

大数据时代算法审计构想

付冉冉

(南京审计大学 法学院, 江苏 南京 211815)

摘要: 在现代语境里,算法是解决问题的自动化程序,在公共管理、司法裁判、金融系统、以及商业领域等都得到广泛应用。算法应用一方面提高了社会整体的运行效率,另一方面也因其复杂的技术性特征导致了社会管理的失序。当前,我国算法规制体系立法较为松散、监管主体不够明晰,而审计作为一种独立而成熟的审查制度,在算法设计监督、算法数据保护以及算法决策法律问责中都大有可为,可解决算法监管体系松散、监管责任不明等问题。在审计监管模式下,可由专业人员制定审计计划,根据特定的审计标准进行风险评估和事实审查,给出审计结论,制作审计报告、进行审计公开,对相关人员实施问责以及开展后续审计质量追踪。由第三方审计、平台自我规制与行业自律相结合,算法监管可迎来新的突破。

关键词: 算法;算法风险;算法治理;算法审计

中图分类号: F239

文献标识码: A

DOI: 10.19358/j.issn.2097-1788.2023.02.008

引用格式: 付冉冉. 大数据时代算法审计构想[J]. 网络安全与数据治理, 2023, 42(2): 48-52.

The conception of algorithm audit in big data era

Fu Ranran

(School of Law, Nanjing Audit University, Nanjing 211815, China)

Abstract: In the modern context, algorithm is an automatic program to solve problems, which is widely used in public management, judicial adjudication, financial system, business and so on. On the one hand, the application of the algorithm improves the operation efficiency of the whole society; on the other hand, it also leads to the disorder of social management because of its complex technical characteristics. At present, the legislation of the algorithm regulation system in our country is relatively loose, and the supervision subject is not clear enough, and audit, as an independent and mature examination system, has great potential in the supervision of algorithm design, algorithm data protection and legal accountability of algorithm decision-making. It can solve the problems such as loose algorithm supervision system and unclear supervision responsibility. Under the audit supervision mode, professionals can formulate audit plans, conduct risk assessment and factual review according to specific audit standards, give audit conclusions, produce audit reports, conduct audit publicity, and implement accountability to relevant personnel and follow-up audit quality follow-up. By the combination of third-party audit, platform self-regulation and industry self-discipline, algorithm supervision can usher in a new breakthrough.

Key words: algorithm; algorithm risk; algorithm governance; algorithm audit

0 引言

在大数据时代,算法得到广泛应用,政府治理和社会运转均离不开算法的参与。当前 chatGPT 的开发使用使人工智能应用迎来了一波新的讨论,使很多问题通过智能化程序得到答案,其程度已远超人们的想象。人们在惊叹它巨大的工具价值之余,同时也会担忧它是否会代替人们工作而造成失业

问题,以及人的创造能力是否会因此退化等问题。

因此,一方面算法的应用为社会治理提供效率价值,另一方面算法复杂的技术特性决定了算法在运行中存在着诸多社会风险。我们亟需构建一套完整而独立的算法规制制度,在算法应用领域、算法开发、算法设计与算法决策等环节进行审计审查,使算法使用合规合法,合乎科技伦理,合乎人类文

明发展的长远利益。

1 算法定义

算法是将输入的数据转化为计算结果的编码程序。计算机对所输入的数据应用特定计算程序,“计算”出特定的算法结果。由于当下“大数据”“机器学习”技术的发展,算法含义的外延变得更加广阔,算法概念不仅包括线性的数学函数模型计算,也体现为利用大数据通过机器挖掘,自主“学习”得到算法决策结果。数据挖掘流程如图1所示。

