风险防范下算法的监管路径研究

张凌寒*

目次

摘要 人工智能时代，由源源不断的数据驱动的算法已经成为新的权力代理人。然而，对算法的监管仍依照传统路径，在结果监管的思路指导下，进行事后的内容审查，由网络平台承担法律责任。机器深度学习和自动决策技术的发展超越了结果监管思路的制度假设，给传统监管路径带来严重冲击。应及时树立以风险防范为目的的监管思路，实行内容与算法并重的双轨审查机制，以及设立平台责任与技术责任双轨并行的责任体系，并对算法的生产性资源数据的收集和使用进行合理限制。

关键词 算法监管 平台责任 风险防范 技术责任

一、问题的提出：假新闻事件凸显算法监管紧迫性

2016 年 Facebook 遭到社会广泛指责，认为其造成了假新闻泛滥进而对本届美国总统选举

* 东北师范大学副教授、法学博士、北京航空航天大学法学院博士后研究人员。本文系中国法学会部级项目“新闻传播中社交媒体的法律规制——以公共利益为视角”[CLS(2016)C14]、国家社会科学基金重大项目“信息法基础”(项目编号：16ZDA075)的阶段性成果。

产生了巨大影响。[1] 然而，2017年数项研究成果显示，Facebook上的假新闻数量仅有1%而已。[2] 原因在于，算法主宰了社交媒体对用户的新闻投放，其推送标准是用户的点击率和转发率而非新闻的真实性。因此颇具噱头的假新闻被用户高频点击或转发，并进一步被算法推送而广泛传播。

美国大选中的假新闻事件引起了对算法如何进行监管的关注和忧虑。算法已经主导和控制了社交媒体上的信息传播。如《纽约时报》数字版使用"非常前沿和复杂的算法"挑选文章推送，将用户点击率提高了38倍。[3] 学者们担心，社交媒体的算法过度利用了用户偏好数据推送信息，制造了信息"过滤泡沫"，造成用户接受的观点越来越极端。[4] 不仅如此，算法通过社交媒体平台对公众日常生活产生巨大影响，人们通过搜索引擎获取知识和商业信息，通过Facebook、微博社交和获取新闻，通过评价类的社交媒体知晓餐馆评价，通过约会类的社交媒体结识伴侣。算法甚至开始逐渐被用于公共部门管理影响公民权利。如何对算法进行有效的监管，是人工智能时代法律制度急需应对的挑战。

算法的定义从Tarleton Gillespie的假设开始，即"算法不需要是软件：在最广泛的意义上，它们是基于指定计算将输入数据转换为期望输出的编码过程。这个过程既指出了一个问题，也指出了解决这个问题的步骤"。[5] 因此，算法是"为了解决一个特定问题或者达成一个明确的结果而采取的一系列步骤"[6]。具体到社交媒体算法，算法的目标是获得和保持用户的数量，并且尽量提高用户的参与度。所以，它收集社交媒体用户数据，预测用户的偏好进行推荐，而用户对于推荐内容的刺激和反馈可以为下一步的推荐提供数据。

本文无意进行人工智能时代法律制度的宏大叙事，而是以社交媒体的算法作为切入点，分析算法技术发展带来的监管挑战，并梳理现有法律制度对算法发展应对的缺陷和不足，提出应对人工智能的技术挑战，应以风险防范为目的，对算法同时进行事先与事后监管，主体与技术监管，以期用法律化解风险。

---

〔1〕 Facebook假新闻影响选举的指责，从2016年9月就开始被多次报道。2016年11月11日，美国《财富》杂志刊文指责Facebook称，尽管Facebook没有散播怂恿选民投票支持特朗普的新闻，但是Facebook上流传着关于美国政治虚假的小道消息使得选举的天平向特朗普倾斜。《赫芬顿邮报》也放出调查报道称，Facebook热门话题板块上针对希拉里的谣言数量尤为突出，并借此质疑Facebook对这类谣言的整治不力。《纽约杂志》也发文称，数千万的Facebook用户都有预谋或者情绪性地分享了针对希拉里的虚假新闻。此外，纽约大学新闻系教授Jay Rosen、社会学家Zeynep Tufekci、尼曼新闻实验室负责人Joshua Benton也在近日分别撰文称，希拉里败选与Facebook上传播的虚假消息关系密切。参见《美国主流媒体与扎克伯格激辩：Facebook的假新闻到底帮没帮特朗普胜选？》，载搜狐网（http://it.sohu.com/20161114/n473138951.shtml，最后访问时间2018-02-25）。

〔2〕 史安斌、王沛楠：《作为社会抗争的假新闻——美国大选假新闻现象的阐释路径与生成机制》，载《新闻记者》2017年第6期，第4～12页。

〔3〕 Shan Wang: The New York Times Built a Slack Bot to Help Decide Which Stories to Post to Social Media(AUG. 13, 2015)(http://www.niemanlab.org/2015/08/the-new-york-times-built-a-slack-bot-to-help-decide-which-stories-to-post-to-social-media/，最后访问时间2017-03-20)。

〔4〕 Eytan Bakshy, Solomon Messing, and Lada A. Adamic, "Exposure to Ideologically Diverse News and Opinion On Facebook", 348 (6239) Science 1130-1132 (2015). David Lazer, "The rise of the social algorithm" 348 (6239) Science 1090-1091 (2015).

〔5〕 Tarleton Gillespie,"The Relevance of Algorithms", in T. Gillespie, P. J. Boczkowski, and K. A. Foot (eds.), *Media Technologies: Essays on Communication, Materiality, And Society* (Cambridge Mass.: MIT Press, 2014), pp.167-194.

〔6〕 Nicholas Diakopoulos, "Algorithmic Accountability", 3(3) Digital Journalism 398-415 (2015).

