钟晓雯. 算法推荐网络服务提供者的权力异化及法律规制[J]. 中国海商法研究,2022,33(4):63-72

算法推荐网络服务提供者的权力异化及法律规制

钟晓雯

(西南政法大学 民商法学院,重庆 401120)

摘要:网络服务提供者基于法定和意定授权取得了私权力,并呈现出利用算法推荐侵犯用户合法权益,构建隐形规训和统治空间的权力异化现象。当前"红旗"规则难以适用于公开型算法推荐,算法的"黑箱"与价值内嵌逻辑构筑的技术壁垒难以打破,导致该权力归化面临困境。注意义务和算法透明义务是破解困境的关键。算法推荐网络服务提供者具有强大的信息管理能力,充当着言论表达和控制的"把关人",与内容生产者对同一内容的利益回报存在价值差,以强化注意义务的方式归化其权力符合法经济学原理。算法透明义务主要在知情权与可问责性上发挥功用,"鱼缸透明"是算法透明应当达到的限度,"理性透明"是其未来的发展方向,实现算法透明应当以算法的"可信解释"为技术支撑。

关键词:算法推荐;网络服务提供者;权力异化;注意义务;算法透明义务

中图分类号:D923 **文献标志码**:A **文章编号**:2096-028X(2022)04-0063-10

Power alienation and legal regulation of algorithm recommending network service providers

ZHONG Xiao-wen

(School of Civil and Commercial Law, Southwest University of Political Science and Law, Chongqing 401120, China)

Abstract: Network service providers have obtained private power based on legal and voluntary authorization, and there is a phenomenon of power alienation that uses algorithmic recommendations to infringe on users' legitimate rights and interests, and build invisible discipline and domination space. Currently, power domestication faces the dilemmas that the "red flag" rule is difficult to apply to open algorithm recommending and the technical barrier constructed by the algorithmic "black box" and value embedded logic is difficult to break. Duties of care and algorithmic transparency are core elements of breaking through the dilemma. The algorithm recommending network service provider has strong information management capabilities and acts as a "gatekeeper" for speech expression and control. There is a difference in value between the algorithm recommending network service provider and the content producer in return of the same content. At the meantime, naturalizing its power in a way of strengthening the duty of care is in line with the principles of law and economics. The duty of algorithm transparency mainly plays a role in the right to know and accountability. The "fishbowl transparency" is the limitation that algorithm transparency should achieve, and "reasoned transparency" is the future development direction. The realization of algorithm transparency should be supported by the "credible interpretation" of the algorithm.

Key words: algorithm recommending; network service providers; power alienation; duty of care; duty of algorithm transparency

一、网络服务提供者的权力取得依据

"权力"是社会科学中的基本概念,其本质体现为"意志自由"和"支配力",其中"意志自由"指向权力主体的内部效力,"支配力"指向权力主体与权力受体的外部关系。从内部效力来看,权力的本质表现为权力主体的意志自由。马克斯·韦伯(Max

收稿日期:2022-04-26

基金项目:2019年度司法部国家法治与法学理论研究项目"网络法基本术语的界定、辨析及规范使用研究"(19SFB2003),2020年度广东省哲学社会科学规划办"制度理论研究"专项项目"信息技术与法治耦合健全社会公平正义法治保障研究"(GD20ZD16)

作者简介:钟晓雯,女,西南政法大学民商法学院民商法专业博士研究生。

Weber)曾描述“权力”为“一个人或若干人在社会行为中实现自己意志的机会,甚至不顾参与该行为的其他人的反抗”。[1]即权力主体能够排除其他反抗意志,以自我决定和自我控制的方式实施自己的意志,但这需要以牺牲权力受体的意志自由为代价。如此一来,就衍生出了权力主体与权力受体的外部支配关系。支配力是权力的核心要素,权力主体的支配力来源于其占有的资源优势,如政府等公权力机关凭借统治资源形成对社会主体的支配力;私人资本凭借经济资源形成对劳动者的支配力等。[2]将这一结论移植于网络空间,即可发现,网络服务提供者凭借技术资源形成了对用户的支配力,但这仅为事实上的权力,尚缺乏行使的正当性依据。从现有法律与网络服务实践来看,立法与网络服务用户协议实际上授予了网络服务提供者资格审查、信息监管和处置,以及规范用户的权力,恰好弥补了网络服务提供者行使私权力缺乏正当性依据的缺陷。

(一)法律授权基础:网络服务提供者的资格审查、信息监管和处置权力

《中华人民共和国民法典》(简称《民法典》)、《中华人民共和国网络安全法》(简称《网络安全法》)和《中华人民共和国电子商务法》(简称《电子商务法》)等法律文件授予了网络服务提供者资格审查、信息监管和处置的权力。例如《民法典》第1195条明确了网络服务提供者在接到权利人提供的构成侵权的初步证据及其真实身份信息的通知时,有义务将该通知转送相关网络用户,并采取删除、屏蔽、断开链接等必要措施(“通知—必要措施”规则);《网络安全法》第47条要求网络服务提供者应当承担信息监管义务,在发现具体违法信息时,应采取停止传输、消除等必要处置措施;《电子商务法》第38条明确电子商务平台经营者负有对平台内经营者的资质资格进行审核的义务,以及在平台内经营者侵害消费者合法权益时,负有采取必要措施的义务等。单从文义解释上看,上述条款表达的是国家课以网络服务提供者的义务,但从反向视角来看,这也可以理解为是一种授权。网络服务提供者负担的资格审查、信息监管等义务,本质上与物理空间中行政机关采取行政许可、接到报案后对受害人进行行政救助等情形一致,是一种借助权力实现社会监管、权利救济的机制。

网络空间以网络服务提供者为辐射中心形成了自上而下的“行政主管部门—网络服务提供者—用户”的权力逻辑关系,法律将本属于“行政主管部门”的权力部分授予了网络服务提供者。这是因为,相较于行政主管部门囿于网络空间的虚拟性而难以追踪或控制相关违法行为的劣势,网络服务提供者能够利用其占有的技术资源优势迅速作出因应,“根据危险控制理论,离危险源越近的人,越容易控制危险的发生”。[3]因此,网络服务提供者基于其技术资源优势,取得了源自行政主管部门的部分权力,可对用户发布的信息等进行审查、监管和处置。

