

# 人工智能辅助量刑的实践回视与理论供给<sup>〔\*〕</sup>

孙道萃

(中国政法大学 国家法律援助研究院,北京 100088)

〔摘要〕人工智能技术法律应用的迅猛发展、司法改革政策红利的强劲输出,为人工智能辅助量刑提供了发展空间。人工智能辅助量刑有独特的实践要素与理论根基,会冲击传统刑事司法观念与运作模式等整套知识体系。现阶段技术水平决定了人工智能是“辅助”量刑的司法交往角色,奠定了理论上的可调和性。量刑规范化改革的本质是提高量刑活动的可预测性,人工智能辅助精准预测量刑功能是理论与应用的融合,通过有序的制度供给,可有效提升量刑规范化的正当性与有效性,为推进量刑规范化注入新的规范潜能与动力。算法及其规则是核心实施要素,应当克服算法认知、应用的正当化以及说理等应用难题,夯实算法运行的正当性根基与规则,遏制不可控的司法异化风险,为公正量刑供给新动能。

〔关键词〕人工智能辅助量刑;理论发展;量刑规范;精准预测;算法正义

DOI:10.3969/j.issn.1002-1698.2023.03.010

在现阶段,人工智能与当代刑事法的深度交互仍存在不少争议、分歧。既为实现司法公正提供新的通道和平台,<sup>〔1〕</sup>也可能引发对技术的过度依赖症或者恐惧症。<sup>〔2〕</sup>实然地看,人工智能的广泛司法应用,需处理价值对立、功能疑问、理论互斥等深层次问题。<sup>〔3〕</sup>人工智能辅助量刑是面向未来的重大司法实践课题。我国正在积极推进的实践探索工作,既取得了积极成效与经验,也面临系列挑战。<sup>〔4〕</sup>回顾我国人工智能辅助量刑的探索,尚需澄清人工智能辅助量刑与传统刑事司法知识的调试、价值取舍、功能设定等重大理论问题,也更需明确人工智能辅助量刑的司法风险、

知识体系、基本原理、实施机制,并输出更加实用的产品端供给等。借此,不断深化人工智能辅助量刑的基本法理与固定司法应用的核心逻辑。

## 一、人工智能辅助量刑的发展供需

在最高司法机关积极部署与推进下,释放了重大政策红利,人工智能辅助量刑呈现出迅猛发展的实践态势,其作为辅助办案工具备受期待,也开辟了全新的量刑改革时代。人工智能辅助量刑系统的探索面临无限机遇与未知风险的混合交错,既为量刑规范化等司法改革提供了发展的契机,也存在一些问题和不足。