二、现有的算法监管路径：法律框架与制度假设

当前，我国并没有任何直接对算法监管的法律规定。应对算法造成的不利法律后果，我国采取结果监管的法律规制路径。具体言之，通过事后的内容审查发现算法造成的不利法律后果，进而将这种不利后果的法律责任分配给开发者或使用者——网络平台。本部分对现有的有关算法的监管路径进行梳理与分析。

（一）算法监管的法律框架：结果监管下的内容审查与平台责任

大数据时代到来后，网络信息传播呈指数级增长，算法被广泛应用。相对于网络平台，国家权力在网络空间影响力日益减弱。为此各国普遍加强了对于网络信息内容的监管，并要求网络平台对网络信息非法内容的传播承担平台责任。

1. 结果监管路径：内容审查范围的扩大

各国对算法造成的不利后果进行内容审查的范围近年来普遍扩大。以社交媒体上的假新闻泛滥为例，各国普遍加强了社交媒体的内容审查，不仅对恐怖、色情、仇恨等侵害公共利益的非法内容严加监管，对诽谤的侵害私权利的内容也加强了监管。〔7〕

我国一直重视互联网传播内容的政府监管。2000 年实施的《互联网信息服务管理办法》作为基础性法规，规定了互联网信息服务提供者不得制作、复制、发布、传播的八项内容，涵盖“反对宪法所确定的基本原则的”“危害国家安全”等非法内容，并设定了一个兜底条款 “含有法律、行政法规禁止的其他内容”。〔8〕 近一两年，新规密集出台，〔9〕将内容审查的范围进一步扩大，如侵害私权利的“侮辱或者诽谤他人，侵害他人合法权益的”，以及模糊的道德性规定“危害社会公德或者民族优秀文化传统的”等。〔10〕 并且，对内容审查也提出了倡导性要求，如“弘扬社会主义核心价值观”等。〔11〕对算法的监管通过事后对内容进行审查的方式进行，审查范围的扩大也意味着对算法结果监管的加强。

2. 法律责任承担：平台监控责任的加强

在互联网发展早期，如算法造成的不利法律后果，如内容分发的算法造成了侵害著作权的法律后果，则由网络平台承担民事侵权责任。一般网络平台可以主张不提供内容或尽到合理注意义务而通过“避风港”规则免责，这一原则也被世界多国立法采纳。〔12〕

---

〔7〕 为加强网络言论管理，德国联邦议会周五通过了新的网络管理法，新法将于 2017 年 10 月正式实施。在此法律下，Facebook 等社交媒体未能在 24 小时内删除“显而易见的非法内容”——包括仇恨性言论、诽谤及煽动暴力等，将面临德国监管部门的罚款（http://finance.sina.com.cn/roll/2017-07-01/doc-ifyhryex5667798.shtml，最后访问时间2017-07-15）。印尼要求社交媒体关闭宣扬极端言论的账户。印尼通信部长警告社交媒体不关闭激进内容账户将被阻止共享（http://finance.sina.com.cn/roll/2017-07-01/doc-ifyhryex5667798.shtml，最后访问时间 2017-07-12）。

〔8〕 《互联网信息服务管理办法》第 15 条。

〔9〕 2016 年和 2017 年，国家互联网信息办公室进入立“法”密集期，“微信十条”“账号十条”和“约谈十条”先后出台。

〔10〕 《互联网文化管理暂行规定》。

〔11〕 《互联网群组信息服务管理规定》。

〔12〕 2000 年，欧盟发布《电子商务指令》移植了避风港原则和红旗标准，其第 14 条规定，除用户受控于平台或依平台指令实施的侵权行为外，平台不知道用户的行为违法，或在知悉违法行为或事实后删除该违法信息或采取措施阻止该违法信息的传输，不承担侵权责任。《指令》第 15 条明确了平台原则上不负有积极查找平台内违法行为和信息的义务，除非基于保护国家安全、国防和公共安全以及为防止、追查、侦破和惩治刑事犯罪的需要而要求采取有针对性、临时性的监控措施。《新加坡电子交易法》进一步细化了红旗标准。《新加坡电子交易法》（转下页）

早在2000年的《互联网信息服务管理办法》就强调平台对用户发布的不法内容有避免传播的义务，处理措施包括停止传输，保存记录与向国家有关机关报告。〔13〕它为网络平台设立了合理注意义务，在其“发现”不法内容时要及时采取措施，否则网络平台要承担不法信息传播的责任。近两年，国家网信办密集立法并明确提出，对于虚假信息采取“强双责”的方针——强化网络平台的主体责任与社会责任。〔14〕不仅要求网络平台在明知不法信息的情况下承担责任，更是将对信息的主动监控义务加诸网络平台。〔15〕在这种制度框架下，网络平台承担了违法内容造成不利法律后果的民事责任，并且开始逐渐承担普遍性的信息主动监控责任。

（二）现有算法监管路径的制度假设

网络技术具有较强的专业性，而网络信息传播的法律规制隐含着立法者对于技术的理解。对于结果进行监管的路径包括内容审查与平台责任，其隐含着如下制度前提：

1. 内容审查：开发者的“全能性”假设

对于网络不法信息传播的内容审查范围日趋扩大化，隐含着立法者假设网络平台等软件的开发者和使用者可以完全控制信息的生产和传播，并具有对网络信息审查的“全能”的能力和权限。因此，对不法信息审查的内容范围不仅局限于违法信息，也包括对私权利侵害的审查，甚至涵盖对道德标准的审查。例如2017年的阿里云案件，虽然阿里云发布声明，称其作为云服务提供者并没有权限和能力审查租用网络空间的用户数据，但法院仍判定其涉嫌共同侵权。〔16〕法院的理由是，

（接上页）第26条：“仅提供接入、存取服务的网络服务提供者对站内第三方制作、发布、传播或散布的侵权信息，除违反合同特别约定，或违反成文法规定的监管要求，或者是违反成文法或法院的删除、阻止、限制访问的要求外，不承担民事责任和刑事责任。”《巴西网络民事基本法》第19条：“除法律另有规定外，网络服务提供者仅在接到法院令后，未在规定时间内删除或屏蔽侵权信息时，才对站内用户侵害他人权益的行为承担侵权责任。”

〔13〕 参见《互联网信息服务管理办法》第16条：“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十五条所列内容之一的，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〕 见《强化网站主体责任正当时》，载中华人民共和国互联网信息办公室官方网站（http://www.cac.gov.cn/2016-12/22/c_1120166441.htm，最后访问时间 2017-07-15）。

〔15〕 参见《互联网用户公众账号信息服务管理规定》第7条：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（http://www.cac.gov.cn/2017-09/07/c_1121624269.htm）。参见《互联网用户公众账号信息服务管理规定》第11条：“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，应当立即采取消除等处置措施，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》第8条：“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。”（http://www.cac.gov.cn/2017-08/25/c_1121541842.htm）

