

算法决策下隐蔽性就业歧视的风险预防与法律合规

苏 畅 魏子钦 周培桓

(南京审计大学, 江苏南京 211899)

摘要: 随着信息技术与大数据算法的深度发展与应用, 现代企业在用工过程中开始大量使用算法自动化决策辅助, 由此使得隐蔽性就业歧视逐步嵌入“人机交互决策”之中, 进而引发就业歧视纠纷等现实问题。本文即在此背景下展开研究, 通过分析算法决策下隐蔽性就业歧视的生成路径, 进而提出隐蔽性就业歧视的风险预防与治理策略, 并由此总结隐蔽性就业歧视的法律逻辑与规制边界。

关键词: 算法决策; 隐蔽性就业歧视; 风险预防; 法律合规

随着社会公平的发展, 就业歧视已经成为用工领域中的重要话题。在“算法社会”环境下, 尽管算法决策以效率为基本导向, 可以有效提升企业的决策效能, 但表面上的合理与中立算法, 在实际执行与实践过程中却容易产生隐蔽性就业歧视风险, 进而为劳动者带来不公平对待后果。2021年, 我国正式通过《中华人民共和国个人信息保护法》(简称《个人信息保护法》), 其中明确提出, 要禁止企业通过自动化决策进而为交易主体造成歧视, 由此说明我国立法者已经认识到自动化算法决策中隐藏的歧视风险, 进而通过法律做出规范与引导, 对于社会公平建设具有重要价值。

一、算法决策下隐蔽性就业歧视的生成路径

在企业用工中, 算法决策应用主要依靠“人机交互决策”的技术路径实现, 即通过人与机器收集和处理劳动者信息数据, 进而通过算法做出雇佣决策。而“人机交互决策”显然有着不可忽视的就业歧视风险。

一方面, “人机交互决策”主要包括“企业的价值判断与行为选择”和“机器算法的自动判断”两部分, 其中企业主要负责问题建构与数据收集。问题建构即在算法技术开发过程中, 针对用人问题制定一个目标, 并将其转化为可衡量的结果变量。而在此过程中, 企业由于存在利益驱动, 由此导致其目标转化变量可能存在抛弃公平价值的现象, 进而损害了劳动者的平等就业权。数据收集是指在算法技术开发过程中, 企业通过各种渠道途径进行数据挖掘, 进而将其作为结果变量中的衡量数据, 也是训练集数据。机器算法则主要负责数据挖掘活动, 指的是在算法技术开发过程中, 企业可以根据收集的结果变量数据, 采用统计学分析工具进行数据分析, 由此寻求统计学层次的相关性因素, 由此将用人标准转化为具体指标, 进而做出用人决策。但该算法分析的相关性并不一定满足因果关系, 因此其中很可能存在具有隐蔽性就业歧视风险。

另一方面, “人机交互决策”算法有着非价值中立、静态技术的特征, 即其中的决策既包含了企业自身的价值判断与行为选择意愿, 也包括了机器算法的自动判断。这就使得决策过程存在一定的“算法黑箱”, 如果要求算法技术开发内容进行公开, 可能涉及侵权或侵害商业秘密的问题, 而不公开就必然面临着就业歧视风险。

二、算法决策下隐蔽性就业歧视的风险预防与治理策略

(一) 技术治理

针对现阶段反歧视相关法律无法应对算法决策所导致的隐蔽性就业歧视风险问题, 可以通过建立算法影响评价制度, 进而从技术完善与优化角度, 从源头进行就业歧视风险防范与治理。

算法影响评价是一种算法规制技术, 指的是通过特定标准对算法的决策过程、内容与结果进行评估, 进而根据评估结果判断

其决策是否符合标准。通过影响评价规制技术的应用, 一方面可以提高算法可解释特性, 敦促相关开发者能够科学阐述算法决策的机理与结果, 由此对算法结果进行客观描述、说明、解释并论证。另一方面, 其不仅可以构建风险防范机制, 在事前评估算法可能存在的算法歧视风险, 而且可以作为算法公平的技术治理方法, 由此定制标准化的影响评价, 确保企业用人算法符合职场多元化与弱势群体保护等公平理念。算法影响评价规制技术可以针对算法决策过程生成评估报告, 进而展示决策过程涉及的模型建构原理, 说明“训练集”数据的来源与处理方式, 由此从基础数据层面检测其是否存在侵犯个人信息权利、是否对特定人群进行歧视和筛选的情况。此外, 基于算法影响评价规制技术, 还可以将就业平等要求嵌入算法公平性设计之中。比如随机性嵌入算法后, 可以避免隐藏的偏见与一致性观念造成的歧视结果。在调试和训练模型时, 应通过修改和遴选数据, 确保其输出结果符合公平特性。比如在训练集数据输入时, 可以删除受保护特征变量信息, 比如性别、民族、健康情况等。

