

# 网络平台监管的算法问责制构建

张凌寒

**【摘要】**算法的广泛应用使得平台运行日益自动化,加剧了网络平台事前的主观过错认定机制困境,导致现有平台监管追责机理模糊化、治理节点滞后、责任设置不符合比例原则。人工智能时代的平台监管,既应符合平台底层的技术逻辑,也应符合主客观相一致、责罚相适应的法律原则。因此,应穿透网络平台运行的技术面纱,将平台监管的触角和追责视角直接指向背后的算法责任。技术原理是中立的,但是技术的应用是包含主观意图的,应将平台算法设计部署的主观过错作为问责依据。平台的主观过错认定需明晰可回溯的问责点,可通过算法评估制度设置,并以算法备案制度事前固定问责点。事后可要求平台根据备案内容作出算法解释,说明设计的目的和预期后果,结合客观损害结果予以归责。平台算法法律责任的设置上,对于本身的设计过错作为实质责任予以归责,如果平台提供虚假备案与解释则承担不真实解释责任。

**【关键词】**算法;平台监管;主观过错;算法备案;算法解释;算法问责

**【作者简介】**张凌寒(1982-),北京科技大学文法学院副教授,法学博士。

**【原文出处】**《东方法学》(沪),2021.3.22~40

**【基金项目】**本文系国家社会科学基金一般项目“网络平台监管的算法问责制构建研究”(项目批准号:20BFX151)的阶段性研究成果。

## 一、问题的提出:平台算法自动化加剧的归责难题

长期以来,网络平台责任的视野一直被动地停留在事后的严格责任或者间接责任式的归责模式,平台责任追究的社会效果与法律效果均不尽如人意。网络平台时常自辩自己处于技术中立的地位,无法承受“海量数据处理”之重。公众虽然是受保护方,但也时常认为平台可能在作为“责任兜底之筐”而承担过重的责任。更重要的是,网络平台监管部门的法律责任设置和事后归责机理模糊化,调整对象不明确,责任设置似乎也不符合比例原则,因此常被诟病是“事故型问责”或“按需求监管”<sup>①</sup>。原因在于,针对平台的监管和法律责任设置,更多在事后根据危害结果要求平台承担责任,事前主观过错认定机制模糊不清,导致必要性、合理性存疑,难免产生平台委屈、公众同情、监管部门底气不足的情况。<sup>②</sup>

然而,平台主观过错认定机制的模糊不清的问

题在人工智能时代进一步加剧。ABC技术(即算法Algorithm、大数据Big Data、云存储Cloud)<sup>③</sup>进一步削弱了认定主观过错的基础:第一,平台对生产资源、用户的控制只提供“轻链接”,地位似乎更加中立。例如,网约车平台、外卖平台只是提供了供需双方的匹配服务,与生产资料并没有太强的黏性。第二,算法使得平台运行更加自动化,人为干预日益减少。例如内容平台传统的“通知—删除”机制、电商平台的定价依赖算法,<sup>④</sup>算法出现的错误常被平台解释为“技术事故”<sup>⑤</sup>。按照传统平台责任理论,基于“平台是技术中立的介质属性”<sup>⑥</sup>,平台责任的“行为人与责任人相分离”<sup>⑦</sup>,平台的自动化似乎必然导向更轻的平台主观过错,进而导向更轻的平台责任。

与此种推论相悖,各国近年来却不约而同地加强了平台责任,平台责任严格化似乎有愈演愈烈的趋势。我国不仅不断加强平台信息管理的主体责任,<sup>⑧</sup>也加强了平台垄断、不正当竞争行为的法律规

制。<sup>⑨</sup>2019年欧盟通过《欧盟版权指令》，其中第13条要求脸书、YouTube等平台积极监测其用户的内容，加强版权审核。<sup>⑩</sup>就连一向主张宽松产业政策的美国，也出现了革新平台中立地位、加强平台责任的趋势。2019年美国参议院的《停止支持互联网审查法案》主张取消大型科技公司在《通信内容端正法》第230条之下所自动享有的责任豁免。<sup>⑪</sup>与此同时，科技企业也按照政策导向展开自我规制，如“剑桥丑闻”后脸书放弃了自称“内容中立平台”的立场开始人工审核内容。<sup>⑫</sup>

理论与现象的撕裂引发思考。显然，传统的平台追责思路无法应对技术与社会的发展。只要监管指向平台的行为，面临的都是自动化的“技术面纱”。如何形成符合主客观相一致的法理和平台技术逻辑的问责思路，尤其是明确平台在主观过错方面的认定机制与以此为基础的归责机制、责任体系，是人工智能时代平台责任基础理论面临的重大议题。

## 二、主观过错认定机制缺失下的平台监管困境

由于平台主观过错认定机制的缺失和模糊，现有的平台监管存在两大缺憾：其一，现有平台追责思路不符合技术逻辑，仍遵循“主体—行为—责任”的传统思路，但平台运行自动化的技术逻辑使其缺乏事前过错的认定机制，导致只能依据事后结果被动追责，治理节点滞后。其二，现有平台责任的设置缺乏对不同层次主观过错的考量，不符合主客观相一致、责罚相当的法理，使得平台责任设置缺乏体系性、合理性。不符合比例原则，既损害法律的权威性，也使得法律法规受到基于功能主义设置、缺乏规范和理论基础的质疑。

### （一）平台主观过错认定缺失下的事后被动监管

由于平台主观过错的认定机制面临着“技术中立面纱”与“行为人与责任人相分离”的困境，监管部门只能在危害结果发生后启动监管和追责，具体追责时却往往找不到、理不清追责的基础和理由，陷于实际上的严格责任或者无过错责任。

从规范论的角度看，主观过错是平台承担民事责任或行政处罚的要件。早在2000年的《互联网信息服务管理办法》就强调平台在“发现”的前提下，对

用户发布的不法内容有避免传播的义务。<sup>⑬</sup>我国原侵权责任法第36条规定，在平台“知道”或者“应当知道”发生侵权行为时，对损害结果承担直接或连带侵权责任。<sup>⑭</sup>我国网络安全法为平台设定的“网络安全保护义务”，也要求平台对违法信息的处置以“发现违法信息”为前提，对于“未被发现”或“未能被发现”的违法信息不承担法律责任。<sup>⑮</sup>换句话说，平台的法律责任是一种过错责任，过错形态包括注意与过失。2020年我国修订的行政处罚法也体现了考量行政相对人主观过错、追求公平正义的精神。<sup>⑯</sup>

损害结果发生之后，法律责任的追究必然要回溯平台的主观过错，过错的认定却遭遇两重障碍。其一，过错认定面临“技术中立”的抗辩。例如快播案中，其创始人王欣在法庭的抗辩理由即为“技术是中立的”<sup>⑰</sup>。今日头条创始人张一鸣在多个场合提到“算法是没有价值观的”<sup>⑱</sup>。其二，面临着“行为人与责任人相分离”的难题。用户违法后平台是否要承担法律责任？平台方一贯主张自己是严格的“传输介质属性”，要区分内容服务提供者和网络服务提供者。<sup>⑲</sup>由此可见，由于“技术中立面纱”与“行为人与责任人分离”两个障碍，对平台追责的基础和理由并不明晰。

这两种过错的认定障碍造成监管部门只能根据危害结果对平台的主观过错作出倒推，导致实践层面和规范层面的严重后果。从实践层面来看，监管部门只能根据违法结果倒推主观过错。例如，2015年国家工商总局发布阿里巴巴“涉嫌在明知、应知、故意或过失等情况下为无照经营、商标侵权、虚假宣传、传销、消费侵权等行为提供便利、实施条件”<sup>⑳</sup>。根据结果倒推主观过错，陷入了实质上的“严格责任”或“无过错责任”式的追责模式。

从规范层面来看，倒推导致法律只能不断扩大平台的事前注意义务范围。二十年来，网络平台的注意义务范围不断扩张。我国不断强化网络平台的主体责任与社会责任，将对信息的主动监控义务加诸网络平台，客观上无限扩张了“过失”的注意义务范围。<sup>㉑</sup>此举使平台追责的主观过错得以合理化，但也必然导致了平台责任的构成要件与法律后果的不明。与其说主体责任适用于平台监管是一种理论革

新,不如说是为了使平台承担责任达到主客观相一致作出的理论妥协。

由于主观过错认定机制的缺失,导致平台责任的追究缺乏正当性,形成了客观上“出事找平台”的事后监管被动局面。监管部门被批评隐藏于平台的幕后,避免自身直接面对网络海量信息监管可能带来的行政复议与行政诉讼;平台责任的设置则被批评呈现“管道化”趋势<sup>②</sup>,承担了过于沉重的私人审查义务。<sup>③</sup>

## (二)平台过错层次不明下的责任体系混乱

平台责任缺乏明晰的主观过错认定机制,导致实践中对平台的追责具有随意性,平台责任难以符合主客观相一致的基本原则,缺乏可预见性。因此,现实的平台执法往往被诟病为“按需求处罚”或“运动式执法”,也导致了平台监管容易走向过度预防而脱离法治框架。

缺乏清晰的主观过错认定机制,导致实践中平台法律责任体系的混乱和随意。第一,实践中的平台责任严重缺乏层次性。网络安全法及其相关法规中平台责任形式包括责令整改、警告、罚款等。<sup>④</sup>在网络平台的监管实践中,责令整改已经成为最主要的行政措施。2020年江苏省网安部门依据网络安全法办理的580起行政案件中,责令限期整改430家次,占执法总量的74%。<sup>⑤</sup>第二,实践中的平台监管存在严重的随意性。以网络安全法第47条平台未尽到“网络信息管理义务”为例,存在个案罚与不罚、责任程度不一的情况。有的平台被顶格罚款从重处罚,有的则不予处罚仅予责令整改。<sup>⑥</sup>平台监管中缺乏对平台是否尽到注意义务的主观过错的论证和考量,导致了从重处罚与免于处罚缺少有力的说理论证,平台责任呈现逻辑的混乱与监管的随意。平台责任呈现实质上的“客观归责”,即仅以结果作为追究责任的原因,难免被诟病“按需求处罚”<sup>⑦</sup>。