〔16〕 参见2017年阿里云服务器一审败诉的案件。北京市石景山区人民法院对阿里云被诉侵权案做出一审判决。法院认定被告阿里云公司构成侵权，需赔偿乐动卓越公司经济损失和合理费用约26万元。2015年8月，乐动卓越公司发现某网站提供的《我叫MT畅爽版》涉嫌非法复制其开发的《我叫MT online》的数据包。乐动卓越公司通过whois域名查询系统、域名备案系统等，均没有查到涉案网站经营人的相关信息。但他们发现《我叫MT畅爽版》的游戏内容存储于阿里云公司的服务器，并通过该服务器向用户提供游戏服务。接着，乐动卓越公司两次致函阿里云，要求其删除涉嫌侵权内容，并提供服务器租用人的具体信息，阿里云并没有予以积极回应。乐动卓越公司便以阿里云涉嫌共同侵权为由，向北京市石景山法院提起诉讼，请求法院判令阿里云公司断开链接，停止为《我叫MT畅爽版》游戏继续提供服务器租赁服务，并将储存在其服务器上的《我叫MT畅爽版》游戏数据库信息提供给乐动卓越公司；赔偿经济损失共计100万元。随后阿里云就此事发布声明，称作为云服务器提供商，阿里云无权审查任何用户数据。

平台的技术开发者和使用者有能力知晓非法内容的存在，不能声称自己无法控制信息的传播。因此对造成的社会危害不能够逃脱责任。同样的立法精神也体现在了2016年轰动全国的快播案中。〔17〕

2. 平台责任：算法的“工具性”假设

网络平台承担民事责任，以及对内容承担普遍的主动监控义务，隐含着立法者对平台与算法关系的假设。即算法是网络平台的工具，网络平台作为算法的开发者和使用者，从对算法的运行和算法的决策，都具有类似对“工具”的控制能力。例如，对于网络平台上的算法进行的新闻分发和推荐，假设平台与传统媒体有相同的管理和控制能力，进而要求网络平台新闻服务提供者与传统媒体一样承担看门人的法律责任。〔18〕

这种算法工具性假设还意味着，网络平台对于算法的运行结果有充分的预测能力。因此监管者对于算法产生不利后果的技术原因并不感兴趣，无论在算法运行的“黑箱”中发生了什么，只要在法律对网络平台能力假设中此种后果可以被避免，平台即需要承担相应的责任。

通过对算法造成的结果进行监管，事后追究责任是传统的法律规制手段。由于算法一直被认为是不应公开的商业秘密，这种监管路径可以避免立法者和司法者介入算法运行的内部结构，陷入司法者并不了解的技术领域。追究事后责任是较为稳妥的规制方法，可以避免司法机构过于依赖专业知识，而仅仅处理纯粹的法律问题。然而，人工智能时代，算法的技术发展和角色转变都给传统监管路径带来了新的挑战。

三、人工智能时代算法的发展及其带来的监管挑战

人工智能时代，算法迅速发展出全新特性，超出了公众对于算法与其开发者和使用者关系的认知，为传统的监管路径带来了巨大挑战。

（一）算法机器深度学习技术：开发者“全能性”的否认

人工智能时代，算法的功能不再局限于处理数据并按照设计的特定目的决策和行动，而是发展出了利用数据进行机器学习，进行自我深化发展的功能。机器学习，多层神经网络等技术的发展，使得人类愈发无法了解和还原算法决策的内在运行机制，进而使得已经饱受诟病的“监管黑箱”的不透明性进一步深化。

社交媒体上算法尤其具有代表性。这种算法本质是推荐算法，即用户不需要提交任何查询的兴趣偏好，系统就可以自动化算法来进行信息推送。比如对于新闻推荐，算法根据文本内容、关键词、用户的点击、收藏和转发行为获得用户的反馈信息进行信息推荐。这种反馈信息包括显性的评分、评论，也包括隐形的停留时间、点击次数等。人工神经网络的研究为社交媒体的推荐算法提供了深度学习完善自身的条件，社交媒体平台的两个关键特征又为机器学习提供了最佳的机会。第一，社交媒体信息的量级和细节程度。社交媒体拥有用户个人的及其与好友互动的大量信息，

---

〔17〕 2016年9月13日，北京市海淀区人民法院对深圳快播科技公司及其主管人员王欣等四名被告人涉嫌传播淫秽物品牟利案进行了一审宣判。深圳市快播科技有限公司成立于2007年12月26日，公司相继开发了“快播”服务器软件和“快播”网页播放器，却被控用来大量传播淫秽色情视频。本案否定了软件开发者“技术中立”地位，强调了软件开发者和运行者对用户上传内容的监管义务。

〔18〕 参见《互联网跟帖评论服务管理规定》第8条：“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。”（http://www.cac.gov.cn/2017-08/25/c_1121541842.htm）

这些信息不仅细节丰富，并且与个人用户的历史具有对应性。第二，社交媒体为算法提供了可以即时反馈的机制，即算法可以通过用户的反馈，及时更改调整，探索有助于其优化目标的策略。[19]

机器深度学习使得算法可以自动将各种低层次特征计算为更加抽象的高层特征。例如将色彩选择，图片上传时间，打字速度等抽象为更高层的特征，基于此而发展更加智能的决策方式。这种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。[20]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开发者都无法理解。

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。大部分深度学习推荐算法将数据变换到一个隐含空间，在这个隐含空间可以计算推荐物品与用户的相似性，但是很难提供直接的决策理由。[21] 例如，Google 的 AlphaGo 项目设计的视频游戏算法，算法通过学习发展的策略不仅胜过了顶级的人类玩家，而且通过数据来优化自身，最后算法学会了利用游戏设计中的弱点，将对手逼入角落赢得胜利。[22] 然而开发者并无法具体解释这一过程是如何达到的。机器学习的发展使得算法的开发者和使用者对于算法完全了解和掌控的假设不再成立。算法的不透明性不再仅仅对技术领域之外的人，其内部机制和决策过程甚至对其开发者和使用者不再完全透明，更遑论“全能性”的掌控。

（二）算法自动决策的应用：“工具性”特质的超越

人工智能时代，算法的角色逐渐超越了网络平台的“工具性”角色，而成为信息配置的基本规则。算法的自动决策逐渐在人类社会中扮演日益重要的角色，开始独立拥有巨大的资源配置权力。最初，策展人算法仅仅是左右了社交媒体上的信息流动，决定了用户的阅读内容，帮助保险公司和招聘单位决定保险价格和筛选雇员。随着人工智能的发展，算法具有智能性，开始主动对用户实施影响展开操纵行为。如 Facebook 的“情感蔓延”项目通过控制对 50 万名用户的新闻投放，用积极或消极的语言来表述新闻，来使用户受其影响而自己发出积极或者消极的消息。[23] 美国电子隐私信息中心敦促联邦贸易委员会要求 Facebook 公布平台推送新闻的算法，并对公众公开算法细节。[24] 现在，算法下的自动化决策已经进入许多国家公共部门，在社会保障、医疗保健、公职人员监督和司法系统等领域发挥重要作用。例如，美国许多法院使用计算机程序来评估重复犯

---

〔19〕 Samuel Albanie, Hillary Shakespeare, and Tom Gunter, “Unknowable Manipulators: Social Network Curator Algorithms”, arXiv preprint arXiv: 1701.04895 (2017).