(二)意定授权基础:网络服务提供者规范用户行为的权力

意定授权指向网络服务用户协议,即网络服务提供者与用户间签订的,载有双方服务关系及其权利义务的格式合同。网络服务用户协议通常为电子合同,传统纸面上的签字盖章也逐渐演化为电子签章。基于电子商务法中的功能等同原则,以数据电文形式订立的电子合同与以传统书面形式订立的纸质合同具有同等法律效力。网络服务提供者规范用户行为的权力依据主要体现在协议内容上。尽管不同的服务协议内容设置各异,但协议内容多以规范用户行为和免除网络服务提供者责任为主。

以微信APP的《软件许可与服务协议》(简称《协议》)为例,《协议》第8条“用户行为规范”针对微信用户行为从五个方面作了详尽规定:“信息内容规范”“软件使用规范”“服务运营规范”“对自己行为负责”以及“违约处理”;其中第8.1条“信息内容规范”以“列举+概括”的方式明确约定了微信用户行为的边界,即约定用户使用微信过程中禁止制作、复制、发布和传播的信息内容范围;第8.5条明确约定腾讯公司在用户违反约定的行为规范时,有权不经通知随时采取删除、屏蔽相关内容及账号封禁等必要措施。网络服务用户协议对网络服务提供者的授权除体现在规范用户行为上,也体现在网络服务提供者的责任免除上。《协议》第10.3条规定,腾讯公司和用户共同承担维护软件安全与正常使用的责任,但腾讯公司不对用户终端设备信息和数据安全的保护承担完全保证责任。由此可见,网络服务用户协议名义上以平等自由为基础,实则充斥着大量约束用户行为与免除网络服务提供者责任的条款。

总体而言,网络服务提供者基于技术资源优势,以法定和意定的方式取得了针对用户的权力,一方面能够较之行政主管部门高效治理网络违法行为或信息;另一方面用户以协议的形式让渡部分权利至网络服务提供者,使其能够在特定范围内对相关行为作出必要规范,确保用户权利有序行使。

二、算法推荐网络服务提供者的权力异化表象

正如霍布斯所总结的,人类有着“得其一思其二、死而后已、永无休止的权势欲”,[4] 网络服务提供者也并未按照预设轨迹运行权力,而是开始扩张权力。算法推荐强化了网络服务提供者占有的技术资源优势,更加剧了其权力扩张与滥用的趋势,具体表现为网络服务提供者利用算法推荐侵犯用户合法权益,以及构建隐形规训和统治空间。

(一)侵犯用户合法权益

网络服务提供者利用算法推荐侵害用户的合法权益,在隐私权、知情权以及人格尊严上表现得尤为明显。

其一,侵害用户隐私权。隐私权是自然人与生俱来的权利,该项权利已被明确列入法律。《民法典》第 1032 条第 2 款对“隐私”作了界定:“隐私是自然人的私人生活安宁和不愿为他人知晓的私密空间、私密活动、私密信息。”因此,对隐私权的侵害既表现为个人私密信息的泄露,亦表现为对“私人生活安宁”的破坏。在个人私密信息的泄露上,由于算法实质上是一种分析利用数据的智能技术,即算法通过分类整理数据并依据计算模型作出算法决策,数据的体量直接决定了算法所能达到的深度,数据的精准度直接影响着算法决策的精确度。因此,算法的运行需要收集海量数据,但这些数据收集可能并未获得用户的事先同意。不同的算法所针对的数据类别存在差异,但未经用户同意的数据收集行为必然侵犯了用户的隐私权。比如搜索引擎的算法推荐应用中,搜索引擎的决策过程与决策结果(搜索结果)所依据的是算法对用户的“个人画像”,用户在搜索时所输入的关键字等数据仅是一个索引,“个人画像”的形成更多的是依赖用户的搜索历史与浏览记录,甚至是聊天记录。倘若在未经用户同意前即对这些数据进行收集,将侵犯用户的隐私权。[5]

算法对个体隐私权的侵犯还表现为对“私人生活安宁”的破坏。“北京百度网讯科技公司与朱烨隐私权纠纷”①即为典型案例。在该案中,朱烨以“减肥”“丰胸”为关键词在百度搜索引擎展开搜索后,其在浏览网页时会频繁出现关于减肥、丰胸的广告。在这一过程中有几点信息是可以确认的:朱烨的搜索关键字被搜索引擎所记录,该搜索记录被搜索引擎以外的广告公司获取(当然也可能是搜索引擎本身,但这种利用已经背离了用户使用搜索引擎的初衷,而且这种推送可能多次发生在不同对象中),广告公司根据该搜索记录对朱烨进行“画像”并根据“画像”向朱烨不断推送广告。在这些过程中,搜索引擎获取朱烨的搜索关键词是被许可的,因为这是算法展开分析与决策的前提条件,但搜索引擎将关键词的相关数据转移至广告公司,以及广告公司向用户持续直接地推送广告的行为,无疑侵犯了用户的“私人生活安宁”。

其二,侵害用户知情权。知情权的行使关乎公平、正义和自由。算法推荐通过收集、筛选并整理个体在网络空间留下的数字痕迹(个体注册信息,用户阅读、搜索、浏览、评论记录等),能够精确勾勒出个体的兴趣和偏好图谱,形成精确的“个人画像”。利用该“个人画像”,各种消费平台能够将契合用户兴趣和偏好的信息精准分发至目标人群。可以说,个体消费者对于购物、用餐、出行、酒店的选择基本全部依赖算法推荐。然而个体却对其“个人画像”形成的逻辑、根据(包括平台抓取的数据类型、特征,形成画像的具体维度等)以及内容毫不知情,这显然侵犯了其知情权。