(二) 法律约束

算法决策所造成的隐蔽性就业歧视问题不仅存在认定与规则困难的问题, 而且目前缺乏详细完善的法律约束机制。对此, 应当从两方面进行完善与规制。

首先, 确立算法影响评价规章制度, 针对评价主体、程序、标准等建立符合我国国情与时代特征的管理制度。第一, 针对评估主体, 应分设评价监管主体、评价实施主体, 同时明确各自的职责。第二, 评价程序应凸显人民立场, 保障公众知情权与参与权。第三, 评价标准要具备平等、公平、公正等特征, 突出社会主义核心价值观的引领作用。第四, 评价报告应进行备案、审查与公开, 内容必须向算法决策相关者强制披露。

其次, 依托现有反就业歧视法, 确认算法就业歧视的“劳动歧视形式”概念, 根据认定标准与程序, 将算法就业歧视进行分类与归责。在认定标准上, 应从差别影响、可替代算法决策、正当抗辩理由三个层面展开, 均衡平等就业与用工自由之间的诉求。针对举证困难的问题, 求职者或劳动者只需提供初步证据, 而企业需要提供招聘标准、就业政策等作为自证材料, 或者提出正当合法的抗辩理由。

(三) 政策均衡

法律规制并不能彻底解决算法决策所导致的隐蔽性就业歧视问题, 而这就需要通过政策均衡进行风险补充治理防范。在大数据与人工智能技术支持下, 数字经济已经成为产业升级转型发展的重要途径, 而政策建设必须既要发挥相关技术的应用优势, 又要均衡与协调技术所带来的负面影响。具体来说, 一要坚持推动数字经济发展, 比如可以参考“平台就业歧视责任豁免制度”, 对算法开发者予以就业歧视责任豁免, 以此确保相关技术开发者

能够不断探索与创新。二要缓解算法决策对少数弱势群体带来的负面影响,通过政策引导并推进弱势群体就业工作,为弱势群体创建均等就业机会。

三、算法决策下隐蔽性就业歧视的法律规制

针对算法决策所导致的隐蔽性就业歧视风险,其法律规制应从“过程”与“结果”两个途径分别落实。现存反就业歧视法是依托“结果”视角实施的规制措施,主要对存在就业歧视的企业予以法律责任惩处,具有滞后性特征。但由于劳动者与企业均存在举证困难问题,因此还需要建立“过程”规制机制,在尊重“算法黑箱”的基础上,通过优化“人机交互决策”,缩减企业与劳动者之间产生的信息鸿沟,同时避免企业通过“算法操纵”等方式达成“就业歧视”目的。因此,针对“过程”与“结果”两种规制途径,应通过“设计与协商”“测试与监督”“制裁与举证”三个层面落实。

(一) 设计与协商

“设计与协商”指的是企业针对算法开发环节为劳动者提供个人信息权利与平等就业权利等基本保障,进而在算法技术中渗透“劳资协商”理念。“设计与协商”的具体内容可以从如下几个方面展开分析:

第一,企业应建立“信息政策”,并设置两方面内容。其中“通知政策”包括“数据”与“算法技术”两类内容。在算法开发与设计过程中,其必须依靠对劳动者个人数据的收集,进而形成训练集,以此为算法构建提供数据基础。而企业或开发者在数据收集与应用过程中,必须向劳动者提供告知义务,说明数据收集的类型、内容与预期用途。由于算法决策标准会直接影响劳动者的筛选条件,因此企业必须明确雇佣环节中是否运用了算法技术。“透明度政策”则指的是企业必须向劳动者解释说明应用算法雇佣决策的事实,是对“通知政策”的深入落实。

第二,尊重劳动者的数据权利。《个人信息保护法》中明确规定了劳动者“查阅权”与“更正权”,这就要求劳动者具备了解企业收集其个人数据的权利,当发现数据存在问题或偏见时,也可以提出修正意见,进而降低算法“训练集”中的错误与偏见。

第三,除正当职业事由外,禁止算法用于可识别性雇佣决策活动中。由此可以回应纠纷后企业针对“直接或间接歧视”提出的“抗辩事由”。除“正当职业事由”之外,企业不允许通过识别要素筛选而雇佣劳动者,而“正当职业事由”有着三个必要条件,其一为“工作实质要件”,其二为“全部或者几乎全部要件”,其三为“合理需要要件”。企业在用人过程中,必须确保“不存在其他合适替代方案”的前提,才能使用算法技术识别劳动者与工作岗位之间的相关性特征,并由此作为雇佣决策依据。

