大数据时代的算法歧视及其法律规制

●刘　璐

[摘要] 由算法引发的歧视问题是人工智能发展过程中产生的新型风险，和传统歧视相比，算法歧视带来的是一种更深层次的歧视和不平等，并引发了很多伦理问题。算法歧视的产生不仅是歧视在算法领域以数字的形式转化，而且含有技术本身的复杂性所带来的风险。因此，本文认为，应当首先理解算法歧视的内涵和本质，即算法歧视不仅包含现实中延伸到算法领域的人们固有的歧视，还包括算法技术的复杂性、隐蔽性以及数据偏差造成的歧视。算法歧视在实践中的表现形式主要有借贷歧视、AI 应用歧视等。分析算法歧视的产生逻辑和国内外算法歧视的法律规制现状，能够为我国构建算法歧视法律规制体系提供思路，从而确立数字经济时代我国的算法法律秩序，确保数字社会的安全。

[关键词] 大数据时代；算法歧视；法律规制

Q 算法歧视的内涵及其本质特征

（一）算法歧视的内涵

结合对“算法”和“歧视”的概念理解算法歧视，其主要是指以算法作为手段，依靠数据挖掘和机器深度学习等技术的自动化决策系统在对数据主体做出分析、决策时，由于数据和算法本身不具有中立性或者隐含错误、被人为操控等原因，对相对人进行差别对待，造成歧视性后果。通常表现为对具有某种不同属性的特定社会群体成员采取的不利行为或实施的不利程序。大数据时代以数据作为基础资源，算法作为核心引擎，算法主宰了数据的收集、处理和输出的全生命周期。

算法歧视在严格意义上不是法学体系中的用语，其概念也没有统一的定义，但是却可以以实践中的具体歧视形式总结出其内涵。算法技术的发展使传统意义上的歧视有了新的载体表现形式，但是却没有改变歧视本身的特性，即与平等相对的差别对待，而且是一种不合理的区别、限制、排斥或侵害人格尊严的歧视，这种歧视还因算法技术的复杂性变得更加具有隐蔽性和不可逆性。

（二）算法歧视的本质特征

一是精准性。算法可以通过对用户精准画像的方式，为用户打上歧视性标签。如性别、年龄、职业、消费能力等都可以成为歧视性因素。二是多元性。传统的歧视通常依据性别、学历等显性特征，但算法能够挖掘出更深层次的隐性特征来作为其歧视处理的依据，包括网页浏览记录、购物记录、行车路线等。三是片面性。社会对于个体的判断通常是综合和动态的，但算法无法获取或处理用户的全部数据，基于样本数据的有限性和可伪造性，算法的准确性和公正性就大打折扣，至此由算法导致的歧视就具有片面性。四是隐蔽性。算法技术自身具有复杂性和黑箱性，于是算法的专业壁垒导致普通人无法理解其运行机理，这就导致其在技术的遮蔽下产生更深度的歧视。

Q 算法歧视的表现形式

（一）信贷歧视

信贷歧视主要存在于金融领域，随着大数据与金融行业的深度融合，金融平台借助算法算力形成了借贷歧视，金融平台通过大数据分析挖掘客户的个人信息，算法根据这些信息对用户的借贷限额及借贷利率等进行差别对待，这样金融借贷平台不仅没有实现借贷公平，反而与线下金融借贷的歧视没有什么区别。例如，加州大学伯克利分校教授对网贷平台的研究“金融科技时代的消费借贷歧视”表明，算法也可能像贷款人员一样，对借款人进行歧视。以美国最大的贷款机构之一的在线平台 Quicken Loans 为例，研究人员发现，在线下贷款机构受到歧视的少数族裔在金融借贷平台上仍然受到歧视，他们在线贷款的申请利率普遍比普通人群高 5.3 个基点。这与其从线下借贷机构借款时需要支付的 5.6 个额外积分没有太大区别。这种利率差别并不是由信誉差异导致的，算法借贷中“大数据”使用的增加可能会进一步加深歧视。

（二）AI 应用歧视

清华大学研究团队发现，AI存在性别歧视，如“重男轻女，爱白欺黑”，原因是数据集带来的偏差、设计者的缺陷和算法本身的不足。如今，随着人工智能的不断发展，AI机器人被广泛使用，AI聊天机器人在情景交互中对信息进行学习、储存来模仿人类对话，进而实现与用户进行交互对话，并具备协助用户对日常工作进行记录和信息搜索等功能。一旦AI聊天机器人在学习过程中学习到了带有歧视性的信息，就会引发算法歧视问题。比如，由韩国研发的一款AI聊天机器人能够通过在交互情景中学习对话而能像真人一样和用户聊天。然而，这款AI聊天机器人在与用户进行交互的过程中，发表了歧视残疾人、孕妇、同性恋等群体的不当言论，而这些观点正是AI机器人在与用户交互的过程中学习到的。AI机器人在未来会被更广泛地普及应用，AI机器人发表言论是否正向积极直接关系用户的使用体验及身心健康，而AI机器人导致的算法歧视问题也可能会引发更严重的社会歧视问题。

