

算法悖论与制度因应

——基于用户算法应用感知的实证研究

许可程华

摘要: 尽管我国算法治理的基本架构已经初步形成,但相关实证研究依然薄弱。我国用户算法应用感知大规模问卷调查与结果分析,反映出民众对算法风险的高度关注,也揭示出不同群体对算法关注的分歧、同一群体算法关注与算法行为的矛盾以及算法推荐便利与算法关注的取舍。面对“算法悖论”的种种表现形式,应根据其内在机理予以制度回应,将“算法向善”嵌入“业务流程”,从“算法透明”迈向“算法素养”,经“过程公平”落实“算法公平”,借“算法问责”铸就“算法安全”,最终形成熨帖公众真实需求、平衡多元目标的中国算法治理体系。

关键词: 算法感知; 算法治理; 算法悖论

DOI: 10.19836/j.cnki.37-1100/c.2022.06.008

如果说2021年被公认为是中国算法治理元年,那么,随着《关于加强互联网信息服务算法综合治理的指导意见》(下称《算法综合治理意见》)、《互联网信息服务算法推荐管理规定》(下称《算法推荐管理办法》)的实施,以及国家网信办“清朗·2022年算法综合治理”专项行动的推进,2022年可以说是中国算法治理落地生根之年。与如火如荼的规范研究与规则制定迥异,系统性地观察、收集经验事实,进而以此为对象开展分析和推论的算法实证研究尚属冷门。为此,本文从“经验事实可作为规范理由”^①的理论出发,采取问卷调查方法,描述我国用户对企业算法应用的具体感知,发现其真实关注和行动困境,探索其内在原理,希冀有裨于我国算法治理体系的反思与完善。

一、用户算法应用感知调查:设计与结果

(一)用户算法应用感知的实证研究回顾

我国用户算法应用感知的既有调查,多集中在“差别化定价”“信息茧房”等具体场景上。例如,中国青年报社社会调查中心对2008名受访者的调查显示,51.3%的受访者遇到过互联网企业利用大数据“杀熟”的情况,59.1%的受访者希望价格主管部门立法规范歧视性定价行为^②。吴志艳等人基于310名美团平台用户的招募和问卷,发现算法价格歧视导致用户感知的价格公平性降低和背叛行为增加^③。喻国明、方可人使用2019年全民媒介接触与使用暨媒介价值观调查数据,发现算法并未导致信

基金项目: 对外经济贸易大学优秀青年学者培育计划(18YQ10);中国人民大学“双一流”跨学科重大创新规划平台“数字经济跨学科交叉平台”支持项目。

作者简介: 许可,对外经济贸易大学法学院副教授(北京100029; xuke_lawyer@foxmail.com);程华,中国人民大学经济学院副教授(北京100044; chenghua1220@ruc.edu.cn)。

^① 张永健:《法实证研究:原理、方法与应用》,台北:新学林出版有限公司,2022年,第69页。

^② 《中国青年报:51.3%受访者遭遇过大数据“杀熟”》, <https://baijiahao.baidu.com/s?id=1594957994108562654&wfr=spider&for=pc>, 访问日期:2022年5月16日。

^③ Wu Z, Yang Y, Zhao J, et al, “The Impact of Algorithmic Price Discrimination on Consumers’ Perceived Betrayal”, *Frontiers Psychol*, 2022, 13, pp.825420.

息茧房,而是为个体提供了多元理性的信息世界^①。在《算法推荐管理办法》出台后,《光明日报》联合中央民族大学调研组就用户对算法推荐的感知开展问卷调查,并提出了提升算法素养的建议^②。

与我国实证研究聚焦于“算法推荐”不同,西方学者多以“算法决策”对用户的影响为研究重点。2018年,皮尤研究中心就大数据信用评级、犯罪风险评估、简历筛选、面试评价等算法决策典型场景中的用户接受性进行了调研^③。其他的一系列研究则进一步表明,用户对算法决策的态度因结果、性质和领域而大相径庭^④。

回顾国内外研究,鲜有对用户算法感知的普遍性调研,亦缺乏对算法治理体系的整体把握。在我国算法律架构初成并持续完善的背景下,立足我国实践、以问题为导向的算法感知实证研究显得尤为迫切而重要。一方面,如果说法律制度以保障人民福祉为鹄的,那么用户对算法应用的真实感知便是“可触、可感之法”的实然基础;另一方面,衡量法律实效是评估算法制度有效性的前提性作业,对公众算法感知的系统性、历时性观察,将成为判断法律预设目标完成程度的最佳指针,进而有裨于我国算法规范之改进。

(二)问卷设计:理论、结构与问题

作为对算法律的回音,问卷从我国算法治理的价值立场、法律原则和监管导向出发,以《算法综合治理意见》《算法推荐管理办法》所确立的“算法透明、算法安全、算法公平、算法向善”为经,以“用户认知—用户态度—用户权利—用户行动”为纬,共设置了27个问题。此外,为了更准确地刻画不同群体对算法的感知,问卷还设计了个人基本信息的6个问题,包括性别、年龄、职业、地域、学历、职业和年收入。

“算法透明”是全球公认的首要算法价值。对全球算法治理文件的梳理发现,84种文件中有73个支持“透明原则”^⑤。我国《算法综合治理意见》第13条亦将“推动算法公开透明”作为重要立法目标,督促企业及时、合理、有效地公开算法基本原理、优化目标、决策标准等信息,做好算法结果解释,畅通投诉通道,消除社会疑虑。以此为基础,并结合《个人信息保护法》第24条和《算法推荐管理办法》相关条款,问卷设计了“您是否知晓企业在提供互联网服务时使用了算法?”“您了解企业使用算法的内容和目的吗?”“您认为企业需要向用户解释算法吗?”“若企业根据个人特征标签(如性别、喜好、地理位置、交易和浏览记录等)向您推荐商品,您的态度如何?”“当您对企业算法结果(如用户星级、信用分等)有异议时,您希望企业进行人工介入重新核查吗?”等问题。

“算法安全”是独具中国特色的算法价值^⑥。《算法综合治理意见》以“建立健全算法安全治理机制和算法安全监管体系”为宗旨,以“算法自身安全、算法安全可控、算法应用安全”为框架,防范算法滥用带来意识形态、经济发展和社会管理等方面的风险,防止利用算法干扰社会舆论、打压竞争对手、侵害网民权益等行为,维护网络空间传播秩序、市场秩序和社会秩序。基于《算法综合治理意见》《算法推荐管理办法》,问卷设计了“您认为企业使用算法对用户权益造成损害的风险状况如何?”“您是否认为算法在向您推荐个性化的广告、视频、新闻时侵犯了您的个人隐私?”“您认为以下哪类企业的算

① 喻国明、方可人:《算法型内容推送会导致信息茧房吗?——基于媒介多样性和信源信任的一项实证分析》,《山东社会科学》2011年第11期。

② 毛湛文、白雪蕾:《如何远离算法之“算计”?——新规之下的网民算法素养调研》,《光明日报》2022年4月7日,第7版。

③ Pew Research Center, “Public Attitudes Toward Computer Algorithms”, 2018.