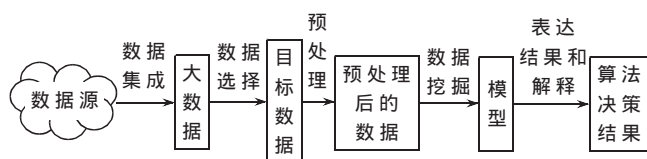


图1 数据挖掘流程图

不同于传统数学建模要求确定变量之间的函数关系,“机器学习”是通过大数据集、新型统计建模技术构建算法模型,识别数据之间的相关性,算法系统在“迭代”“学习”中不断校正决策结果,在输入和输出之间提供最强的预测关系。在大数据时代,数据越多,数据质量越高,算法结果越精确^[1]。

2 算法应用与算法风险

2.1 算法的应用

当前,算法在公共医疗卫生领域、司法系统、金融领域以及商业领域已有广泛应用。同时,算法的技术壁垒与商业化应用使人们难以厘清其运行逻辑,难以认识到其中暗含的风险。

2.1.1 算法在医疗卫生领域的应用

新冠疫情爆发之际,健康码以算法为依托得到了广泛的应用。健康码的诞生是政府基于治理需求和企业合作的典范,企业设计出算法工具以配合政府行政工作的开展,得到了广泛的应用^[2]。根据《关于依法科学精准做好新冠肺炎疫情防控工作工作的通知》,健康码是用于个人健康风险等级评定,获得出行、复工资格的法定证明,这说明健康码的运用是由政府主导的行政行为。在传统行政中,合法性行政程序要求政府决策公开透明,而算法的使用将原本人为的行政决策变为自动化不透明决策。但在这个过程中,算法应用的合规性需要得到审查。如在2022年底防疫新规出台之际,健康码屡屡被爆出作假,阳性患者显示阴性,甚至出现随便输入姓名也能得到阴性的健康码证明的情况,这种混乱的算法

程序应用无疑扰乱了防疫秩序,使行政行为的科学性与合法性遭受质疑,影响了政府公信力。

2.1.2 算法在司法系统中的应用

在司法环节中,无论是在发现犯罪嫌疑人,还是在案件取证审查环节以至于监狱管理系统都依赖算法进行治理。例如,在追逃工作中,人脸识别系统以算法技术作为支撑,通过对人脸数据的获取和识别进行算法处理,从而发现人员特征与目标吻合性,能够在逃犯追踪中锁定目标。实践证明,这一技术在客观上提高了办案效率,但是在这一过程中,个人身份认定信息从文本数据到图像数据一览无余,存在极大的个人数据泄露风险,因此,这一领域的个人信息以及隐私保护问题值得关注。此外,在日常的治安管理中,公安部门会利用算法系统识别并制作可视化的犯罪风险地图,从而做出加强特定区域社会治安监管力度的决策,起到预防犯罪和保护公民生命财产安全的作用。但是,如果在划定犯罪高风险地区时发生数据偏差,会影响到决策的合理性,反而事与愿违^[3]。

2.1.3 算法在金融领域的应用

算法在金融领域的应用也非常广泛。以蚂蚁金服为例,其借贷产品会根据大数据精准计算出客户需求、自动计算客户信用等级,向客户开放一定额度的超前消费额度或贷款。当每个商家和用户和蚂蚁金服产生借贷合同,蚂蚁金服会利用这些借贷合同对外进行抵押。因其业务量庞大,影响领域众多,这种模式下的借贷审查弹性很大,具有极大的坏账风险及社会伦理风险,长此以往可能引发金融风险^[4]。因此,对涉及重大社会风险的领域,算法审查是必要的。

除此之外,算法技术在其他公共领域和商业领域都有广泛应用,也普遍暴露出许多问题,比如媒体传播领域出现的“信息茧房”,电商平台“大数据杀熟”“价格操纵”以及外卖平台暴露出的“被困在算法里的骑手”等问题,都在警醒社会亟需关注算法应用问题。