Q 算法歧视产生的原因

（一）算法的价值负载性

算法不是客观中立的，而是包含预设的价值立场。也就是上面提到的“偏见”，如果算法在设计之初就已经存在一些价值偏差，并且嵌入到信息的收集、标注和处理中，那么算法的运行结果也会体现出价值偏见，形成的社会效果就是算法歧视。如传统歧视中的性别偏见、种族偏见经过算法的广泛应用进一步转化为算法歧视，这种由算法的价值负载性造成的算法歧视反映的正是一种社会权力关系中深层次的不平等。

（二）数据输入的局限性

导致算法歧视的一个重要原因是向算法提供的训练数据或输入数据本身存在偏见，即所谓的“偏见进，偏见出”。如在数据收集过程中以社会对某一地区或群体的歧视为例，或者数据收集的范围是具有偏见意识的人群样本，那么算法产生的“训练数据”也会受到影响，从而影响算法的输出结果。除此之外，采样的数据来源集中于部分群体，而忽略其他群体，那么算法仅能捕捉到被采样群体的特征，其输出的信息也仅适用于被采样群体，而那些未被采样的群体就有可能成为被歧视的对象。

数据集来自人群，一旦带有敏感因素或错误的数据在不脱敏的情况下被过度收集和利用，就会导致机器学习在处理特定问题时自动识别敏感数据，从而导致算法自动决策结果出现歧视。正如有的学者指出：数据是人类社会生活的表现形式，如果社会文化具有歧视，那么算法决策就会在网络世界映射人类的歧视。

（三）“黑箱”带来的算法权力

学界普遍认为算法权力是算法歧视发生的重要原因，依照马克斯·韦伯对权力的解释，当强势一方可以贯彻自己的意志而无需顾及弱势一方的反对时，即存在权力关系。算法技术有着很强的专业性，这导致算法从接受输入数据到输出决策结果的整个过程，会形成一个外界无法知晓的“黑箱”。但是受算法影响的群体最多仅能获知算法运行的结果，而算法使用的数据、逻辑分析等关键过程则被算法“黑箱”隐藏，算法的非透明性使得歧视更加隐蔽。在此情况下，算法系统的开发者认为这些技术属于“商业秘密”或“专有特权”，无需向公众披露。这就导致进一步助长了某些算法设计和经营者为追逐经济利益滥用算法这种技术权力。

Q 我国算法歧视法律规制的现状及其完善

（一）我国算法歧视法律规制的现状

目前，我国还未有专门针对以算法歧视为对象的立法，只有在《消费者权益保护法》《电子商务法》《反垄断法》《价格法》及《个人信息保护法》等法律文件中有针对算法歧视规制的零星条款。不仅如此，我国的司法实践在算法的技术壁垒下，举证的难度和请求权基础缺失都直接导致了维权困难，这就使算法歧视的风险在事后不仅得不到规制，甚至可能会扩大权利损害。《个人信息保护法》第二十四条赋予了个人对算法的拒绝权和要求解释的权利，但是总体来说我国对于算法歧视的重视以及应对尚存在问题，还未形成具有整体性、系统化的法律规制体系。

（二）我国算法歧视法律规制的完善设想

第一，建立算法公正的基本原则。平等在我国宪法中是一项基本原则和权利，而算法公正则是这一法律原则在数据智能领域的进一步延伸和细化。在算法公正原则的指导下，算法的设计者以及算法控制者应自觉维护公民的平等权利，避免对特定群体或地区的不合理差别对待。例如，在招聘算法中不应存在基于种族、性别、年龄等个人特征的隐形歧视。这还要求在具体使用算法的过程中，算法的决策过程和结果应当遵循透明原则，可被理解和解释。一旦发现算法存在不公正的行为，需要通过建立全面的日志记录和监控系统实现可追溯。除此之外，还必须明确人是算法的立法者和控制者，法律的算法和算法的法律不应成为一个闭环，它们之间，必须由人作为起点和终点。

第二，规范对算法的利用。算法的使用应当始终秉持以人为中心，在利用过程中努力维护人的主体价值。目前，我国已经出台的相关制度中规定了信息处理者要推进算法公开和透明、公民合理合法行使算法解释权以保证个人的“知情同意”，以及从制度层面要求反对算法歧视和行使算法拒绝权等。由此可以看出，我国针对算法的规制已经开