④ Logg J., Minson J., Moore D., “Algorithm Appreciation: People Prefer Algorithmic to Human Judgment,” *Organizational Behavior and Human Decision Processes*, 2019, 151, pp.90-103. Yalcin G. et al, “Thumbs Up Or Down: Consumer Reactions to Decisions By Algorithms Versus Humans”, *NA-Advances in Consumer Research*, 2022, 48, pp.1155-1159.

⑤ Anna J., Ienca M., Vayena E., “The Global Landscape of AI Ethics Guidelines”, *Nature Machine Intelligence*, 2019, 9(1), pp.389-399.

⑥ 许可:《算法规制体系的中国建构与理论反思》,《法律科学(西北政法大学学报)》2022年第1期。

法存在较严重的损害用户权益的现象?”“您是否了解国家有关算法安全的相关规定和公民自身所拥有的合法权利?”“为避免算法可能带来的不利影响,您通常会如何做?”“您认为可以如何避免算法可能带来的安全隐患”等问题。

“算法公平”亦是广为接受的算法价值。作为一个宽泛的概念,《算法综合治理意见》和《算法推荐管理办法》下的算法公平包含着避免个体歧视、实现结果公平和保护弱势群体的多重内涵。鉴于本次调研对象是一般用户,问卷主要考察了民众对算法结果公平尤其是“差别化定价”的感知。为此,问卷设计了“您经历过企业对相同产品或服务对您和其他人收取不同价格吗?”“您感觉企业在定价上是如何进行差别化对待的?”“您对企业按照收入水平高低、新老用户、接受服务频率的高低或是否会员进行差别化定价的行为,是何态度?”“您感觉哪种类型的企业差别化定价的现象比较严重?”“当您察觉到自己遇到差别化定价对待后,您会怎么做?”等问题。

“算法向善”同样有着鲜明中国特色。《数据安全法》第 28 条要求,新技术研究开发应有利于促进经济社会发展,增进人民福祉,《算法综合治理意见》《算法推荐管理办法》亦将“引导算法应用向上向善”作为立法愿景,并进一步细化为“算法为信息内容服务”“算法为良好生活服务”“算法为老人和未成年人服务”等具体规范。基于此,问卷设计了“您有过受企业推送广告的影响而购买了不必要的产品或服务的经历吗?”“您收到不良信息(谣言、低俗、色情等)的频次是多少?”“您认为目前企业利用算法进行人为的信息扭曲情形多吗?”“依赖企业自动排序进行阅读或观看,您在获取信息、知识方面的感觉如何?”“您认为哪一类企业在‘适老化’‘未成年人模式’等人性化设计方面做得比较好?”“面对不良信息您会主动向企业举报吗?举报后,不良信息出现的频次会减少吗?”等问题。

(三)问卷的发放与回收

本调查于 2021 年 12 月,通过在线问卷平台“问卷星”和“支付宝”应用程序发放,共发放问卷 1500 万份,收到有效问卷 6941 份。尽管应答率偏低,但有效应答的参与者在性别、年龄、地区、收入、学历等方面分布广泛,反映了我国网民总体的多样性(见表 1),相关应答率与“无应答偏差”(nonresponse bias)并无直接关联^①。更重要的是,鉴于本问卷以“对算法有真实感知的人群”为中心^②,因此参与人“选择性回复”以及所导致的样本在性别、年龄、地域的偏差,可被“弱应答样本代表性”(R 指标)所解释^③,其统计结果具有有效性和可信性。

表 1 调查对象整体情况

性别	男性(67%)	女性(33%)			
年龄	18 岁以下(3%)	18—25 岁(31%)	26—40 岁(39%)	41—60 岁(23%)	60 岁以上(4%)
地域	直辖市(19%)	省会城市(23%)	地级市(24%)	县级市或县(28%)	其他(6%)
学历	高中以下(37%)	专科(25%)	本科(29%)	研究生以上(9%)	
职业	学生(19%)	机关事业单位人员(13%)	企业工作人员(27%)	自由职业者(24%)	其他(17%)
年收入	2 万以下(31%)	2 万—5 万(21%)	5 万—20 万(38%)	20 万—50 万(6%)	50 万以上(4%)

① Groves R., Heeringa S., “Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs”, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2006, 169(3), pp.439-457.

② 例如,18 岁以下未成年人在线上购物、本地生活、游戏等领域受到无独立收入及未成年保护因素的限制,不是主要的消费人口,而 60 岁以上人群对电商、本地生活、线上社交等服务的接受也有限。

③ 任莉颖、邱泽奇、丁华、严洁:《问卷调查质量研究:应答代表性评估》,《社会》2014 年第 1 期。

(四)调查结果分析

1.算法透明方面

在认知方面,用户对互联网服务算法了解水平有限。超过半数的受访者对企业是否使用算法不清楚,六成受访者不了解企业使用算法的内容和目的。其中,16.90%的受访者完全不清楚企业是否使用算法,39.08%的受访者则不太清楚;20.34%的受访者完全不了解企业使用算法的内容和目的,44.64%的受访者则不太了解。这说明多数受访者对企业算法应用有一定程度的感知,但处于了解浅显、认知模糊的阶段。

在态度方面,对于基于算法的广告推荐,受访者表现出接受态度,只有7.17%的受访者认同不使用算法的统一广告推送。但另一方面,当问及使用性别、喜好、地理位置、交易和浏览记录推荐商品时,更多的人表达了忧虑,28.99%的人希望关闭该功能。

在权利方面,用户对算法解释权的诉求强烈,只有7.22%的受访者认为不需要解释,其余都认为需要不同程度的解释。

在行动方面,如用户对算法结果有异议时,超过3/4(76.48%)的受访者希望人工介入,只有11.55%的受访者认为人工不如算法准确。

2.算法安全方面

在认知方面,近八成的受访者认为算法应用存在可能损害用户权益的风险,20.28%的人认为风险很高,接近60%的人认为存在一定风险,只有不到5%的人认为不存在风险。关于个性化推荐是否侵犯隐私这一更具体的风险感知问题进一步佐证了该结论,超过60%的人表示赞同,而反对的比例只有12%。