2.2 算法风险种类辨析

在众多应用场景中,存在着相当复杂的算法风险,总结下来主要有以下几种。

2.2.1 算法黑箱

所谓“算法黑箱”指的是在数据的输入到数据的输出过程中,其技术复杂性带给人们的认知壁

垒。人们只能观测到它的输入与输出,而无法知道它如何将输入的内容变为输出的结果,至此形成黑箱,造成了算法决策的不透明^[5]。在这个过程中,平台通过算法对海量的数据进行收集和处理,如果数据出现瑕疵或被人为干预,算法决策会出现无法预知的偏离,算法决策的科学性与合理性无法得到保证。与此同时,数据安全也是重要衍生问题。对数据质量与算法设计进行审查是解决算法风险的第一步。

2.2.2 算法歧视

算法歧视是指算法设计者在编程时,有意或无意地将自身偏见、社会风气、制度体制以及文化差异嵌入算法之中,经过算法运算,产生了不公平、不合理的歧视性结果,进而不同程度地损害算法使用主体的利益。算法歧视种类主要包括算法设计者的偏见,先天训练数据的偏见以及技术性算法偏见,在算法涉及的所有领域中,几乎都会存在算法偏见的可能^[6]。国外学者伊格纳西奥指出,歧视性结果在刑事诉讼和就业领域尤为明显,例如 COMPAS 犯罪预测算法引发了种族歧视,Amazon 公司的招聘算法则涉及性别歧视。在我国,算法歧视也逐渐暴露,例如“大数据杀熟”就是典型的例证,不同消费习惯的人在同一平台消费,却价格不一,这瓦解了人们对于公平秩序的梦想,减少了对算法平台的信任。

2.2.3 算法霸权

网络空间领域存在技术先占的自我赋权,倒逼政府“承认”和作出制度变革^[7]。正如,各类平台凭借其运营模式、资金和技术优势创建了平台规则,赋予网络互动复杂的身份构建,形成了技术权力化倾向。在公共领域,公共部门在建设智慧政务、智慧司法、智慧检务中,算法技术担任重要角色,例如智能监控、人脸识别等智能辅助办案系统的应用。如今,算法绝不仅拘泥于技术领域,更在社会管理领域产生了重要影响。应当关注算法的权力化倾向以及商业的逐利特征导致的社会主体权利与义务的失衡问题。

2.2.4 算法合谋

2017 年,欧盟委员会针对电子商务行业的调查报告显示,超过一半的受访经营者表示会跟踪竞争对手的在线价格,其中约 2/3 已开始使用自行创建的算法程序或借助第三方提供的算法软件进行搜索跟踪,抓取竞争对手的产品价格或其他商业信

息。例如,在亚马孙公司的电商平台上,大多数的供应商已经在使用 Buy Box 算法自动定价。经营者使用计算机算法虽可改进定价模型、定制特色服务和预测市场趋势,提高竞争效率,但在缺乏监督的情况下,也极有可能出现竞争对手之间实施价格共谋,损害消费者利益,使其共同利润最大化。破解共谋经营者的责任问题也需要成为重点关注的领域。

2.2.5 算法伦理

在 chatGPT(chat Generative Pre-trained Transformer)迭代发展的今天,人们甚至已经可以用人工智能程序写出判决书。2023 年 1 月 30 日,哥伦比亚司法部卡塔赫纳区巡回法院第一劳工法庭就一桩未成年入疾病费用给付案件的相关问题使用 chatGPT 进行了回答,并以其回答文本作为说理依据,这加快了诉讼案件的处理效率。与此同时,法律职业已经遭到了人工智能的渗透,科技力量对法律职业进行了颠覆性改造,法律人何去何从,或许在不久的将来也会成为重要问题。

3 我国算法法律规制现状

目前,我国对于算法缺少体系化的监管,规制主体尚不统一。在法律层面,《民法典》从原则上规定“自然人的个人信息受法律保护”。《中华人民共和国消费者权益保护法》则具体规定了商业主体的义务,即未经消费者同意,不得擅自收集使用消费者个人信息以及商业主体保密以及信息泄露的补救义务。《中华人民共和国网络安全法》设立了网络运营者的责任和义务以及公民对个人信息的删除和更正的权利。在数据安全领域,《中华人民共和国数据安全法》规定在中华人民共和国境外开展数据处理活动,损害中华人民共和国国家安全、公共利益或者公民、组织合法权益的,依法追究法律责任,并提出“统筹协调国家数据安全的重大事项和重要工作,建立国家数据安全协调机制”,明确了数据安全的工作协调和统筹机制。这些法律规范在一定程度上明确了算法源头个人信息收集规范。