第四,确保劳动者在算法决策下的参与权,缩短劳动者与企业之间的信息鸿沟。首先,企业在运用电子监控技术、算法技术等收集劳动者个人数据的过程中,均对劳动者的劳动条件或环境实施了实质性改变,其必须与工会进行协商,以此获取工会“同意”。在该问题中,工会“同意”可以取代劳动者个体的“同意”,本质是为了保护劳动者的劳动权益正确行使。同时,应针对“明确同意”与“默示同意”建立不同的数据收集权限,前者可以涉及“尊严”相关的数据内容,后者不能涉及“尊严”相关数据。其次,企业应面向全体劳动者公示其规章制度,并积极听取劳动者意见。在算法技术正式应用前,企业必须根据劳动者异议进行新一轮审视,确保算法决策不会产生就业歧视风险后再全面应用落实。最后,还应防止劳动者正当行使权力后遭受企业的打击报复。劳动者可以授权工会,由此借助代理人参与起诉,进而通过隐匿真实身份

而避免打击报复。

(二) 测试与监督

第一,应坚持“预防原则”,实施“算法影响评估”规制技术。算法决策可能带来就业歧视风险,而“预防原则”应当具备主导功能,以此通过预警系统建设,提前解决算法决策可能带来的歧视风险。“预防原则”的核心在于要求企业在算法技术应用前,必须对算法技术进行多方测试,并通过“算法影响评估”,判断其算法对劳动者是否存在歧视问题,以此督促开发者增强算法开发责任意识,并能够积极监督算法的运行与应用结果。通过“算法影响评估”,还可以生成对应的评估报告,报告中应通过系统性描述,详细说明企业算法决策中技术应用的“正当职业事由”;通过风险评估,说明算法决策可能存在的安全隐患、危害性质以及严重程度;通过比例评估,说明算法技术的预期目标,说明可能产生就业歧视风险的概率,并评估算法决策应用的必要性。

第二,依托独立机构实施监督作用。独立机构参与监督可以有效缩短信息鸿沟,一方面其可以独立披露企业的算法技术开发信息,另一方面可以增强信息真实性,获取劳动者信任。国家应借鉴该经验,通过行业工会、行业协会等组织建立独立机构,由此披露完整的算法技术开发信息与“算法影响评估报告”。

(三) 制裁与举证

“制裁与举证”是立足“结果”视角的规制措施,主要在于当企业对劳动者造成就业歧视事实后,要求其履行对应的法律责任。针对目前我国立法与司法实践情况,我国应从以下几个方面进行完善。

第一,修订并完善《反就业歧视法》,不仅要适用传统就业与雇佣领域,也要适用当前信息化环境下的多元雇佣模式与体系。具体来说,一要扩大就业歧视行为范围,并针对算法决策引起的就业歧视后果予以惩戒。二要扩大“受保护特征”范围,针对“年龄、户籍、学历、身体特征”等因素的歧视,也要纳入法律规定之中。同时,将“劳动争议”修订为“特殊侵权行为”纠纷,并建立统一的纠纷处理规则,脱离就业歧视行为的“录用阶段”与“实际雇佣阶段”的时间因素限制。三要针对企业故意造成就业歧视问题的情况,实施“补偿和惩罚性赔偿”机制。

第二,减轻劳动者的举证责任。在算法决策下的就业歧视行为不仅具有较高隐蔽性,而且难以提供可靠的证据,因此劳动者只需要提供前期证据,同时执法机关单位可以证明企业所造成的就业歧视行为并不具备“正当职业事由”,便需要其承担歧视责任。

四、结语

综上所述,在“数据驱动”社会环境下,算法技术为社会形态与企业模式带来了颠覆性改革。在企业雇佣与用人环节中,算法决策技术展现出一定的高效化、智能化优势,但同时也带来了隐蔽性就业歧视风险,进而造成侵犯劳动者平等就业权的问题。对此,需要相关部门从技术、法律与政策方面提出风险治理方法,并从“设计与协商”“测试与监督”“制裁与举证”三个层面落实法律规制,以此科学、公正地利用算法与大数据技术。

参考文献:

[1] 黄脂,谭金可.算法就业歧视的法律规制[J].山东工会论坛,2023,29(03):77-84+100.

[2] 赵婵.AI招聘的算法歧视风险与治理之道[J].湘潭大学学报(哲学社会科学版),2023,47(03):96-102.

本文系:“江苏省省级一般创新训练项目,”项目编号为202211287109Y