否给人们的权益带来了负面影响<sup>[19]</sup>,如风险评估或信用评分等<sup>[20]</sup>。就面部识别而言,算法审计不仅关注偏见而且关注算法是否被滥用<sup>[21]</sup>。Koshiyama 则把算法审计细分为开发、评估、缓解以及保证四个部分,算法审计是针对算法安全性、合法性以及伦理性展开的评估、缓释和保证的研究与实践<sup>[22]</sup>。Vijayakumar 指出,算法透明性是审计研究的一个重要但非唯一的内容<sup>[23]</sup>。Lee 提出,算法审计可以被视作是对前件、算法与后果三个相互紧密关联元素的检查<sup>[24]</sup>,能够建立起算法和结果两者之间的因果关系<sup>[12]</sup>。Shen 等则提出日常算法审计是用户持续地对算法的边界进行检测,受到算法侵害的团体会形成共同意志的集合,一起验证并抵御算法偏见<sup>[25]</sup>。

#### 2. 算法审计的作用

社会对算法的依赖增加了对算法审计的需求<sup>[11]</sup>,越来越多的算法决策研究要求对算法的实践活动进行公共审查<sup>[26]</sup>,以实现算法公平性、透明性和问责性<sup>[27]</sup>。荷兰审计院通过对其中中央政府使用的算法进行审计后表示,算法对于独立审计师来说并不是一个“黑匣子”,审计可以且有必要检查和评估所有算法<sup>[28]</sup>。算法审计已成为从不透明平台算法中收集见解的一种重要方法,可以有效防止算法滥用、管理算法运行、打开算法黑箱<sup>[29]</sup>,同时平衡自动化决策中各方的利益冲突,增强公众信任<sup>[30]</sup>。

对于如何应用算法审计,Shen 等强调在日常行为中算法审计的情境性,用以理解用户与算法系统的日常互动<sup>[25]</sup>。Brown 等从利益清单、伦理道德和相关性维度构建了一套针对算法性能评价的审计框架<sup>[19]</sup>。Koshiyama 主张通过数据和任务设置、特征预处理、模型选择、稳健性处理及报告、生产和部署五个阶段充分审查算法的可解释性、稳健性、公平性和隐私保护这四个关键领域<sup>[22]</sup>。Baxter 指出,进行算法审计不仅是检查代码,而且要充分了解算法的意图并彻底审查业务流程,确保控制措施的覆盖程度与风险关键程度保持一致<sup>[13]</sup>。Lee 提出了一个反事实算法审计框架,以确定可能导致某些理性或非理性决策的算法决策因素<sup>[24]</sup>。Bandy 综合 62 项面向公众的算法系统的审计研究,展示了算法审计如何澄清并帮助解释技术与社会之间的关系<sup>[31]</sup>。

### (三) 研究述评

综上所述,数字平台作为数字经济的关键主体,肩负着巨大的社会责任。算法作为数字平台的核心技术,其在政治传播、社会治理和科技伦理等多个领域所形成的治理困境不容小觑。正所谓无审计不治理,算法审计对于实现数字平台算法公正具有重要意义,国外对于算法审计的研究处于前沿阶段,数字平台算法审计仍是国内学术研究的“富矿地带”<sup>[32]</sup>。从现有文献来看,以下几点有待进一步研究:第一,理论研究尚需完善,算法审计适

应当前我国数字经济健康发展理念,但算法审计作为一个交叉学科融合的新领域,如何与现有审计理论相衔接还需深入探讨,本文将着力构建数字平台算法审计的理论框架;第二,算法审计促进数字平台健康发展的机制还需明确,算法审计如何应用到数字平台的日常监管中,如何兼顾审计活动的全面性和算法的机密性,本文将深入阐释算法审计的作用机制;第三,数字平台算法审计实务路径有待明确,如何开展数字平台算法审计以及形成规范的算法审计实务操作都需要进一步厘清,本文将从多个维度提出数字平台算法审计的实务难点、关键审计问题和程序。

### 三、数字平台算法审计构成要素

#### (一)数字平台算法审计本质

数字平台作为数字经济的核心载体,其算法不仅是技术工具,更嵌入了平台的商业逻辑、用户关系与社会影响。数字平台算法审计的本质并非单一的技术验证,而是具有鲜明的场景特征。从微观角度来看,数字平台的双边市场属性导致算法信息不对称具有双向性,既存在用户对算法推荐逻辑的未知,也存在商户对平台流量分配规则的模糊。因此,数字平台算法审计要同时消除用户层与算法层以及商户层与算法层这两类信息不对称,增强算法的透明性和可解释性,保障算法安全与技术伦理。从宏观层面看,数字平台的公共属性使算法风险具有外溢性,可能引发市场垄断与舆论极化等系统性问题。算法审计是推动国家治理现代化和法治化进程的一项重要工具,不仅要针对技术漏洞采取行动,而且要抵御数字平台权力滥用给市场秩序和社会伦理带来的冲击。数字经济健康发展的关键在于多元主体共同参与治理,算法审计的主要目的在于预防、揭示与抵御数字经济运行过程中出现的各类风险,发挥监督作用提升数字平台在运行期间的抗风险能力,从而确保数字经济安全且高质量地运行,进而维护国家及公众利益。综上,数字平台算法审计的本质是技术验证与伦理审视所构成的双螺旋结构,发挥预防、监督、评价及抵御作用,制衡数字平台算法权力,防范可能出现的算法系统性风险,重塑公众信任,促进数字平台开展负责任的创新和数字经济健康发展。数字平台算法审计并非限制数字平台及其算法技术的发展,而是发挥数字经济“看门人”角色,维护数字时代人的尊严和权利。

#### (二)数字平台算法审计目标

数字平台算法审计的关键目标是系统评估算法从设计到训练、部署、输出乃至其社会影响,确保算法契合公平性、透明性、安全性以及合规性等要求,促使数字平台切实履行算法治理的主体责任,开展负责任的创新,平衡技术进步、社会伦理以及公共利益,进而实现可持续发展。数字平台凭借自算法权力,往往会引发诸如大数据“杀熟”、算法歧视、信息茧房和算法垄断等一系列问题,严重侵害公众的平等权。所以,数字平台算法审计充当“看门人”的角色,借助技术增信与行为监督方式,遏制算法权力滥用,重构公众与数字平台之间的权利平衡。