〔20〕 赵鑫：《深度学习在推荐算法上的应用进展》，中国人工智能学会通讯第 7 期（http://www.yunzhan365.com/basic/25489680.html，最后访问时间 2017-08-19）。

〔21〕 Adnan Masood：《深度神经网络的灰色区域：可解释性问题》（2015-08-18 08:15，http://www.csdn.net/article/2015-08-17/2825471，最后访问时间 2017-07-20）。

〔22〕 David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al.,“Mastering the Game of Go with Deep Neural Networks and Tree Search”, 529(7587)Nature 484-489 (2016).

〔23〕 G. S.McNeal, Controversy over Facebook Emotional Manipulation Study Grows and Timeline Becomes More Clear, Forbes(2014, June 30, http:// www.forbes.com/sites/gregorymcneal/2014/06/30/controversy-over-facebook-emotional-manipulation-study-grows-as-timeline-becomes-more-clear/，最后访问时间 2017-05-17）。

〔24〕 Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock,“Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks”, 111(24) Proceedings of the National Academy of Sciences 8788-8790 (2014).

罪的风险，这已被证明是“偏向黑人”。[25] 甚至在我国的大学食堂，学生家庭情况的识别和贫困生补助工作也得到了算法的技术支持。[26] 正如劳伦斯·莱斯格告诉我们，“代码是一种法则”——系统的架构，使之运营的代码和算法，将会对自由造成深远影响。

由于人工智能技术的专业性，人们容易认为算法更为客观并日渐依赖算法的自动决策。算法既会审查并撤下越南战争的经典新闻照片，[27]也可能拒绝求职者或者健康保险。用户暂时仍没有权利要求算法给出合理解释，或者说即使有权利算法的开发者可能也无法给出解释。尤其值得注意的是，社交媒体上信息流动的规则由算法控制，而算法由私人公司创设。这些规则主宰了网络空间的信息流动，却不像法律一样接受公众的质询和监督。算法的地位早已超越了平台的“工具”。我们生活在一个算法越来越对公民生活享有裁决权的世界，由源源不绝的数据驱动的算法俨然已成为新的权力代理人。

具有智能的算法是否应该成为法律主体留待哲学家们讨论，但人工智能时代带来的法律挑战却是迫在眉睫的。尤其是在全球各国百舸争流发展人工智能的时候，制定相关法律政策既要保护产业创新机制不受到抑制，又要保护网络用户权利不受到过度损害。当前算法技术的发展亟待法律规制思路的转化与具体规则的完善。

四、人工智能时代算法监管的完善：风险防范下的双轨制监管路径

2016 年联合国发布《人工智能政策报告》，表达了对人工智能的关注，针对人工智能技术发展带来的各种问题提出了新的思考方式和解决途径。人工智能时代技术的迅猛发展并未给法律制度的应对留下足够的时间。对于当前算法发展带来的法律风险和监管挑战，匆忙应对的立法难免存在缺陷与不足，而行政法规与政策则具有较强的灵活性。短期内应对法律风险最为关键的是转化结果监管的基本思路，并以此为指导进行基本规则的设计。

（一）算法监管基本思路的转化：从结果监管到风险防范

由于网络传播的特性，算法带来的不利法律后果往往十分严重。一方面，算法在人类社会日益扮演重要角色，其造成的损害往往十分严重；另一方面，网络传播速度使得算法造成的损害难以挽回，而且这种损害无法用传统方式进行计算，更无法移转至加害人。正如王泽鉴先生指出，“对于损害，传统侵权行为法系采取移转方式，而现代侵权法系采分散方式，其所关心的基本问题，不是加害人之行为在道德上应否非难，而是加害人是否具有较佳之能力分散风险”。[28] 因此偏重于结果监管往往不能达到预想的目的，对算法监管的基本思路应在结果监管的同时，更将重点转移至风险的防范。诚如吴汉东教授所言，“对于现代各国而言，人工智能发展的政策考量，其实是基于风险的制度选择和法律安排，我们应通过法律化解风险，通过法律吸纳风险，将风险社会置于法

---

〔25〕 Anupam Chander, “The Racist Algorithm”, 115 Mich. L. Rev. 1023 (2016).

〔26〕 据报道，中科大采用算法，根据学生的消费频率、消费金额来识别贫困生并进行隐形资助。而未曾在学校食堂经常用餐却消费很低的学生也由算法判断不符合资助标准。参见《暖心！这所大学竟用这种方式，偷偷资助“不舍得吃饭”的学生……》(2017 年 7 月 10 日，http://www.sohu.com/a/157397381_252526，最后访问时间 2017－08－20)。

〔27〕 张耀升：《制造假新闻删越战照片政要媒体围攻 Facebook 算法》，见搜狐新闻(http://media.sohu.com/20160913/n468359527.shtml，最后访问时间 2017－08－15)。

〔28〕 王泽鉴：《民法学说与判例研究》(第二册)，中国政法大学出版社 1998 年版，第 165 页。

治社会的背景之中，即对智能革命时代的法律制度乃至整个社会规范进行新的建构。”〔29〕对于算法带来的风险，除了对结果进行监管，更应注重建立一套完整的以风险防范为目的的制度。

首先，风险防范始于预防。结果监管指向内容审查，而对预防性监管则必然指向对算法本身进行审查。这就意味着监管审查的对象，从单轨的内容监管应转化为内容与算法的双轨制监管。这种监管以算法安全为目的，以预防技术产生的负面效应或副作用。算法的研发与设计应遵循道德标准和设计的规则、运行阶段应充分对用户公开透明其设计目的和策略议程，以上均应当接受监管部门的审查，建立标准审核流程。归根结底，以算法为审查对象目的在于增强算法对公众的透明度，以预防风险的发生。

其次，算法的风险防范制度，还应建立一套算法的责任制度。平台责任假设算法作为工具，而技术责任则将算法作为法律客体进行规制，围绕算法建立法律责任与伦理责任。以风险防范为目的的责任体系，应从单轨的平台责任，转化为平台责任与技术责任的双轨责任制。此类责任对应的是用户的权利，使利害关系人知晓算法的决策过程，既是对算法设计者的制约，也是赋予算法利害关系人免于受到算法不利结果反复损害的权利。算法的开发者和使用者承担双轨制的技术责任和平台责任，实质上是通过双轨制对算法造成的责任进行合理的分配。法律这样分配风险有助于形成人对算法自动化决策的基本信任，而工业的发展、科技的研发等都需要一种信任模式的建立和良性运作。〔30〕