在态度方面,用户对不同互联网行业的算法风险感知度存在一定差异。55.86%的受访者认为电商存在严重损害用户权益现象,认为社交平台、视频媒体、金融机构和搜索平台存在这一现象的受访者也分别达到47.51%、44.72%、44.5%和44%。

在权利方面,受访者对算法安全法规及自身合法权益了解有限。受访者中仅有7.95%的人亲自阅读过相关法规,36.14%的受访者只听说过部分内容,25.08%的受访者仅听说过名字,高达30.83%的受访者完全不了解我国立法。

在行动方面,受访者表现出较强的自我保护意识,并对企业自我主动管理寄予厚望。为规避算法带来的不利影响,受访者主动管理企业数据收集请求、拒绝不合理请求的比例高达47.27%,选择关闭个性化推荐的比例也超过了45%。同时,有60.12%的受访者希望对企业加强安全算法评估,位居各项选择之首。此外,用户也提出了强化企业使用算法的法律责任、加强企业监管以及限制企业数据搜集范围等明确主张。

3.算法公平方面

在认知方面,受访者认为差别化定价普遍存在,仅22.16%的受访者表示没有经历过差别化定价,分别有51.46%和26.38%的受访者偶尔或经常经历差别化定价。在经历过差别化定价的受访者中,最多经历的是新老客户定价不一致的情形(59.26%),这既包括“回馈老客户”活动,也包括“吸引新客户”活动;其次是依据是否活跃客户、是否会员而定价不一(56.19%和53.54%),以及根据收入水平的差别化定价(34.88%)。此外,用户对互联网不同行业的差别化定价情况有不同认知,其中,受访者认为电商企业最常采用差别化定价(60.96%),而音乐娱乐平台则最少采用(22.30%)。

在态度方面,当遭遇差别化定价后,受访者总体反映比较负面,有55%的受访者表示很在意差别化定价,不过,也有超过30%的用户表示理解和忍受。同时,用户对不同的差别化定价行为态度也不相同,其中最被认可的是根据是否是会员进行定价,69%的受访者表示认同或中立;最不被认可的是按照收入水平定价,有44%的受访者表示不太认同或完全不认同。不过,即使是认同度最高的差别化定价行为(根据是否是会员定价),也有31%的受访者表示反对,而在不认同度最高的差别化定价行为(根据收入高低定价)中,也有30%的受访者表示认同,说明公众尚未就判断差别化定价的实质合理性标准达成共识。

在权利和行动方面,绝大多数用户会积极维护自己的权益,一半以上(55.06%)的人会选择“减少或暂停对该企业产品或服务的使用”,有14%的人会选择曝光或投诉相关企业。

4.算法向善方面

在认知方面,用户对不良信息感受明显,过半数受访者在使用App的过程中收到过不良信息(15%经常收到,37%偶尔收到,32%收到比较少)。同时,用户对企业利用算法故意造成信息扭曲的行为感知强烈。超过80%的受访者认为企业利用算法人为扭曲信息的情况很多或比较多。此外,用户认为企业面向“老人和未成年人”等网络弱势群体开展的人性化设计不如人意。在行业细分领域中,视频媒体和社交平台评价相对较高,而在线旅游平台和搜索平台评价相对最低,金融机构也只得到了14.58%的受访者认可,更有甚者,有35.18%的受访者认为所有行业都难称满意。

在态度方面,用户对“算法推荐信息和诱导消费”比较中立,算法诱导过度消费问题并不突出,“信息茧房”因人而异。根据问卷结果,85%的受访者没有或很少在推荐算法的引导下购买不必要的产品或服务。需要追问的是,这一态度是否因用户“有限理性”被“套路”而不自知?对比剩余15%的受访者情况,在逐一分析年龄、职业、收入水平等变量后,未发现年轻人、学生、低收入群体更容易被诱导或陷入错误感知的证据。此外,受访者在获取知识和信息方面的感受存在明显差异,40%的人认为推荐算法有助于获取丰富的信息、节省信息搜寻时间等;25%的人认为算法推荐的信息较为同质,限制了知识获取的多元性;18%的人认为推送内容过于娱乐化,容易导致上瘾;还有17%的人对算法推荐有效性存疑,认为不如自己查找信息。

在权利和行动方面,用户对不良信息侵入的权利意识不强,53%的受访者表示在面对不良信息时没有举报,而在举报过的受访者中,2/3的人表示虽然举报过但不良信息出现的频次并未减少,说明企业对用户反馈信息的重视程度低,投诉机制有待完善。

二、理论意蕴:算法悖论的提出

(一)用户的算法认知、态度及行为背离

通过对调查结果进行逻辑分析和分组交叉研究,我们从数据中发现了民众关于算法认知、态度和行为之间的背离,本文称之为“算法悖论”。

首先,用户对算法的认知与态度间存在背离。如表2所示,有55.98%的用户表示对企业使用了算法不太清楚或完全不清楚,高达64.98%的用户对企业使用算法的目的表示不太了解或完全不了解,但当被问到对企业使用算法的态度时,极高比例的用户表示企业使用算法会对用户权益造成损害、侵犯隐私及操纵信息等,对算法表示出风险很高的负面评价。图1给出了相关调查结果的结构,可以看到,79.39%的用户认为算法侵犯了用户权益,60.33%的用户认为推荐性算法侵犯了自己的隐私。这意味着,一方面,有相当比例的用户对企业使用算法的行为不了解,但另一方面,对算法使用却抱有明显的负面态度。

表2 受访者对算法的基本认知

您是否知晓企业在提供互联网服务时使用了算法?			
完全清楚	比较清楚	不太清楚	完全不清楚
21.58%	22.44%	39.08%	16.90%
您了解企业使用算法的内容和目的吗?			
完全了解	比较了解	不太了解	完全不清楚
13.04%	21.98%	44.64%	20.34%