这种责任体系的混乱造成了平台与监管部门双输的局面。从平台的角度看,是否承担法律责任并不取决于是否履行了注意义务,而取决于监管部门是否开展了相关的专项治理及其治理力度。互联网行业常常担忧不可预测的运动式执法带来的行业寒冬,长此以往极易因恐惧处罚而失去创新动力。从

监管部门的角度看,主观过错层次考量的缺失使平台责任陷入扁平化、缺乏体系性的特点,客观上呈现“要么约谈,要么关停”的两极化趋势。这使得法律责任的设置和监管执法行动并不符合比例原则,具有“事故型问责”特征而丧失执法的权威性。更严重的后果是监管部门缺乏体系性和层次性的治理策略,而又有避免风险的强烈动机,因此只能不断加强对平台的行为控制。这种预防型的行为管制具有与生俱来的“越严越好”和“越早越好”的内在扩张逻辑,极易脱离法治的框架走向过度预防。

由此可见,平台监管和追责中缺乏对主观过错层次的考量,使得平台法律责任的设置不符合权责一致、责罚相当的法律原则,具有了不确定和难以捉摸的特性。这既损害了监管部门的权威性,又使得平台法律责任的设置缺乏合比例性和体系性,平台监管日趋走向严格监管。

## 三、平台问责原理:主客观相一致的算法责任

平台算法自动化运行造成了“主体一行为一责任”传统理论下的“责任鸿沟”,其实质是以缺失控制权作为理由主张“无过错则无责任”。为此,人工智能时代的平台责任设置必须既符合平台的技术逻辑,也符合法律责任主客观相一致的法律原理。技术原理是中立的、没有主观过错的,但技术的应用如算法的设计与部署是包含价值观和主观意图的,这是法律追责之根本指向,也是平台承担法律责任的根本依据。平台监管的需求不是严格监管而是精准监管,在平台责任的设置中加入主观过错的考量,可以使平台责任设置具有体系性与合比例性。

### (一)平台算法问责的适用场景:弥合“责任鸿沟”

平台算法问责主要适用于如下场景:当损害(危害)是由算法造成时,平台主张算法自动运行“技术中立”而不应承担法律责任时,应对平台算法问责并考察平台在算法设计或运行方面的主观过错,用以弥合平台算法自动化运行造成的“责任鸿沟”。

“责任鸿沟”是2004年由安德利亚斯·马提亚提出的概念,所指的是算法自主学习使得算法的设计者和运营者未能预测未来及其运行产生的后果,人对机器的行动并没有足够的控制权,因此无法要求其承担传统的机器制造者和操作者的过错责任。<sup>⑧</sup>

依照传统的归责原则,一个法律主体只有在其知晓自己的行为及其后果,并且可以自由选择是否作为的时候,才应该承担法律责任。<sup>②</sup>因此,对于机器运行造成的损害结果,传统责任理论一般追究制造者的责任,如果操作者按照制造者设计的具体意图操作,就意味着其对责任的继承和接受。“责任鸿沟”的本质是算法的设计部署者主张因对算法运行缺乏控制力,因此无主观过错则无责任。在平台算法自动运行造成损害结果时,判定应启动算法问责制一般应符合如下场景:

第一,算法运行而非人的行为造成了损害结果。这是因为,如果人的行为在损害结果的产生中占据了主要地位,则一般会落入“主体一行为一责任”的法律框架。而一旦算法的自动化决策对结果呈现占据主导地位,则会产生“责任鸿沟”。举例而言,大众耳熟能详的“大数据杀熟”背后是算法的差异化定价,因此会产生是否合法的讨论。而如果由人来进行“消费越多,价格越高”的价格歧视行为则显然是不合理的。

第二,算法的形式多为平台的“基础服务”或“一般规则”。“基础服务”是指平台算法搭建的平台交易架构与流程,包括但不限于数据收集、存储,认证与信息安全,纠纷解决机制<sup>③</sup>,行为控制激励机制<sup>④</sup>等。从表面上看算法是全自动的决策,实际上算法根据预先设定的输入会自动触发后果。算法也经常以平台运行“一般规则”的形式出现,包括但不限于算法的分配、派单、排序、推荐等功能。在平台设定的目标内,算法的规则对平台内的个体普遍适用,自动生成结果。如电子商务中的全自动定价、叫车平台上的司机与乘客匹配、外卖平台上的用户信用评价、短视频平台上的内容推荐等。

第三,损害常表现为由多方因素造成而难以确定责任来源。有时数据的来源和数据质量都可能造成算法运行的损害结果。例如,在德国前第一夫人贝蒂娜·武尔夫起诉谷歌诽谤责任一案中,谷歌提出的抗辩理由就是“搜索联想词由算法根据实时数据等自动生成,客观无法控制”。<sup>⑤</sup>算法运行的不利后果并非孤立、静态、割裂的,而可能是数据、算法和平台相互联结聚合而成的结果。

以上三种情形,都可被平台用来抗辩并未存在主观过错,因而无须承担法律责任。面对算法带来的平台“责任鸿沟”,平台责任理论以扩张安全保障义务来应对,这导致平台被不恰当地类比为提供经营场所的商家,面临着平台责任被无限扩大化的风险。从平台和消费者的角度来看,这些平台责任分配上的不确定性都是创新和产业发展的障碍。明确平台法律责任追究的理论基础与依据,是保障平台发展与消费者权利的共同需求。

## (二)平台为算法承担过错责任的理据

算法平台即利用算法构建商业模式、争夺用户流量,又利用算法的自动性试图减轻甚至避免监管与责任。技术原理虽然是中立的,但是平台对算法的设计和部署是包含价值观和主观意图的,对算法运行的结果是有基本预见的。这是法律追责的根本指向,也是对平台算法问责的根本依据。将算法设计部署的主观过错作为平台追责的根本依据有三项理据:

### 1. 平台作为算法设计者嵌入了主观意图

首先,算法隐含了平台的价值观和主观意图。例如,十年前有学者即提出“设计隐私”理念,主张平台在算法设计时就应将保护用户隐私的价值嵌入平台架构。我国在2019年开展的App治理专项行动中,也将“嵌入代码、插件”和“提供注销功能”等设计层面的算法作为治理对象,违反者则认定为具有不当收集个人信息的故意。<sup>⑥</sup>由此可见,平台算法对基本交易规则和基础服务的默认设置并不能抹杀其嵌入了设计者主观意图的实质。

事实上,平台很多时候也承认对算法设计和部署的主观意图。在搜索王与谷歌的案件中,搜索王公司声称谷歌的恶意篡改网页排名算法导致其访问量急剧下降。<sup>⑦</sup>谷歌工程师直言这是谷歌故意编辑的结果,因其享有“作为言论者的权利”<sup>⑧</sup>。滴滴出行科技有限公司某高管也曾定义涉及强奸杀人恶性案件的“顺风车”服务是“性感的社交场景”。<sup>⑨</sup>可见,平台算法的设计部署即蕴含着平台的主观意图和价值判断。

其次,算法深度学习的“技术黑箱”也无法抹杀平台设计算法的主观意图。深度学习使得算法输入

可以为提取得到的多种特征,输出可以为目标任务的标签或者数值,本质上构建了一种复杂的非线性变换。算法的输出目标一旦被设立,就会自动通过深度学习接近这一目标。<sup>⑤</sup>换句话说,算法自动学习向深层架构发展,它的规则和运作方式很可能让开发者都无法理解。

最后,实质上,深度学习的平台算法同样适用算法问责制,因为算法只是借由深度学习获得了“异质自治权”。归根结底,算法只是在人类设定的目标范围内,起到了替代人类的作用。从哲学的角度来看,责任只能归因于具有自主行动能力的道德主体(人)。现阶段的弱人工智能算法可以被理解为在复杂的环境中无需人工监督,具有自动收集、处理和输出数据的能力。<sup>⑥</sup>这种自动化的能力确切地说应被称为“异质自治权”,也就是在目标范围内为算法提供的最高程度的自治权,而非无限的自治权。<sup>⑦</sup>换句话说,平台系统中,自动运行的算法被人类赋予要完成的任务,由算法自行决定完成这些任务的方式。深度学习造成的算法黑箱内的技术细节,并无法掩盖设计开发算法的主观意图。以“剑桥丑闻”为例,监管部门即使不知晓“剑桥分析”的算法源代码,也不妨碍其了解算法运行目的在于定向投放政治广告操纵选举。<sup>⑧</sup>

## 2. 平台作为算法使用者具备运行控制力与合规义务

当平台不是算法的设计者,而仅为算法的使用者时,平台也需因其具备对算法运行的控制力而承担算法问责的后果。平台屡屡将算法运行中“控制力缺失”作为抗辩法律责任的理由,最为常见的是平台以“行为人与责任人相分离”<sup>⑨</sup>作为抗辩理由。

“行为人与责任人分离”是指对平台法律追责的原因,往往是由用户或者第三方行为引发的,平台并不“生产”违法内容,因此不应为此承担法律责任。例如,微博上的用户发布侵犯他人名誉权的内容,行为人是提供“内容信息提供”的用户而非平台。这一观点源自1996年美国《通信内容端正法案》第230条交互式计算机服务者不被视作提供内容信息的出版商,因此为网站和互联网服务提供商提供了强大的豁免权。<sup>⑩</sup>之后,2000年的美国数字千禧版权法案

