

欢迎按以下格式引用:徐前权,郭伦.我国智能辅助量刑系统风险防范[J].长江大学学报(社会科学版),2023,46(1):98-102.

我国智能辅助量刑系统风险防范

徐前权¹ 郭伦²

(1.长江大学 人文社科处,湖北 荆州 434023;2.长江大学 法学院,湖北 荆州 434023)

摘要:智能辅助量刑系统的应用对于解决目前的司法困境,推动我国量刑规范化改革有重要作用,但是,在“弱人工智能”的今天,其带来的“算法黑箱”“算法歧视”等机械性风险也需要警惕。基于智能辅助量刑系统的运行原理,可以从明确定位、优化资源、公正保障、强化说理、完善立法五个方面予以完善,实现人工智能与司法裁判的更好衔接。

关键词:人工智能;司法大数据;风险治理;量刑规范化

分类号:D925.2 **文献标识码:**A **文章编号:**1673-1395(2023)01-0098-05

法院信息化建设与量刑规范化改革是当下司法改革的两大任务,加之 AI 技术的逐渐成熟,三者共同促成了人工智能辅助量刑系统的研发推广。所谓人工智能辅助量刑,是指在大数据、云计算的技术支撑下构建法律图谱,通过提取法律文书要素,进行类案识别和模型训练,用机器学习等人工智能方法形成量刑算法,在此基础上实现量刑预测和偏离度测算。^[1]

人工智能辅助量刑是大数据时代司法改革的尝试,与传统的刑事诉讼相比,其核心有两点。第一,以大数据技术为突破口,以人工智能为载体,借以解决人类思维局限性、司法压力大、量刑公信力低等问题,将科技价值与人类情感价值相结合,互补短板,以人工智能辅助量刑推动法院量刑规范化、精细化,最大限度发挥科技的法律价值。第二,通过在量刑程序中引入人工智能,以人工智能量刑的独立性,为量刑程序的相对独立保驾护航,这对于定罪程序与量刑程序的相对分离有极其重要的作用。其本质都是为了更好地实现量刑规范化改革与法院信息化建设。

在司法政策的支持下,各地法院积极将人工智能引入审判体系,涌现出许多人工智能辅助量刑系统,如北京“睿法官”系统、广州“智审辅助量刑裁决系统”、上海“206”系统等,这些系统尽管不尽相同,但大都具有卷宗识别提取、理论量刑建议、实证量刑分析三大底层功能。然而,相较于刑事政策与司法实践的积极主动,学界在对人工智能辅助量刑的应用表示欢迎的同时,更多了一份谨慎。现阶段,人工智能辅助量刑系统正在全国推广应用,其风险还未充分暴露。人工智能辅助量刑系统实现其价值主要依靠逻辑算法与数据库,本文将以二者为出发点,探讨人工智能辅助量刑系统的应用可能造成的风险,进而为规避其风险提出对策。

一、人工智能辅助量刑的风险预判

(一)机器价值驱逐人的价值

马克斯·韦伯曾说过“司法不是自动售货机”,这句法律人耳熟能详的话经常被用来批判司法过程中的机械主义。人工智能辅助量刑系统的应用,固

收稿日期:2022-09-22

基金项目:湖北省人民检察院重点课题“检察机关服务保障乡村振兴战略研究”(HJ2021A06)

第一作者简介:徐前权(1963—),男,湖北仙桃人,教授,主要从事社会法学研究。

通信作者:郭伦(1998—),男,山西运城人,主要从事刑事诉讼法研究,E-mail:2957386086@qq.com。

然对量刑规范化改革有很大益处,但同时我们也应当警惕机器对人的价值的过度排斥,防止量刑规范化改革向机械主义量刑转变。在刑事量刑过程中,我们应当注意到,刑罚的裁量不仅仅是对已有犯罪行为的惩罚和报复,还有对人身危险性、再犯可能性的判断。人工智能可以通过对司法数据库的检索分析得出基本犯罪事实的类同量刑结果,但是对于人身危险性、再犯可能性的判断,人工智能无法全面涵盖和考量,因为后者除了通过对犯罪人已有犯罪行为以及行为后的表现来推断,很多方面还需要法官当面交流、切实感受,即所谓的“察言观色”、个人体会,这里需要的是经验、情感,而不是大数据与算法。一个合格的司法裁判,不仅要立足于案件事实和法律规定硬性因素,更要结合社会伦理、价值判断、社情民意等软性因素进行综合考量,自由心证、排除合理怀疑等制度设计,就是为了缓解机械司法所带来的弊端。^[2]证据裁判规则对法官的每一项决定都作出了证据要求,这似乎为人工智能介入量刑提供了很好的理论支撑,但是,在实践中人的因素不可避免,这就导致了机器与人之间的博弈。