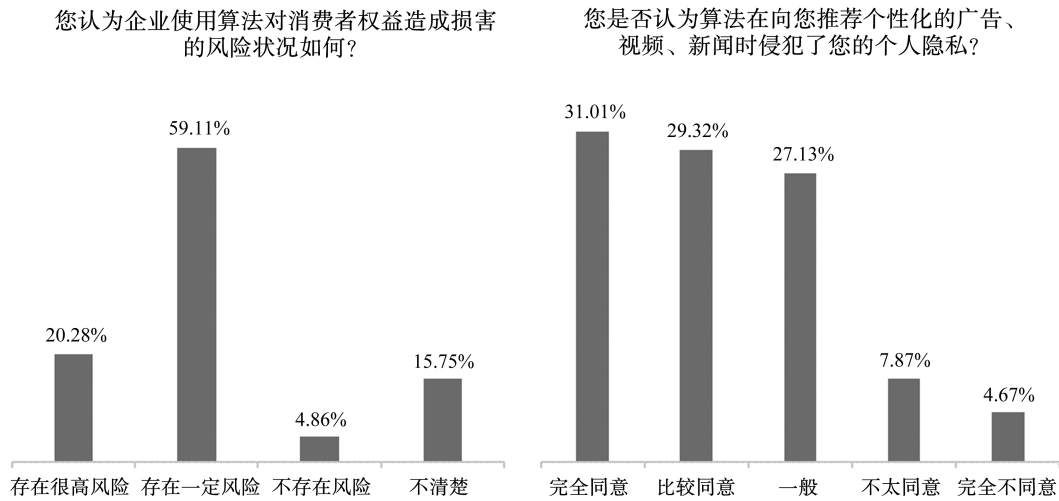


图 1 受访者对于算法风险的态度

特别地,我们对问卷数据做进一步分析,把用户分为两组,一组对企业使用算法的内容和目的的认知度较高,另一组认知度较低。我们发现,前一组用户关于风险和隐私侵犯的负面态度弱于后一组,意即有相当一部分用户虽然对算法“无知”,但有先入为主的负面评价(表 3)。

表 3 认知度不同用户的态度差异

	认为算法使用对用户权益造成很高风险或一定风险	认为推荐性算法侵犯个人隐私
高认知度组	75.55%	55.67%
低认知度组	82.39%	62.33%

其次,用户对算法的态度与行为之间存在背离。其一,虽然用户在态度上认为算法会侵犯用户权益,但在行为上却倾向于选择算法带来的便利。约 80%的受访者认为,企业使用算法可能会损害用户权益、侵犯个人隐私或造成人为信息扭曲,但相当比例的用户并不抵触用算法获得服务、产品和体验。例如,只有 7%的用户赞成对所有人推送相同广告,有 1/4 的用户赞成“千人千面”的推送,其他用户也在一定程度上接受企业使用个人信息分类推荐。此外,当被问到企业使用算法给个人造成的影响时,用户的实际感受也呈现出相对正面的结果,85%的用户表示不会因为自动推荐产生过度购买的行为,40%的用户感觉自动信息推送精准、对己有益。一般而言,相当多的用户对算法抱有负面评价,但涉及具体的场景,用户行为表现出对推荐算法的容忍和接受。其二,用户虽然认为企业使用算法会引发普遍风险,但行为上表现比较被动消极,缺乏保护自身权益及掌握算法规定的主动性。调查显示,85%的用户受到过不良信息的侵扰,但其中超过一半的用户从未主动举报。

最后,在认知、态度和行为的背离上,60 岁以上老年人相比于年轻人,表现更加明显。年轻人表现出比老年人更高的算法认知度,对算法更多怀有好奇和包容的态度,对算法持有相对信任的态度。在接受个性化推荐、差别化定价方面,年轻人更加理性,在自我权益保护方面,年轻人也更加积极。该事实说明,“数字原住民”的算法素养较高,而“移民一代”则存在一定程度的认知和行为偏差。

(二)算法悖论:渊源与类型

算法悖论并非全新事物。实际上,与之密切联系的概念——“隐私悖论”(privacy paradox)早已被

人们揭示^①。2006年,Barnes首次使用“隐私悖论”指代社交网站上青少年轻易披露个人隐私与成年人对网络隐私泄露顾虑的差异,其实质是不同人群对待隐私的不同处理方式^②。之后,人们从“隐私关注”(privacy concerns)暨“人们针对隐私泄露、隐私侵害相关联的意识和主观感受”出发,将所有隐私关注与隐私相关行动不符的情形称为“隐私悖论”。研究发现,用户隐私行为并不总是与他们的陈述相吻合,在特定场合下,用户常常忘记或者降低隐私关注的级别,有时甚至会随意地、没有任何理由、不加以任何防范地披露隐私^③。大量的实证研究已经证实了这一现象。2001年,施皮克曼(Spiekermann)模拟了线上购物产品咨询的情境,通过聊天机器人与用户的互动观察用户披露信息的行为。研究结果令人惊讶。在声称极其担忧隐私的“隐私原教旨主义者”(privacy fundamentalists)、策略性对待隐私的“实用主义者”和隐私“略微关注者”中,24%—28%的“隐私原教旨主义者”在与机器人交互前就自愿提供了家庭地址,30%—40%的“实用主义者”在未做任何隐私保护措施的情况下就提供了家庭地址^④。这并非孤例。2019年,CIGI-Ipsos《全球互联网安全与信任调查》的研究表明,人们担忧隐私泄露风险与信任互联网从而愿意分享个人信息比例存在显著偏差^⑤。以至于有学者直陈“人们对隐私关注与他们的隐私行为无关”^⑥。近年来,随着移动互联网、用户画像(profileing)技术以及推荐算法的应用,“个性化—隐私悖论”(personalization-privacy paradox)成为重要的隐私悖论形式。在信息过载和注意力日益稀缺的背景下,用户既享受个性化带来的更好服务,又担忧隐私遭到侵犯;相反,企业既获取通过个性化服务提高竞争力所创造的利益,又面临隐私顾虑所造成的用户流失。就此而言,“个性化—隐私悖论”产生于企业个性化服务与用户隐私之间的拉锯^⑦。

根据上述研究,我们不妨将“算法关注”(algorithm concerns)暨“人们针对算法歧视、算法滥用等算法风险相关联的意识和主观感受”与算法相关行为的扞格,统称为“算法悖论”,进而将之具体化为三种情形:(1)不同群体“算法关注”的分歧;(2)同一群体“算法关注”与算法行为的矛盾;(3)算法推荐便利与算法关注的取舍。

(三)算法悖论的成因阐释

1.作为一种发展偏好的“算法关注”

算法关注系对科技风险的后天感受,其并非源于个体风险规避的先天偏好,而是被人生经历、知识结构、生活习惯和文化背景所塑造,这解释了为何不同人群对算法持有不同的态度。对支付宝用户隐私感受的研究发现,隐私担忧可能是在使用数字服务的过程中形成的一种偏好,当用户逐渐从使用App提供服务中产生收益和乐趣时,他们也开始对App数据收集与共享的潜在风险产生更多担忧。简言之,用户的隐私关注可能会随着其数据的积累而提升^⑧。将此理论运用于算法中,可以将接受算法服务的频率和对算法的了解程度作为变量,去解释不同群体在“算法关注”方面的分歧。

① “隐私悖论”系在英文语境下的理论,其“隐私”概念系广义,包括了《民法典》下的隐私、个人信息等相关内容。

② Barnes S., “A Privacy Paradox: Social Networking in the United States”, *First Monday*, 2006, 11(9), pp.86.