最后，算法的风险防范制度，应以算法运行机制为基础。数据是算法发挥作用的必备要素，或者说，人工智能的本质就是算法和数据的利用。以社交媒体为例，其算法带来的风险简要来说基于社交媒体提供的关键资产——数据。这些数据既包括用户自身提供的内容也包括用户根据社交媒体的反馈机制产生的数据。对用户数据收集进行合理程度的保护和限制，这也将会对隐私权保护、机器学习数据来源、用户操纵行为等产生系列影响，如算法收集的敏感数据数量下降会进而减少机器学习的深化程度和用户操纵行为的精准程度。

算法监管机制应以风险防范为基本思路，实行包括内容审查与算法审查并重，平台责任与技术责任并行的双轨制监管思路。同时，应基于算法的运行机制，对算法的生产资源——数据进行管理，对算法造成的消极后果进行风险防范。以下部分将予以详述。

（二）风险防范下监管对象的双轨制：内容监管与算法监管

算法由编程者设计，进而给网站带来巨大的商业价值，因此其本质上是具有商业秘密属性的智力财产。基于此理念，早期各国均不主张对算法进行直接监管，而由技术优势而生的先进算法是互联网公司在商业竞争中的法宝。〔31〕尤其计算机程序的复制非常容易且难以证明侵权行为，所以互联网公司一直坚持发布算法可能会造成更多的危害。〔32〕算法的直接审查既存在着商业秘密方面的顾虑，也存在着技术上的壁垒。

然而，算法已经不仅是导致非法内容的原因，而且已经成为配置资源的规则的时候，就应该像

〔29〕 吴汉东：《人工智能时代的制度安排与法律规制》，载《法律科学》2017年第5期。

〔30〕 龙卫球：《我国智能制造的法律挑战与基本对策研究》，载《法学评论》2016年第6期，第1～13页。

〔31〕 See “Trade Secrets: 10 of the Most Famous Examples”: “Google developed a search algorithm and continues to refine it. Some changes are announced but many are not. Google continues to modify its top secret algorithm to keep businesses and people from gaming the system. It is the top search engine today and shows no signs of giving up its place. ”Google公司开发和不断改进的排名算法是其在商业竞争中制胜的主要原因(https://info.vethanlaw.com/blog/trade-secrets-10-of-the-most-famous-examples，最后访问时间2017-08-20)。

〔32〕 Maxime Cannesson, and Steven L. Shafer, “All Boxes Are Black”, 122 (2) Anesthesia & Analgesia 309-317(2016).

法律一样具有公开性和可预测性，接受公众的审查和质询。而算法开发者和使用者主动接受公众的质询也是自我澄清和保护的手段。在引言中谈到的 Facebook 在美国大选中受到舆论的广泛指责，甚至参议员致信要求 Facebook 澄清“热门话题”中的新闻是否存在主观性时，Facebook 的创始人扎克伯格不仅撰文驳斥，甚至发布了一个 28 页的内部编辑指南，详细介绍了编辑者和算法如何在网站上选择“热门话题”的过程。〔33〕对算法的审查和质询将成为必然的发展趋势。

1. 算法监管的内容和限度

应对算法的哪些方面进行审查？限度在哪里？算法的审查应以用户权利的保护为必要限度，不应要求源代码和算法内部运算透明，否则殊无必要且抑制私营公司的创新效应。算法应该向公众予以公开的是在编码决策过程中，设计者设置的算法的目的和政策议程。

按照设计者设置的算法是否有特定的目的和政策议程，我们可以将算法分为“中性算法”和“定向算法”。〔34〕举个简单的例子，如果 Facebook 在设置“热点新闻”的时候仅仅根据点击率进行排名，那么这种算法就是中性算法。而如果 Facebook 的新闻排名算法优先推送不利于某个候选人的新闻，就是定向算法。仅仅根据数据进行运算得出结果的算法，即使产生了某些不公平的后果，往往是由于现实社会的偏见造成的。例如，美国著名的约会社交软件 Buzzfeed 利用算法对用户进行自动匹配。由于算法追求更高的匹配成功率，最后匹配的结果绝大部分是将相同种族的人进行匹配推荐。而这并不是由设计者的种族主义偏见造成的，而是人们潜意识地喜欢与自己同一种族的人约会。〔35〕而如果算法特意设置了政策导向，为了推进种族之间的融合而忽略这一偏好事实，向用户推荐不同种族的约会对象，这样的算法就是具有目的的定向算法。尤其需要指出的是，在技术上区分二者较为容易，但在法律上势必无法给予二者明确的界定。一个算法是否带有定向性不是泾渭分明的二分法，而是呈现一个频谱，即使是中性算法可能也不免受到设计者价值观的影响，而定向算法的目的性和策略也有强弱之分。

2. 算法监管应建立信息披露义务

针对算法的监管，应为算法的开发者和使用者建立信息披露义务。当算法的设计目的是为了推进预定义的策略议程，而有针对性地设计策略定向算法时，算法的设计必须遵循道德标准和设计的规则，应当接受监管部门审核，审核内容包括设计目的和策略议程。如果搜索引擎决定清除搜索结果清楚的明显的偏见和歧视，应该让用户知道他们正在看到一个修剪整齐的版本。事实上在具体案件上监管部门已经开始实行这一政策。例如，在魏则西事件后，联合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，严格限制商业推广信息的比例对其逐条加注醒目标识，并予以风险提示。〔36〕在社交媒体上，算法编辑必须对于用户来说可见，可审阅和发布明确声明。用户有权知道他们在社交媒体上看到的内容是否政策中立。否则，即使意图是高尚的，社交媒体也会因没有充分的透明度而陷入困境。

---

〔33〕 Facebook, “Trending Review Guidelines”, Released May 11, 2016 (https://fbnewsroomus.files.wordpress.com/2016/05/full-trending-review-guidelines.pdf，最后访问时间 2017-08-20)。

〔34〕 Omer Tene, and Jules Polonetsky, “Taming the Golem: Challenges of Ethical Algorithmic Decision Making” (2017).

〔35〕 Katie Notopoulos, “The Dating App That Knows You Secretly Aren't Into Guys From Other Races”, BUZZ FEED, Jan. 14, 2016 (http://www.buzzfeed.com/katienotopoulos/coffee-meets-bagel-racial-preferences 7l5zKPb2).