数字平台作为数字经济中的一种新型责任主体,其算法行为具有显著的外部性和公共属性,因此数字平台算法审计目标要切实考虑数字经济发展要求,评估和验证受托责任履行情况。《“十四五”数字经济发展规划》提出,我国数字经济的发展目标之一就是数字经济治理体系要更加完善,建立协调统一的数字经济治理框架和规则体系。宏观层面,数字平台算法审计以国家数字经济发展战略为根本,依靠审计监督、评价与鉴证作用,健全算法治理体系,维护国家安全以及社会公共利益,推动数字经济健康发展,实现审计全覆盖。微观层面,数字平台算法审计立足场景化及精细化分类原则,通过对算法全生命周期各阶段开展审计监督、评价及鉴证,给出审计意见与建议,明确算法责任归属,提升算法信息质量,推动数字平台切实履行治理义务。公平性目标聚焦数字平台双边市场出现的差异化对待或歧视,着重检测算法对不同群体产生的歧视性结果,确保算法决策公平性;透明性目标在于评估和验证算法的可解释性,判断算法的决策逻辑能否被人类所理解,推动数字平台进一步完善信息披露;安全性目标则是检测算法是否存在技术漏洞和业务风险,防止出现恶意攻击或意外情况;合规性目标是要验证数字平台算法是否满足国家法律法规、行业规范和伦理标准。值得一提的是,数字平台算法具有动态迭代的特性,这就要求审计目标具备全生命周期追踪的动态特征,而非一次性评估。当然,我国数字平台算法审计不能照搬国际通用的算法透明度单一目标,而须结合我国《网络数据安全条例》和《算法推荐服务管理规定》等相关制度对于“安全与发展并重、合规与创新平衡”方面的要求,把合规性、安全性以及公平性都纳入数字平台算法审计的核心目标,这满足了我国数字治理语境要素扩容的现实需求,丰富了中国审计学自主知识体系的理论内涵,使数字平台算法审计的理论框架更加契合我国数字经济的监管需求。

### (三)数字平台算法审计主体

国家审计、社会审计和内部审计三类审计之间既存在共性又彼此独立。算法审计作为数字经济的重要治理手段,更要发挥三大审计协同优势和联动效应,形成服务于数字经济发展目标的跨部门、跨地区的协同监管机制,建立以国家审计为主导,社会审计和内部审计共同参与的多审计主体治理架构,但不同主体涉及的审计侧重点和审计作用有所不同。

国家审计在数字平台算法治理中具有权威性和强制性优势,尤其适用于具有系统重要性的大型数字平台。因为大型数字平台类似社会基础设施,提供的服务具有公共物品属性,其算法行为关系公共利益与国家安全,国家审计应主动介入,牵头制定审计标准、组织安全评估、督促平台配合调查,从而满足保密性与合规性双重要求。国家审计需要时刻关注数字经济环境的发展变化,根据国家数字经济发展战略创新审计方式以适应算法审计的宏观目标。社会审计凭借其专业性和独立性,可接受委托对数字平台开展第三方算法审计,出具算法审计报告,增强结果的公信力,并且由于社会审计的社会关注度高,因此可以起到肃清算法披露环境的作用。数字平台内部审计熟悉算法设计、数据流程与业务逻辑,具有对算法实施全过程监督的天然优势。国家审计和社会审计可以充分利用数字平台内部审计资源有效节省审计时间和成本,提高审计效率,形成协同监督与问责合力<sup>[33]</sup>,数字平台算法审计更强调多元共治。对数字平台算法审计协同与多元共治进行理论提炼,正是构建中国审计学自主知识体系的重要环节之一。

### (四)数字平台算法审计客体

算法不同于传统的审计事项,问责主体模糊一直都是算法治理的难点。算法使数字平台运行日益自动化,平台倾向以“技术中立”和“算法黑箱”为由逃避法律责任,从而导致问责困难。但是,当技术造成了损害后果时,如果不能穿透技术“面纱”,明确责任承担的主体,解决“谁负责”的问题,那么就会存在潜在的责任缺口,所有的治理努力也将失之空泛<sup>[34]</sup>。2022年颁布的《互联网信息服务算法推荐管理规定》明确要求平台承担算法主体责任,各类提供算法推荐服务的互联网公司几乎都在监管范围内,并通过算法备案制度将算法列为直接规制对象,在关键节点对数字平台进行主观过错认定,判断数字平台是否有效评估、控制和纠正算法带来的危害。

作为数字经济的主要组织形态和治理焦点,数字平台算法导致歧视性后果、造成垄断或者不正当竞争行为、传播虚假或不良信息等情况,数字平台及其算法开发设计者、部署者、运营管理者等均应被纳入审计范围。因为人工智能等数字技术从应用到构建的每一步都暗含着构建者的感知范围和价值取向。与传统信息系统审计聚焦系统是否按预设规则运行不同,算法审计需要深入到算法决策逻辑的内核,验证算法决策公平性、安全性、合规性、伦理性以及社会影响等,审计内容不仅包括数据输入的偏差问题,还包括模型的特征权重分配、算法迭代中的逻辑漂移以及输出结果对不同群体的差异化影响。

### (五)数字平台算法审计依据

数字平台算法审计依据主要来自国家法律法规、部门规章、国家标准及行业规范等。针对算法风险问题,各国各地区开始立法进行审计治理。如,2016年欧盟通过《通用数据保护条例》(GDPR)对个人数据保护的权利赋予审计强制力,2023年欧盟通过《人工智能法案》要求对生成式人工智能进行专项审计,2023年美国国家标准与技术研究院(NIST)正式发布了《人工智能风险管理框架》及配套使用手册,旨在帮助人工智能系统的开发者、用户和评估人员更好地管理人工智能风险以提高人工智能系统的可信度。

近年来,我国人工智能审计制度经历了“基础立法-伦理建构-标准突破”三阶段。如表1所示,我国相关法律、规章规范和标准三位一体的人工智能治理框架为算法审计提供了制度依据。

### (六)数字平台算法审计模式

数字平台算法审计需采用适应算法复杂性和动态性的多元审计模式。现有算法审计模式至少有三种:一是源于社会科学的分类审计,Sandvig等根据研究者、算法系统与用户三者的关系,将算法审计方法归纳为代码审计、非侵入式用户审计、抓取审计、马甲审计以及众包或协作审计五类<sup>[11]</sup>;二是委托第三方专家开展算法评估<sup>[24]</sup>,评估要素包括算法利益相关者的利益列表、算法伦理特征指标以及前两者的相关性矩阵三个方面;三是用户驱动的日常算法审计<sup>[25]</sup>,旨在描述与分析用户通过与算法系统的日常交互来检测、理解和质询可能存在问题的机器行为,并根据其遇到的行为采取一系列行动的过程。张超将算法审计分为依托平台的算法审计模式和依托用户的算法审计模式<sup>[35]</sup>,前者直接从平台获取源数据,主要分为代码审计、审计工具审计和爬取审计,后者主要包含非侵入用户审计、代理审计和众包审计。