③ Günther O., Spiekermann S., “Privacy in E-Commerce: Stated Preferences vs. Actual Behavior Bettina Berendt”, *Communications of the ACM*, 2005, 48(4), pp.101-106.

④ Spiekermann S., Grossklags J., Bettina B., “E-privacy in 2nd Generation Ecommerce: Privacy Preferences Versus Actual Behavior”, *Proceedings of the 3rd ACM conference on Electronic Commerce*, 2001, pp.38-47.

⑤ CIGI-Ipsos, “Global Survey on Internet Security and Trust”, <https://www.cigionline.org/cigi-ipsos-global-survey-internet-security-and-trust>, 访问日期:2022年5月16日。

⑥ Tufekci Z., “Can you See me Now? Audience and Disclosure Regulation in Online Social Network Sites”, *Bulletin of Science Technology & Society*, 2008, 28(1), pp.20-36.

⑦ LEE C. H., Cranage D., “Personalisation-privacy Paradox: the Effects of Personalisation and Privacy Assurance on Customer Responses to Travel Web Sites”, *Tourism management*, 2011, 32, pp.987-994.

⑧ Chen L., Huang Y., Ouyang S., et al., “The Data Privacy Paradox and Digital Demand (2021)”, <https://www.nber.org/papers/w28854>, 访问日期:2022年5月16日。

调研结果显示,在各年龄段的群体中,对算法接触最少的60岁以上受访者对算法可能带来的风险感知度最低,只有12%的人认为算法存在高风险。在汇聚个人信息用于“差别化定价”的场景下,针对算法使用个人特征标签的问题,“研究生及以上学历水平的受访者”更加倾向于约束企业行为,选择“允许企业自主选择标签及企业如何使用标签无所谓”的比例,高中及以下群体为23%,而本科及以上学历群体大致为13%,低10个百分点。在针对算法可能产生的负面影响的应对方面,高学历受访者更为主动和坚定,相对更有章法。随着学历的升高,为避免算法可能带来的不利影响,更高比例的受访者选择关闭个性化推荐,而非尽量多样化地浏览企业内容,或选择直接设法规避算法对个人的个性化认知与计算,而非在更多的浏览中让算法越来越了解自己。对比来看,在高中及以下学历的受访者中,没有使用过所列示的任何做法来主动规避算法潜在不利影响的最高(16%)。

2. 陷入扭曲操纵的“算法行为”

“行为扭曲和操纵理论”试图从不恰当影响算法行为的主客观因素出发,揭示算法关注和算法行为的背离。就外部因素而言,企业往往通过设定菜单默认值、强制注册、晦涩语言、影子档案等应用交互界面的“暗黑模式”(dark patterns),将互联网的架构环境转化为武器,给用户选择提供假象,通过心理操控和变相形式影响用户决定。可事实上,如果用户完全知情,且有能力选择替代方案,其可能不会作出这些选择^①。作为一种微观权力形态,企业可以通过算法强化对用户的影响力和控制力,借由评级、分类、预测,潜移默化地左右着用户的选择^②。

从内在因素出发,用户的算法行为将面临信息不完整/不对称、有限理性和系统性认知偏差的困境^③。所谓“信息不完整/不对称”,意指用户对算法存在和其性质缺乏充分了解,难以准确判断算法风险的大小,同时也不清楚是否存在其他的替代技术和保护性方案。而当用户倾向于采用一种相当近视的观念时,其实际决策就被扭曲。“有限理性”指的是用户根本无法处理与算法成本和收益相关的所有非确定性信息,也难以预计其策略选择的回报程度。为此,个体不得不求助于直觉、常识、猜测等“简化的心智模型、近似策略和启发法”^④,由此引发了“系统性认知偏差”。行为经济学的研究表明,乐观偏见、影响式启发、双曲贴现、框架效应等均会导致判断错误^⑤。其中,“乐观偏见”是指个体倾向于表现出对自己算法保护技能和知识的过度自信,认为自身面对的算法风险较小;“影响式启发”即个体倾向于低估他们喜欢事物带来的风险,高估他们不喜欢事物带来的风险,这部分解释了未成年人对算法的不同态度。调研显示,较诸其他年龄段的人群,数字原住民对算法的风险感知度很低,对算法错误也持相对开放包容的态度。“双曲贴现”意味着人们以不一致的方式评估遥远和临近的事件对其造成的影响,当被问及个人是否打算采取保护策略时,个人可能认为算法风险更为重要,但是当面临使用算法所带来的好处时,个人的偏好改变,而选择获得眼前既得利益。“框架效应”使得人们的算法决策因信息呈现方式的变化而变化,即便是客观风险不变,只要以不同的方式描述算法风险和收益,用户就会作出不同的选择。作为引导人们感知和重现现实的认知结构,当前社会对算法广泛的批评构成了算法认知的“基本框架”,对算法的普遍性未知进一步放大了算法忧惧,加之大多数受访者不了解国家有关算法安全法律法规,无法通过法律救济途径保障自身权益,往往只能通过拒绝算法应用以减少算法风险,陷入“要么全盘接受,要么彻底拒绝”的“假两难困境”(false dilemma)。

① Ari E. W., “Cognitive Biases, Dark Patterns, and the ‘Privacy Paradox’”, *Current Opinion in Psychology*, 2020, 31, pp. 105-109.

② 周辉:《算法权力及其规制》,《法制与社会发展》2019年第6期。

③ 该表述借鉴了隐私悖论相关研究,参见 Holland B., “Privacy Paradox 2.0”, *Widener Law Journal*, 2010, 19, pp.893-932.

④ Acquisti A., Brandimarte L., Loewenstein G., “Privacy and Human Behavior in the Age of Information”, *Science*, 2015, 347(6221), pp.514.