(CDMA)沿袭了这一对互联网产业宽松包容的政策,确立“通知—删除”规则,允许网络平台被动中立无需主动发现和介入违法事实。在我国平台民事责任的认定中,也经常将平台认定为“帮助侵权”行为,而主张平台承担较轻的法律责任。

实际上,算法赋予网络平台对信息流的自动处理能力,使得平台的法律责任早已超越了“网络服务提供者”而独立化。平台貌似并不生产具体内容,只是通过算法对具体网页(商户、内容)进行排序与呈现。然而,比起内容生产者,平台算法所起到的屏蔽、推送、排序对用户接收的信息具有更强大的干预作用。从信息论的角度来说,平台算法显著降低了海量网络信息的混乱程度,将用户需要的信息以“关键词——店铺排名”或者“关键词——搜索建议”的形式提供给用户,这本身就是信息,平台也是此种意义上的“信息生产者”。无论是搜索引擎、新闻网站,还是电闪管平台,都扮演着这种信息生产者的角色。例如,淘宝网站虽然不直接出售商品,但其算法决定了商家是否能呈现于消费者的页面中。淘宝搜索规则在2010年7月8日的调整将小商家排序靠后导致流量急剧下降,导致淘宝网众多商家的聚众抗议事件。<sup>⑪</sup>

因此,即使平台仅为算法的使用者,也需要为算法运行承担责任。以谷歌为例,2016年谷歌发现“希特勒”的自动补足是“希特勒是英雄”,以及“犹太人”自动补足为“犹太人是邪恶的”之后,出手对搜索的自动联想结果进行了修改限制。<sup>⑫</sup>早在2009年,我国监管部门因谷歌算法的自动联想功能涉嫌传播淫秽色情内容要求其整改,谷歌随即表示“将彻查所有的服务,并采取一切必要措施来解决搜索结果中存在的问题”,此举与被认为谷歌完全有能力控制搜索算法的结果,而并非其一贯宣称的由算法自动生成。<sup>⑬</sup>“责任鸿沟”只是平台规避法律责任的工具,随时可以根据需要拿出或者收起。

退一步说,即使不从平台对算法控制力的角度论证,世界各国的实践也要求平台承担算法运行(数据处理活动)的合规义务。无论是欧盟一般数据保护条例(GDPR)第35条,还是我国个人信息保护法(草案)第54条,均要求平台对算法运行对“个人的影响

及风险程度”，以及“所采取的安全保护措施是否合法、有效并与风险程度相适应”承担合规义务。这就要求平台在对涉及个人数据处理的重要问题上先进行自我评估，将有可能侵犯用户数据的行为作出相应的决策以及应对数据侵犯的措施，进而将数据处理影响评估内容报告给相关机构。因此，即使平台不是算法的设计者而仅为部署和运行者，也需承担相应的注意义务与合规义务。

### 3. 平台对算法结果输出负有注意义务

实时数据成为算法输出结果的重要因素。即使平台在算法设计、运行上均无过错，也需要为因数据问题造成的算法损害结果承担责任，理由如下：

第一，平台作为算法设计者时，具有自主性的算法对外部实时网络环境作出的反应方式并非完全不可预测，其对不良数据的反应能力仍在算法设计范围内。假定算法在上线后就有了修改自身的能力，这种能力并非从天而降，仍是算法设计者为其产品提供的能力。并且，这种对实时数据的反应能力并非没有限度，而是在事先划定的范围之内进行的。很显然，平台如果作为算法设计者，应将算法对实时数据的反应输出能力设定在安全范围内，以免发生不良后果。

第二，平台作为算法使用者时，应设计算法的干预和嵌入安全机制。对于更复杂的算法，可能需要远程监控，以便在人们认为即使有潜在危险时（即使在远程）加强安全措施。以2020年4月判决的“蚂蚁金服诉企查查案”为例，企查查推送的涉及蚂蚁微贷的清算信息，因算法运行推送方式的设置问题，引发公众将历史清算信息误认为即时信息，发布了“蚂蚁金服破产清算”的算法错误结果。这一案例显示，即使算法自动抓取数据得出结果，平台仍应负有审查算法结果的注意义务。正如判决书所言：“大数据企业对于收集、发布的数据信息仍具有基本的注意义务，对于发布的重大负面敏感信息，应当通过数据过滤、交叉检验等数据处理，确保数据质量。”

第三，平台算法投放前要充分进行线上环境的安全测试。算法应经过数据的“训练”，确保它能够应对现实生活中的数据风险，否则不应将其投放市场。例如，微软的AI聊天算法Tay在推特上运行了

几个小时后，就发表了同情纳粹的言论。<sup>④</sup>微软显然并没有设计这一结果出现，但是其作为设计部署者应该能够预见到将一个机器人暴露在推特这个不乏骚扰和仇恨言论平台上的危险，也应设计程序记录这些恶意的影响来源于哪里。算法从环境中习得适应性行为，因此平台保证算法在设计、部署阶段进行足够的训练，在训练中需模拟实时环境，给算法足够的犯错机会，这是系统运行和改善性能的可行路径。

平台的算法运行造成的危害后果可能是多方作用的结果，但这并不足以使平台逃避其算法责任。在这种情况下，对于任何算法的最终运行结果，可能有多个潜在的责任方。<sup>⑤</sup>但不论算法通过深度学习如何演变，平台作为设计者和部署者都对算法演变有一定的义务，包括监督记录其运行和防止危害结果出现。<sup>⑥</sup>这也是巴尔金所提出的“制造者义务体系”，涉及封闭式和开放式的算法中要求嵌入审计日志，或提前考虑到开放式算法的权限控制。<sup>⑦</sup>

在法律应以平台设计部署算法的主观意图为追责之指向的前提下，需要澄清的是平台算法责任是一种过错责任，但并不意味着过错责任是算法责任的唯一归责原则。如同在侵权责任领域，过错责任是默认的责任形态，但也同时存在着特殊侵权形态的无过错责任或公平责任形态。算法的应用形态并不局限于商业平台，已经被广泛应用于信用评分、政府管理甚至是司法审判中，不同场景下的算法责任不可能适用同样的归责原则，必然会随着算法所处理的数据所涉利益的性质发展出不同样态的归责原则。某些应用于公民自由、重要权利等特定部门的算法，就应如同巴尔金指出的那样，承担公共责任，以类似环境侵权的严格责任要求算法避免对社会公平的污染效应。<sup>⑧</sup>即使在笔者所限定的平台算法论域，算法可能用来导航、新闻推荐、投放竞选广告和动态定价，除基本的过错责任原则外，不应否认其他归责原则适用的可能性。有学者提出，平台某些领域的运行应嵌入公共义务，承担一定的公共责任，这既可能通过扩大平台注意义务的形式实现，也可能以改变归责原则的方式落地。<sup>⑨</sup>

### （三）平台算法问责的合理性与可行性

平台算法问责，从客观方面符合平台底层的技

术逻辑,使平台监管的触角穿透平台的技术面纱而直触其底层技术——算法。平台算法问责符合主客观相一致的法理,可使平台责任合理化、精细化、体系化。

### 1. 平台算法问责符合平台技术逻辑

从底层技术逻辑的角度来看,算法在网络平台的硬件基础设施(架构)、应用程序均扮演着核心角色,决定了平台的运行模式,塑造了用户的习惯与价值观。将算法直接作为平台监管的对象,可以改变现有监管层次过浅的现状,升级传统的“主体—行为—责任”的监管路径。算法应用给平台力量带来了根本性变革,从“生产关系适应生产力”的角度看,原有的平台监管对象应相应革新,以适应算法技术带来的平台角色和地位的演变。

党的十九届四中全会提出“数据作为生产要素参与分配”<sup>①</sup>,算法就是收集处理数据,挖掘数据价值的生产工具。平台类型的演进实质是算法技术的演进。无论何种类型的平台,掌握的算法越先进,就越能在数字经济的时代取得和巩固竞争优势。例如,2006年美国视频平台奈飞推出了百万美元大奖,奖励打败原有个性化推荐算法的队伍,后因此取得长足发展,算法是平台“竞争必备工具”的理念深入人心。<sup>②</sup>同时,算法也可维护平台巨头企业的商业优势,例如谷歌新闻以算法自动化推荐代替人工编辑,由于新闻领域舆情更新频率极高,要求算法不断累计数据重建架构,谷歌为此投入了大量基础设施,价值高达数十亿美元,使得其他平台企业无力与之抗争。<sup>③</sup>

平台运行的技术逻辑是算法,平台商业价值和竞争优势的基础是算法,但是,损害国家利益、社会公共利益和公民个人权益的原因也是算法。尤其需要注意的是,平台通过算法以“作为”的形式作出了某些不当行为。今日头条旗下“内涵段子”被关停是因其推荐算法“只要价值不要价值观”。<sup>④</sup>近年来饱受诟病的“大数据杀熟”事件本质是算法的差异化定价策略。哈佛大学教授朱波夫的《监控资本主义》揭示了算法是平台商业逻辑之根本,它不仅引导用户行为、侵害消费者隐私,更影响市场秩序与社会公平。<sup>⑤</sup>这与“主体—行为—责任”逻辑下平台未尽到