表1 法律-规章-标准三位一体的数字平台算法审计制度依据

	发布时间	法规标准名称	发布单位	核心审计要求	
法律	2015.07.01	《中华人民共和国国家安全法》	全国人大常委会	建立国家安全审查机制,对影响国家安全的关键技术、网络产品等进行审查	
	2016.11.07	《中华人民共和国网络安全法》	全国人大常委会	建立网络安全风险评估机制,加强安全信息分析与通报	
	2021.06.10	《中华人民共和国数据安全法》	全国人大常委会	建立数据安全审查制度,对影响国家安全的数据处理活动进行审查	
	2021.08.20	《中华人民共和国个人信息保护法》	全国人大常委会	必须定期进行合规审计(可委托第三方),监管部门可要求高风险活动强制审计	
规章	2021.09.17	《关于加强互联网信息服务算法综合治理指导意见》	国家网信办等九部委	健全算法安全治理机制,政府、企业、社会协同共治	
	2021.10.29	《互联网平台落实主体责任指南(征求意见稿)》	国家市场监督管理总局	超大型平台需定期委托第三方审计主体责任履行情况	
	2021.12.31	《互联网信息服务算法推荐管理规定》	国家网信办等四部委	算法推荐服务提供者应当定期审核、评估、验证算法机制机理、模型、数据和应用结果等;具有舆论属性的需进行安全评估;配合监管部门提供技术数据	
	2022.03.20	《关于加强科技伦理治理的意见》	中办、国办	科技活动需进行伦理风险评估或审查	
	2023.07.10	《生成式人工智能服务管理暂行办法》	国家网信办等七部委	具有舆论属性的生成式人工智能服务需开展安全评估;配合监管说明算法机理及训练数据	
	2023.09.07	《科技伦理审查办法(试行)》	科技部等十部门	科技活动需进行伦理审查与风险评估	
	2024.09.24	《网络数据安全条例》	国务院	国家建立数据安全审计制度;大型平台定期委托第三方审计数据安全、个人信息保护等	
	2025.09.01	《人工智能生成合成内容标识办法》	国家网信办等四部门	细化并强制要求对 AIGC 进行显式或隐式标识,建立内容的可追溯管理体系	
	2025.02.28	《关于推动网络交易平台企业落实合规管理主体责任的指导意见》	国家市场监督管理总局	平台企业建立合规组织、加强合规管理,主要负责人有合规管理责任	
	2025.12.18	《网络交易平台规则监督管理办法》	国家市场监督管理总局、国家网信办	平台规则中需明确信息安全条款,不得大数据"杀熟"	
	标准	2019.06.17	《新一代人工智能治理原则》	国家新一代人工智能治理专业委员会	新一代人工智能治理的八条原则
		2020.03.06	《信息安全技术 个人信息安全规范》	国家市场监督管理总局、国家标准化管理委员会	开展个人信息安全影响评估与审计,确保措施有效性
2021.09.25		《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	将伦理道德融入人工智能全生命周期,6项基本伦理要求和18项具体伦理要求	
2023.08.06		《信息安全技术 机器学习算法安全评估规范》	国家市场监督管理总局、国家标准化管理委员会	算法需通过对抗攻击测试,确保鲁棒性达标	
2025.01.24		《人工智能深度学习算法评估》	国家市场监督管理总局、国家标准化管理委员会	建立深度学习算法可靠性评估指标体系	
2025.01.26		《人工智能安全标准体系(v1.0)》(征求意见稿)	全国网络安全标准化技术委员会	覆盖基础共性、安全管理、关键技术、测试评估、产品与应用等5个部分的人工智能安全标准体系	
2025.09.15		《人工智能安全治理框架(v2.0)》	全国网络安全标准化技术委员会、国家计算机网络应急技术处理协调中心	系统性地提出基于风险分类的敏捷治理和分级管理方法论,是未来统一立法的重要理论基础	

数字平台算法的动态性和复杂性使其审计模式更需构建动态适配的方法论体系,形成静态解析+动态追踪+多主体协同的复合模式。静态解析模式通过分析代码的结构、语法、语义、数据流、控制流等信息,理解算法的行为、发现潜在问题或提取关键特征;动态追踪模式要求审计从事后追溯转向动态监测,通过嵌入数字平台数据接口进行实时风险捕捉;跨主体协同模式根据数字平台的多方参与特性,采用国家审计主导+社会审计技术支持+内部审计日常监督的复合模式。在实际操作当中,可以引入现代风险导向审计模式,首先从制度层面对数字平台算法治理相关文件实施抽样审核,借助控制测试确认关键控制点在设计及执行方面的有效性,随后凭借实质性程序针对高风险模块或敏感输出开展穿行测试。还可以通过跨平台对比分析算法输出情况,根据国家标准判别和揭示系统性偏见;在采用对抗性测试方式时,精心设计特定的输入去主动“攻击”算法,验证算法是否存在偏见、安全漏洞或非预期行为。数字平台算法审计模式的选取应基于具体场景,综合考虑审计目标、算法类型及信息可获取性等因素,形成能够针对数字平台的高适应性审计路径。

#### 四、数字平台算法审计的作用机制

##### (一) 预防机制

预防机制是指数字平台算法审计依照科学性、专业性以及重要性原则,在事前审计阶段及时介入、识别并预警设计环节中的伦理风险和技术漏洞,推动算法可解释性基础框架的构建与源数据治理。算法审计的预防机制增强了数字平台算法的可解释性,能够有效降低算法决策出现的失误,打破人工智能系统的算法“黑箱”,提高用户对算法的信任度和交互程度。具体而言,算法审计应当首先结合数字经济所处的时代大背景,立足数字平台具体业务场景以及庞大的用户群,运用全面且实用的评估与分析算法的工具,对算法设计是否嵌入合规控制以及伦理约束展开系统的评估,判别在推荐、定价、信用评估等核心功能当中是否存在隐含的歧视性或者操纵性

逻辑,既要评估算法设计的安全合理性,也要保障算法运行的有效性。接着,算法审计人员需要依靠专业技术工具,针对训练数据集的代表性、平衡性以及合法性做出审慎的判断,尤其是要防范因为数字平台自身数据闭环而产生的偏见固化以及回声室效应的发生。

数字平台算法审计的预防机制强调从设计源头切入,为数字平台构建算法免疫系统。其一,强调算法设计的可理解性,从设计源头管控可能存在的价值风险,以保证算法后续运行的实质性价值理性;其二,确定源数据选择范围的合理性,数字经济时代的用户数据安全是亟须重视的问题,确保算法的源数据在安全边界内是预防算法隐私风险的重要一步。在算法设计阶段,审计提前介入,植入“风险抗体”,确保数字平台不利用技术不对称实施大数据“杀熟”、不滥用算法推荐等。预防机制不仅能有效降低算法上线后的修正成本,更在数字平台与用户之间建立初始信任基础,是从源头抑制算法权力滥用的关键治理手段。数字平台算法审计的预防强调商业伦理嵌入,即通过审计倒逼数字平台将公平性、安全性等目标写入算法开发手册,而非单纯依赖技术补丁。