作者简介:孙道萃,中国政法大学国家法律援助研究院副教授,中国刑法学研究会副秘书长。

〔\*〕本文系中国政法大学“青年拔尖人才”(2023-2025)资助项目的阶段性研究成果。

### (一)前所未有的改革政策红利

在日益蓬勃的智慧法治建设进程中,智慧(智能)司法是重要的实施内容。<sup>[5]</sup>人工智能在司法改革中的地位和作用显著提升,特别是对刑事诉讼的深度渗透与嵌入越发全面和系统。在刑事法领域,智能化应用快速推进,在刑事司法智能化的专业性、具体性上迈出了坚实的步伐,甚至在事实认定等传统刑事司法中最“坚硬”的环节也有突破和进展。<sup>[6]</sup>人工智能辅助量刑(又称“智能量刑”“人工智能辅助量刑系统”等,下同)应运而生。《最高人民法院关于深化人民法院司法体制综合配套改革的意见——人民法院第五个五年改革纲要(2019—2023)》(2019年)明确要求加强智能辅助办案系统建设,建设智能辅助审判系统,完善类案推送、结果比对、数据分析、瑕疵提示等功能,促进裁判尺度的统一。最高人民法院院长周强在“世界互联网大会智慧法院暨网络法治论坛”(2019年7月)上指出,研发与应用各类人工智能办案辅助平台,利用大数据实现案件的标准化认定,有助于推进司法公正与司法的现代化。我国审判机关正在加快智能技术推动下的全流程、全场景与集约化裁判尝试,<sup>[7]</sup>量刑的智能化推进已纳入议程。2020年1月,最高人民检察院正式上线试点运行全国检察机关统一业务应用系统2.0版,智能辅助量刑系统是重要功能模块。<sup>[8]</sup>按照最高司法机关的部署,运用大数据、人工智能等先进技术辅助量刑,才能真正实现量刑的智能化与现代化,深入推进量刑规范化工作,确保量刑的公正。<sup>[9]</sup>最高人民法院发布《人民法院在线诉讼规则》(法释[2021]12号,以下简称《规则》),宣告迈向在线诉讼2.0时代。<sup>[10]</sup>随后,最高人民法院发布《人民法院在线运行规则》(法发[2022]8号,以下简称《运行规则》),在世界范围内首次构建全方位、系统化的互联网司法规则体系。<sup>[11]</sup>而且,诸如积极构建“全域数字法院”的重大设想与实践也在逐步实现过程中。<sup>[12]</sup>但是,也存在通用性过重而專屬性应用不足等结构性问题,数字司法在

公正与效率上的兼顾不周,甚至可以完全由类案系统来替代人工智能辅助量刑的独立地位与功能等问题。<sup>[13]</sup>《最高人民法院关于规范和加强人工智能司法应用的意见》(法发[2022]33号,以下简称《意见》)在基本原则、加强应用、系统建设、综合保障等方面作了规定。

与此同时,《刑事诉讼法》(2018年修订)第15条正式确立了认罪认罚从宽制度。这宣告建立起有中国特色的认罪认罚案件控辩量刑从宽协商机制,同时也对量刑规范化改革提出了更高的质量要求。2019年4月,全国检察机关贯彻落实认罪认罚从宽制度电视电话会议强调,各级检察机关应加强量刑规范化建设,进一步细化常见罪名的量刑标准。<sup>[14]</sup>同月,全国检察机关“量刑建议精准化、规范化、智能化”网络培训会议强调,深入加快推进量刑建议工作。充分发挥大数据智能辅助系统的作用,适时组织研发可以普遍适用的量刑建议辅助系统,有效提升量刑建议的精准度。<sup>[15]</sup>最高司法机关已经充分意识到智能量刑的发展契机及其对认罪认罚案件的积极意义。以量刑的一般原理及其规定为前提,立足司法大数据及其潜能,充分整合司法大数据与人工智能技术,可以实现智能预测量刑,提升量刑建议的精准度。<sup>[16]</sup>未来的量刑活动,需要大数据智能辅助量刑办案系统的支撑。人工智能辅助量刑系统整合了现代信息技术、大数据技术等,为传统量刑规范化改革注入以“智能(精准)预测”为核心功能的全新赋值。加快量刑智能辅助系统研发和推广,是提升量刑建议的客观性、精准度和公信度的必由之路。<sup>[17]</sup>此后,疫情下的在线诉讼制度得到快速发展。全面推动构建中国特色、世界领先的互联网司法模式,对人工智能辅助量刑的政策支持与保障力度更是空前的。