〔36〕 参见《国信办联合调查组结果：百度竞价排名影响魏则西选择百度：从 6 方面整改》(2016-05-09 http://www.guancha.cn/economy/2016_05_09_359617.shtml，最后访问时间 2017-08-20)。

算法监管的内容、流程，以及加诸互联网企业的强制披露义务，均为专业性和技术性较强的领域。因此应设立专业的监管部门进行算法监管，根据算法的功能进行层次不同的信息披露机制的设计。这种内容与算法的双轨制监管，是建立平台、数据和算法的监管体系的核心制度。现在网络治理的模式以平台责任作为起始点，数据和内容作为平台责任的支撑，其中包括监管平台在数据的收集、使用过程中的责任，也包括平台在内容发布和监管中的责任。现行监管体系中缺失对于算法的监管。算法以数据作为资料生成内容，实际上是网络治理模式中的核心要素。及时建立对于算法的监管审查制度，对平台责任以及网络空间治理具有关键意义。

（三）风险防范下责任承担的双轨制：平台责任与技术责任

对算法的责任应从单轨的平台责任，转化为平台责任与技术责任并重的双轨责任体制。技术责任意指机器伦理，法律责任的设置应围绕技术的运行，而非不论网络平台与算法的关系，一味由平台承担责任。在算法的技术发展下，平台与算法的关系已远非“工具”一词可以概括，故应该对责任承担的制度设计予以调整。

1. 技术责任之一：赋予相对人的算法解释权

算法的技术责任即意味着相对人的权利。赋予相对人算法解释权即是加诸算法开发者和使用者的技术责任。算法解释权是指在因算法的机器学习和自动决策而认为自己可能或已经遭受损害的人，并有权要求知晓任何自动个人数据处理的逻辑，可以向算法做出的决定提出异议，并要求算法更正其错误。算法解释权是典型的人工智能时代新型的救济权利。它与行政相对人请求复议的权利相似，只不过行政复议请求的对象是掌握公权力的政府，而算法解释权请求的对象是掌握着技术权力的算法设计者和使用者。

算法日益渗入日常生活不可避免会出现错误。例如美国航空公司的一位资深驾驶员称，由于计算机算法将他与一位爱尔兰共和军领导人混淆，使得他先后80次在机场遭到拘禁。〔37〕而由于人们对于算法客观性和可靠性的盲目信任，这种对算法决策结果的挑战和更正尤为艰难。各国逐渐展开了和算法解释权相关的立法工作。如英国的《数据保护法案》(Data Protection Act)规定，人们有权对人工智能做出的决定进行挑战。欧洲2016年通过的新的《欧洲通用数据保护条例》(GDPR)提出，“每个数据主体应该有权知道和获取通信，特别是关于任何自动个人数据处理中涉及的逻辑，至少是在基于分析的情况下，这种处理的后果”。〔38〕《欧洲通用数据保护条例》(GDPR)中也提出，“应该采取适当的保障措施，其中应包括数据主体的具体信息和获得人为干预(而非数据干预)的权利，以表达其观点，已获得对此类评估之后达成的决定的解释，并对决定提出质疑。”〔39〕

---

〔37〕 见《应建立第三方机构以管控作出糟糕决定的人工智能》，载搜狐网(http://www.sohu.com/a/125322861_465915，最后访问时间2017-08-21)。

〔38〕 “Every data subject should ... have the right to know and obtain communication in particular with regard to ... the logic involved in any automatic personal data processing and, at least when based on profiling, the consequences of such processing.” Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

〔39〕 《欧洲通用数据保护条例》(GDPR)中在第71条明确提出了解释权，表述为被自动决策的人应该具有适当的保护，具体应包括数据主体的特别信息和获得人类干预，表达自己的观点，并且有权获得评估决定的解释，并对决定提出质疑。“Recital 71, a person who has been subject to automated decision-making: should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, *to obtain an explanation of the decision reached after such assessment* and to challenge the decision.”

从技术层面来说，有很多研究力量投入到了和决策策略有关的深层神经网络的理解中，如对随机扰动技术、[40]不变形分析、可视化和维度降低，[41]但是显然这种技术上的可解释性并非法律上的可解释性所指的“有法律意义的信息”。而对权利人有法律意义的解释内容可能分为两个层面：其一，解释的对象是系统功能，例如该自动决策系统的逻辑、意义、算法设定的目的和一般功能，包括但不限于系统的需求规范、决策树、预定义模型、标准和分类结构等。其二，解释的对象是具体决策，及特定自动决策的理由、原因、产生决策结果的个人数据，例如每种指标的功能权重，机器定义的特定案例决策规则，起参考辅助作用的信息等。[42]这些信息都可以解释自动决策产生的原因。举例而言，很多互联网信贷公司借助用户的社交媒体和购物记录等信息来对用户进行信用评级发放小额信贷。这种自动的信用评分系统，用户可以申请信贷公司或算法提供者解释算法的功能，和通用的逻辑（比如参与决策的数据类型和特征，以及决策树的类别），算法的目的和意义（进行信用评分以发放贷款），和设想的后果（可能影响信用记录，影响利率）。在第二个层面，用户可以要求解释具体决定的逻辑和个人数据的权重，例如用户的信用评分结果参考了哪些数据以及这些数据在决策树或者模型中的权重。

第一个层次的信息类似行政复议中对行政决定的合法性审查，通过对算法的决策基本情况的了解，用户有权知晓算法是否合法，是否包含歧视因素，等等。而第二个层面的审查类似行政复议中对行政决定的合理性审查，否则，如果一个人被互联网信贷公司拒绝，他被告知，算法充分考虑了他的信用记录、年龄和邮政编码，但此人仍然不清楚申请被拒的原因，解释权便形同虚设。虽然欧盟有法律提出相对人应享有“拒绝接受数字化决策，要求人为干预”的权利，但其仍停留在学理讨论阶段。[43] 不仅应建立风险防范的事前审查机制，[44]同时应重视类似算法解释权等应对损害的事后解决机制，算法解释权意在如何分配风险以求得全社会利益最大化。

2. 技术责任之二：机器伦理的发展

算法解释权的行使面临着重大限制，原因是各国政府更加倾向于保护科技产业的发展而避免发展技术责任。一直到现在，对于数据控制者必须对用户披露的信息类型，以及适用于自动化决策的算法访问权限的限制，尚未在欧洲各地的法院的判例中得到普遍的明确范围。[45]例如德国数据保护法规定，数据控制者必须在决策的“评估”中向用户通报其所考虑的因素，但不必揭示给予

---

〔40〕 Matthew D Zeiler and Rob Fergus, “Visualizing and Understanding Convolutional Networks”, in *European Conference on Computer Vision* (Springer, 2014), pp.818－833.