注意义务,以“不作为”的方式实施违法行为的认识迥然不同。算法应当成为监管问责的核心对象。

“主体—行为—责任”思路下,法院疲于将层出不穷的平台算法与传统服务类型比对,以确定网络服务提供者的类型。这种思路忽视了平台算法对不同数据的实质控制力,可能造成平台不当逃避责任。以“微信小程序案”为例,杭州互联网法院认为微信小程序类似于《信息网络传播权条例》(以下简称《条例》)中规定的“接入服务提供者”,仅提供单纯的互联网接入服务。而实际上,《条例》对平台功能区分为存储服务 and 接入服务,是基于台式机时代网页 HTTP 协议跳转的链接行为,在移动互联网 App 时代早已发生改变。<sup>⑥</sup>用户可以基于微信与小程序直接链接的算法和内置于微信内的搜索引擎、排序算法进入微信小程序。因此,平台通过算法对侵权的微信小程序具有控制力。法院甚至在传统平台主体类型外被迫创新平台服务主体类型,也使得平台责任的设置具有了不可预见性。例如,在2017年的“阿里云案”二审中,北京知识产权法院认为,阿里云公司提供的云技术服务既不是自动接入、传输服务提供者,也不是储存、缓存服务,或是连接搜索链接服务提供者,而属于侵权责任法第36条的其他网络服务提供者。<sup>⑦</sup>由此可见,法院知晓云服务与传统存储、接入平台主体类型的本质区别,但又苦于现行法律中没有对应的主体类型,只能通过解释的方法将其归类为“其他网络服务提供者”。传统的“主体—行为—责任”路径必然造成司法实践中越来越多类似的困境。

近两年来,世界各国逐渐重视算法在平台监管的核心地位,将算法作为监管对象的立法动作频出。2019年美国国会议员提出《算法问责法案》,拟对用户在一百万人以上的平台公司进行算法审查。<sup>⑧</sup>2019年澳大利亚竞争与消费者委员会宣布计划设立专门分支机构“主动监视”平台算法运行,赋予其要求披露算法详细信息的权限。<sup>⑨</sup>我国网络监管部门也意识到平台言论呈现的意识形态,其很大程度是算法对网络信息内容排序、推送、屏蔽的结果,2019年的立法中要求平台“优化个性化算法推荐技术”<sup>⑩</sup>。可以预见的是,将算法作为平台监管的直接

对象是人工智能时代的大势所趋。

## 2. 考察主观过错可使平台责任精细化

主观过错既是我国行政监管考量的基本要素,也是我国平台责任立法的已有实践,只有对算法设计部署中存在的主观过错进行网络平台监管,方可实现法律预防违法行为、促进网络产业积极发展的意义。要求平台因不同程度的主观过错而承担不同程度的法律责任,符合法律责任主客观相一致的基本原则。

第一,从现有规范角度来说,无论是侵权责任的过错责任原则,还是刑法的罪责刑相适应原则,主观过错均对法律责任的存在与轻重具有决定性作用。以我国原侵权责任法第36条为例,平台在接到通知后主观状态转为“故意”,相应的法律责任也升格为了“连带责任”。在行政监管中,主观过错亦是责任要件,行为人若其决定从事违反行政法上义务之行为,即得对其非难而要求其承担责任。<sup>②</sup>主观过错的不同程度决定了法律责任的轻重不同。根据行政处罚法第25、26条的规定,在决定是否对违法行为人从轻处罚、减轻处罚甚至是不予处罚时,相对人违法时的主观状态是重要的考量要素之一。在我国行政法体系中对风险防控要求最高的药品监督管理法中,对药品生产者、药品销售者的责任原则也是过错责任原则。<sup>③</sup>

具体到网络监管领域,网络安全法的立法目的之一在于“促进经济社会信息化健康发展”。因此,将主观过错程度纳入法律责任追究的考量中,既可以避免客观归责的严厉制裁带来的打击平台守法积极性的后果,又可以为平台提供接受法律追责的心理基础,提升平台对处罚的可接受程度。将主观过错作为平台责任考量要素也符合我国一贯的立法实践。最高人民法院的司法解释曾就平台主观过错如何认定给出过详细指引,对网络服务提供者是否“知道”设置了一系列考量因素。<sup>④</sup>

第二,从价值考量角度来说,主观过错作为平台追责要素方可体现公平正义的基本理念。行为人的主观过错、危害性与责罚相当,方才具有法律责任的正当性。依据惩罚程度的不同实现对恶性违法行为的预防功能,从而尽可能地增强行政处罚预防功能

的有效性。从我国网络安全法第二章可以看到,其立法目的之一在于“促进经济社会信息化健康发展”。将主观过错纳入平台责任的考量,使平台监管从严格监管走向精准监管,可以促进网络产业积极发展。

从反面论证,如果平台反对法律考察算法设计部署的主观过错,可能导致严格责任的适用,使得平台面对更为严厉的监管来分散平台运行的风险。与严格责任(适用于某些产品责任案件)相比,更宽松的合理性标准可能会鼓励使用机器,从而促进创新和提高安全性。<sup>⑤</sup>同样,考察主观过错使得责任可以在设计者、使用者和其他主体之间合理分配。当算法过程导致有害的结果时,主观过错具有可追溯性,可促使监管部门发现故障的根源,为其分配责任并惩罚或至少教育责任者成为可能。<sup>⑥</sup>

将算法作为平台底层技术逻辑作为监管对象,并在监管中将平台主观过错作为考量,不是为了平台严格监管而是为了精准监管。这不仅符合主客观相一致、责罚相当的法理,也符合我国法律的一贯精神,更加符合奉法强国、良法善治的理念和精神。那么下一个需要回答的问题是,如何在算法的自动化运行中判定和度量主观过错呢?

## 四、平台算法问责路径:平台的主观过错认定机制

平台算法问责的制度构建目的在于促进理想或有益的结果、防止不良或有害结果的产生,并确保适用于人类决策的法律可以有效应用于算法决策。<sup>⑦</sup>清晰的主观过错认定机制对于平台的算法问责至关重要,这意味着平台设计、部署和应用算法的主观过错有着具体界定和表述,应该是在生产和生活实践中可观测、可区分的,在技术上可度量、可标定的。<sup>⑧</sup>平台算法的主观过错应包含哪些内容?平台自动化运行的主观过错如何固定?

### (一)过错的事前固定:算法备案制度

平台因算法被问责的原因是其未履行法律和道德的义务,并无法向利益相关方解释和证明其无主观过错。合理的平台算法问责有赖于清晰的主观过错认定,需通过算法评估、备案与审计对平台事前和运行中的问责点进行固定,以有效评估平台是否可

以评估、控制、纠正算法带来的危害。

### 1. 平台算法备案的内容: 算法评估与审计

第一, 平台算法备案的目的在于设置问责点, 相关具体制度包括算法评估、算法运行监管以及算法审计。这样既能避免事后平台隐瞒与错误披露的可能, 也可预防风险, 避免错误的计算反复被适用于海量主体而使损害弥散化。以风险预防为目的的事前监管已经成为立法实践的选择。例如, 欧盟一般数据保护条例的数据处理评估制度(DPIA)即是针对高风险数据处理活动对数据控制者设置的预警自查义务。<sup>⑧</sup>在事前问责点的设置中, 算法审计也发挥着重要作用, 往往与评估制度被同时使用。例如我国个人信息保护法(草案)第53条规定, 个人信息处理者应定期对个人信息处理活动、采取的保护措施等是否符合法律、行政法规的规定进行审计; 第54条明确提出利用个人信息进行自动化决策应进行事前的风险评估。纽约大学的AI Now研究院也提出了类似的算法影响评估(AIA)制度, 要求在重要的公共部门使用算法前, 应采取的关键步骤来确定何时以及如何产生潜在危害, 包括增加其专业知识和能力以有效实施和评估算法系统, 并允许第三方审计公共部门的算法。<sup>⑨</sup>

第二, 需要进行评估与审计的算法范围。应以算法处理的数据所涉及的利益、对用户行为的干预程度、社会动员程度为标准, 确立不同的算法风险等级并设定不同的监管标准。国家应围绕高敏感、高风险算法建设“关键算法系统监督网络”, 而对于相对低风险算法降低监管强度。例如, 欧盟一般数据保护条例的数据处理评估制度, 要求受到评估的算法系统从事“高风险数据处理活动”; 又或德国数据伦理委员会提出的算法风险评估方案, 主张对数字服务企业使用的算法进行五级风险评级的制度, 对不同级别的算法采取不同强度的监管。<sup>⑩</sup>我国立法将某些算法应用定义为“具有舆论属性或社会动员能力的互联网信息服务”, 并要求其进行自行安全评估。

第三, 算法问责点考量与记录的内容。日后的算法问责需要清晰的问责点, 因此相关制度应着重平台设计部署算法的目的、风险与过程控制能力。

其一, 为考察算法设计部署目的设置的问责点, 应包括对算法设计的目的、与算法设计与应用必要性相称的评估。当一个算法系统有多个价值目标时, 则必须将目标之间的优先级透明化。例如, 自动驾驶汽车可能旨在减少交通事故死亡人数, 降低其对环境的影响, 减少严重伤害, 缩短运输时间, 避免财产损失并提供舒适的乘车体验。当这些目标冲突时哪个优先? 在2018年优步公司的自动驾驶汽车事故中, 优步公司则被指责过于追求运输速度, 因其系统一直判断行人为“不明障碍物”而不减速, 最后一秒识别出行人后则为时已晚。<sup>⑪</sup>人与算法系统之间的主要区别之一是: 人能够协商有冲突的价值或规则, 而算法系统需要在设计时就考虑这些价值的优先顺序。<sup>⑫</sup>

其二, 为考察算法设计部署风险(影响)设计的问责点, 具体内容应包括对算法可能产生的公共利益、公民权利等风险的评估, 包括风险的来源、性质、特殊性和严重性等。有他国实践主张算法的设计部署者应提供更为广泛的评估, 包括算法应用对人权<sup>⑬</sup>、隐私<sup>⑭</sup>和数据保护<sup>⑮</sup>的影响等; 有的学者主张算法设计部署应提供“社会影响声明”“歧视影响评估”<sup>⑯</sup>, 甚至“人类影响声明”<sup>⑰</sup>等, 对此应根据平台与算法的功能进行场景化的界定。