综上,人工智能介入量刑确实会实现类案类判这一价值目的,但是,人工智能的介入不是毫无界限的,不能因为人的因素不可控,就将法官的自由裁量权无限压缩甚至排除在量刑之外。量刑不仅具有法律价值,还具有情感价值,在“弱人工智能”的今天,量刑不可能完全机器化、数据化,机器无“人性”这一特性导致机器无法实现量刑中的情感价值,在考虑罪犯的人身危险性和再犯可能性时,人的因素还是占主导地位。对于法官滥用自由裁量权的问题,解决的关键在于意识培养与监督制约,而不是剥夺权力。即便在“强人工智能”时代,机器可以发挥情感价值,拥有了人的属性,这时候对其进行监督,无非是今天对法官监督的一个轮回。再小的权力都有滥用的风险,我们需要尊重这一现实,做好对权力的监督制约,而不是剥夺法官的自由裁量权。

(二) 过度追求司法效率,忽略司法公正

目前,司法机关普遍存在对司法效率的过度追求,忽略司法公正。以“于欢辱母案”为例,于欢因不堪母亲受辱,持刀刺伤四人,最终导致一人死亡。在整个事件中,不论是警察执法还是法院司法,效率优先都体现得淋漓尽致。警察在执法过程中一直秉持“大事化小,小事化了”的理念,未及时采取相关强制

措施,以致悲剧最终发生。在法院审判过程中,于欢一审被聊城市中级人民法院判处无期徒刑,舆论发酵后,山东省高级人民法院在二审中撤销了一审判决,二审判处于欢有期徒刑五年。聊城中院做出一审判决的理由是,当时的情况下,于欢及其母亲的生命健康权受到的危险性较小,不存在防卫的紧迫性。^[3]事实上,法官还是没有站在被告人当时的地位考虑问题,没有从人性角度出发,没有考虑当时环境下被告人于欢的判断力和控制力。更多的底层逻辑依然是司法效率,对于命案依然是稳妥且迅速办理,缺乏对案件的开创性思考,本质上就是缺乏对司法公正的执着追求,这种效率优先的思维在智能辅助量刑的实践中更需警惕。

效率和公正都是司法所追求的目标,二者不可偏废。在人工智能迅猛发展和司法改革的背景下,全国各地法院都掀起了“智慧法院”建设,大力发展司法人工智能,想要借此打破“案多人少”和“类案别判”的司法困境。司法人工智能的应用无疑有巨大潜力,但是,过于追求解决目前司法困境的功利性目的,势必会造成对司法公正的忽略。利用人工智能辅助法官办案会提高办案效率,但同时也加强了对法官的监督,限制了法官自由裁量权的行使,类案毕竟有别,法官的自由裁量权的行使有利于个案正义的实现,这就需要寻找一个契合点,找准人工智能与法官之间的界限。“司法是维护社会公平正义的最后一道防线”^[4],司法人工智能的发展应当同时关注司法效率与司法公正。我们也应当明白,公正是司法的第一追求,当司法人工智能的运用导致司法效率与司法正义二者冲突时,应当坚持“公正优先,兼顾效率”的原则,^[5]始终把公平正义作为根本目标和价值追求。

(三) 算法歧视冲击量刑公正

算法歧视将会是一种最根本、最彻底、最统一的歧视。^[6]人工智能辅助量刑系统在对犯罪人既存的犯罪事实进行分析的基础上,结合已经存在的类似的犯罪历史数据,通过不掺杂私人感情的、客观公正的算法计算,得出一个最终的量刑结果,这减少了人的主观因素等不确定因素对案件量刑的影响。但是,作为计算基础的数据库是已经存在的,如果数据库存在数据偏差,那么,算法将被迫接受这样的规律,最终导致统一的、大规模的量刑偏差。例如,无业盗窃犯往往更会引起法官的厌恶,在法官自由裁量权范围内可能会遭受更重的刑罚;住所地与法院所在地是否一致,成为法官是否决定适用缓刑的重