⑤ 申琦、邱艺:《打开隐私悖论背后的认知黑箱》,《西南政法大学学报》2021年第5期。

3. 基于成本收益的“算法行为”

与“行为扭曲和操纵理论”下“非理性人”的假设迥异,面对个体行为和感受的背离,“隐私演算”(privacy calculus)理论旨在从“利益最大化”的理性人角度,认为个体会在潜在隐私损失和预期隐私收益间进行计算,其最终行为取决于隐私权衡的结果^①。就其损失而言,包括安宁侵扰、社会歧视、身份盗窃、网络诈骗、人肉搜索等;就其收益而言,包括娱乐、便捷、个性化、自我呈现、维系社会关系、获取社会资本等。当收益大于损失时,人们就会用隐私换取更高的利益^②。对隐私的“陈述偏好”(stated preferences)和“显示偏好”(revealed preferences)的区分进一步表明,尽管人们经常声称关切隐私,但在具体场景下的行为才能揭示其真正的偏好^③。事实上,人们愿意通过隐私披露来获取企业的数字内容和数字服务,正如2015年欧盟《关于数字内容提供合同部分问题的指令议案》(EU Proposal for a Directive Concerning the Supply of the Digital Content)前言所指出的那样,在数字经济中,用户“支付个人数据”(pay by data)和“支付金钱”(pay by money)具有同等意义。面对纷繁芜杂的互联网信息和产品,算法推荐通过“千人千面”的页面设置、精准匹配的搜索结果、反映用户偏好的内容推送,帮助用户降低搜寻成本和决策成本。

不仅如此,基于算法的差别化定价具有快速、动态特征,能够在提高产品服务质量的的同时,稳定地满足多变的市场需求^④。算法所提供的种种便利,使得用户在权衡之后,愿意在明知算法风险的前提下,作出接受算法服务的选择,这或许解释了调研结果中“算法推荐便利与隐私”的矛盾。用户对算法的排斥实质上是对企业过度利用算法行为的排斥,而非排斥算法技术和算法应用本身。当算法满足更便利的信息搜集渠道、更精准的信息推荐、更有质量的信息内容等用户期待时,用户对算法并不抵触。相反,当算法无法实现上述期待时,人们对算法沉迷、信息茧房、算法侵权等算法滥用的担忧就自然会凸显。

三、制度回应:算法悖论的化解

(一)理解发展偏好:将“算法向善”嵌入“业务流程”

作为一种发展偏好,“算法关注”随着服务和产品的使用而上升,因此,企业应未雨绸缪,将“算法向善”引入到服务、产品设计和使用的各个环节中,在系统设计之初就将“向善”需求嵌入其中,成为系统运行的默认规则,而非事后补救。这一“经设计的算法治理”遵循“以人为本”的设计原则(Human-Centred Design, HCD)。HCD主张将“人”放在任何系统的中心,从用户的需求、兴趣和能力出发,通过直接与人们接触来评估和理解人类,以提供可用、易于理解和符合社会价值观的产品和服务^⑤。HCD本质上是跨学科实践,为此,企业首先应当设立一个由伦理专家、技术专家、法律专家、公众代表组成的算法伦理委员会,在对企业或社会可能产生重大影响的算法部署之前,委员会应率先启动伦理审查,基于“算法向善”的原则,识别、预防、消除相关应用对基本价值观的背离^⑥。另一方面,HCD还

① Dinev T. and Hart P., “An Extended Privacy Calculus Model for E-Commerce Transactions”, *Information Systems Research*, 2006, 7(1), pp.61-80.

② Krasnova H., Spiekermann S., Koroleva K., Hildebrand T., “Online Social Networks: Why We Disclose”, *Journal of Information Technology*, 2010, 25(2), pp.109-125.

③ Cooper J. C., “Lessons from Antitrust: The Path to a More Coherent Privacy Policy(Feb. 26. 2017)”, <https://www.us-chamberfoundation.org/reports/lessons-antitrust-path-more-coherent-privacy-policy>, 访问日期:2022年5月16日。

④ 张江莉:《论相关产品市场界定中的“产品界定”——多边平台反垄断案件的新难题》,《法学评论》2019年第1期。

⑤ Donald N., “The Four Fundamental Principles of Human-Centered Design and Application(August 2019)”, <https://jnd.org/the-four-fundamental-principles-ofhuman-centered-design/>, 访问日期:2022年5月16日。

⑥ 许可:《算法规制体系的中国建构与理论反思》,《法律科学(西北政法大学学报)》2021年第1期。

需要整合信息科学、心理学、认知科学、人类学等知识,企业有必要汇聚用户体验设计师、视觉设计师、交互设计师和信息设计师,使之通力协作,保证将用户的权益置于任何设计的最前沿。

算法向善的内涵十分宽泛,为达致监管实效,理应有所侧重。从调研结果观察,民众对于算法故意造成的信息扭曲行为、不良信息的算法推送、网络弱势群体服务水平不高反映强烈,但算法诱导下的过度消费问题并不突出,“信息茧房”也不尽相同。为此,监管机构可以“信息内容向善”和“网络弱势群体保护”为监管重心,落实《互联网信息服务管理办法》《网络信息内容生态治理规定》等信息内容治理的法律法规,加强违法信息内容禁令的执行,规范企业进一步优化过滤算法和推荐算法以防范和抵制不良信息,禁止算法实施流量劫持、虚假注册账号、非法交易账号、操纵用户账号等破坏网络生态秩序的行为。同时,企业应积极响应民众诉求,畅通投诉—举报—反馈机制,并为未成年和老年人量身定制适合其认知特点的服务产品。

(二)避免扭曲操纵:从“算法透明”迈向“算法素养”

面对可能陷入扭曲操纵的算法行为,首先要破除企业的“暗黑模式”。阳光是最好的消毒剂,旨在打开“算法黑箱”的算法透明机制由此成为各种路径中最直接有效的方式^①。我国既有监管法律为算法透明提供了算法备案、算法审计、算法检查、算法解释等多种工具,而在民众对算法了解水平有限的情形下,面向用户的“算法解释权”成为化解算法悖论、落实算法问责、实现算法公正的关键所在^②。调研结果表明,用户对算法解释的诉求强烈,但不同用户之间也有着显著差异,有54.28%的人需要对算法的简单解释,38.49%的人则需要详尽解释。同时,在算法处理结果引发用户异议时,超过3/4的受访者希望人工介入,以矫正可能的算法错误。因此,尽管学界对算法透明原则存在争议,但从用户出发,监管机构仍应坚持算法透明要求,设定分层次的算法透明规则。一是“算法服务基本情况的透明”(简单解释规则),即企业应以显著方式(包括但不限于通过算法备案系统)告知用户与服务直接相关的核心算法名称、应用领域、算法类型和算法目的。二是“算法服务基本原理的透明”(详细解释规则),即企业可以通过易见、易读、易懂的方式告知用户与服务直接相关的核心算法的基本原理和主要运行机制。三是“算法服务处理结果的透明”(全面解释规则),即在用户就算法决策结果提出异议时,应当告知个人信息收集和处理、个人特征参数模型选择及其与决策结果的逻辑关系。四是“算法服务异议的人工介入”(用户拒绝规则),即在用户就算法决策结果提出异议时,企业应当通过人工方式复核决策结果。