其三, 为考察平台对算法风险与过程控制能力而设置的问责点, 应包括处理可能发生风险的预案与措施。具体包括相关技术信息的留存措施、处理相关风险的技术措施、与相关部门沟通与协助的措施等。比如, 欧盟一般数据保护条例第35条要求数据控制者实施数据安全影响评估。正如联合国国际治理论坛专家报告所言, 平台应采取“持续性、前瞻性和反应性”的步骤来保证相关技术满足其风险评估的技术需求。<sup>⑱</sup>以上问责点的设置为确保准确性和可审计性提供了重要的信息——通过进行敏感性分析、有效性检查和纠错过程, 敦促平台仔细调查错误和不确定的领域, 并在需要的情况下启动第三方的算法审计。

### 2. 平台算法备案制度具体规则

算法评估与审计的内容需要通过算法备案的形式予以固定。算法备案是行政机关作出的一种存档

备查的行为,其目的在于获取平台设计部署的具有潜在危害和风险的算法系统的相关资讯,以固定问责点为今后的行政监管提供信息基础。算法备案可根据算法风险等级,内部自行备案或向监管部门备案,用于日后对平台主观过错的考量。

第一,算法备案的性质。平台算法问责的目标在于问责点的固定,而非设置行政许可或前置的审批程序。其对本身报备的事项(算法应用)并不产生直接的法律效果,其目的不像行政许可那样在于“解禁”,而在于通过对其事后进行的审查、监督来保障备案事项于公益的无害。在这种情况下,备案本身即存在着监督效应。要求平台经过行政审批方可部署算法,势必会降低平台对于创新的追求动力,也会大大减缓平台应对市场竞争的速度。因此,仅要求平台对于一定风险级别的算法进行信息披露,而不以行政部门审批作为部署条件,才能在保证事后监管问责清晰的情况下,不阻碍产业的发展。

第二,算法备案的对象。算法备案应根据不同算法风险等级,要求平台进行自我备案或向监管部门备案。将监管部门作为备案对象,可避免向公众披露而造成商业秘密泄露与算法被不当使用。例如,网络用户在掌握搜索引擎排名权重等信息后,可能通过搜索引擎优化(SEO)技术对排名和搜索建议进行操纵。在各国制度中,也一般将监管部门作为信息披露的对象。诸如美国的《算法问责制(草案)》拟要求平台对监管部门(FTC)披露信息<sup>⑧</sup>,澳大利亚反垄断部门的平台算法监管措施也仅限于对监管部门的披露<sup>⑨</sup>。需要指出的是,尽管备案没有许可的强制效力,信息留存或向监管部门披露本身对于平台来说就形成了监管的威慑力。

第三,算法备案的形式。算法备案的形式可通过官方发布模板,通过平台填写模板的方式备案算法部署的目的、风险,评估过程控制方案。备案内容可用于日后一旦发生客观损害结果后对平台主观过错的考量,并可供监管部门在算法生命周期中持续使用及定期审核。

参考各国算法规制的模式,采取平台就核心算法向监管部门的备案制度,可达到政府监管与行业自律的平衡、信息披露与商业秘密的平衡、源头治理

与事后追责的平衡。过度依赖事后追责难免造成损害结果的弥散化,而对风险预防的过度追求必然导向严格的事前规制,难免抑制产业发展也使监管负担过重。不同的技术机理意味着需要不断从技术源头寻找最新、最关键的治理节点和工具,并将其纳入治理机制之中,以实现治理的可持续性。<sup>⑩</sup>算法备案制度正式从源头设置治理节点,以备案方式为事后追责提供工具。同时,源头技术治理还有一个重要内容,就是在技术底层赋予人工智能“善用”的基因。要求平台在部署算法之前自我评估可能带来的风险,就是将伦理嵌入技术的具体过程。

### (二)过错的事后认定:独立的算法解释环节

如何在危害结果发生之后对平台的主观过错进行认定?监管部门事前备案内容可判断平台是否具有故意或者过失的主观过错,并以此为依据,结合客观的损害结果直接进行行政处罚与问责。在此过程中平台需进行算法解释。算法解释是认定主观过错进行平台问责的必经程序。

第一,算法解释在平台责任认定中具有程序上的必要性,并应在平台算法问责中成为独立设置的环节。算法解释在平台责任认定中具有行政程序上的必要性,这是平台接受行政处罚前的不可或缺的陈述、申辩,是行政法原则保障的行政相对人的基本程序性权利。平台因算法的设计部署而接受行政处罚是一种法律上的不利后果,如果缺乏陈述与申辩环节将是严重的程序瑕疵。听证、陈述、申辩的主要功能是确保当事人参与到行政程序中进行意见表达,算法解释则是这一环节在平台算法问责的行政活动中的具体体现。

第二,算法解释在平台责任的认定中具有功能上的必要性,但其过去一直作为具体案件事实查明、行政处罚程序的一部分呈碎片化状态,而未引起注意。具体案件的算法解释以认定(排除)平台的法律责任为限,程序上多由法庭调查发起,并未与审理过程分离形成单独的程序。如个性化推荐算法向用户推送隆胸广告,用户以百度侵害其隐私权提起诉讼。承办法官撰文详细介绍原告获得算法利用其Cookie数据投放个性化推荐广告的过程,也是基于在审理过程中获得的自动化决策的算法解释。<sup>⑪</sup>在

监管部门的要求下,平台也需进行算法的体系性解释。例如,“魏则西事件”后,联合调查组进驻百度,其调查和监管过程包含着要求百度对竞价排名算法进行解释。最终调查组要求百度采用以信誉度为主要权重的排名算法并落实到位,实际上审查的就是其设计考量的价值取向而非内部的技术细节。因此,算法解释并非创制的制度,而是一直存在于我国司法与行政的实践中。

第三,从监管成本、监管对象的角度,算法解释都应成为平台算法问责制对主观过错考察的独立环节。事后的算法解释行为本身是监管考察的对象。算法解释发生在客观的损害结果发生之后,此时平台对监管部门的算法解释也会受到法律的评价,解释中的隐瞒、虚假解释等行为可能构成相应的法律责任。为监管部门提供虚假解释将承担不实解释的法律责任。事后的算法解释是对平台分配的合规的举证责任。在监管部门启动的有关算法的调查和行政处罚中,平台等互联网企业承担着自证合规的举证责任。算法解释的内容包含要求对算法本身合法性和合理性的审查,类似于因具体行政行为提起诉讼后,对抽象行政行为的附带性审查。应考虑将证明算法合法、合理、无歧视的责任分配给算法使用者和设计者。

### (三)责任设置:平台的实质责任与虚假解释责任

平台法律责任的设置,应根据主观过错的程度轻重分层次设置。对于故意和过失的平台算法的主观过错,施以主客观相一致的实质责任。对于提供虚假算法备案与虚假解释的平台,则施加类似于证券虚假陈述责任的虚假备案与虚假解释责任。

#### 1. 主客观相一致的平台实质责任

平台的实质责任,是指平台由于算法设计、部署与运行中的故意或过失,承担的法律后果。平台算法责任的设置,应符合主客观相一致、责罚相当的法理。根据平台的主观过错形态设置不同层次的法律后果。

第一,平台的过错范围应具有多重层次和丰富内涵。算法责任的主观过错包括故意、过失形态。其一,过错的范围应涵盖现行法律规定,并将其明确化。其二,过错的内涵应吸纳不得对用户行为操纵、

保障用户自治性等技术伦理内容。其三,应鼓励平台算法通过主动承诺的形式,扩大自身注意义务的范围。以美国联邦贸易委员会对脸书的5亿美元罚款事件为例,其罚款的依据是脸书于2012年自身作出的企业隐私政策。当企业自己主动提供了隐私政策和承诺之后,监管部门方有理由对其违背隐私政策的行为以“欺骗性贸易”的名义予以处罚。其四,对待“关键算法”,国家应直接制定相关技术标准,违反则为存在主观过错。

第二,平台的算法责任应实现动态治理,即不仅包括在设计部署时的责任,也包括算法运行过程中的责任。当平台使用算法违反现有的法律或法规时,监管机构应首先检查平台是否能够及如何有效地证明他们有控制权,以确保算法按预期行事。如果发生重大损害并且不存在此类控制措施,或者技术人员在满足此标准方面存在疏忽与过失,则技术人员行为应有可能受到处罚。如果这些控制措施是彻底的并得到适当实施,监管机构可能会确定平台没有采取疏忽行为或故意伤害的行为。

第三,平台实质责任的设置应与主观过错相符合。从社会角度来看,由于公司的疯狂、疏忽、故意忽视或无能而伤害消费者的错误与由于公司努力创新而伤害消费者的错误之间存在显著差异。同样,公司违反规定并对消费者或竞争对手造成重大损害的行为,以及造成很少或根本没有伤害的行为应该加以区分。<sup>⑧</sup>

#### 2. 平台虚假备案与解释的虚假陈述责任

如果平台提供虚假备案与解释,则承担不真实解释责任,类似于公司法、证券法中的虚假陈述责任。虚假备案与解释的形态具体包括作出违背真相的虚假备案、误导性解释、干扰性披露、信息重大遗漏、不当的备案与解释行为等。平台需要对虚假备案与解释承担法律责任,是因为在平台算法问责制的制度框架下,平台的信息披露义务是具有强制性的,披露目的具有公益性。欧盟一般数据保护条例的实践经验表明,如果没有一个明确的定义,平台就有可能以最无害的方式来解释它们的算法。<sup>⑨</sup>