其三,侵害用户人格尊严。个体人格尊严的侵害与算法推荐衍生的算法歧视密切相关。算法歧视已经成为算法时代下的常见风险形式,其充分证明了算法并非是技术中立的活动。算法歧视的产生主要来源于两方面:一是算法抓取数据中产生的歧视。数据对于算法推荐结果具有基础性作用,如果数据本身存在偏差,或者算法对于数据的选取出现偏差或侧重,则可能导致推荐结果呈现歧视性。例如亚马逊招聘系统将高薪职位更多推送给男性求职者的歧视现象,其根源在于该算法选用的原始数据是亚马逊公司过往员工的相关数据。在亚马逊公司的用工历史上,男性员工明显多于女性,由此导致算法推荐的结果偏重于男性。二是算法的设计/部署产生的歧视。算法在很多时候并非是一种完全价值中立的科学活动或数学活动,相反,其总是与特定的价值立场相关,蕴含着价值判断。[6] 算法设计/部署者往往通过算法指令掺杂自己的价值观,从而导致歧视现象产生。

(二)构建隐性规训和统治空间

算法推荐能够通过收集分析用户在网络空间中留下的数字痕迹(注册信息、浏览日志、历史评价记

① 参见江苏省南京市中级人民法院(2014)宁民终字第 5028 号民事判决书。

录等),挖掘用户的兴趣偏好,精准勾勒出用户的“个人画像”,并对用户进行个性化的信息推荐与分发,从而影响算法相对人,由此算法也具有了一定的权力(“算法权力”)。算法权力的表象是技术权力,其背后隐藏的是资本权力。[7]当算法权力渗透到人类生活的方方面面,即会在某种意义上形成一种隐性规训和统治空间(“算法规训”和“算法统治”)。

算法规训首先表现在对用户价值观的影响上。人的价值观的形成受到多种因素影响,其中新闻媒介的影响力不容小觑。在纸质传媒时代,新闻的内容是经过专业编辑选择的,其中的意识形态与价值观念也当然经过了选择。具有普世意义的价值观更容易得到采纳与传播,并影响终端用户,促进用户正确价值观的塑造。此外,纸质传媒内容的丰富性,有利于读者全面获取信息。依据信息论的观点,一个人掌握的信息越充分,其在进行个人价值实现方面的决策时就越能处于有利地位并趋于理性,但在算法分发新闻的情况下会出现截然相反的情形。

算法分发新闻会根据用户的初始选择对用户主动推送相关新闻,在算法分发已成为网络新闻最主要的分发方式的情况下,其优势与劣势同时出现。尽管算法分发新闻只是对同类、偏好新闻的推送,并不创造新的内容,但同类推荐会形成封闭的信息空间,隔绝多元信息穿透传播的可能性,致使用户接收到的信息趋于同质化,逐渐形成信息茧房并衍生回音室效应。“信息茧房”和“回音室效应”的概念由凯斯·桑斯坦(Cass R. Sunstein)提出,他认为尽管互联网技术可以促使人们逃离地理学上的茧房和回音室,但互联网中的个性化信息服务过滤了多元化的观点,不同网络群体更倾向于选择和获取自己感兴趣或与自己观点相同/相似的信息,并忽视与外部世界的交流和沟通。这种持续衍化的群内同质和群际异质现象会产生“信息茧房”效应,即如同置身于蚕茧般作茧自缚,进而形成“回音室效应”。[8]“回音室效应”意指信息过滤机制使人们只获取到他们感兴趣或认同的信息,长期以往,当这些信息中所蕴含的观点被不断重复并加深时,人们在这个信息封闭的圈子中将只能听到自己的“回声”,变得容易固守偏见甚至误将偏见视为真理,进而排斥和拒绝其他的合理性观点和意见。[9]“信息茧房”及其衍生的“回音室效应”会导致用户的阅读范围受限,尤其是致使塑造人们价值观的信息范围越来越窄,进而导致个人可能在非理智的情况下作出决策。例如某用户可能只是因心情烦闷而点击了有关抑郁症的新闻,却导致类似新闻大量推送,最终可能致使该用户逐渐失去生活的希望并走向极端。此外,分发新闻的算法也可能被嵌入算法设计/部署者的价值观念或主观意图,类似于剑桥分析公司干预美国总统选举一样,会在潜移默化中影响用户选择。算法推荐网络服务提供者对用户的算法规训也就此形成。

算法规训还表现为推荐算法对用户的行为选择的干涉。网络服务提供者可以利用算法收集、分析用户的行为痕迹从而逐渐掌握用户习惯和偏好,并依据用户习惯和偏好定制契合或足以诱导用户行为的推荐。用户可能无法拒绝这种具有诱惑力且似乎尚未超出自己能力范围的推荐,于是消费习惯开始发生变化。这种变化实则是算法推荐在人的行为领域中产生的规训作用。

当算法规训大量出现时,算法推荐相当于依靠自身影响力形成了一套行为准则。它对个体行为进行的预测与干预,成为现实中强有力的行为规范。[10]当个体成为算法下可被预测和计算的客体,算法由此可以控制人的生活,消除个体行为的不确定性,肆意赋值个体,并形成国家权力的替代性权力。[11]拥有算法推荐的网络服务提供者也因此具备了在网络中构建隐性规训和统治空间的权力。

三、算法推荐网络服务提供者的权力归化困境

网络服务提供者当前呈现出利用算法推荐侵犯用户合法权益,构建隐形规训和统治空间的权力异化现象。存在权力异化即需要进行权力归化,但当前算法推荐网络服务提供者的权力归化面临着法律与技术困境。

(一)法律困境:规制算法推荐网络服务提供者的制度漏洞

网络服务提供者的主要职能是利用自身占有的技术资源为用户提供信息传播中介服务,即为公众搭建一个信息交流平台。鉴于网络空间中存在海量数据和信息,由网络服务提供者对用户发布的全部信息负主动审查义务既不可能也不合理,故立法明确规定,当网络服务提供者已尽到合理注意义务,在接到侵权通知后采取了必要措施的,可免除相关法律责任。但这一规范在规制算法推荐网络服务提供者时存在罅隙,此处需要结合算法推荐的类型作进一步分析。