### (二)多元化的探索动态

积极推动人工智能辅助量刑是最高司法机关正在密切关注的重要战略部署。正在如火如荼进行的人工智能辅助量刑探索,立体地勾勒了这场盛宴的境况。

1. 相对专门或者特定的辅助应用系统。各地先后主要有以下做法:(1)智审辅助量刑裁决系统。2014年起,广州市中院建立智审辅助量刑裁决系统,解决盗窃、抢劫类案件数量大、量刑尺度不统一的问题。法官输入案件要素后,系统自动推送相似案例,自动进行比对和运算,自动推送量刑幅度,并辅以图形分析、数据参照等形式。<sup>[18]</sup>(2)量刑建议辅助分析系统。贵州省检察院运用算法模型,建立案件“要素—证据—量刑”的关联模型,创建案件智能研判系统,对案件风险、要素与量刑的偏离度等,进行全流程的数字“画像”、数据分析和类案数据对比,为案件质量评查提供精准依据。<sup>[19]</sup>杭州市检察机关研发运行危险驾驶案办案辅助系统。它具备自动提取、智能关联、量刑研判等功能,依据系统预置的量刑计算规则,结合以往同类案件判决情况,自动计算、智能对比、自动输出量刑建议,可视化展现计算的过程,辅助准确量刑。<sup>[20]</sup>2018年,杭州市西湖区人民检察院发布《认罪认罚案件审查起诉规则》。对犯罪情节和刑期较难量化的罪名,充分利用大数据为量刑建议提供技术支持和参考。<sup>[21]</sup>(3)智能研判系统暨“睿法官”。2016年底,北京市高级人民法院推出智能研判系统“睿法官”。<sup>[22]</sup>它依托北京市法院统一的审判信息资源库、法律规则库和语义分析模型,运用大数据与云计算等技术,充分挖掘、分析数据资源,自动推送案情分析、法律条款、相似案例、判决参考等信息,为法官裁判案件提供统一、全面的规范指引。(4)量刑规范化智能辅助办案系统。2017年,海南高院成功研发量刑规范化智能辅助办案系统,具有智能识别提取犯罪事实和量刑情节、自动推送关联法条和类案、自动依据历史量刑数据推荐量刑、自动生成程序性法律文书和框架性裁判文书、多维数据统计等功能,以及深度学习与自主优化能力。<sup>[23]</sup>量刑预测的试点效果良好。<sup>[24]</sup>

2. 以类案智能推送系统为主的辅助办案系统。2017年,上海研发并试运行刑事案件智能

辅助办案系统,具有量刑参考功能和类案推送功能。可以根据案件的事实、情节等,通过语音识别,经过大数据处理输出量刑建议,智能推送全市案件数据库中与本站证据相似的案件。<sup>[25]</sup>2017年,最高人民法院推出“类案智能推送系统”,实现精准和有效推送类案,并可以作为定罪量刑的重要参考。2017年底,最高人民法院正式启动开发建设量刑规范化智能辅助办案系统工作,构建完善的全国法院量刑智能辅助系统,辅助法官办案。<sup>[26]</sup>2018年,最高人民法院正式运行“类案智能推送系统”。<sup>[27]</sup>该系统以海量的司法大数据资源为基础,构建司法案件知识图谱与案例智能服务产品。该系统作为类案推送系统的“智慧大脑”,可实现案件的智能认知和分析,使类案剖析更专业、类案识别更智能、推送结果更精准。中国司法大数据服务网作为最高人民法院指定的官方平台,目前已全面上线“类案智能推送”功能。

3. 最高司法机关的官方做法。主要有:(1)全国检察机关统一业务应用系统2.0版。2019年12月,最高人民检察院召开有关统一业务应用系统2.0版的智慧检务座谈会,旨在着力提升流程管理、智能辅助、数据应用、知识服务,初步实现为检察办案提供辅助量刑等功能。<sup>[28]</sup>2020年1月,统一业务应用系统2.0版正式试用,扩大了智能辅助量刑探索的覆盖面。而且,《人民检察院办理认罪认罚案件开展量刑建议工作的指导意见》(2021年)第20条规定,人民检察院可以借助量刑智能辅助系统分析案件、计算量刑,并依法提出量刑建议。(2)最高人民法院有关部门研发人民法院量刑智能辅助系统。该系统具有智能导入案件信息、智能提取量刑情节、智能辅助量刑并生成量刑表格等功能,可以智能地辅助法官进行量刑。目前,该研发系统正在部分法院进行试点,适时将在全国法院推广应用。<sup>[29]</sup>特别是在《运行规则》的第二章“系统建设”中,还着重规定了智慧法院信息系统的建设要素与构成等。其中,第6条规定,智慧审判系