除此之外,《中华人民共和国电子商务法》等法律法规也为算法应用提供者规定了行为规范。例如,第十八条规定了有关个性化推荐的内容,要求对用户进行个性化推荐的同时也应当提供没有经过算法处理的非个性化的商品;第四十条规定了商品搜索排名的规范,要求商品排名应当存在多种方式,并提出应当对竞价排名类商品进行标识等。

《互联网信息服务算法推荐管理规定》规定了算法服务提供者的责任,为算法服务提供者划定了法律的红线。

由此可见,我国已经逐渐重视算法领域的相关立法,但是当前立法更多地关注于算法大数据以及商业行为,着眼于具体问题,缺乏体系化制度管理,存在监管主体分散的问题,需要多部门进行协调,管理成本较高。

4 审计参与算法治理

4.1 审计参与算法治理辨析

目前,对于算法的规制大体存在两种模式:一种是个体赋权模式,另一种是外部问责模式。个体赋权模式以欧盟为代表,例如《通用数据保护条例》(GDPR)赋予了个人在一定条件下免于算法自动化决策和退出算法自动化决策的权利;外部问责模式以美国为代表,以2018年算法问责法的通过为标志,该法律规定了自动化决策工作组为算法问责制的实施主体,工作组由具有专业知识的技术人士组成,其职责包括在成立后的18个月内向市长提交算法评估标准、救济程序、确保公众知情权、参与权的相关程序和技术要求,类似一种审计方式^[8]。在当前算法应用已经普遍渗透到了生活的方方面面,算法风险也无处不在的前提下,个人赋权模式往往效果有限,必须以强有力外部问责模式为主进行算法监管,兼顾个人赋权。在外部问责模式下,我国的算法监管主体是分散的,但审计体系成熟且独立,工作流程完善,可由审计机构主导进行算法审计审查,有效回应算法风险产生的社会管理问题。审计作为重要的监管方式,在国有资产和社会财会秩序监管中发挥了重要作用,具有极强的工具价值,同时在算法治理中也具有广阔的发展空间。

在法律依据上,我国《个人信息保护法》第五十四条规定:“个人信息处理者应当定期对其处理个人信息遵守法律、行政法规的情况进行合规审计”。该条规定了在算法系统的数据端,平台应当利用审计的方式进行自我规制,这为研究算法审计制度在法律层面提供了启发和指引。

4.2 算法审计之设计

4.2.1 标准制定

审计标准即审计依据,是据以作出审计结论、提出处理意见的客观尺度,具有权威性、层次性等特征,按照性质、内容的不同可以分为法律、法规、

规章等。此外,审计准则作为审计人员进行审计工作的行为规范和专业标准,也是审计标准的重要组成部分。算法审计不属于传统的审计范围,与传统财务审计相比在审计属性上也有较大区别,需要制定新的审计标准。在法律层面,立法者可建立普适大部分领域的算法审计原则和规则,在行政规章、准则中制定各领域的算法审计规范。国家应当关注算法审计在国家治理中的价值,由专业人士参与制定相关标准,出台相关文件,促进算法审计的开展与实施。

4.2.2 人员设置

由于审计领域的特殊性,要求参与算法审计的审计人员具备相应的专业知识。当前,从事常规审计的审计人员难以满足这一要求,因此,针对算法审计应当作出不同的人员设置与安排。比较可行的做法是审计部门引进从事算法研究的研究人员,或聘请社会第三方专业人士参与算法审计工作,由专业人员对整个应用算法进行治理和决策的系统进行评估。例如,在推荐算法中,不同影响因素不同程度地决定实际业务价值,模型的影响面大概是10%,产品设计占40%,数据占30%,领域知识的表示和建模占20%^[9],如图2所示。审计人员需要在重点领域进行重点审查,降低由商业逐利行为引发的社会伦理风险。