要考量因素等。如果一个地区大部分法官都带有这种非法律因素的偏见,那么,人工智能辅助量刑在对司法大数据进行分析时会被迫接受这种原始歧视,并且这种原始歧视将得到系统强化,甚至发展成为统一的、百分之百的司法错误。在法官主导司法的过程中,即便人的种种不确定因素会影响量刑结果的准确性,但不可否认的是,仍然会有一部分法官坚守对量刑公正的追求,这会不断纠偏,并且随着相关监管措施的落实,法官单独量刑的准确性必然会得到提高,只是需要时间来达到这一目标。如果盲目地使用智能辅助量刑系统,可能会导致人工智能对已判案件量刑情节和结果的机械性适用,一旦原始数据出现上述量刑歧视的情况,而人工智能的机械性又注定其不具有实现个案正义的开创性,系统会从司法数据库中寻找类同于“住所地决定缓刑适用”这样的潜规则并指导实践,实践案例又将成为系统学习的“母本”,最终这条潜规则将架空法律。因此,在量刑这样具有价值判断的领域,应当保证法官的主导地位。

(四)算法黑箱影响司法透明化

算法作为人工智能的核心决定着智能化系统的行为,对于多数人(包括司法人员和当事人)来说,算法的隐秘性决定了它是一个黑箱般的存在,人们只知其果,不知其因,尤其是对大数据中类同案件的分析取舍,人们无法得知,或者说,也难以明白其算法逻辑,这对需要接受量刑结果的当事人而言,有悖于司法透明化原则。美国威斯康星州诉卢米斯案是人工智能辅助量刑中关于算法黑箱的典型案列,惩戒署官员制作量刑前调查报告时,利用 COMPAS 风险评估工具对卢米斯进行累犯风险评估,因为 COMPAS 背后的评估算法属于商业秘密,所以惩戒署提供的报告中仅有累犯风险的评估数值。法院在量刑时参考了该评估结果,并部分基于此判处了卢米斯 6 年有期徒刑和 5 年社区监督(extended supervision)。卢米斯主张 COMPAS 以性别作为其量刑分析的考量因素违宪,法院依靠 COMPAS 系统做出的量刑判决侵犯了自己的正当程序权利,但最终威斯康星州最高法院维持了原判。^[7]COMPAS 风险评估工具所使用的算法由于涉及商业秘密是不公开的,过程不公开必然导致最终量刑结果的说理不充分,直接影响司法透明化甚至是司法公正,要想解决算法黑箱,需要从源头入手,相关法学工作者需要介入算法的初始设计阶段,进行监督和学习,算法设计企业也需要做出妥协,目前公开算法是解决该问题最可行的办法。

二、人工智能辅助量刑的防范建议

(一)明确定位

人工智能辅助量刑系统,“辅助”二字就表明了系统的工具属性,其价值是帮助法官更好地进行司法裁判,而不是侵蚀甚至取代法官的独立裁判地位,其所追求的是人工智能和法官都能实现自己对于案件的价值。具体而言,就是将案件的情感价值与法律价值相分离,法官解决案件的情感问题,人工智能解决案件的法律问题,二者之间是配合而不是替代关系,并且法官将一直占据主导地位。这是因为目前的人工智能和人脑差距巨大,人工智能所依靠的核心是算法,机器深度学习是在算法不断更新、数据不断录入的情况下进行的,这是一个十分艰难的过程。而人脑基于天然的优势,每次摄入数据都会带来人们思维上的长进,只有人脑可以应对新情况、新问题,而机器对数据的解读能力是以算法能够覆盖数据为前提的。^[8]因此,人工智能只是技术工具,无法代替人类进行主体性思考。将人工智能引入量刑是为了提升司法效率和推进量刑规范化改革,这与保留法官自由裁量权并不冲突,在现阶段,二者之间有一个契合,即法官占主导地位,人工智能起辅助作用。这样做的考量有二:一是目前的社会观念无法接纳机器深度介入量刑,法官的权力来源于公共授权,法官以事实为依据,以法律为准绳,独立行使司法权,不受任何人与机构的制约,如果人工智能取代法官的主导地位,将人类置于机器人的司法管辖之下,不仅违背了公众对于法官的期待,还有悖于伦理与法理;^[2]二是目前的智能量刑算法模型、机器深度学习等核心技术尚未成熟,暂时还没有深度介入司法裁判全过程的能力。人工智能还远远不能实现量刑过程中的情感价值,诸如在对犯罪人人身危险性、再犯可能性进行考量时,法官可以通过长时间生活、工作形成的经验和人类情感上的共鸣来切实感受,而机器是在对犯罪人既存的犯罪事实和表现综合分析的基础上得出结论,相较于法官的切实感受而言,机器分析存在明显的覆盖瑕疵。因此,现阶段人工智能只可能是法官量刑的一种技术性辅助工具。