1. 算法推荐的主要类型

学理研究通常从技术角度对算法推荐进行分类,分为基于内容的算法推荐和协同过滤的算法推

荐,其中基于内容的算法推荐是根据用户先前的消费、浏览或阅读过的内容之相似度来推荐内容(例如,推荐的图片 X 会具有与用户先前查看的图片 Y 和 Z 相似的标题);协同过滤的算法推荐则根据相似用户消费的内容进行推荐(例如,A、B、C 喜欢 E;D 作为与 A、B、C 相似的个体,可能也喜欢 E)。但笔者对算法推荐的分类是基于非技术层面的因素,即以算法推荐所依赖的数据集的来源为依据,将其类型化为三种:一是公开型算法推荐;二是精选型算法推荐;三是封闭型算法推荐。[12] 公开型算法推荐所依赖的数据集主要由用户或广告商直接提供或通过其他来源渠道自动引入/聚合。这些数据集并未由平台进行专门选择(尽管平台内部也会产生数据且用于算法推荐)。目前谷歌、YouTube、Facebook、Reddit、Instagram 和亚马逊采用的都是公开型算法推荐。例如,YouTube 的用户上传的任何视频通常都被默认带入推荐系统。区别于公开型算法推荐,精选型算法推荐所依赖的数据集来源于经由平台策划、批准或许可而形成的内容池,其主要用于涉及行政许可的传统媒体领域,如音乐、电影或电视节目等,Netflix 即是采用精选型算法推荐的典型示例。区别于前两类算法推荐,封闭型算法推荐是由平台本身或运营该站点的组织生成欲推荐的内容。例如,新闻机构向用户提供的个性化故事和文章提要即是由组织本身制作或委托制作的。

2.“红旗”规则难以适用于公开型算法推荐

《民法典》第 1195 条确立了有关网络服务提供者法律责任的“通知—必要措施”规则(也被称为“通知—删除”规则),明确了网络服务提供者在接到权利人关于网络用户利用网络服务实施侵权行为的通知时,应当采取包括删除、屏蔽、转通知等在内的必要措施(“转通知”在理论研究中是否属于“必要措施”的范围尚存在争议)。在《民法典》颁布之前,《电子商务法》为加强对网络知识产权的保护,已经在第 42 条中规定了知识产权权利人与网络服务提供者间的“通知—必要措施”规则。《电子商务法》第 42 条与《民法典》第 1195 条是一脉相承的,《民法典》第 1195 条相当于扩大了权利主体范围,将“通知”的权利主体从《电子商务法》中的“知识产权权利人”扩张至所有“权利人”。早期对网络服务提供者采用“通知—必要措施”规则主要有两方面原因:一是在危险控制理论下,网络服务提供者类同于物理空间中的公共经营者,对网络侵权危险的控制能力较强,由其履行注意义务具有经济学上的成本优势;二是出于促进互联网产业经济发展的考量,立法以“必要措施”的履行作为免责条件,旨在减轻网络服务提供者的注意义务。故“通知—必要措施”规则看似为网络服务提供者的法律责任规范,实则将其视为网络服务提供者的免责条款更为恰当。因此,该规则也被称为“避风港”规则。

在“避风港”规则的保护下,网络服务提供者接到侵权通知后及时采取了必要措施的,可免除其法律责任。但在以下情形中,网络服务提供者仍然要承担侵权责任:网络服务提供者知道或者应当知道侵权行为存在且未采取必要措施的,需承担侵权责任。此即“红旗”规则。“红旗”规则规定于《民法典》第 1197 条,根据该条规定,只有在侵权行为显而易见如“红旗飘扬”时,网络服务提供者才需要与实施侵权行为的用户共同承担连带责任。当前司法实践将“红旗”规则的适用标准限定于包含定位信息的特定侵权内容的知道(应知/明知)上,例如《最高人民法院关于审理侵害信息网络传播权民事纠纷案件适用法律若干问题的规定》(简称《规定》)第 12 条①列举的应认定网络服务提供者为“应知”的相关情形,都暗含了网络服务提供者必须知悉侵权内容所在位置。因此,网络服务提供者仅宽泛地知悉其平台内存在侵权行为并不会触发“红旗”规则。

总的来说,当前中国对网络服务提供者采用的是“避风港”规则与“红旗”规则并行的规制模式,但该模式在规制算法推荐网络服务提供者时存在“真空地带”。封闭型算法推荐因其依赖的数据集来源于平台本身,当其造成侵权结果时,依据侵权责任法的一般规则即可解决。精选型算法推荐依赖的数据集需要经由平台选择,即经由网络服务提供者审查编辑,此种情形下,一旦造成侵权结果而网络服务提供者未采取必要措施的,可以认定网络服务提供者明确知悉侵权内容所在的定位信息(《规定》第 12 条中亦列举了这一情形),从而归入“红旗”规则的调整范围。但在规制公开型算法推荐服务提供者时,“避风港”规则和司法系统宽松的责任豁免倾向

① 《规定》第 12 条规定:“有下列情形之一的,人民法院可以根据案件具体情况,认定提供信息存储空间服务的网络服务提供者应知网络用户侵害信息网络传播权:(一)将热播影视作品等置于首页或者其他主要页面等能够为网络服务提供者明显感知的位置的;(二)对热播影视作品等的主题、内容主动进行选择、编辑、整理、推荐,或者为其设立专门的排行榜的;(三)其他可以明显感知相关作品、表演、录音录像制品为未经许可提供,仍未采取合理措施的情形。”

会导致“红旗”规则被实质性架空。其一,在“避风港”规则下,公开型算法推荐网络服务提供者被免除了对于用户或广告商直接提供或通过其他来源渠道自动引入/聚合的数据集的事先审查义务,此时公开型算法推荐服务提供者将缺乏主动发现和处理侵权行为的制度动因。其二,如前所述,司法实践中援引“红旗”规则的前提是网络服务提供者明确知悉平台内侵权内容的定位信息,而公开型算法推荐网络服务提供者因毋须对来源数据集进行审查编辑,自无法直接认定其明确知悉平台内侵权内容的定位信息。