〔41〕 Aravindh Mahendran and Andrea Vedaldi, “Understanding Deep Image Representations by Inverting Them”, In 2015 IEEE conference on computer vision and pattern recognition (CVPR), IEEE, 2015, pp.5188－5196.

〔42〕 Bryce Goodman, and Seth Flaxman, “European Union Regulations on Algorithmic Decision-Making and a Right to Explanation”, in 2016 ICML Workshop on Human Interpretability in Machine Learning, New York, NY: ArXiv e-prints.

〔43〕 Ibid.

〔44〕 司晓、曹建峰：《论人工智能的民事责任：以自动驾驶汽车和智能机器人为切入点》，载《法律科学：西北政法大学学报》2017 年第 5 期，第 166～173 页。

〔45〕 “for instance, debate in the UK House of Lords concerning the meaning of ‘logic involved’ and ‘trade secrets’ in the 1998 Data Protection Act: Grand Committee on the Data Protection Bill, ‘Official Report of the Grand Committee on the Data Protection Bill [HL]’ (*Hansard*, 23 February 1998)” (UK Parliament－House of Lords 1998) (http://hansard.millbanksystems.com/grand_committee_report/1998/feb/23/official－report－of－the－grand－committee#S5LV0586P0_19980223_GCR_1，最后访问时间 2017－08－02)。

每个因素的精确重量(即在自动化决策过程中使用的版权保护算法)。[46] 根据几位评论家的观点,[47]德国SCHUFA59判决[48]显示,用户没有权力彻底调查自动处理系统(在判例中是信用评分)的准确性,因为基础公式受到商业秘密的保护。

面对算法解释权的限制,发展机器伦理作为替代规则成了必然选择。解释权不是在算法决策中实现问责制和透明度的唯一途径。[49] 立法者可能需要通过对算法通过道德审查标准,来防止对用户的操纵或产生不公平的后果。在敏感性和风险较高的算法决策领域更适用伦理性的审查要求。当然,目前这种方法面临着许多挑战。在短期的具体规则中,可以采用人工筛查的手段,预见性地将算法可能出现的歧视问题等进行解决。算法本质上是"以数学方式或者计算机代码表达的意见",设计者很容易将自己的偏见嵌入算法系统中。例如,美国一些法院使用的一个犯罪风险评估算法COMPAS被证明是对黑人造成了系统性歧视。发展机器伦理要求社会学家、法学家等共同设计伦理框架,在算法设计阶段就为算法的研发和应用提供道德准则。

以算法解释权为核心的技术责任必然将成为人工智能时代亟待深入研究的主题。算法解释权制度中,权利人的资格要求,行使的前提,行使的程序等,以及机器伦理的道德标准等都需深入研究,本文篇幅所限不再展开。需要指出的是,平台责任与技术责任并非一定泾渭分明,某些情况下可能存在交叉。平台在很多情况下本身就是算法的开发者与算法运行的维护者,技术责任的承担主体可能就是平台,例如算法解释和机器伦理的考察可能都指向平台。例如,上文中提到的Facebook的情感蔓延项目,当Facebook收集较多的用户数据用情感操纵算法进行处理后,美国FTC要求Facebook平台公布算法,这同时体现了公权力通过对平台的监管来实现内容和算法的双重监管。然而,平台责任和技术责任评价对象不同,算法的设计者和使用者也并不必然重合,这两点决定了平台责任和技术责任需做出区分。

(四)风险防范下的生产性资源保护:数据资产收集与使用的限制

进入人工智能时代,数据的体量和质量决定了算法是否足够"聪明"。数据对于算法造成的风险防范具有至关重要的作用:一方面,从数据体量入手,限制数据的收集可以防范算法造成的用户隐私泄露的风险;另一方面,注重数据收集的质量,可以避免算法以歧视性的数据作为自动决策的依据,造成歧视性后果。需要注意的是,数据的收集以及使用无法回避平台在数据收集和管理方面的责任。算法监管的体系性和有效性,亟待完善数据管理方面的平台责任制度。

1. 敏感数据收集的合理限制

数据的范畴和形式在不断进行扩张。以社交媒体为例,其数据量大,更新快,且具有难得的多

---

〔46〕 Douwe Korff, "New Challenges to Data Protection Study - Country Report: United Kingdom"(European Commission DG Justice, Freedom and Security 2010), 48(http://papers.ssrn.com/sol3/papers.cfm? abstract_id=1638938,最后访问时间2017-07-15)。

〔47〕 Bräutigam and Schmid-Wudy (n 35) 62; Jens Hammersen and Ulrich Eisenried, "Ist Redlining in Deutschland erlaubt? Plädoyer für eine weite Auslegung des Auskunftsanspruchs" [2014] ZD Zeitschrift für Datenschutz 342.

〔48〕 Judgment of the German Federal Court Bundesgerichtshof 28 January 2014 - VI ZR 156/13. LG Gießen 6 March 2013 - 1S 301/12. Also, AG Gießen 11 October 2014 - 47 C 206/12.

〔49〕 For additional discussion of transparency and the GDPR, see: Dimitra Kamarinou, Christopher Millard and Jatinder Singh, "Machine Learning with Personal Data"(https://papers.ssrn.com/sol3/papers.cfm? abstract_id=2865811,最后访问时间2017-08-22)。

样性，用户体量、使用语言、表达方式极为丰富。[50] 在人工智能的竞争中，社交媒体平台成为企业发掘个人数据的“金矿”。在这座数据“金矿”中，算法对于数据的分析和利用能力，使得有关个人隐私的敏感数据范畴在不断扩大。例如，在2016年两位科学家发布研究结果，可以通过用户在Instagram上发布的照片的内容、配色方案等用算法进行分析，预测用户的抑郁症倾向。其预测指标可以实现在医生确诊前识别抑郁症患者，诊断准确性甚至优于普通医师。[51] 独立于提供的内容的其他特征，如可以由任何人使用的任何计算机实时捕获的自然打字模式，也被确定为早期识别微妙的精神运动障碍的可能患者，具有帕金森病或痴呆早期诊断的潜在应用。[52]