### (二) 监督机制

数字平台算法在运行中存在动态异化风险,根源在于数字平台的“算法黑箱”具有双重隐蔽性,既包括技术层面的模型不透明,也包括业务层面的规则模糊。数字平台算法审计的监督机制旨在遵循及时性、充分性和独立性等原则,通过监督数字平台算法运行的全过程,规制算法权力,构建算法运行中的“数字瞭望塔”,实现从事后追责到事中熔断的转变以及技术与业务的双穿透。算法审计通过制定审计工作计划,全面考虑商户与用户权益保障,监督算法在运行过程中是否存在人为操控,是否存在用隐私安全换取技术红利的隐形规制等。

数字平台算法审计监督机制主要体现在以下几点:其一,监测算法在实际部署运行中的合规性。算法审计通过监督算法运行过程的规范性,防止数字平台出于商业利益进行隐蔽操纵,遏制算法歧视、算法沉迷和茧房强化等行为。其二,审查算法程序执行的一致性,明确规则的一体适用性,防止针对特定用户、区域或群体进行差异化处理。数字平台算法审计不仅涉及对算法的源代码进行静态分析,更需要结合日志审计、输出比对与实时性能监控,动态监测程序运行时的状态,严格稽查以排除算法歧视、大数据“杀熟”等问题,确保不存在偏离设计初衷的异常决策。与通用算法的审计监督侧重合规性检查不同,数字平台算法审计监督更强调透明性与公平性。

### (三) 评价机制

数字平台算法行为兼具技术性与社会性,其复杂的决策流程与广泛的外部效应导致主体责任难以清晰界定。评价机制是算法审计以客观性、谨慎性和准确性为原则导向,通过分析评价,确认并落实主体责任,对算法运行体系的安全有效性提出专业的审计意见,数字平台的公共属性决定了其算法评价不能仅以商业效益为标准,还需要构建治理、技术与社会层面的评价体系。评价机制的作用在于追溯技术缺陷与管理失责的因果链条,明确责任边界,实现主体责任锚定。

算法审计评价机制的主要内容包括以下三方面。第一,评价数字平台算法治理体系的成熟度。数字平台作为算法研发、部署与收益获取的主体,须承担起与其技术能力相匹配的管理责任。尽管目前算法的风险大多聚焦在技术层面和道德层面,但是企业管理层和治理层的责任是对算法存在的潜在风险及时采取积极的治理措施,保障算法运行的安全性和有效性,并设置自动预警机制,这对作为数字经济核心技术的算法发挥积极作用至关重要。第二,评估算法技术的稳定性与可解释性。不稳定的算法可能因输入的细微差异而导致输出剧烈波动,降低结果的可信度,高可解释性的算法能帮助用户信任结果、排查错误,并满足监管要求。第三,评估算法的社会效果。在数字经济时代,算法正在由工具、实践层面抵达社会安排和制度规范层面,人类在享受数字化带来便捷生活的同时,也将个人对于生活的诸多决定权让渡给算法,算法已经与人们生活形成密切的关系。因此,要跳出纯技术指标,从用户满意度、社会公平性、市场竞争秩序和隐私保护水平等角度,评估算法运行实施后的社会效果,考察算法是否真正实现技术向善。数字平台算法审计的评价强调在商业目标与社会责任之间找到平衡点,让算法既可控又可理解,既服务于平台增长,又兼顾用户、商家与社会的长期利益。

### (四) 抵御机制

数字平台融入社会经济越深,算法介入人类生活就越频繁。抵御机制体现了算法审计的治理闭环,遵循针对性、普适性、可行性的原则,借助问题诊断、成因分析、问题整改以及结果公开等方式,抵御算法系统性风险,构筑起算法社会的“免疫盾牌”,凸显了算法审计推动数字经济健康发展的重要意义。数字平台算法审计的抵御作用不单指具体技术问题的修复,还重视推动数字平台内部治理体系与外部监管环境的协同改进,所以有更强

的影响力。其一,审计结论能够为行业标准修订以及立法完善提供参考,推动制度变革;其二,审计发现可促使数字平台建立偏见矫正机制,调整商业策略,重视算法伦理;其三,审计整改建议有清晰的目的性、指向性以及针对性,有助于落实算法责任。这种从技术修复到社会治理的延伸,彰显了算法审计在影响广度方面的独特价值,具备风险隔离、损失控制与修复优化的递进功能。风险隔离在于识别数字平台算法的高危模块并设置熔断;损失控制是针对已发生的算法风险,量化影响范围,监督数字平台采取补救措施;修复优化体现在形成“发现-整改-治理”的闭环体系,体系的关键在于治理建议的可行性,不论是改进建议还是算法问责,其价值在于落实,发挥算法审计抵御机制的整体治理合力。如针对电商平台因算法漏洞导致的虚假促销,算法审计不仅要督促平台修复漏洞,还需通过模拟交易验证整改效果,并跟踪后续用户投诉率变化。

数字平台算法审计作为具有较强综合性的治理手段,把促进数字经济健康发展作为重要目标,发挥数字平台、社会和国家在算法审计中的协同作用,构建算法社会的信任生态。这有别于国外第三方审计主导的机制设计,更贴合中国审计制度情境,进一步增强了中国审计学自主知识体系中协同审计方法论的实践适配性<sup>[36]</sup>。如图1所示,四大机制形成正向循环,预防机制降低算法的基础风险,监督机制提供算法运行的全过程证据链,评价机制输出治理清单,抵御机制起到优化算法设计标准的作用。

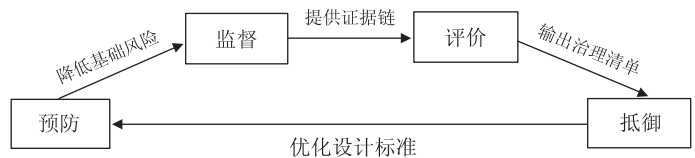


图1 数字平台算法审计四大机制正向循环作用逻辑

### 五、数字平台算法审计的主要维度

数字平台算法风险主要来自治理、数据、算法以及应用等四个维度。其中,治理维度是影响其他各维度的关键核心组织因素。本文从治理维度、数据维度、模型维度以及应用维度等出发,明确算法审计的重点方向以及实务操作路径,并针对每一个维度分别指出在开展算法审计工作时需要重点考虑的审计难点与挑战、关键审计问题、关键审计程序以及实务要点等,为算法审计人员更好地理解和评估数字平台算法提供参考,如图2所示。