人工智能辅助量刑作为一种技术性辅助工具可以很好地发挥提升司法效率的价值,同时,法官的主导地位又可以最大限度地兼顾个案的正义,这对于解决目前的司法困境和推进量刑规范化改革都具有重要意义。为了更好地保障法官的主体地位,防止对人工智能的过度依赖,需要在立法层面对人工智

能辅助量刑的辅助属性进行确定，对此，可以参考国家卫生和计划生育委员会颁布的《人工智能辅助诊断技术管理规范（2017 年版）》，该规范明确规定在医疗活动中，人工智能辅助诊断技术属于临床诊断决策的参考，不能直接作为临床最终诊断的结论，最终的诊断必须经由具备资格的临床医师决定。^[9]此外，需要加强对法官使用人工智能辅助量刑的培训学习，明确相关的操作流程和使用风险，切实发挥好机器辅助法官的科技价值。

（二）优化资源

数据库是人工智能辅助量刑的根基，司法大数据应用的前提是数据客观、真实、准确。要保证量刑结果的客观公正，首先需要保证数据的客观公正。

1. 保证数据客观完整

第一，注重对入库案件的质量审查。对已经入库的案件要做质量筛选，将文书语言不规范、法律适用不合理、量刑失衡的案件剔除，在初期阶段对准备入库的案件也要审查相关的文书语言与量刑内容，保证原始数据的准确可靠。第二，整合数据库，保证数据的全面性。使用智能量刑辅助系统的地方司法机关应做好案件公开工作，整合当地的司法数据库，解决数据库局限的问题。还要扩大司法数据的收集范围，不仅仅是案件的裁判文书收集，还需将讯问笔录、庭审笔录等相关信息收集录入，保证案件信息的完整性。

2. 保证数据精准合理

第一，及时更新数据库。随着经济、社会的不断发展，很多法律条文也在不断更新，以废止掉的法律为依据量刑的案件也就不再适宜作为新案件量刑的参考数据，这就需要定时对司法数据库进行鉴别筛选。新出台的相关法规也需要及时录入，并及时补充以其为依据的案件。对此，需要建立一套完整的筛选更新机制，保证数据库的时效性。第二，加大指导案例的参考价值。经过司法人工智能产出的量刑数据属于“二手数据”，法官对其的控制权并不大，在目前这样一个适用人工智能辅助量刑的初级阶段，限制人工智能辅助量刑产出数据的录入，提高指导性案例的参考价值，保证法官对司法核心数据的控制权，是十分必要的。

3. 保证数据分类科学

数据完整与数据精准是对司法数据的内容优化，科学分类则是对司法数据的系统优化。第一，按照司法裁判数据中的基本信息进行分类，包括但不限于裁判年份、法院层级、地域及文书类型、案由。

第二，细化裁判文书关键内容分类，如故意犯、偶犯、共同犯罪、自首、立功、人身权利、财产权利等。第三，对法官说理部分合理分类，法官说理是法官情感价值在裁判文书中的集中体现，可以按照法官在说理中所引述的法条、原则和裁决关键词进行分类。通过多方面、多层次的科学分类，做到类案检索的快、全、准。

（三）公正保障

算法歧视和算法黑箱是司法人工智能绕不过去的问题，算法的隐秘性和数据库的复杂性是造成二者的原因所在，也是影响量刑公正的原因所在。解决算法歧视的关键在于，保证司法数据库中的数据客观、完整、合理，属于前面优化资源所追求的目标。

对于解决算法黑箱，笔者有两点建议。

1. 完善审查机制，强制算法公开

算法黑箱的本质是算法的隐秘性，解决问题的关键也在于此。在算法构建引入的初期，司法机关就需要做好相关的审查工作，做好对算法模型构建的监督工作，评估其算法的客观性、合理性，会同社会相关专家研讨，并公布相关的审查报告，做好对算法构建的内部监督。人工智能辅助量刑是为了维护和促进司法公正，但算法往往因为涉及商业秘密而无法公布，反过来损害了司法公正，这有悖于司法人工智能的设计初衷。为了维护司法的透明化和司法公正，算法设计的相关企业机构需要做出让步。