然而，法律关注和保护的数据仍局限于涉及个人隐私的敏感数据。如可以定位个人的姓名、身份证号码、银行卡号密码等，或者体现个人社会活动的通话记录、短信记录、位置信息等。如2017年公安部破获的贩卖公民个人信息达50亿条的特大案件，数据形式包括物流信息、交易信息、个人身份等。[53] 在人工智能高度发展的时代，非专业人士无法合理预期个人数据被算法使用的方式和后果，即使可以，也由于生活和网络的高度融合无法拒绝提供个人数据。这些数据无意识地贡献了包含着精神健康、身体健康、经济状况等敏感的可识别的特征。例如，用户的精神健康状态可以通过其社交媒体发布的信息被识别和诊断，甚至被医疗保险公司用来确定用户的保险费用。在这种情况下，用户被剥夺了个人决定是否受到诊断的权利，患者卫生信息保密的法律也被彻底架空，而这些健康隐私有可能被私营公司购买以确定营销目标，或者被用来避免高风险的医疗保险客户。用户在使用服务的需要下被裹挟前行，往往并没有其他选择权。用户应被赋予切实存在的选择权利，比如在社交媒体通过按键选择的方式增加用户对数据被收集和使用的自主性。

2. 歧视性数据资产的使用限制

数据是现实社会的反应，因此很可能包含着现实社会的歧视现状。算法决策是在用过去预测未来，而过去的歧视可能会在算法中得到巩固并在未来得到加强，因为错误的输入形成的错误输出作为反馈，进一步加深了错误。数据的有效性、准确性，也会影响整个算法决策和预测的准确性。因此，应在对数据资产的管理中，充分认识到数据资产质量的重要性，避免歧视性数据资产的适用。

例如，2013年，波士顿采取了创新方案解决普通市政道路坑洼问题。Boston Street Bump程序可以使用智能手机运动感应功能，向市政府报告用户驾驶时遇到的街头坑洼引起的震动，以确

---

〔50〕 参见清华大学电子工程系信息认知与系统智能研究所副所长黄永峰于2015年12月23日在清华RONGv2.0系列论坛之“社会关系网络与大数据技术”专场上所做的题为《网络社交媒体的情感认知与计算》的演讲(http://www.sohu.com/a/61999021_308467，最后访问时间2017-08-20)。

〔51〕 M. De Choudhury, M. Gamon, S. Counts, & E. Horvitz, “Predicting Depression via Social Media”, Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM) (2013), pp.128-137(https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/download/6124/6351，最后访问时间2017-08-20)。

〔52〕 L. Giancardo, A. Sánchez-Ferro, I. Butterworth, C. S. Mendoza, & J. M. Hooker, “Psychomotor Impairment Detection via Finger Interactions with a Computer Keyboard during Natural Typing”, Scientific Reports 5, 9678(2015).

〔53〕 参见《黑产团伙长期贩卖公民信息，泄露数据达50亿条》：“近期，公安部破获了一起盗卖公民信息的特大案件，50亿条公民信息遭到泄漏，而嫌疑犯被传是京东网络安全部员工，与黑客长期相互勾结。据公安部披露，犯罪嫌疑人郑某鹏利用京东网络安全部员工这一身份，长期监守自盗，与黑客相互勾结，为黑客攻入网站提供重要信息，包括在京东、QQ上的物流信息、交易信息、个人身份等。”(http://www.leagsoft.com/news/p/1864，最后访问时间2017-09-07)

定要修复的街道上的问题。有趣的是，数据中心统计城市的富裕地区比贫困地区有更多的坑洞。这是由于智能手机在贫富地区不均等分配造成的。[54] 同理，如果将算法应用在犯罪评估、信用贷款、雇佣评估等关切人身利益的场合，歧视性数据资源带来的群体、种族的利益影响更为巨大。应以风险防范为目标，对数据资产的质量进行充分管理。具体手段应包括建立数据评估机制，数据去噪筛查机制，以及由专业人士研究如何对数据进行复原。

人工智能发展带来的不仅有法律监管的风险和挑战，更有监管技术和手段的丰富和提高。许多过去不可能的监管措施随着物联网、大数据等技术的发展都将变为可能。监管的思路转化为风险防范是关键，具体措施的发展可以留待人工智能和算法技术的发展来逐步解决。未来风险可控的前提下，人工智能极大地造福于人类的愿景值得期待和憧憬。

五、余　　论

从社交媒体的新闻推送开始，本文梳理了传统的算法监管的制度框架及其隐含假设，并介绍了人工智能时代算法通过社交媒体海量数据和机器学习取得的发展。这些发展对传统监管路径提出了巨大挑战。面对算法发展带来的风险，有必要采取预防性的行为和因应性的制度。“法律制度的价值和意义就在于规范和追寻技术上的可以管理的哪怕是可能性很小或影响范围很小的风险和灾难的每一个细节。”[55]对风险进行法律控制是人工智能时代法律对社会发展做出的积极回应。对于个人数据收集、使用行为的限制，设立具有不同层次的算法强制披露义务以增强算法透明度，创设算法解释权给利害关系人以救济，都是在通过不同角度去应对科技发展带来的风险。

本文仅仅是对人工智能时代法律应对风险具体规则的抛砖引玉，许多具体制度设计仍有待法学学者的深入研究。而在这些研究中有一点应予明确：现今的制度框架总体来说是有利于科技发展和进步的，法律的适当滞后并不必然是缺陷。面对科技的发展，法律应有一种保守克制的态度，而这种克制态度的重要体现之一，就是并不必然将所有问题归咎于人工智能和算法的发展，而急于对算法发展进行不必要的限制。对于算法控制的信息流动和进行的自动决策的批判不绝于缕，而这种算法发展带来的“黑箱”可能并不是唯一原因。事实上，太阳底下无新事，过去的股票内幕交易，某些行政决策又何尝不是所谓的“黑箱”呢？相比人的决策，算法决策更不容易受到固有偏见的影响而更加公平客观。《大数据：打击歧视和授权团体的工具》一书中提到了多个案例，展示了企业、政府和民间社会组织如何利用数据分析来保护和授权弱势群体，数据驱动的决策可以有效减少歧视，促进公平和机会。[56] 对算法带来的风险进行法律规制时，制度达到目的衡量的标准应是现实世界的一般标准，而非人人大同的乌托邦。秉承这一原则，有利于在创制法律制度时，充分衡平科技发展与权利保护。

（责任编辑：蒋红珍）

---

〔54〕 Phil Simon, “Potholes and Big Data: Crowdsourcing Our Way to Better Government” (https://www.wired.com/insights/2014/03/potholes-big-data-crowdsourcing-way-better-government/, 最后访问时间 2017-09-02)。

〔55〕 [德] 乌尔里希·贝克：《从工业社会到风险社会——关于人类生存、社会结构和生态启蒙等问题的思考(上篇)》，王武龙编译，《马克思主义与现实》2003年第3期。

〔56〕 Future of Privacy Forum & Anti-Defamation League, Big Data: A Tool for Fighting Discrimination and Empowering Groups, 2014 (https://fpf.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-Report1.pdf).