(二)技术困境:算法的“黑箱”与价值内嵌逻辑

1. 算法的“黑箱”逻辑

控制论中对“黑箱”作了定义:“所谓黑箱是指这样一个系统,我们只能得到它的输入值和输出值,而不知道其内部结构。”[13]换言之,对于“黑箱”,人们只能观测到输入值和输出值,但无从了解其内部的复杂性工作。“黑箱”一词也被控制论者用以表示“任何一部过于复杂的机器或者任何一组过于复杂的指令”。[14]科技实践中的“黑箱”指代的是已经被广泛接受为真实的、有用的科学理论或技术产品(也被称为“‘自然’的因素”),它将复杂且抽象的科学理论或技术产品作为一个不需解析的整体进行讨论,阻止人们对其内部的复杂性工作进行探究、质疑或争论,从而有助于科技的广泛传播。此外,科技的实践应用一旦成功,也会进一步固化其“黑箱”状态,盖因科技企业会将此技术作为商业秘密而避免向外界扩张适用或解释。算法作为人工智能时代的新技术,天然具有“黑箱”性质,这与算法的技术运行原理密切相关。从技术原理层面观察,算法是一种计算机程序,其是在数据采集和数据模型训练的基础上,依据已设定的数学规则进行运算并输出结果。简单来说,算法包括了“输入—运算—输出”三个步骤,是“程序设计+数学规则”的集合。即便人们能够获悉算法的输入数据和输出结果,但对于算法内部的运算过程也无从得知,由此便形成了算法“黑箱”。

算法所形成的“程序设计+数学规则”的高度技术性和复杂性,导致非算法专业人士尤其是普通公众,无法掌握或理解算法的运行和决策原理。这掩盖了网络服务提供者在运行算法过程中的一些潜在风险,同时也导致社会个体难以发现算法技术运行过程中的问题并参与到对其的质疑和争论中。换言之,算法“黑箱”的技术壁垒导致非算法专业人士无法探究算法推荐网络服务提供者是否存在过错,这足以阻碍对算法推荐网络服务提供者过错责任的归责。即便是采用无过错责任原则,算法“黑箱”的技术壁垒也会导致司法实践难以认定损害结果与算法运行间是否存在因果关系,不仅当事人举证与论证困难,作为非算法专业人士的裁判者也无法作出正确判断。

2. 算法的价值内嵌逻辑

当前不少研究提出了算法的“技术中立”的原假设,但实践中一旦算法走向应用,各种利益与价值取向的交织都使其难以保持中立。技术中立原则最初确立于“环球电影诉索尼案”中,即只要一项技术构成“实质性非侵权使用”,无论这种技术是否被用于合法或有争议的目的,技术服务提供者都不必对用户实施的或可能实施的侵权行为承担责任。但此后的实践表明,部分技术服务提供者明知或应当知道自己的用户会利用自己提供的技术实施侵权行为,仍持放任态度。此时,以“实质性非侵权用途”为标准免除技术服务提供者的法律责任既不公平也不合理。“Napster 案”的判决遂引入了“帮助侵权”的判断标准,即如果明知一种行为构成侵权,仍然“引诱、促成或实质性帮助他人进行侵权行为”,就应当作为“帮助侵权者”承担侵权责任。在“索尼案”①中,少数派法官指出:“没有人会单独为非侵权的目的去购买产品,制造商明显是有意从侵权行为中获利,要求其承担责任是合适的。”确立“引诱侵权”规则的“Grokster 案”②判决也明确提出:第一,被告的目的在于吸引之前的 Napster 用户;第二,被告未尝试开发建立一定的过滤软件来防止使用者下载、传播受版权法保护的作品,同时也没有尽到相应的监管责任,证明被告主观上有为其用户提供侵权之便的意图;第三,被告是通过出卖广告空间来获利的,所收取的广告费与软件的用户数量成正比。[15]

算法作为技术勃兴的产物,在推动人类生产和生活方式颠覆性变革的同时,其中立性同样受到质疑。算法在实践应用中会无可避免地呈现出偏向性,这与其运行过程中的价值嵌入逻辑密切相关。代码的内嵌规则决定了算法会隐藏人的自主意识,算法部署或使用者在算法部署/使用过程中或多或

① 参见 *Sony Corporation of America er al. v. Universal City Studios, Inc., et al.*, 464 U. S. 417 at 498 (1984)。

② 参见 *MGM Studios Inc. v. Grokster Ltd.*, 545 U. S. (2005)。

少会嵌入自身的价值观念或主观意图。例如,数据集选取、变量选择、权重设定、架构设计等均是算法部署者在设计算法任务时需考量的因素,且可能受到个人价值观念或主观意图的影响。部分算法推荐网络服务提供者可能出于增强用户粘性、达到流量最大化、攫取商业利益的目的,肆无忌惮地在运用算法技术时嵌入利润导向的价值倾向,更甚者不惜制造"脏数据",污染网络生态场域。[15]值得注意的是,当算法部署者或使用者在算法中嵌入自身的价值观念或主观意图后,算法在其后续的技术更新迭代过程中仍会沿袭且不断深化和放大此种价值倾向。因此,在算法价值内嵌的运行逻辑下,算法推荐网络服务提供者能够通过设计/部署算法,赋予算法以特定价值,并通过算法运行达到约束甚至支配他人行为的效果。此时,在算法"黑箱"的加持下,非算法专业人士难以甚至无法获悉算法推荐网络服务提供者是否在设计/部署算法时嵌入了其自身的价值观念或主观意图,以及嵌入的价值观念或主观意图是否合法,更无从提及对此种行为的规制问题。