统要为审判人员提供阅卷裁判辅助等在线服务,智慧审判系统包括智能裁判辅助系统、量刑规范化系统等,智慧审判系统应当具备类案智推、量刑辅助等功能。这些都是人工智能辅助量刑的具体实践,并且进入了全面建设阶段。同时,《意见》也对人工智能辅助量刑的基本原则、系统研发、安全保障等作了规定。

4. 民间市场的产品建设也在加速。例如,“小包公”智能系统。以刑事定罪与量刑预测为核心的办案辅助系统,可以提供精准量刑预测及说理过程,具有国内首创实用量刑协商工具等优势。理论量刑预测系统提供理论预测,实际量刑分析系统依托大数据提供案件的可视化分析,双系统相互印证,最大程度实现精准量刑。该系统在检察机关的试用效果良好。<sup>[30]</sup>

人工智能辅助量刑系统在研发与应用方面取得了非常喜人的成绩。这些大胆的积极探索,不仅初步搭建起我国智能量刑的实践模式,累积了丰富的经验理性;也通过“百家争鸣”的方式,实现了“优胜劣汰”的良性竞争效应,还为深度透视智能量刑模式的基本理论与展望其未来的发展趋势,提供了有益的认识前提。

### (三)人工智能辅助量刑的共性归结

最高司法机关联动人工智能与辅助量刑的政策改革导向,使人工智能辅助量刑迎来迅猛发展的历史契机。不同的人工智能辅助量刑应用系统在内容与形式上虽有一些差异,但核心功能与目的基本一致。这折射出探索与试点的积极成效。

1. 系统内容与基础。主要包括:(1)主要表现为智能推送类案、量刑建议参考等形式。目前较为普遍的做法是,通过对司法大数据进行大数据提纯或智能化分析,建立类案检索、类案推送等机制,或对已有判决中的量刑规律进行整理、归纳和总结规律,并将其投放到相同或相似案件中,用于预测个案的量刑“估值”,实现“同案同判、类案类判”效果。(2)司法大数据是重要基础。从技术可行性看,可以在短时期内实现智能

辅助量刑工作的跨越式发展,主要得益于司法大数据所积累的海量公开案件信息。司法大数据蕴含“活着的”量刑规律与经验,最接近实际量刑生态。对司法大数据的深度智能分析,为智能量刑探索提供了海量的“粮草”。通过向智能系统输入“活着的”量刑逻辑,可以生成、分析与计算相关变量、指标体系等要素,与个案进行精准匹配后,智能输出“预测”的量刑建议。但数据库的不完整、智能处理能力等相对不足,使类案量刑的推送、量刑建议参考等成为主要的工作方式,导致在“(精准)预测”上的着力明显不足。

2. 功能与优势。表现为:(1)达致量刑规范化是终极目的。推进智能司法的初衷是为了更好地重塑量刑公正理念。智能量刑系统与量刑规范化在目标上殊途同归,前者以后者为基本前提和基础,而不是脱离或者互斥。从人工智能辅助量刑系统的操作要求和基本功能等内容来看,在方法上可以修复量刑规范化不足的问题,纠偏传统经验式量刑估堆等做法。通过结合量刑原理与司法大数据中的量刑规律,遵循演绎法、归纳法的司法逻辑,可以对个案进行智能精准预测,相互验证并自动输出最佳的量刑建议,更好地实现“同案同判、类案类判”目的。(2)基于“算法”的智能精准预测量刑是本质特征与优势。任何具体的量刑活动都旨在获得最公正的量刑结论。对所有已决案件进行“穷尽”参照是最好的保证。智能量刑探索完美地提供可靠的预测方式,以司法大数据等为前提,按照“算法”运行规则,根据已有案件的量刑趋势或规律等,对新出现的类似案件或相同案件进行匹配、核查与验证,能够实现“智能精准预测”。这是最核心的特质与功能,不再是经验式的量刑做法或基于人类智力的一般性量刑模式等。“智能预测”正在取代经验、智力以及其他规范化方法,可以最大限度地通过智能技术实现量刑预测的“最适当性”。算法及其司法化的优势,是最关键的支撑点与制度保障。