算法透明并非意味着算法可知,受限于民众的技术能力、算法的复杂化、机器学习和干扰性披露(信息混淆)等问题^③,算法透明可能无法实现提升用户理性的目的。隐私悖论研究表明,增加人们对隐私技术、威胁的了解,培养科学的隐私风险意识,帮助人们获取如何保护隐私的信息,能够有效减少隐私悖论行为^④。当前,民众对算法存在显著的认知偏差。在算法应用的第一阶段,民众充分享受算法便利,但对算法一无所知;在算法应用的第二阶段,随着隐私侵害、大数据杀熟等负面信息不断传播,引发了民众强烈忧虑,但由于羊群效应,亦同时出现了对算法的错误认知。调查显示,民众尚没有建立起基本的算法知识结构和自身逻辑一致的算法认知,由此导致用户面对算法时难以作出最符合其真正利益的选择。调研进一步发现,对算法的认知与态度因用户的学历背景、成长环境、收入水平而异,不同群体之间存在明显鸿沟。为此,有必要系统性提升公众的“算法素养”(Algorithmic Literacy)。

所谓“算法素养”,即“意识到算法在网络平台和服务中的使用,了解算法的工作原理,能够批判性

① 汪庆华:《算法透明的多重维度和算法问责》,《比较法研究》2020年第6期。

② 许可、朱悦:《算法解释权:科技与法律的双重视角》,《苏州大学学报(哲学社会科学版)》2020年第2期。

③ 沈伟伟:《算法透明原则的迷思》,《环球法律评论》2019年第6期。

④ Weinberger M., Bouhnik D., Zhitomirsky-Geffet M., “Factors Affecting Students’ Privacy Paradox and Privacy Protection Behavior”, *Open Information Science*, 2017, 1, pp.3-20.

地评估算法决策,以及拥有应对甚至影响算法操作的技能”^①。作为一项系统工程,算法素养就像更宽泛的“数字素养”一样,需要国家、企业、科研机构、社会组织等主体共同参与^②。为此,各方可以彼此协力加强用户教育,开展广泛、持续的日常性算法教育,帮助用户提高对算法服务的认知能力和自我保护能力。

具体而言,各方从算法态度、算法知识、算法技能三方面开展如下工作:(1)普及算法知识。监管机构可以通过专门网站向公众提供相关的、界面友好的维权法规以及自我保护措施;企业应在算法透明的基础上,全面披露算法对用户正反两方面的影响。(2)改善算法态度。监管机构可以定期公布算法治理案例和执法活动成果,企业则可以通过算法应用研究、算法合规审计报告的形式,影响民众对算法收益和风险的感受、偏好与评价情况。(3)强化算法技能。社会各界可以在学历教育和社会教育中广泛开展算法教育,培养民众用算法工具创建和编辑文字、图像和视频内容的能力,利用算法原理选择推荐信息、影响算法服务内容的能力,保护隐私、个人信息、数字身份的能力,以及维护算法权利、解决算法纠纷的能力。

(三)优化成本收益演算:经“过程公平”落实“算法公平”

互联网海量内容呈现与可及数据极大提升了算法筛选、匹配、推送的必要性和准确率,算法推荐由此成为数字经济时代信息传递的重要途径。调研结果也体现出民众对算法推荐的普遍认可。正如以成本收益为基础的“算法演算”理论所洞见,“得不偿失”是最重要的算法不公平。在差别化定价这一可能破坏算法公平的典型场景中,调研初步窥见了用户对得失之间的轻重权衡:根据收入水平(实质是“用户支付意愿”)的差异化定价令人反感,而根据“会员或非会员”的差异化定价则被很多人认可,至于根据用户接受服务频率和新老用户的差别化定价(即“大数据杀熟”)却并未达成一致。这一分歧的背后,是纷繁多样的差异化定价类型,以及对用户权益正反两面的影响:被收取较高价格一方的消费者剩余可能转移至被收取较低价格的一方,使得后者能够享有统一定价场景中无法享受的服务或产品,从社会福利的角度,这有助于减少“哈勃格三角”的“无谓损失”(deadweight loss),实现卡尔多-希克斯效率(Kaldor-Hicks Principle)。因此,试图通过类型化方式来厘清差别化定价的“合理”与“不合理性”的边界,难以达成共识。另一方面,差别化定价规则的不透明、信息与地位的不对等是造成用户感知价格欺诈、胁迫,进而引起强烈反感的主要原因。职是之故,未来的监管不妨从强调“算法应确保有价值事物在各方之间平等分配”的“结果公平”转向“算法应平等对待所有参与者,各方享有平等的机会、条件和权利”的“过程公平”^③。

算法的“过程公平”首先要求企业在差别化定价时充分告知其理由。例如,在开展区分新老用户的“拉新活动”中,可明示“首单减免、新用户专享优惠”等;在开展区分接受服务频率的“促活留存”活动中,可明示“尊敬的用户,您很久没有光临小店了,特送您一张优惠券”等;在区分不同群体的营销活动中,可明示“学生专享价格”“60周岁以上专享价格”等。同时,“过程公平”还应保障用户对差别化定价的“选择退出”权利。调研表明,若用户对差别化定价不满时,大部分会通过用脚投票和曝光投诉等途径维护自身权益,从而给企业行为施加有力的外部制约。我国《个人信息保护法》第24条第二款在《电子商务法》第18条规制个性化推荐的基础上,特别增加了“向个人提供便捷的拒绝方式”的企业义务,有效避免了“选择加入机制”的实质不知情同意。据此,若用户认为遭到定价歧视,可以随时选择

① Dogruel L., Masur P., Joeckel S., “Development and Validation of an Algorithm Literacy Scale for Internet Users”, *Communication Methods and Measures*, 2021, 16(3), pp.1-19.

② 蒋敏娟、翟云:《数字化转型背景下的公民数字素养:框架、挑战与应对方略》,《电子政务》2022年第1期。

③ Reuben B., “Fairness in Machine Learning: Lessons from Political Philosophy”, *Proceedings of Machine Learning Research*, 2018, 81, pp.149-159.