#### (一) 治理维度

##### 1. 实务挑战

数字平台的治理结构是算法风险的第一道防线,是约束算法权力防范系统风险的首要制度保障。由于深度学习等主流算法的不可解释性,算法的设计者也无法完全预测自动化决策的结果,数字平台可能以技术“黑箱”为借口掩饰主观意图,导致算法权责不清。因此,在治理维度,数字平台算法审计面临明确算法责任的挑战。一方面,算法决策的自动化程度使得责任追溯链条断裂,如外卖平台有“骑手困于系统”的现象,当配送算法引发骑手安全事故时,技术团队、运营部门与管理层常相互推诿;另一方面,数字平台商业目标与公共利益存在天然张力,一些电商平台为提升转化率设计的价格歧视算法,即众所周知的大数据“杀熟”,其治理机制可能沦为形式化合规。不论是基于技术难点的客观算法风险还是基于盈利的主观算法动机,数字平台都面临算法技术安全和伦理双重风险问题。数字平台公司治理作为预防和隔离算法风险的关键屏障,是算法审计首先要考虑的维度。治理维度算法审计的实务重点和难点在于实现治理责任的可溯源与

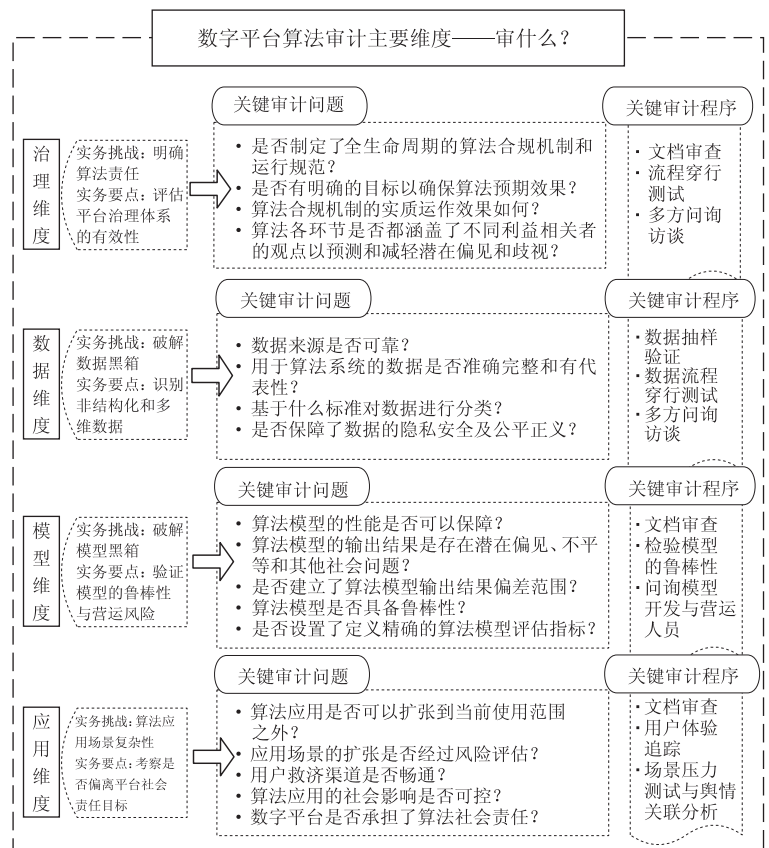


图2 数字平台算法审计的主要维度的实务挑战、关键审计问题与程序

流程的可评估,数字平台的治理实践是否体现了科技向善的价值理念和负责任的愿景目标,具体包括数字平台组织建设、风险控制机制建设、文化建设、生态建设和治理工具建设。从治理维度开展算法审计的核心价值在于制度预防,体现了预防机制的价值锚定特性,本质是通过预防性制度设计构建数字平台的算法章程,推动科技向善。

## 2. 关键审计问题

治理维度的关键审计问题包括:其一,数字平台是否制定了完备的算法运行规范? 算法治理章程是否明确商业目标与伦理约束的优先级? 其二,数字平台算法目标是否明确、可测量且合法合规? 算法的目标是否符合相关的法律法规和政策? 其三,数字平台是否组建了统筹推进算法伦理治理工作体系建设的专门机构? 是否组建了负责个人信息隐私数据保护和算法安全评估工作的数据安全与个人信息保护委员会? 其四,数字平台是否为算法的全生命周期明确了相关部门与人员的职责及授权? 算法各环节所出现的问题是否都可问责,算法伦理委员会或类似机构是否能直接否决可能引发舆论风险的推荐规则? 其五,数字平台算法各环节是否都涵盖了多元利益相关者的观点以预测和减轻潜在偏见和歧视? 是否切实做到了向其他利益相关者兑现价值观承诺?

## 3. 关键审计程序

基于上述问题,审计人员可以考虑执行以下相关审计程序。首先,文档审查。收集相关审计证据,包括但不限于数字平台算法治理的章程或政策性指导文件等、数字平台算法开发设计的战略计划及目标等、数字平台岗位设置的描述性文件等、涉及利益相关者在算法整个生命周期中参与的文档等。其次,流程穿行测试。基于所收集的文档,追踪从提案到上线的全流程审批记录,验证伦理审查的有效性;具体而言,数字平台规范准则是否合法可执行,算法的目标文件是否明确可计量,算法相关人员的角色职责及授权相关文件是否清晰可理解。最后,多方问询访谈。核实制度落地与沟通机制实际效果,以确认治理体系能否有效预警防控算法风险。访谈技术负责人,了解算法设计逻辑;访谈平台管理层,了解规范准则是否切实执行;访谈用户代表,了解数字平台对用户问题的反馈处理实际效果,以及用户是否理解数字平台对算法目标及效用的阐释。通过执行上述审计程序进一步厘清数字平台算法是否具备完整的治理结构与流程从而可以及时预防算法事前的潜在风险。

## 4. 实务要点关注

治理维度的算法审计是实现公平性、合规性和透明性等审计目标的基石,要求基于风险导向审计模式,依据相关法律和规章,运用代理审计、协作式审计和对抗性测试等方法评估数字平台的组织架构、决策机制及激励制度等,以期在事前阶段就能以完备的风险治理结构预防算法风险。治理维度要求算法审计聚焦数字平台算法治理体系的有效性而非形式性,一方面,审查数字平台是否具有明确的治理结构、治理机制和流程来管控算法风险,并在此过程中向算法利益相关者展示数字平台建立的价值观和道德原则的承诺,确保数字平台及其算法遵守相关法律法规;另一方面,要特别关注数字平台治理结构中的决策权,即谁有权做关键决策以及决策流程是否体现透明度、问责性与多方制衡。