综上所述,算法的"黑箱"与价值内嵌逻辑所构成的技术壁垒成为了算法推荐网络服务提供者权力归化的困境之一。

四、算法推荐网络服务提供者权力归化的应然路径

(一)强化算法推荐网络服务提供者的注意义务

算法时代下的相关利益关系已经发生了变化,因此,不应一味减轻网络服务提供者的注意义务。相反,应当强化其注意义务,理由如下。

1. 算法推荐网络服务提供者极强的信息管理能力

在民法的语境下,网络服务提供者的注意义务实乃物理空间内公共场所经营者的安全保障义务在网络空间的延伸。立足民法安全保障义务的相关理论,"善良管理人"这一概念可以为网络服务提供者的注意义务之高低提供基准线。"善良管理人"的注意义务要求具有特定职责的行为人应当承担与其专业思维、认知能力相匹配的注意义务。[16]引申至网络空间中,网络服务提供者信息管理能力的强弱将直接决定其注意义务的高低,而算法推荐的运行机制赋予了网络服务提供者极强的信息管理能力。算法推荐的运行机制遵循以下四个步骤:首先,智能推荐算法对推荐内容进行编辑、审查、识别、筛选和分类(编辑、审查推荐内容仅出现在封闭型推荐算法和精选型推荐算法中);其次,智能算法收集、分析用户的行为痕迹后,对用户进行标签化并勾勒用户的"个人画像";再次,以形成的用户标签和"个人画像"为根据向其推送相应的信息;最后,结合用户的行为与反馈,对此前形成的用户标签和"个人画像"进行修正,并根据修正后的用户标签和"个人画像"再进行内容推荐。[17]在这一闭环式的运行机制中,网络服务提供者作为推荐算法的设计/部署者,实际上直接控制了用户所能接收到的信息范围和内容,具有极强的信息管理能力。因此,算法推荐网络服务提供者极强的信息管理能力决定了其应当承担较重的注意义务。

2. 算法推荐网络服务提供者在言论表达和控制中的"把关人"角色

网络服务提供者运用其掌握的算法推荐技术,可以干预和控制言论表达的所有环节。有学者认为网络服务提供者充当着数字环境中的"结构性角色",将其权力归结为四个方面:一是管理网络空间公共表达与参与平台,决定信息流向;二是管理用户活动空间与数据,知悉传播主体的行为;三是控制用户网络活动;四是突破国家边界。[18]上述四个方面实际上基本涵盖了"信息获取—言论表达—言论传播"的全过程。在算法推荐技术的加持下,网络服务提供者的言论控制和保护权力被放大,盖因算法推荐技术的设计标准、运行规则等均由服务提供者自行制定。掌握了算法推荐技术的网络服务提供者不仅能干预言论表达的所有环节,甚至能够决定用户得以发表的言论范围、内容及其影响射程。此外,网络服务提供者还部分承担了类同于言论"裁判"的司法功能,尤其是社交网络平台,往往会制定相关言论规则,在纠纷发生时通过聆听各方的主张、申辩和解释,进而作出具有实质意义的"裁判"结论(删除相关言论、禁言等)。[19]因此,基于算法推荐网络服务提供者在言论表达和控制中的"把关人"角色,有必要强化其注意义务。

3. 网络服务提供者与内容生产者对同一内容的利益回报存在价值差

当前已进入流量经济时代,海量的内容生产者诞生并相互竞争,这致使内容生产者欲获得合理利益回报的难度加大。相反,流量经济使得呈现内容的网络服务提供者能够通过精准分发和传播内容获取高额利益。概言之,权利人的利益回报与网络服务提供者的利益回报处于不对等状态。但社会发展依赖的是创作者的高质量创作内容,而非网络服务

提供者对内容的精准分发和传播。此外,强化算法推荐网络服务提供者的注意义务已有立法雏形——《规定》第11条第1款即明确了网络服务提供者从网络用户提供的作品中直接获得经济利益的,要承担较高的注意义务。总体而言,对于同一内容,内容生产者与网络服务提供者间的利益回报存在价值差,其中网络服务提供者属于资源优势方,为平衡二者间的利益关系,有必要对网络服务提供者课以较高的注意义务。

4. 强化算法推荐网络服务提供者的注意义务符合法经济学原理

1947年美国法官勒尼德·汉德(Learned Hand)在"美国诉卡洛尔·波音拖船公司案"中提出了"汉德公式"。根据汉德公式,当B(Burden)<P(Probability)×L(Loss),即避免侵权事故发生的成本(B)小于预期侵权事故发生的概率(P)和预期侵权事故造成的损失(L)的乘积时,行为人需要承担侵权责任①。对于算法推荐网络服务提供者而言,首先,由于网络服务提供者出于设计/部署算法推荐的目的,已经投入大量资金并形成较为成熟的内容识别机制,此时由网络服务提供者直接在内容识别机制中增加侵权内容识别模块,能够大幅度降低避免侵权事故发生的成本(B)。其次,在流量经济时代,网络服务提供者往往出于谋取巨额商业利益的目的而利用算法推荐大肆精准分发和传播侵权内容。这将提高预期侵权事故发生的概率(P)。最后,当前的网络服务提供者已经突破国家边界的限制,在全球形成了一个整体的虚拟空间,在算法推荐的加持下,侵权内容的传播范围更广、传播速度更快,预期造成的侵权损失(L)也必然更高。[20]因此,从法经济学的角度出发,强化算法推荐网络服务提供者的注意义务能够实现效益最大化。

(二)廓清算法推荐网络服务提供者的算法透明义务

算法推荐网络服务提供者的私权力之所以难以归化的主要技术困境是,算法技术天然具有"黑箱"特性。解决"黑箱"问题的最直接方法就是打破"黑箱",让算法在阳光下运行。由此,学者们纷纷提出了算法透明理论。目前关于算法透明的解释有广义与狭义之分,辨析的关键在于算法透明与算法解释的关系。在算法透明的狭义解释下,算法透明与算法解释是相对独立的治理维度,二者在自身功能和关注点上存在差异:狭义的算法透明关注源代码等信息的公开,是算法解释的前提和基础;算法解释则更侧重可理解性,即着重以通俗易懂的方式向算法相对人或社会公众解释算法决策的原理、决策树和逻辑关系等。[21]在算法透明的广义解释下,算法透明和算法解释并非两个独立的治理维度,而是二者融合成为"可理解的算法透明",即广义的算法透明涵括了算法解释,不仅关注源代码等信息的公开,同时侧重以通俗易懂的解释方式达致算法决策的可理解性。[22]153-154 笔者采用的是算法透明的广义解释。