退出服务,从而激励其挑战企业不合理的算法推荐^①。

放宽视野看,更准确、更全面、更便捷的算法成本收益演算,有助于用户形成真正的“计算性信任”(Calculated Trust),将个人和企业均置于数字经济的利益共同体中,立足于规制激励而非规制威慑,有效弥合因差别化定价而可能减损的“用户—企业”的数字信任关系,最终实现双赢^②。

(四)回应算法风险:借“算法问责”铸就“算法安全”

尽管算法安全是《算法综合治理意见》中确立的首位监管目标,但从消解算法悖论的角度,算法安全应居于底线,只有在算法向善、算法透明、算法公平等措施力有不逮之时,才有适用的空间。这是因为,算法滥用固然可能危及网络空间传播秩序、市场秩序和社会秩序,但上述秩序均有着自我修复与完善的潜力,算法安全规制应保持谦抑,避免安全泛在化戕害了既有秩序的自我发展与演进。不过,这并不意味算法安全与算法悖论无关。正如隐私悖论研究所揭示的那样,隐私保护是一个庞大、复杂和永无止境的项目,只是将知情权和选择权赋予个人,使之自我管理是严重不够的,监管者必须为个人信息的收集和使用划定界限,以保护个人免受侵害^③。算法悖论亦是如此。在伦理导向的算法向善、权利导向的算法透明、程序导向的算法公平外,还应将问责导向的算法安全作为监管底线,回应民众关切。这恰恰印证了本文的调查发现,即有 58.12% 的受访者选择通过“强化企业使用算法的法律责任”消除算法风险。

“算法问责”要求在算法应用引起侵害或负面后果时应能向算法服务提供者追责。据此,一方面,对于因场景特殊性和关涉法益重要性而引发的风险,监管者应将算法看作侵权行为或违法行为的工具,采取“结果导向”和“实质主义”进路,在危害后果发生后予以问责。另一方面,问责并不限于“事后救济和惩罚”,对于使用强化学习、无监督学习、深度学习的算法,法律难以透过算法模型全然把握其内部逻辑和决策过程,因而无法认定行为过错和因果关系。从“事后”向“事前”的拓展,成为算法问责的发展方向。例如,2022 年,美国推出新版《算法问责法(草案)》,将“算法影响评估”作为核心内容,要求企业使用算法作出决策时,应对偏见、有效性和相关因素进行系统化的影响分析。

就此而言,算法问责已成为管控算法侵害的全程机制,通过对数据收集、特征提取、算法设计的记录留痕、认证与审查机制和侵害发生后的纠正机制,积极预防算法可能产生的负面影响^④。覆盖事前、事中、事后的算法问责需要市场(代码)、社群(规范)、政府(法律)的共同参与^⑤。其中,市场中的企业以代码为基础的自我规制是成本最小、效率最高的治理方式。调查亦显示,超过 60% 的人赞同“企业加强算法的安全评估”,成为各项措施中的首选。认证机构、专家代理等社会组织则有助于执法和司法过程中确定争议事实、认定过错和因果关系,是算法问责的重要构成性力量。另外,政府应通过主动监管追究企业民事、行政和刑事责任,实施有效威慑。

四、结语

法律实证研究具有发掘规范论证的实然基础和衡量法律实效的功能^⑥。本次用户算法应用感知调查,一方面反映出民众对算法风险的高度关切,为我国算法立法和执法提供了坚实的事实基础,另

① 李丹:《算法歧视消费者:行为机制、损益界定与协同规制》,《上海财经大学学报》2021 年第 4 期。

② 谢尧雯:《网络平台差别化定价的规制路径选择》,《行政法学研究》2022 年第 1 期。

③ Solove J. D., “The Myth of the Privacy Paradox”, *George Washington Law Review*, 2021, 89(1), pp.1-51.

④ 王莹:《算法侵害类型化研究与法律应对——以〈个人信息保护法〉为基点的算法规制扩展构想》,《法制与社会发展》2021 年第 6 期。

⑤ 许可:《驯服算法:算法治理的历史展开与当代体系》,《华东政法大学学报》2022 年第 1 期。

⑥ 张永健、程金华:《法律实证研究的方法坐标》,《中国法律评论》2018 年第 6 期。

一方面亦揭示出人们的算法关注与算法行为之间的“算法悖论”。作为一个经“隐私悖论”而发展的概念,算法悖论颇具理论深意。在监管者积极回应算法风险时,有必要审慎对待用户的真实感受和欲求,依循数字经济与算法社会的内在规律,从而推动我国以算法安全、算法透明、算法公平、算法向善为架构的算法治理体系日臻完善。

当然,作为我国对用户算法感知的首次大规模实证研究,其可能无法反映事实全貌,更重要的是,它是算法法规尚未实行之时的切片式观察。基于此,我们在不久的将来会更新问卷,继续开展新一轮调查,以期以时间为维,描述算法法规的实效,发现民众认知、态度、权利和行动的变迁,进而弥合法律应然与实然的鸿沟。

Algorithmic Paradoxes and Institutional Responses

—An Empirical Study Based on Users' Perceptions of Algorithmic Applications

Xu Ke Cheng Hua

(School of Law, University of International Business and Economics, Beijing 100029, P.R.China;
School of Economics, Renmin University of China, Beijing 100044, P.R.China)

Abstract: Although the basic structure of algorithmic governance in China has been initially laid, the relevant empirical studies are still weak. The analysis of the results of a large-scale questionnaire survey on users' perceptions of algorithmic applications in China reflects the high level of public concern about the risks of algorithms. It also reveals the differences between the “algorithmic concerns” of different groups, the contradiction between the “algorithmic concerns” and algorithmic behaviour of the same group, and the trade-off between algorithmic recommendations and “algorithmic concerns”. In the face of the various manifestations of the “algorithmic paradox”, institutional responses should be made in accordance with its internal mechanism: embedding “algorithmic goodness” into “business processes”; moving from “algorithmic transparency” to “algorithmic literacy”; implementing the “algorithm fairness” through “process fairness”; implementing “algorithmic security” through “Algorithmic accountability”. By doing this, it will ultimately lead to a system of algorithmic governance in China that is responsive to the real needs of the public and the balancing of multiple goals. Last but not least, the questionnaire survey is only a sliced observation when the algorithm regulations have not yet been implemented, and we will carry out a new round of surveys in the future to describe the effect of the implementation of algorithmic regulations and to discover changes in people's cognition, attitudes, rights and actions towards algorithms.

Keywords: Algorithmic perception; Algorithmic governance; Algorithmic paradox

[责任编辑:李春明]