平台进行算法的虚假备案与陈述可能有多重表现形式。除了故意的虚假备案,平台可能通过混淆

的方法提供干扰性披露。平台可能通过披露大量冗余干扰性数据,混杂在关键数据中,以此妨碍解释关键数据内容。<sup>⑥</sup>《黑箱社会》的作者帕斯奎尔称这种行为为“混淆”,就是指刻意增加冗余信息,以此来隐藏算法秘密带来混淆。<sup>⑦</sup>此外,也需关注在技术条件局限作用下导致的信息披露不能。正如詹娜·伯勒尔所提醒的,算法相关信息披露不实的原因可能涉及故意隐瞒、信息的复杂性或者专门机器学习本身的复杂性。<sup>⑧</sup>具体平台算法虚假备案与解释的制度超出了本文篇幅范围,将是未来的研究方向。

### 结语

平台治理与算法规制均是人工智能时代的重要的法学议题。平台借助算法技术链接资源、自动运行,取得和巩固市场优势地位。算法嵌入平台搭建架构、调配资源,实现技术对社会的深刻变革,智能平台就是两者深度融合的具体体现。人工智能时代,法律如何应对这种生产方式与社会关系的深刻变革,是本文的关切所在。

法律相对平台技术的滞后性导致了平台监管中的问题。传统的“主体—行为—责任”的平台追责框架受到平台运行自动化的冲击,主观过错的认定和责任承担理论出现错位,导致了现有监管中主观过错认定机制不清、责任设置不符合比例原则的困境,理论层面也要面对“平台责任中行为人与责任人相分离”“平台本应技术中立所以承担的是无过错责任”的诘问。监管层次过浅、监管时点滞后、监管机制模糊的问题,需要构建新的平台算法问责机制来解决。

一方面,笔者主张将平台责任的理论基础深化至平台底层逻辑——算法层面,从而使得平台责任独立化,解决“平台责任中行为人与责任人相分离”的理论误区;另一方面,笔者主张考察算法设计部署的主观过错,解决“平台技术中立”的理论误区。这需要建立平台算法备案制度来事前固定问责点,并设置独立的算法解释制度来认定平台主观过错,解决监管时点滞后与监管机理不清的问题。平台从提供“连接”服务走向智能化与自动化后,网络平台应当基于“算法责任”的主观过错,以“客观损害结果”和之前已经事先存在的“算法责任”相结合,建立权

责一致、责罚相当的监管框架与法律责任体系。

如何打造“负责任”的人工智能?平台的算法问责应该是制度建设迈出的第一步。人工智能算法营造了运行自动化、损害风险化的假象,深度学习算法、人工智能主体性被反复言说,实际上为平台(或人工智能、机器人生产企业)规避责任提供了理论基础。<sup>⑨</sup>将人工智能法律制度落地细化为平台算法监管,实际上是将人工智能应用与监管场景化,一方面可以避免对人工智能不求甚解的恐惧,另一方面也可以消除理论界一些夸大其词的判断。技术带来的风险必须由人来负责,如果不能解决“谁负责”的问题,那么所有的治理努力最终都将毫无意义。当然,政策制定者必须认识到,平台算法问责制的目标不是要实现完美无误的算法,而是要将风险降至最低,正如汽车安全标准不要求汽车百分之百安全,而是尽可能合理地安全一样。创新技术与固有思维之间的冲突与激荡,必将伴随人工智能技术发展的过程。如何建立具备足够的包容之心和适应能力的动态治理机制,是人工智能时代的学者和政策制定者必须思考的问题。

### 注释:

①有学者认为,平台承担的主动监控义务是基于“功能主义”的考量而非规范主义的进路。参见赵鹏:《超越平台责任:网络食品交易规制模式之反思》,载《华东政法大学学报》2017年第1期。

②参见魏露露:《互联网创新视角下社交平台内容规制责任》,载《东方法学》2020年第1期。

③人工智能是继移动互联网之后的互联网又一发展阶段,其以物联网技术为基础,以平台型智能硬件为载体,以算法为核心,结合云计算与大数据应用,产生的三维符合、虚实互动的网络化平台技术。参见中国信通院:《互联网趋势发展报告(2017-2018)》第2页前言部分,载 [http://www.cac.gov.cn/wxb\\_pdf/baipishu/fazhanqushi020171213443448958139.pdf](http://www.cac.gov.cn/wxb_pdf/baipishu/fazhanqushi020171213443448958139.pdf), 2020年3月20日访问。

④网络平台利用算法执行“通知—删除”机制,算法监测内容并确定其是否与数据库中的内容相匹配。参见万勇:《人工智能时代的版权法通知——移除制度》,载《中外法学》2019年第5期。

⑤美国亚马逊网站的定价算法曾为一本二手书开出天价,后证实是定价算法的设置出现问题。See John D. Sutter: Amazon seller lists book at \$23,698,655.93—plus shipping, <http://edition.cnn.com/2011/TECH/web/04/25/amazon.price.algorithm/index.html>, 2020-02-10 accessed.

⑥杨乐:《网络平台法律责任探究》,电子工业出版社2019年版,第146页。

⑦“行为人与责任人相分离”所指的是平台侵权的内容都是由用户上传的,如商标侵权的商品、诽谤他人的言论,而平台却要为此承担不利法律后果。参见尹培培:《网络安全行政处罚的归责原则》,载《华东政法大学学报》2018年第6期;梅夏英、刘明:《网络侵权归责的现实制约及价值考量——以〈侵权责任法〉第36条为切入点》,载《法律科学》2013年第2期。

⑧2016年以来网信办相继出台了《互联网信息服务管理规定》《移动互联网应用程序信息服务管理规定》《互联网直播服务管理规定》;2017年又出台了《互联网新闻传播管理条例》《互联网论坛社区服务管理规定》《互联网跟帖评论服务管理规定》《互联网群组信息服务管理规定》《互联网用户公众账号信息服务管理规定》《互联网新闻信息服务许可管理实施细则》等政策性文件。2020年出台了《网络信息内容生态治理规定》,2021年出台《互联网用户公众账号信息服务管理规定》等。

⑨《国务院反垄断委员会关于平台经济领域的反垄断指南》,载国家市场监督管理总局网站, [http://gkml.samr.gov.cn/nsjg/flfdj/202102/t20210207\\_325967.html](http://gkml.samr.gov.cn/nsjg/flfdj/202102/t20210207_325967.html), 2021年2月8日访问。

⑩ Directive on Copyright in the Digital Singles Market (2019), Article 13.

⑪ S.1914—Ending Support for Internet Censorship Act 116th Congress (2019–2020), <https://www.congress.gov/bill/116th-congress/senate-bill/1914>, 2020-04-24 accessed.

⑫ Rachel Gutman: The 13 Strangest Moments From the Zuckerberg Hearing, <https://www.theatlantic.com/technology/archive/2018/04/the-strangest-moments-from-the-zuckerberg-testimony/557672/>, 2020-03-22 accessed.

⑬《互联网信息服务管理办法》(2000年)第16条规定:“互联网信息服务提供者发现其网站传输的信息明显属于本办法第15条所列内容之一的,应当立即停止传输,保存有关记录,并向国家有关机关报告。”

⑭侵权责任法(2010年)第36条规定:“网络用户、网络服务提供者利用网络侵害他人民事权益的,应当承担侵权责任。网络用户利用网络服务实施侵权行为的,被侵权人有权通知网络服务提供者采取删除、屏蔽、断开链接等必要措施。

网络服务提供者接到通知后未及时采取必要措施的,对损害的扩大部分与该网络用户承担连带责任。网络服务提供者知道网络用户利用其网络服务侵害他人民事权益,未采取必要措施的,与该网络用户承担连带责任。”

⑮网络安全法(2016年)第47条规定:“网络运营者应当加强对其用户发布的信息的管理,发现法律、行政法规禁止发布或者传输的信息的,应当立即停止传输该信息,采取删除等处置措施,防止信息扩散,保存有关记录,并向有关主管部门报告。”第68条规定:“网络运营者违反本法第47条规定,对法律、行政法规禁止发布或者传输的信息未停止传输、采取删除等处置措施、保存有关记录的,由有关主管部门责令改正,给予警告,没收违法所得;拒不改正或者情节严重的,处十万元以上五十万元以下罚款,并可以责令暂停相关业务、停业整顿、关闭网站、吊销相关业务许可证或者吊销营业执照,对直接负责的主管人员和其他直接责任人员处一万元以上十万元以下罚款。”

⑯“行政处罚中将处罚人的主观过错纳入考虑将确保行政处罚案件更加公正。”参见王静:《行政处罚法修订增加主观过错考量的重大意义》,载“城管执法之家”微信公众号, 2021年2月4日。

⑰白龙:《用法治方式读懂“快播案”》,载《人民日报》2016年1月11日,第5版。

⑱崔文佳:《价值观引领算法才有更多优质“头条”》,载《北京日报》2018年4月13日,第3版。

⑲参见同前注⑥,杨乐书,第148页。

⑳《关于对阿里巴巴集团进行行政指导工作情况的白皮书》。转引自赵鹏:《私人审查的界限——论网络交易平台对用户内容的行政责任》,载《清华法学》2016年第6期。

㉑参见《互联网用户公众账号信息服务管理规定》(2017)第7条规定:“互联网直播服务提供者应当落实主体责任,配备与服务规模相适应的专业人员,健全信息审核、信息安全管理、值班巡查、应急处置、技术保障等制度。提供互联网新闻信息直播服务的,应当设立总编辑。互联网直播服务提供者应当建立直播内容审核平台,根据互联网直播的内容类别、用户规模等实施分级分类管理,对图文、视频、音频等直播内容加注或播报平台标识信息,对互联网新闻信息直播及其互动内容实施先审后发管理”,载 [http://www.cac.gov.cn/2017-09/07/c\\_1121624269.htm](http://www.cac.gov.cn/2017-09/07/c_1121624269.htm), 2021年2月4日访问;《互联网用户公众账号管理规定》(2017)规定:“信息服务提供者应加强对本平台公众账号的监测管理,发现有发布、传播违法信息的,应当立即采取删除等处置措施,防止传播扩散,保存有关记录,并向有关主管部门报告。”参见《互联网跟帖评论服务管理规定》:

“跟帖评论服务提供者对发布违反法律法规和国家有关规定的信息内容的,应当及时采取警示、拒绝发布、删除信息、限制功能、暂停更新直至关闭账号等措施,并保存相关记录”,载 [http://www.cac.gov.cn/2017-08/25/c\\_1121541842.htm](http://www.cac.gov.cn/2017-08/25/c_1121541842.htm), 2020年2月1日访问。

②参见丁道勤:《电子商务法平台责任“管道化”问题及其反思》,载《北京航空航天大学学报》(社会科学版)2018年第6期。

③参见同前注②,赵鹏文。

④参见网络安全法第47条、第50条和第68条,《互联网新闻信息服务管理规定》第16条等。责令整改是否是行政处罚尚在学界存在争议。

⑤江苏省网安部门依据网络安全法等法律法规共办理行政案件580起,对突出问题乱挂挂牌整治2批次。其中,责令限期整改430家次,警告538家(人)次,罚款79万元,没收违法所得0.5万元,行政拘留29人;关停或下架违法违规移动互联网应用17款。参见《2020年江苏网络安全行政执法典型案例发布》,载 <https://www.secrss.com/articles/18130>, 2021年4月5日访问。

⑥广东省网信办于2017年8月11日对腾讯公司微信公众平台平台存在用户传播暴力恐怖、虚假信息、淫秽色情等危害国家安全、公共安全、社会秩序的信息问题依法展开立案调查。几乎与此同时,北京市网信办也依据网络安全法就新浪微博对其用户发布传播“淫秽色情信息、宣扬民族仇恨信息及评论信息”未尽到管理义务以及百度贴吧对其用户发布传播“淫秽色情信息、暴力恐怖信息帖文及相关评论信息”未尽到管理义务的违法行为作出从重罚款这一处罚决定。而在另一起行政处罚中,北京市网信办、北京市规划国土委就违法违规发布“大棚房”租售信息一事,联合依法约谈58同城、赶集网、百度等网站。根据网络安全法第47条及《互联网新闻信息服务管理规定》仅仅责令网站落实整改。参见尹培培:《网络安全行政处罚的归责原则》,载《东方法学》2018年第6期。

⑦参见章剑生:《现代行政法基础理论》,法律出版社2015年版,第363页。

⑧Matthias, A. The responsibility gap: Ascribing responsibility for the actions of learning automata, *Ethics and information technology*, Vol.6:3, p.175-183(2004).

⑨J.M. Fischer and M.S.J. Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, Cambridge, 1998.

⑩例如,前文所提到的淘宝对用户违约的举证和合理解

释的规定。

⑪例如,共享经济平台上对网约车司机的身份认定、追踪、报酬发放方式等。参见胡凌:《平台视角下的人工智能法律责任》,载《交大法学》2019年第3期。

⑫Frederic Lardinois, *Germany's Former Foreign First Lady Sues Google for Defamation Over Autocomplete Suggestions*, *TECH CRUNCH*(Sept.7, 2012), <http://techcrunch.com/2012/09/07/germanys-former-first-lady-sues-google-for-defamation-over-autocomplete-suggestions/>.

⑬《App违法违规收集使用个人信息行为认定方法》(2019)。

⑭Search King, Inc. v. Google Technology, Inc., Case No. CIV-02-1457-M(W.D. Okla. May.27, 2003)

⑮See Blackman, Josh. What happens if data is speech. 16 *University of Pennsylvania Journal of Constitutional Law Heightened Scrutiny* 25(2014).

⑯滴滴出行科技有限公司某主管曾在公开场合说:“(顺风车)是一个非常具有未来感、非常sexy(性感的)场景,我们从一开始就想得非常清楚,一定要往这个方向打。”参见张鑫:《深扒滴滴免职两位高管一位定义顺风车“sexy的场景”》,载《法制晚报》2018年8月26日。

⑰参见赵鑫:《深度学习在推荐算法上的应用进展》,载《中国人工智能学会通讯》2016年第7期。

⑱需要注意的是,弱人工智能的算法也有不同的自治程度,如无人机、自动驾驶汽车与定价算法、自动武器等具有不同的自治程度。

⑲Gutman M, Rathgeber B, Syed T. Action and autonomy: A hidden dilemma in artificial autonomous systems, 3 *Robo-and Information ethics. Some Fundamentals* 231(2012).

⑳剑桥分析公司及其母公司战略通讯实验室曾参与过世界各地举行的200多场选举活动,其成果包括利用脸书数据影响2016年美国大选、2013年和2017年两次帮助肯尼亚总统赢得选举,以及影响乌克兰大选等。参见佚名:《“脸书”个人用户数据被滥用?“剑桥分析”在全球有何影响》,载 <https://www.bbc.com/zhongwen/simp/world-43482767>, 2019-04-09 accessed, 2021年2月1日访问。

㉑赵鹏:《网络平台行政法律责任边界何在》,载《财经》2016年第12期。

㉒See Jeff Kosseff, *The Twenty-Six Words that Created*, Cornell University Press, 2019, p.12.

㉓淘宝搜索规则在2010年7月8日的调整曾引发了巨大争议,导致淘宝网众多小商家的聚众抗议事件。尽管淘宝方

面宣称此次调整是为了削减人气流量的比重而给更多商家展现机会,但被很多小商家指责此举目的为“逼迫”卖家投放竞价排名广告。参见《新淘宝搜索规则》,载 <https://baike.baidu.com/item/新淘宝搜索规则/869491?fr=aladdin>, 2018年10月31日访问。

④ FRANK A UGUGLIARO: GOOGLE AUTOCOMPLETE STILL MAKES VILE SUGGESTIONS, [https://www.wired.com/story/google-autocomplete-vile-suggestions/amp?\\_\\_twitter\\_impresion=true](https://www.wired.com/story/google-autocomplete-vile-suggestions/amp?__twitter_impresion=true), 2018-11-22 accessed.

⑤ KENNETHA. CUTSHAW ET. AL., CORPORATE COUNSEL'S GUIDE TO DOING BUSINESS IN CHINA §26:33.

⑥ See Daniel Zwerdling: Internet Trolls Turn A Computer Into A Nazi, <https://www.wnyc.org/story/internet-trolls-turn-a-computer-into-a-nazi/>, 2018-06-04 accessed.

⑦ David C. Vladeck, Machines Without Principals: Liability Rules and Artificial Intelligence, 89 Washington Law Review 117, 150(2014).

⑧ Floridi, Luciano. Distributed morality in an information society. 19 Science and engineering ethics 727(2013); 李磊:《论共享单车平台公司对低龄骑行人的安全注意义务》,载《东方法学》2020年第1期。

⑨ See Jack M. Balkin, 2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data, 78 Ohio State Law Journal 1229(2017).

⑩ Ibid.

⑪ 参见刘权:《网络平台的公共性及其实现——以电商平台的法律规制为视角》,载《法学研究》2020年第2期。

⑫ 《中共中央关于坚持和完善中国特色社会主义制度、推进国家治理体系和治理能力现代化若干重大问题的决定》,2019年10月31日中国共产党第十九届中央委员会第四次全体会议通过。

⑬ Matthew Hindman, The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy, Princeton University Press, 2018, p.43.

⑭ Ibid., at 54-55.

⑮ 参见王頔:《国家广播电视总局责令“今日头条”网站永久关停“内涵段子”等低俗视听产品》,载 [http://www.xinhuanet.com/2018-04/10/c\\_1122661804.htm](http://www.xinhuanet.com/2018-04/10/c_1122661804.htm), 2018年11月20日访问。

⑯ See Zuboff S, Big Other: Surveillance Capitalism and the Prospects of an Informal Civilization, Palgrave Macmillan UK, Journal of Information Technology, Vol.30, pp.75-82(2015).

⑰ 参见同前注⑬,胡凌文。

⑱ 北京知识产权法院民事判决书,(2017)京73民终1194号。

⑲ See H.R.2231- Algorithmic Accountability Act of 2019, <https://www.congress.gov/bill/116th-congress/house-bill/2231>, 2019-05-22 accessed.

⑳ Rob Taylor, Facebook and Google Algorithms Are Secret-but Australia Plans to Change That, <https://www.wsj.com/articles/facebook-and-google-algorithms-are-secret-but-australia-plans-to-change-that-11564134106>, 2020-4-20 accessed.