### (二) 数据维度

#### 1. 实务挑战

数据是算法的基础输入,数据质量直接影响决策公正与社会公平。数字平台数据存在的来源偏差、标注主观和隐私泄露等问题极大地增加了算法审计的难度。第一,数据是数字平台算法的“原材料”,算法通过大量数据形成的数字虚拟映像不断训练及自我学习,以理解和分析现实复杂系统的运行行为、状态和规律,只有高质量的数据才是算法得以分析运用的基础,数据质量直接决定算法输出;第二,数据在数字平台算法系统中承载着价值判断功能,决定算法的公平性、安全性与社会影响等。数据不具有中立性,在数据采集和处理过程中必然带有偏见,甚至侵犯隐私和加载有害内容,反馈回路就会强化这种偏见与危害;第三,数据是数字平台最重要的资产,具有资产属性和权力属性。数字平台常以商业机密为由拒绝披露数据来源与处理逻辑,或以数据匿名化为由规避责任,造成“数据黑箱”壁垒。

#### 2. 关键审计问题

数据维度的关键审计问题可以分解为以下四个具体方面。第一,数字平台所获取的数据来源是否可靠可信? 训练数据是否具备足够的代表性,能否涵盖边缘群体? 数字平台所用数据是否适用于算法系统? 第二,用于算法系统的数据是否足够准确且完整? 数字平台具体采取了哪些举措保障和提升数据质量、数据准确性和代表性? 数据预处理环节有无出现偏见被放大的情况? 第三,数字平台依据何种标准对数据加以分类? 所采用的

分类标准是否完整可靠,是否契合算法所设定的目标?在实际应用过程中,数字平台是否确实执行了该标准?第四,数字平台是否采取相关措施切实保障了数据的隐私安全及公平正义?数字平台是如何识别、评估以及减轻与算法有关的数据隐私及歧视风险的?

### 3. 关键审计程序

基于上述审计问题,审计人员可以考虑执行以下审计程序:其一,数据抽样验证。收集数据全生命周期的相关审计证据,包括但不限于描述数据收集方法及管理计划的文件,体现数据可靠性评估、纠正措施的相关文件,描述数据分类标准和规则的相关文件,算法系统使用数据的安全计划、安全评估及隐私保护评估性文件,抽取数据不同生命周期的原始数据、合成数据、测试数据、输入输出数据,检测各类数据的匹配性,识别数据风险。其二,数据流转穿行测试。基于上述原始文件和数据进一步审查算法系统数据的可靠性和适用性,确定用于算法系统的数据是否准确、完整、可靠和具有代表性等,实行穿行测试以确定数据流通过程的完整性,审查数据样本及其测试运用过程中是否侵犯用户隐私安全或存在负面影响等。其三,多方询问访谈。询问算法开发者以了解数据定义、评估、选择、合成、测试等各环节的基本原理及其所采取的相关措施,同时询问算法开发者和管理层以确定数字平台对数据安全和隐私保护的程序,数据中的潜在偏见是否得到识别和缓解等。

### 4. 实务要点关注

数据维度的算法审计旨在提高数字平台用于训练、测试和验证算法的数据质量,降低数据处理时的风险,破解“源头污染”,提高下一阶段模型训练的效果,保证算法系统产生一致和准确的结果,体现了算法审计监督机制的动态制衡作用,是实现合规性与安全性审计目标的重要环节。传统数据审计工具在应对数字平台的非结构化数据时存在困难,此时就需要借助自然语言处理工具,并采用诸如抓取审计、代理审计、非侵入式审计以及对抗性测试等一系列方法,来合理保证算法系统全生命周期所涉及数据的质量、可靠性及代表性,进而达成公平性、安全性、合规性以及透明性等审计目标。数字平台算法审计工作要提前识别数据偏差,并且要求数字平台补充多维特征,还要关注数据质量的时效性,如外卖平台因为节假日数据突然增多而出现训练数据偏差的情况。

## (三) 模型维度

### 1. 实务挑战

算法模型是算法系统中数据到决策转换的关键环节,是数据同算法相互融合的综合体。模型维度的算法审计主要目的在于对数字平台算法模型的技术正义予以评估,保证算法系统经过算法模型的运算之后,所输出的结果能够与预先设定的目标保持高度一致。通过囊括模型设计、运行以及演化的整个流程,模型维度的算法审计可切实有效地提高数字平台算法的透明性、合规性及公平性,是深入了解算法行为、防范算法风险的关键途径。算法模型自身架构较为复杂且参数繁多,动态迭代兼具黑箱特性,这无疑给算法审计工作带来了极为严峻的挑战。一方面,模型是数据偏见的“放大器”,而模型的架构情况、参数设置以及训练机制等会直接对决策逻辑起到决定性作用;另一方面,数字平台算法模型因机器学习的自主性正变得日益复杂,模型的黑箱特性很可能掩盖决策歧视,大幅增加算法调试以及优化工作的难度,致使算法偏见、安全漏洞以及非预期行为等诸多风险隐藏其中难以被察觉;此外,算法模型始终处于持续不断的迭代更新中,随着新数据被纳入、模型架构调整或者算法优化升级,算法模型的行为都可能发生改变,算法模型的迭代速度远超审计周期的时长。

### 2. 关键审计问题

模型维度的关键审计问题包括:第一,数字平台算法模型的性能能否得到保障?算法模型运行过程中是否高效可靠且稳健?第二,数字平台对于算法模型输出结果是否设定了可接受的偏差范围?可能的偏差范围是否合理合规?数字平台又采取了哪些措施解决算法模型输出存在的偏差?第三,数字平台是否设置了定义精确的算法模型评估指标?数字平台使用这些度量指标的理由是什么?在多大程度上可以与算法预期目标和约束相一致?第四,数字平台算法模型是否具备鲁棒性?对潜在意外或者极端边缘情况的反应程度,算法模型是否具备足够的抗风险及抗漏洞的能力?第五,数字平台算法模型运行过程中是否存在相关人员主观嵌入或算法黑箱客观生成的不平等问题,当使用不同类别或不同性质的数据时算法输出是否表现出差异?数字平台是否可以明确解释差异的产生并及时采取措施减少差异?

### 3. 关键审计程序

算法审计人员可以考虑从以下几个方面执行模型维度的相关审计程序:首先,文档审查。收集与算法模型

相关的审计证据,包括但不限于用以描述算法模型设计开发计划、具体参数选择、运行机理、操作规范、输出评估检测等的相关文档或数据,算法模型评估指标设置及管理类文档,关于可接受差异范围建立方式和实际评估的描述性文档,通过审查上述审计证据用以明确算法模型设计运行机理、参数选择及评估指标与算法目标的适用情况、日常运行的具体情况、面对潜在问题时算法模型的自适应性及数字平台的应对情况,保障算法系统的透明性及可问责。其次,检验算法模型的鲁棒性。审计人员可以基于保密原则在审计与被审计双方达成共识的情况下获取数字平台关于算法模型的访问权限及源数据等,一是进行算法压力测试,输入边缘数据用以测试算法模型在不同应用场景中的表现情况,观察模型输出是否存在系统性偏差;二是进行参数敏感性分析,调整模型关键参数,测试输出结果的波动幅度,识别参数设置的合理性。最后,问询算法模型开发及运营人员,了解模型迭代机制与风险应对策略,分析算法优化方向是否存在牺牲公平性换取加载速度的唯效率论倾向。