2. 构建问责机制，保障权利救济

任何权力都有被滥用的风险，不论是算法设计还是算法应用，都需要构建相关的问责机制。在对算法设计进行监督的同时需要明确相关的责任，不论是设计者还是监督者，都需要对自己行为的结果承担责任。法官对辅助量刑系统的应用也需要受到责任制约，否则，法官将量刑责任推给人工智能辅助量刑系统，那么，群众的权利救济就相当困难。如果二审同样适用人工智能辅助量刑，那么，只要法官在量刑辅助系统里填入相同的信息，一审二审的量刑结果将完全一致，上诉也就失去了意义。为了保障犯罪人得到公正裁决，必须在人工智能辅助量刑的新模式下构建相应的权利救济途径和相关的问责机制。

（四）强化说理

人工智能辅助系统的算法模型拥有案件量刑情节提取分析、文书生成的功能，方便法官量刑的同时也削弱了量刑程序对法官的绝对依靠，法官在摆脱体力劳动的同时也减轻了脑力劳动，这就可能导致

法官量刑责任意识的情化。法官在智能辅助量刑中的情感价值无法得到固定量化之前,法官凭借辅助系统便可完成案件的量刑分析甚至是部分说理,这就会导致法官逃避甚至是无需说理,造成法官办案能力的下降。笔者认为,即便智能辅助量刑可以起到释法的功能,但是说理的功能仍然应当由法官独立完成,并且在智能辅助量刑的背景下得到加强。由机器发挥法律价值,由法官发挥情感价值,都应该在判决书中得到体现。在智能辅助量刑系统应用的初级阶段,强化判决书的释法说理是十分必要的,这不仅是对个案权威的追求,也是对司法数据库原始数据的优化,可以为人工智能深度学习提供良好的基础。具体实践上,可以明确规制法官对判决书中分析说理的独立书写权,限制智能系统对判决书释法说理的自动生成,并且明确法官对量刑结果尤其是说理部分的责任。

(五)完善立法

人工智能辅助量刑是我国正在进行的一项司法智能化改革,目前正处在初级实践阶段,还有很多技术需要完善。在这样的背景下,关于人工智能辅助量刑的立法几乎没有,仅有最高人民检察院、最高人民法院和政府的一些规划书,这就导致了各省都有自己的具体实施计划,其所利用的人工智能辅助量刑系统也不尽相同,并且这些系统缺乏国家的统一管理。智能辅助量刑系统的定位、算法风险、机械性风险都是目前暴露出来的问题,不论是明确定位还是公开算法,都需要国家出台相关法律,在全国进行统一规划,需要明确各省政府采购和使用智能辅助量刑系统的具体事项和限制,明确法官对智能辅助量刑所承担的法律 responsibility,明确智能辅助量刑系统设计厂商所应遵守的具体规定和相应的法律责任,明确智能辅助量刑系统的监管和权利救济途径。在人工智能辅助量刑技术逐渐成熟,系统拥有和使用的

司法权力越来越大的趋势下,立法以明确人工智能辅助量刑的相关内容,能够很好地缓和智能辅助量刑系统与法官和现行相关法律之间的冲突,有利于更好地实现司法的程序正义与实体正义。

三、结语

在大数据技术迅猛发展的背景下,刑事司法智能化是必然趋势,在司法改革的背景下,智能辅助量刑系统是必然结果。智能辅助量刑系统的应用符合目前的司法需求和社会现状,其对法院的量刑规范化改革有时代赋予的契合性和重要性,当前全国各地智能量刑辅助系统的普及应用也表明了国家政策的肯定与支持,在此背景下,警惕智能辅助量刑所蕴含的机械性风险和算法风险并作出相应的完善,对智能辅助量刑系统的良好运行和司法正义的实现具有重要意义。

参考文献:

[1]王禄生.司法大数据与人工智能开发的技术障碍[J].中国法律评论,2018(2).
[2]张卫东.我国人工智能司法裁判问题研究[D].安徽大学,2020.
[3]王康.网络媒介审判研究[D].广西大学,2018.
[4]马小清.刑事和解协议的履行问题研究[D].扬州大学,2019.
[5]“六个并重”全面体现刑事诉讼理念[N].检察日报,2012-10-21(3).
[6]张燕龙.美国的版权犯罪刑罚制度及对我国的启示——兼论人工智能时代我国的量刑规范化改革[J].中国政法大学学报,2019(3).
[7]朱体正.人工智能辅助刑事裁判的不确定性风险及其防范——美国威斯康星州诉卢米斯案的启示[J].浙江社会科学,2018(6).
[8]江隐龙.人工智能可以为司法人员做些什么[N].检察日报,2017-10-31(3).
[9]人工智能辅助治疗技术管理规范(试行)[J].中国药房,2010(16).

责任编辑 叶利荣 E-mail:yelirong@126.com