1. 算法透明的功用

透明原则初始是WTO法律制度的基本原则,主要用以要求政府在市场监管过程中公开其管理和服务市场的相关信息,旨在克服因政策变动导致的市场风险,以及防止市场歧视、为市场运行提供可预见性规则和审议依据。[23]算法透明的提出亦是出于防止算法妨害、建立算法信任,以及为算法审议和问责提供可视化依据之目的。然而,对于算法透明的功用,学者们提出了不同的观点。持算法透明反对论的学者们主要基于三大类原因:一是算法透明损害相关利益,其中涉及损害商业秘密制度的效能、为他人制造规避或"戏耍"算法的机会、泄露个人隐私或诱发个人信息安全问题等;[22]155 二是算法透明不可能,这是一种源于不理解技术产生的遐想;[24]165 三是算法透明无意义,因算法技术极其复杂,即便公开算法代码和数据等信息,普通公众也难以理解其中的运行原理。

然而,算法透明不仅可能而且可行。首先,算法透明的实现方式不仅仅局限于对算法源代码、数据等信息的公开。从当前的立法例来看,从标示义务到向监管部门报备参数,一直到向社会公开参数都是算法透明的重要方式。[24]173 其次,普通公众难以理解算法技术中蕴含的复杂原理并不意味着算法透明无意义,而仅仅反映出当前算法设计/部署者缺乏以传统直观、通俗易懂的文字、文法和语言来解释算法的能力。最后,算法透明在技术上具有可行性,反向工程学即为实现算法透明的重要技术手段。反向工程学可通过输出结果逆向测试的方式打开算法"黑箱"。目前国际上已有相关示例:前谷歌公司工程师尧姆·沙洛(Guillaume Chaslot)在2018年创办了算法透明网站(algotransparency.org),并于该网站上公布了其利用反向工程学破解了YouTube如何利

① 参见 *Unite States v. Carroll Towing Co.*, 159 F.2d 169, (1947)。

用算法扭曲事实的真相。

算法透明主要在知情权和可问责性两个维度发挥作用。其一,算法透明能够赋予算法服务对象相当程度的知情权。此种知情权既可以是事前的,也可以是事后的:通过事前知情,能够起到预防算法妨害之功用;通过事后对算法决策提出公平性和合理性之质疑,能够为被侵权者提供救济途径。其二,算法透明可以让算法设计/部署者具有可问责性,即在算法透明的要求下,当算法运行过程中出现精确性与公平性之偏差,监管者可以依据所披露的算法相关信息来判定算法设计/部署者的责任。此外,由于算法是一种技术机制,即便其天然具有“黑箱”性质,但相较于人类决策者的内在偏见和私念,技术工具仍然更具有可视化的透明标准,也更易于监督。

需要注意的是,从表面上看,算法透明义务的明确似乎是一种显而易见的破除算法“黑箱”的方案,盖因可借助该义务要求算法设计/部署者告知用户算法的工作原理,并在适宜情形下赋予用户相应的权利。但应当警惕陷入“透明度谬误”。算法透明仅仅是一种补救措施,其实际上只是向用户个体提供了不具有实体权利的相关算法信息,却无法提供赋权和控制的预期效应。算法透明并不能真正解决算法“黑箱”问题,但它可以作为一个潜在的有用因素,即将其视为促进其他规则和监管的一般原则,用以支持诸如用户算法解释权、算法问责机制等相应的制度建设。

2. 算法透明的限度

鉴于算法透明涉及多方利益——商业秘密、个人隐私和个人信息安全等,其在践行中应当设置相应的限度。行政法中的两种透明类型或可为算法透明的限度提供参考:“鱼缸透明”和“理性透明”。

“鱼缸透明”侧重于公众获取政府掌握的信息和政府工作的信息,即公众能够窥视政府内部,获取有关官员正在做什么的信息。它的具体实现方式包括举行听证会、公开档案和记录以及网上公开等,其目的是确保受政府决策影响的人能够监测官员正在做什么,并在知情的基础上作出回应。与“鱼缸透明”强调公众获取有关政府正在做的事情之信息相比,“理性透明”更强调了这些信息的有用性,即政府是否揭示了它采取行动的原因。“理性透明”侧重的是行为正当性,即政府需要通过给出理由来解释其行动的正当性。[25] 从表征来看,“鱼缸透明”和“理性透明”似乎分别保障的是形式透明和实质透明,但实质上“理性透明”涵盖了“鱼缸透明”的要求,盖因政府要公开解释为何采取某一特定行动,就必须披露它实际上采取了什么行动,以及它采取这一行动所收集到的事实和进行的相关分析。对于以机器学习为基础的算法技术而言,算法设计/部署者可能能够实现“鱼缸透明”,即确保受算法决策影响的人能够监测算法正在做什么,并在知情的基础上作出回应,但却难以实现“理性透明”,主要原因在于算法设计/部署者难以充分解释算法决策的内在逻辑及其决策结果的来源。但“理性透明”是算法透明的追求目标,也是解决算法“黑箱”实质难题的关键。因此,“鱼缸透明”可以说是算法透明目前应当达到的限度,而“理性透明”则是算法透明未来的发展方向。

3. 算法解释的标准

“算法解释能够直面算法设计并破解算法黑箱,从而在法律问责过程当中发挥作用。”[14] 因此,算法透明需要以算法解释为技术支撑。算法解释有“完全解释”与“可信解释”之分。具体而言,“完全解释”要求解释算法从输入到输出的全过程,即人工智能系统如何运行、如何与数据进行交互的背景信息等,包括要求算法设计/部署者公开数据、参数甚至是源代码。“可信解释”与“完全解释”侧重点不同,“可信解释”侧重对决策结果的解释,即关注决策时对人产生具体影响的相关因素之解释,而非对算法整个运行过程和技术模型的解释。倘若说“完全解释”侧重于对“系统功能”的解释,那么“可信解释”则侧重于对“特定决策”的解释,其不纠结于算法模型的可解释性,而是将重心转移至向个人解释特定决策结果,证明决策结果的合理性即可。