始试图构建体系，但是针对算法歧视的问题还未有相应的规范，因此在接下来出台的相应法律文件或者制度构建中，可以考虑对算法利用的范围、方式及限度进一步做出更加细致的规定。

第三，强化法律对算法的归化。简而言之，即以算法技术本身治理算法歧视，曾有学者指出，为了保障人的权益，应当加强法律对技术的归化。即研究具有反算法歧视意识的技术，将法律化的技术伦理内化于算法技术之中。为了使技术人员理解法律定义下的非法歧视，应加强法学家与计算机科学家的合作，研发出合乎法律伦理和技术可能性的公正算法。除此之外，还可以加强算法科技人员的伦理意识教育，使其形成反算法歧视的意识，可以考虑在国内高校开设技术伦理课程，以使公正价值由算法设计者进入人工智能。在此思路下，增强研发人员及训练数据的多样性亦可有效纠正算法偏见和歧视，实现以算法规制算法歧视。

第四，明晰算法主体的权利和义务。这是保障算法治理透明度和公正性的关键。如明晰其著作权、知识产权以及其他相关权益。著作权是算法主体的重要权益，包括但不限于算法的设计、实现、运行和维护过程中产生的所有思想、方法、步骤等结晶，这些成果应受到法律保护，并为算法主体的创新提供必要的法律保障。进行知识产权保护则是确保算法主体创新成果独占性的重要手段。通过制定详细的知识产权法律法规，明确算法主体对其作品所享有的各项权利，包括但不限于专利权、商标权、著作权等，可以有效地防止他人未经许可擅自使用或侵犯算法主体的知识产权。此外，算法主体的数据权益也需得到有效保护。在大数据时代，算法主体对数据的挖掘、分析、利用等方面的权益得到了法律的有力保障，包括确保数据安全、保护隐私权、禁止大数据杀熟等行为。

第五，建立专门的算法歧视问责制度。当预防措施不能有效遏制算法控制者的歧视行为时，就需要对算法歧视造成的侵权行为进行问责。具体而言，责任追究制度的建立包括两个层面的内容：一是明确侵权主体，设计算法歧视的责任原则和因果关系的确定规则，明确算法歧视的责任主体，一般涉及算法设计者和算法运营者之间的责任划分，然后在确定责任主体的基础上，明确各主体的民事或行政责任。二是畅通权利受害者的救济渠道。一方面，算法的操作者应通过适当的手段确保受害者有足够的查询、反馈和投诉渠道，对受害者提出的问题积极回应和补救；另一方面，应建立相应的集体诉讼制度，确保受害人能够通过司法渠道维护自己的合法权益。

第六，形成多元协同治理格局。由于算法背后的利益主体具有多元性，过多的限制与规制又会损害算法技术所蕴含的经济利益，再加上算法技术有很强的专业性，为实现效益与权利保护的平衡，就需要立法机关、主管部门、算法平台、行业组织等多方主体共同防范算法歧视风险。如在企业内部建立算法安全合规体系，既可以从源头上解决算法歧视问题，又可以提升算法治理的监管效率。另外，可以赋予算法行业协会对企业的审查权，针对算法技术所遵循的伦理、价值以及是否合规进行审查。对于立法机关而言，则要将算法审查、算法解释、算法问责等规范细则化，从而使公民权利救济落到实处。同时，在执法层面对滥用算法的企业等做出行政制裁，在司法层面可以引入针对算法歧视带来损害后果的侵权规则和惩罚性赔偿规则，发挥司法审查的功能，由人民法院对算法歧视案件做出审查认定及评价，并由此形成相应典型案例以供参考。

结束语

算法虽然带来算法歧视等风险性问题，但相信只要各界人士共同努力为算法风险的良性规制达成共识，引导算法向善、科技向善，一定可以实现以算法技术为核心的人工智能可持续发展，成为驱动数字经济时代的新动能。

参考文献

[1]刘朝.算法歧视的表现、成因与治理策略[J].人民论坛,2022(02):64-68.

[2]张玉宏,秦志光,肖乐.大数据算法的歧视本质[J].自然辩证法研究,2017,33(05):81-86.

[3]沈伟伟.算法透明原则的迷思——算法规制理论的批判[J].环球法律评论,2019,41(06):20-39.

[4]姜野.算法的规训与规训的算法:人工智能时代算法的法律规制[J].河北法学,2018,36(12):142-153.

作者简介：

刘璐(1995—),女,汉族,河南商丘人,硕士研究生,江西理工大学法学院,研究方向:民商法。