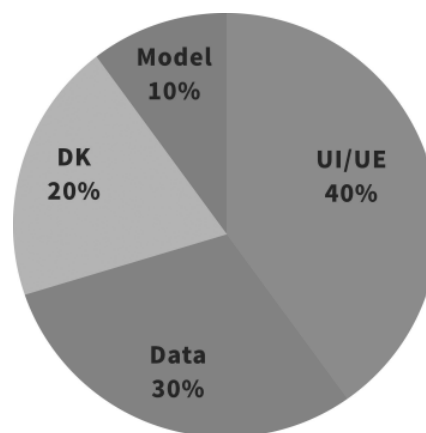


图2 不同因素在实际业务价值中的影响比例

4.2.3 实施全过程审计

由于算法风险发生在算法运行的各个阶段,需要进行算法运行的全过程审计,包括应用领域、数据端、算法设计以及算法结果运用。应当由专业人士对相应领域进行审计,遵循审计的工作流程,

制定审计计划,开展审计调查,得出审计结论,制作审计报告,进行审计公开以及审计质量跟踪,并实现对相关人员的问责^[10]。具体工作规定还应有进一步的讨论。但是在审计已有完整成熟的工作流程的背景下,实施算法审计并不是“空中楼阁”,应当有勇气对审计制度进行进一步的创新。

开展算法审计将会对公共领域治理以及企业内部控制、风险管理起到重要作用,化解算法黑箱、算法歧视、算法霸权、算法合谋、算法伦理等算法风险,提高国家在算法自动化决策领域的治理能力和治理水平。在公共领域,算法审计可以由审计部门及相关内部审计部门组织实施,在其他商业领域,国家可进行社会审计指导,制定相应的审计政策,推动有资质的社会第三方机构参与商业领域的算法审计,使其成为企业合规的一部分。同时,推动行业自律协会的发展,在人工智能算法领域实现互相监督,由国家、社会、行业三方合力,共同维护算法应用领域的公平正义。在这一构想下,审计制度势必要实现新的发展,拓展新的内涵,而审计制度的技术性特征也决定了其应具有广泛的应用场景。在算法技术飞速发展的今天,算法应用日新月异,规模不断扩张,形成一套完整而独立的算法审计审查制度已经刻不容缓,算法审计将成为现代社会治理的重要一环。

参考文献

[1] House of commons science and technology select com-

mittee.Algorithms in decision making[R].London:House of Commons.

- [2] 许可.健康码的法律之维[J].探索与争鸣,2020(9): 130-136.
- [3] YEUNG K.Algorithmic regulation:a critical interrogation[J].Regulation & Governance,2018,12(4).
- [4] 张金茹.蚂蚁金服融资业务的风险分析[J].中外企业家,2020(19).
- [5] [美]弗兰克·帕斯奎尔.黑箱社会掌控信息和金钱的数据法则[M].赵亚男,译.北京:中信出版社,2015.
- [6] BATYA F,NISSENBAUM H.Bias in computer systems[C]//ACM Transactions on Information Systems,1996.
- [7] 马长山.智慧社会建设中的“众创”式制度变革——基于“网约车”合法化进程的法理学分析[J].中国社会科学,2019(4):75-97.
- [8] 周围.算法共谋的反垄断法规制[J].法学,2020(1):40-59.
- [9] 马长山.数字法治概论[M].北京:法律出版社,2022.
- [10] 刘旺洪.审计法学[M].北京:高等教育出版社,2020.

(收稿日期:2022-12-13)

作者简介:

付冉冉(1997-),女,硕士研究生,主要研究方向:审计法学。



版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.cn