在算法时代(特别是自主学习形态),“完全解释”不具有技术上的可行性和应用上的现实性。就技术层面而言,算法是以机器学习为基础的,机器学习作为一种自动化决策方法,其决策规则由数据驱动,源代码的公布只是暴露了使用的机器学习方法,而无法解释决策过程。[26] 有效的学习算法模型包含海量数据和代码,且会根据输入的数据动态调整内在决策模式并迭代代码内容,加之算法模块在运行过程中会持续复杂化和深度拟人化,故算法设计/部署者通常难以说明算法的决策过程。就应用层面而言,算法研究和应用的主体主要为大型互联网企业,驱动它们开发应用算法的根本原因正是算法技术应用背后的巨大商业利益。因此,算法技术也被这些大型互联网企业视为商业秘密,而算法“黑箱”很大程度上则作为一种防止商业秘密外泄、自我防护的

技术手段存在。倘若采用“完全解释”的标准,必然会增加互联网企业的成本,削弱其竞争优势。从长远来看,合规成本的增加和获利的减少必然会打击企业投资算法的积极性,不利于科技发展。

算法透明抑或算法解释并非为了挤压算法的发展空间,而是为了实现算法权利人与相对人间的利益平衡,故算法解释应以“可信解释”为标准,当算法设计/部署者通过源代码公开、运算原理与过程解释、信息输入与反馈及校验等方式,并借助可视化分析、特征关联等技术消解了相对人的疑虑,可认定为实现了“可信解释”。此外,“可信解释”不仅是技术标准,也是算法责任认定的证明标准。为实现“可信解释”,算法部署者还可通过在软件框架内增加可解释技术接口,或提供指标评估模型等方法增强算法可信解释能力,这不仅有利于算法追责,也有利于算法风险预防与运行监管。

五、结语

算法推荐的治理是算法时代的重大命题,算法推荐网络服务提供者基于技术资源优势取得了私权力,但技术并不必然具有中立性。基于代码内嵌规则,算法设计/部署者或多或少会在算法中嵌入其价值观。因此,算法推荐网络服务提供者的权力呈现异化态势,法律应当对此权力异化现象及时作出因应。规制算法推荐网络服务提供者的目的并不在于彻底消除其私权力,而是为了廓清算法推荐的技术运行逻辑和价值负荷序列,通过系统化的思考和制度化的规范,促使算法推荐以技术工具的形态造福人类。

参考文献:

[1]丹尼斯·朗. 权力论[M]. 陆震纶,郑明哲,译. 北京:中国社会科学出版社,2001:6.

[2]谭九生,范晓韵. “算法权力”的异议与证成[J]. 北京行政学院学报,2021(6):13.

[3]司晓. 网络服务提供者知识产权注意义务的设定[J]. 法律科学(西北政法大学学报),2018,36(1):79.

[4]托马斯·霍布斯. 利维坦[M]. 黎思复,黎廷弼,译. 北京:商务印书馆,1997:72.

[5]弗兰克·帕斯奎尔. 黑箱社会:控制金钱和信息的数据法则[M]. 赵亚男,译. 北京:中信出版社,2015:21-22.

[6]丁晓东. 算法与歧视 从美国教育平权案看算法伦理与法律解释[J]. 中外法学,2017,29(6):1622.

[7]陈鹏. 算法的权力:应用与规制[J]. 浙江社会科学,2019(4):53.

[8]凯斯·R·桑斯坦. 信息乌托邦:众人如何生产知识[M]. 毕竟悦,译. 北京:法律出版社,2008:104-110.

[9]薛永龙,汝倩倩. 遮蔽与解蔽:算法推荐场域中的意识形态危局[J]. 自然辩证法研究,2020,36(1):53.

[10]李晟. 人工智能的立法回应:挑战与对策[J]. 地方立法研究,2019,4(5):71.

[11]柳亦博. 人工智能阴影下:政府大数据治理中的伦理困境[J]. 行政论坛,2018,25(3):97-99.

[12]COBBE J,SINGH J. Regulating recommending:motivations,considerations,and principles[J]. European Journal of Law and Technology,2019,10(3):6.

[13]王雨田. 控制论、信息论、系统科学与哲学[M]. 北京:中国人民大学出版社,1988:93.

[14]布鲁诺·拉图尔. 科学在行动:怎样在社会中跟随科学家和工程师[M]. 刘文旋,郑开,译. 北京:东方出版社,2005:4.

[15]何培育,刘梦雪. 技术中立原则在信息网络传播权保护领域的适用[J]. 重庆邮电大学学报(社会科学版),2017,29(3):44.

[16]武豹. 算法推荐时代主流意识形态传播面临的挑战及其应对[J]. 中国石油大学学报(社会科学版),2021,37(4):99.

[17]王泽鉴. 民法研究系列:侵权行为[M]. 北京:北京大学出版社,2009:242.

[18]任安麒. 网络服务平台算法推荐的著作权侵权认定规则[EB/OL]. (2021-12-15)[2022-03-07]. https://doi.org/10.13766/j.bhsk.1008-2204.2021.0801.

[19]张小强. 互联网的网络化治理:用户权利的契约化与网络中介私权力依赖[J]. 新闻与传播研究,2018,25(7):88-90.

[20]齐延平,何晓斌. 算法社会言论自由保护中的国家角色[J]. 华东政法大学学报,2019,22(6):10-11.

[21]魏远山. 算法解释请求权及其权利范畴研究[J]. 甘肃政法学院学报,2020(1):148-151.

[22]魏远山. 算法透明的迷失与回归:功能定位与实现路径[J]. 北方法学,2021,15(1).

[23]蔡莉妍. 论航运领域滥用市场支配地位行为的判定标准及其法律规制[J]. 中国海商法研究,2019,30(4):104-105.

[24]汪庆华. 算法透明的多重维度和算法问责[J]. 比较法研究,2020(6).

[25]COGLIANESE C,LEHR D. Transparency and algorithmic governance[J]. Administrative Law Review,2019,71(1):32-38.

[26]KROLL J A,HUEY J,BAROCAS S,et al. Accountable algorithms[J]. University of Pennsylvania Law Review,2016,165(3):638.