#### 4. 实务要点关注

模型维度的算法审计重点在于突破技术壁垒,需要针对数字平台算法模型训练过程的提示注入风险、鲁棒性不足、决策偏见以及模型运营等风险实施算法审计,运用包括代码审计、对抗测试、代理审计和反事实解释等多种技术方法,对模型的各项指标进行审计,提高模型在部署运营过程中的透明度,尤其要关注迭代过程中的规则漂移与偏见嵌入。如,外卖平台的配送费计算算法,若对新商户设置隐性惩罚系数,可能构成不正当竞争,审计需通过代码解析提前揭示此类规则漏洞,敦促平台定期披露排名算法的核心特征与权重分析。在实务中,由于算法模型的复杂性与动态性,更加要求三类审计之间的协同和资源共享;此外,还可以采用专家协作模式,聘请外部专家审查算法模型技术层面的风险问题,维护算法利益相关者的权益。

#### (四)应用维度

##### 1. 实务挑战

算法的价值最终体现于应用场景,数字平台的算法应用直接关系用户权益、社会秩序与商业伦理。算法应用的场景泛化与风险外溢效应显著,给算法审计带来巨大挑战。其一,应用场景的复杂性可能放大数据与模型缺陷,跨场景算法可能引发连锁风险;其二,算法应用直接作用于用户权益,应用结果直接影响用户的获得感与公平感;其三,数字平台的用户反馈机制可能形同虚设,实际投诉处理率虚低。总之,应用维度是数字平台算法治理的“最后一公里”,也是算法审计发挥评价与抵御作用的重要场景。

##### 2. 关键审计问题

应用维度关键审计问题需要聚焦算法对社会的影响,可以分解为以下几个方面。第一,算法应用是否可以扩张到当前使用范围之外?算法应用场景的扩张是否经过风险评估?是否制定了对算法持续监督的管理计划并确保其按预期执行?第二,用户救济渠道是否畅通?数字平台是否采取措施来应对算法应用过程中的问题?受到算法影响的利益相关者是否有明确的渠道或路径反馈算法问题?对用户申诉是否建有24小时响应机制,是否设置了纠偏机制,是否可以通过持续监督及时发现问题并解决问题,数字平台是否可以明确溯源问题的起因并问责?第三,算法应用的社会影响是否可控?数字平台为承担算法社会责任做出了哪些努力?数字平台是否具备足够的内生动力用以保障算法应用对社会影响的自主可控、安全可靠?

##### 3. 关键审计程序

算法审计人员可以考虑从以下几个方面执行审计程序:首先,文档审查。收集相关审计证据,包括但不限于数字平台关于算法系统的监督计划、进度表及相关跟踪文档,持续监督活动结果或日志等相关性文档,关于扩展算法应用范围或停止使用算法的条件或指标的描述性文件,关于数字平台履行算法社会责任的描述性文件等。其次,用户体验追踪。招募不同群体用户完成平台核心操作,询问算法利益相关者对数字平台履行算法社会责任的具体感受,以确定算法在应用过程中的影响程度、所产生的确切影响以及数字平台所采取的具体措施,验证数字平台是否切实执行了纠偏措施,停止或扩展算法应用的假设条件是否可行且合适,数字平台是否切实履行了社会责任,数字平台在多大程度上具备在算法应用过程中持续监督的内生动力。最后,场景压力测试与舆情关联分析。模拟极端场景,检测算法响应是否符合公共利益,通过舆情监测工具抓取平台相关的负面新闻,分析是否与算法应用缺陷相关。

##### 4. 实务要点关注

应用维度的算法审计重点考察算法应用是否偏离数字平台的社会责任与公序良俗,旨在促进数字平台根据

用户体验和监管要求持续优化算法系统以实现算法系统应用过程中的安全合规与公开透明。算法是动态且自适应的,其性能及其影响会随着时间的推移而变化。如针对社交平台“信息茧房”问题,审计可结合用户行为数据与算法推荐结果,验证平台是否履行多样性推荐承诺,并将监督结果纳入平台信用评级;对短视频平台的推荐算法,审计可以预设极端内容传播阈值,通过模拟测试验证算法是否具备自动过滤机制,从事前预防群体性风险。实务中,可以引入社会学家、心理学家参与审计,应特别认识到算法应用过程中与社会链接的广度和深度,督促数字平台自觉承担起算法社会责任,完善数字平台算法社会责任生态体系。

## 六、结论及展望

数字经济时代,数字平台承载着前所未有的社会价值,算法审计可以为算法治理落地提供新的思路,缓解数字平台发展与规范的现实冲突,平衡算法技术创新与异化的矛盾。本文构建了较为完整的数字平台算法审计理论框架,力求夯实具有中国特色、体现当代社会主义核心价值观的审计学理论基础<sup>[37-38]</sup>,并在此基础上提出了相应的实务路径,创建适配中国情境的审计方法与工具。本文分析认为数字平台算法审计具有以下五个主要特征:其一,算法审计是对算法本身进行审计而非利用算法技术进行审计;其二,数字平台算法审计可由国家审计主导,内部审计和社会审计共同参与,实现多元共治;其三,数字平台算法审计是对算法的全生命周期进行审计治理;其四,数字平台算法审计不仅要发现和识别算法的风险问题,还要督促相关责任方从源头规制整改;其五,数字平台算法审计更强调技术安全-社会价值的平衡,显著区别于非平台算法审计与信息系统审计。

数字平台算法审计是完善我国审计监督体系、优化国家治理能力、保障经济社会健康发展的一个新兴议题,也是中国审计学自主知识体系的重要构成内容。当前各国都在积极开展算法审计治理实践,不断推出算法监管和审计指引,算法审计正在实现“从无到有”的关键突破,未来有待进一步基于算法审计的理论探索、学科融合、实际应用、人才培养等方向拓宽算法审计研究的深度和广度,将算法审计制度化、准则化、规范化和程序化,切实满足算法审计治理数字技术风险、确保算法透明度和可解释性的需要,以应对日益复杂的技术挑战。作为数字经济大国,我们需要通过数字平台算法审计促进数字经济健康发展,保障用户权益,维护市场公平,推动科技向善,进而实现算法审计“从有到优”的进一步跨越,促进我国审计监督制度更加完善且全面。