㉑ 《网络信息内容生态治理规定》(2020)第12条规定:“网络信息内容服务平台采用个性化算法推荐技术推送信息的,应当设置符合本规定第10条、第11条规定要求的推荐模型,建立健全人工干预和用户自主选择机制。”

㉒ 主要系指行为人的主观犯意,若其决定从事违反行政法上义务之行为,即得对其非难而要求其承担责任,此种责任要件有两种,即故意或过失。参见翁岳生:《行政法》(上册),中国法制出版社2009年版,第847页。

㉓ 在国务院2002年颁布的《药品管理法实施条例》第81条规定:“药品经营企业、医疗机构未违反《药品管理法》和本条例的有关规定,并有充分证据证明其不知道所销售或者使用的药品是假药、劣药的,应当没收其销售或者使用的假药、劣药和违法所得;但是,可以免除其他行政处罚。”即能够证明自己确实没有过错的药品经营企业可以免除行政处罚。药品监管即药品风险防控,可以说是整个产品质量监管领域最为严格的,但在这一最为严格的领域中,我们看到了责任主义的贯彻。在销售者如法地履行了注意义务的情况下,仍然有可能出现假药、劣药,这对销售者来说就可以称为一种“法所容许的风险”。参见杨利敏:《论我国行政处罚中的责任原则——兼论应受行政处罚的过失违法行为》,载《华东政法大学学报》2020年第2期。

㉔ 最高人民法院《关于审理利用信息网络侵害人身权益民事纠纷案件适用法律若干问题的决定》第9条对网络服务提供者是否“知道”侵权行为的存在作出了规定:一是考虑网络服务提供者是否以人工或者自动方式对侵权网络信息以推荐、排名、选择、编辑、整理、修改等方式作出处理;二是考虑网络服务提供者管理信息的能力、提供服务的性质、方式引发侵权的可能性;三是网络信息侵害人身权益的类型及明显程度;四是侵权网络信息的社会影响程度或者一定时间内的浏览量;五是网络服务提供者采取预防侵权措施的技术可能性及其是否采取了相应的合理措施;六是网络服务提供者是否针对同一网络用户的重复侵权行为或者同一侵权信息采取了相应的合理措施;等等。

㉕ See Abbott, Ryan, The reasonable computer: disrupting

the paradigm of tort liability, 86 *Geo. Wash. L. Rev.*1(2018).

⑥ See Mittelstadt, Brent Daniel, et al. *The Ethics of Algorithms: Mapping the Debate*, 3 *Big Data & Society* 2(2016).

⑦ See Kroll, Joshua A., et al. *Accountable algorithms*, 165 *University of Pennsylvania Law Review* 633(2016); 参见江溯:《自动化决策、刑事司法与算法规制——由卢米斯案引发的思考》,载《东方法学》2020年第3期。

⑧ 参见傅莹:《人工智能的治理和国际机制的关键要素》,载《人民论坛》2020年2月(上)。

⑨ 欧盟一般数据保护条例下数据处理评估制度的规定集中在第35、36条,和前言条款第89到96项。另外,根据欧盟一般数据保护条例的原则,第29条工作组于2017年10月也更新了数据处理评估制度指南。

⑩ AI Now institution, *AI Now 2019 Report*. [https://ainowinstitute.org/AI\\_Now\\_2019\\_Report.pdf](https://ainowinstitute.org/AI_Now_2019_Report.pdf), 2020-01-20 accessed.

⑪ 根据算法处理的数据所涉及的利益大小,可对算法进行风险评级,对不同类型的算法采取不同强度的监管。以德国数据伦理委员会提出制定算法评估方案为例,其主张对算法分为五类:(1)对于具有较低潜在危害的系统例如饮料制作机,不应监管;(2)对于具有潜在危害的系统,例如电子商务平台的动态定价机制应该放宽管制,可以采用事后控制机制,加强披露义务等来降低其潜在危险;(3)对于具有一般或明显危害的系统,应考虑以发放许可证的方式,促使审批、监管常态化;(4)对于具有相当潜在风险的系统,例如在信用评估方面具有准垄断地位的公司,应公布其算法细节,包括计算所参考的因素及其权重,算法所使用的数据,以及对算法模型的内在逻辑进行解释;(5)对于自动化武器等具有潜在不合理危险的系统,则应该“完全或者部分”禁止。See *Opinion of the Data Ethics Commission* 159(2018), [https://www.bmjv.de/SharedDocs/Downloads/DE/Ministerium/ForschungUndWissenschaft/DEK\\_Empfehlungen\\_englisch.html; jsessionid=C4CE6C88B9310034A97B42CD67553FC7.2\\_cid289?nn=11678512](https://www.bmjv.de/SharedDocs/Downloads/DE/Ministerium/ForschungUndWissenschaft/DEK_Empfehlungen_englisch.html?jsessionid=C4CE6C88B9310034A97B42CD67553FC7.2_cid289?nn=11678512), 2020-04-20 accessed.

⑫ See Timothy B. Lee. *Report, Software bug led to death in Uber's self-driving crash*, May 2018. <https://arstechnica.com/tech-policy/2018/05/report-software-bug-led-to-death-in-ubers-self-driving-crash/>, 2020-03-20 accessed.

⑬ James McGrath and Ankur Gupta, 2018. *Writing a Moral Code: Algorithms for Ethical Reasoning by Humans and Machines*. 9 *Religions* 8, 240-259(2018).

⑭ 在联合国人权委员会的文件中,提出了对人权影响的评估框架。See united nations, human rights council, office of the high comm'r, *guiding principles on business and human rights*

23-26(2011), <https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHREN.pdf>[<https://perma.cc/R3PC-BW5H>]. 同样,纽约大学AINow研究所提出的算法影响评估框架中也包括对人权的影响,See dillon reisman et al., *al now inst., algorithmic impact assessments: a practical framework for public agency accountability* 5(2018), <https://ainowinstitute.org/aiareport2018.pdf>[<https://perma.cc/JD9Z-5MZC>]. 欧盟专家委员会提出的《关于先进数字技术的人权影响框架》也提出了新技术的使用前应考虑对人权的评估, Yeung K. *A study of the implications of advanced digital technologies(including AI systems) for the concept of responsibility within a human rights framework*. Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence MSI-AUT Council of Europe, 5.(2018).

⑮ See *Privacy Impact Assessments*, fed. trade commission, <https://www.ftc.gov/site-information/privacy-policy/privacy-impact-assessments>, 2020-04-23 accessed.

⑯ *Data Protection Impact Assessments*, ICO, <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments>[<https://perma.cc/Q2NL-9AYZ>]; see also REISMAN ET AL., *supra* note 355.

⑰ Barocas, Solon, and Andrew D. Selbst. *Big data's disparate impact*. 671. *Calif. L. Rev.*169(2016).

⑱ Marc L. Roark, *Human Impact Statements*, 54 *washburn l. j.*649(2015).

⑲ As part of fulfilling this responsibility, private actors should take on-going, proactive and reactive steps to ensure that they do not cause or contribute to human rights abuses and that their innovation processes are human-rights friendly... The scale and complexity of the means through which they meet their responsibilities may vary, however, taking into account their means and the severity of potential impact on human rights by their services and systems. See *Draft Recommendation of the Committee of Ministers to member States on human rights impacts of algorithmic systems*, Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence(MSI-AUT), Council of Europe, 10(2018). [https://www.intgovforum.org/multilingual/sites/default/files/webform/msi-aut201806\\_eng\\_draft\\_recommendation\\_12\\_november\\_2018.docx\\_\\_0.pdf](https://www.intgovforum.org/multilingual/sites/default/files/webform/msi-aut201806_eng_draft_recommendation_12_november_2018.docx__0.pdf), 2020-04-10 accessed.

⑳ See H.R.2231- *Algorithmic Accountability Act of 2019*,

[https://www.congress.gov/bill/116th-congress/house-bill/2231](https://www.congress.gov/bill/116th/congress/house-bill/2231), 2019-05-22 accessed.

⑪ Rob Taylor, Facebook and Google Algorithms Are Secret-but Australia Plans to Change That, <https://www.wsj.com/articles/facebook-and-google-algorithms-are-secret-but-australia-plans-to-change-that-11564134106>, 2020-4-20 accessed.

⑫ 参见傅莹:《人工智能的治理和国际机制的关键要素》,载《人民论坛》2020年第4期。

⑬ 参见张晓阳:《基于 cookie 的精准广告投放技术及其法律边界刍议以朱烨诉百度公司隐私权纠纷为视角》,载《电子知识产权》2015年第5期。

⑭ New, Joshua, Daniel Castro, How Policymakers Can Foster Algorithmic Accountability. Information Technology and Innovation Foundation, Washington DC, <https://itif.org/publications/2018/05/21/how-policymakers-can-foster-algorithmic-account>

ability(2018).

⑮ 欧盟一般数据保护条例的条文中包含了很多这种逻辑漏洞,数据控制者很容易通过各种方式规避对于算法的严格审查。参见张凌寒:《自动化决策与人的主体性》,未刊文。国外也有学者撰文指出欧盟一般数据保护条例条文可规避性强的问题,See Tal Z. Zarsky, Incompatible: The GDPR in the Age of Big Data, 47 SETON HALL L. REV.995(2017).

⑯ 参见沈伟伟:《算法透明的迷思——算法规制理论的批判》,载《环球法律评论》2019年第6期。

⑰ See Frank Pasquale, *The Black Box Society*, Harvard University Press 2015, pp.6-8.

⑱ See Burrell, Jenna, How the machine 'thinks': Understanding opacity in machine learning algorithms, 3*Big Data & Society* 1 (2016).

⑲ 参见同前注⑮,胡凌文。

## Algorithm Accountability in Platform Regulation

Zhang Linghan

**Abstract:** The platform increasingly operates automatically as algorithms playing more important role. However, it leads to blurriness of platform's subjective fault. As a result, legal responsibility setting of platform does not conformed to the principle of proportionality. The governance of platform should be consistent with both the underlying technical logic of the platform and the legal principle of subjective and objective consistency. Therefore, the "technical veil" of the operation of the network platform should be penetrated, and the tentacles of platform supervision and the perspective of recourse should be directed to the algorithmic responsibility behind. The subjective fault of the platform algorithm design and deployment should be taken as the basis of accountability. The subjective fault of the platform needs to be identified as a clear and traceable accountability point, which can be set by the algorithm evaluation system and fixed in advance by the algorithm record system. Afterwards, the platform can be required to make algorithmic explanation according to the record. In the setting of the legal responsibility of the platform algorithm, the fault of its own design is imputed as the substantive responsibility, and the platform providing false record and explanation bears the responsibility of untrue explanation.

**Key words:** algorithm; platform regulation; subjective fault; algorithm filing; algorithm explanation; algorithm accountability