### 参考文献:

- [1] 习近平. 不断做强做优做大我国数字经济[J]. 求是, 2022(2): 4-8.
- [2] 李三希, 黄卓. 数字经济与高质量发展: 机制与证据[J]. 经济学(季刊), 2022(5): 1699-1716.
- [3] 朱太辉, 林思涵, 张晓晨. 数字经济时代平台企业如何促进共同富裕[J]. 金融经济研究, 2022(1): 181-192.
- [4] 张晨颖. 公共性视角下的互联网平台反垄断规制[J]. 法学研究, 2021(4): 156.
- [5] 魏巍, 刘贝妮. 算法管理能提高数字零工劳动者的平台承诺吗? ——“控制主义”和“决策主义”的双刃剑效应[J]. 经济管理, 2023(4): 116-132.
- [6] Gorwa R. What is platform governance? [J]. Information, Communication & Society, 2019, 22(6): 854-871.
- [7] Lurie E, Mustafaraj E. Opening up the black box: Auditing google's top stories algorithm[C]. The Thirty-Second International Flairs Conference, 2019(32): 376-382.
- [8] 袁康. 可信算法的法律规制[J]. 东方法学, 2021(3): 5-21.
- [9] Helberger N, Pierson J, Poell T. Governing online platforms: From contested to cooperative responsibility[J]. The Information Society, 2018, 34(1): 1-14.
- [10] 张健, 周旺旺. 数字平台权力与监管动力机制: 一个整体的分析框架[J]. 深圳社会科学, 2024(3): 116-128.
- [11] Sandvig C, Hamilton K, et al. Auditing algorithms: Research methods for detecting discrimination on internet platforms[J]. Data and Discrimination: Converting Critical Concerns into Productive Inquiry, 2014, 22(5): 4349-4357.
- [12] Gaddis S M. Audit studies: Behind the scenes with theory, method, and nuance[M]. Berlin: Springer, 2018.
- [13] Baxter C. Algorithms and audit basics[J]. ISACA Journal, 2021(3): 6-12.
- [14] 王天恩. 人工智能应用“责任鸿沟”的造伦理跨越——以自动驾驶汽车为典型案例[J]. 哲学分析, 2022(1): 15-30+196.
- [15] 于霄. 算法民事归责的理性标准[J]. 河南大学学报(社会科学版), 2023(2): 39-46+153.
- [16] Kaminski M E, Urban J M. The right to contest AI[J]. Columbia Law Review, 2021, 121(7): 1957-2048.
- [17] 徐凤. 人工智能算法黑箱的法律规制——以智能投顾为例展开[J]. 东方法学, 2019(6): 78-86.
- [18] Bitzer T, Wiener M, Cram W A. Algorithmic transparency: Concepts, antecedents, and consequences - a review and research framework[J]. Communications of the Association for Information Systems, 2023, 52(1): 293-331.
- [19] Brown S, Davidovic J, Hasan A. The algorithm audit: Scoring the algorithms that score us[J]. Big Data & Society, 2021, 8(1): 1-8.

- [20] Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations[J]. *Science*, 2019, 366(6464): 447 – 453.
- [21] 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*, 2019.
- [22] Koshiyama A, Kazim E, Treleaven P, et al. Algorithm auditing: Managing the legal, ethical, and technological risks of artificial intelligence, machine learning, and associated algorithms[J]. *Computer*, 2022, 55(4): 40 – 50.
- [23] Vijayakumar S. Algorithmic decision-making[J]. *Harvard Political Review*, 2017, 44(2): 9 – 11.
- [24] Lee S C. Auditing algorithms: A rational counterfactual framework[J]. *Journal of International Technology and Information Management*, 2021, 30(2): 120 – 145.
- [25] Shen H, DeVos A, Eslami M, et al. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors[C]. *Proceedings of the ACM on Human – Computer Interaction*, 2021.
- [26] Raghavan M, Barocas S, Kleinberg M, et al. Mitigating bias in algorithmic hiring: Evaluating claims and practices[C]. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January, 2020.
- [27] Diakopoulos N. Accountability in algorithmic decision making[J]. *Communications of the ACM*, 2016, 59(2): 56 – 62.
- [28] Netherlands Court of Audit. Understanding algorithms[EB/OL]. [2021 – 01 – 26]. <https://english.rekenkamer.nl/publications/reports/2021/01/26/understanding-algorithms>.
- [29] 王玉凤. 模型算法审计: 理论内涵、国际经验与审计框架[J]. *审计研究*, 2023(3): 11 – 18.
- [30] 张涛. 通过算法审计规制自动化决策以社会技术系统理论为视角[J]. *中外法学*, 2024(1): 261 – 279.
- [31] Bandy J, Diakopoulos N. Auditing news curation systems: A case study examining algorithmic and editorial logic in apple news[C]. *Proceedings of the International AAAI Conference on Web and Social Media*, 2020, 14(1): 36 – 47.
- [32] 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.
- [33] 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.
- [34] 张凌寒. 《个人信息保护法(草案)》中的平台算法问责制及其完善[J]. *经贸法律评论*, 2021(1): 36 – 46.
- [35] 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.
- [36] 罗岭, 卫振炎. 数字经济发展、二元审计质量与企业创新持续性[J]. *南京审计大学学报*, 2024(1): 21 – 31.
- [37] 孙甲奎. 面向中国式现代化的审计学科功能定位与创新发展[J]. *南京审计大学学报*, 2025(1): 1 – 11.
- [38] 董必荣. 关于在交叉学科门类下设置审计学一级学科的构想[J]. *南京审计大学学报*, 2025(3): 1 – 9.

[责任编辑: 刘 茜]

## Algorithm Audit of Digital Platforms: Theoretical Framework and Practical Suggestions

HE Yong<sup>1</sup>, LIU Xiaoyi<sup>1</sup>, ZHANG Jun<sup>2</sup>

(1. School of Accounting, Hunan University of Technology and Business, Changsha 410205, China;

2. School of Economics and Management, Hunan Institute of Science and Technology, Yueyang 414006, China)

**Abstract:** The digital platform is the most successful organizational model of the digital economy, occupying an absolutely important position. As the core driving force and underlying architecture of the development of the digital economy, algorithms are like a double-edged sword. While promoting the rapid development of digital platforms, they also come with many algorithmic risks. Algorithmic auditing is an important tool for digital governance and plays a crucial role in promoting the healthy development of digital platforms. Guided by the healthy development of the digital economy, this paper constructs a theoretical framework and practical path for digital platform algorithm auditing. Algorithm is taken as audit object, and the prevention, supervision, evaluation, and resistance mechanisms of digital platform algorithm auditing are systematically analyzed. Practical challenges, key audit issues and procedures of digital platform algorithms are explained from the dimensions of platform governance, data, models, and applications. Digital platform algorithm auditing expands the field and scope of auditing, provides new directions for the diversified integration and development of auditing disciplines, and enriches the independent knowledge system of digital economy governance and auditing disciplines in China.

**Key Words:** digital economy; digital platform; algorithm risk; algorithm auditing; digital governance; algorithm governance; audit supervision; independent knowledge system of China's auditing