3. 应用维度。包括:(1)认罪认罚案件是最

迫切的应用场域。认罪认罚案件已经成为我国刑事诉讼的主要办案对象,针对认罪认罚案件的专属办案机制正在形成。在认罪认罚案件中,自愿认罪认罚使定罪问题基本解决或无争议,量刑从宽协商成为办案核心。量刑从宽协商是从宽处理的决定性因素,量刑建议的提出与采纳是决定公诉权行使效果与办案质量的终极指标,倒逼检察机关提出精准的量刑建议,提高从宽处罚的可预期性。<sup>[31]</sup>认罪认罚案件在整体上是典型类案,具备实现类型化办理的条件。通过充分挖掘司法大数据蕴含的“活着的”量刑经验与规律,结合量刑规范化的基本原则与实施细则,不仅可以智能预测认罪认罚案件量刑,而且可以高效精准预测。(2)角色定位目前主要是辅助办案系统。受限于智能技术及其司法应用的水平,人工智能辅助量刑系统仍不够成熟、可靠,传统量刑理论体系仍发挥重要作用。人工智能辅助量刑系统主要定位为辅助办案系统,旨在辅助司法人员实现更加智能预测量刑,使量刑结论更精准,并缩小量刑失衡问题。法官、检察官是绝对的核心使用群体,律师也是重要的使用主体。

智能量刑探索正在如火如荼地进行中,为量刑规范的智能化改革提供了诸多新期许,为认罪认罚案件的精准量刑开辟了新天地,也将人工智能辅助量刑的可行性与可信赖性推向了新的阶段。但是,日益拓展的人工智能量刑应用系统探索也开始进入制度供给的瓶颈期,包括调整与传统量刑模式的不适并谋求共识性的达成,明确算法的地位、作用并遏制其运行的司法风险,妥善处理与量刑规范化的结构协同关系,充分发挥辅助功能等重大问题。应当启动反思性的理论解构,对智能量刑系统予以结构性的完善,从法理层面予以升华,从制度供给上予以巩固。

## 二、人工智能辅助量刑的理论廓清

刑事司法智能化改革必须以促进与推动司法体制改革为基本导向,否则,扭曲、阻碍甚至消解司法体制改革的负功能可能随之而来。人工

智能辅助量刑系统对传统司法体制的冲击是显而易见的。在新旧刑事司法知识的交替过程中,二者应当积极融合,更好地发挥人工智能辅助量刑系统应有的功能期许与司法意义。

### (一)与传统量刑的理论交锋界面

当前,智能量刑之于传统司法观念颇为“离经叛道”。但从量刑规范化改革的终极需求看,却赋予了丰富的可能性与技术优势上的期待。这种价值悖论及其衍生的关联问题,是当前人工智能辅助量刑系统需要直面的冲突与发展困境。

1. 司法基础的动摇。具体而言:(1)司法权的争夺。在近现代刑法中,司法权是国家权力,具有高度的垄断性、专属性、法定性以及权威性。司法权只能由国家行使,并只有符合法律规定的司法人员才能行使司法权。量刑权是司法裁判权的具体内容,只能由“人”来行使。如若转由“智能系统(主体)”行使,无疑是对司法权配置的重大冲击。而司法权的争夺首当其冲,“司法人员”与智能系统主体性的地位对峙是无法避免且难以调和的。(2)司法权威的旁落。司法活动的法定性、专属性以及程序化,是司法权威的重要形成基础和维护力量,也是形成司法信任的重要基础,更是司法专业化被反复强调的根源所在。在“算法”支撑下,智能量刑系统取而代之,同时导致量刑的公开性、透明性以及可视化、可沟通等正当程序的优点相继流失,量刑活动所应建立的司法信任以及司法权威难有“土壤”。这折射了智能量刑系统的接受度与有效性都是各界担忧的焦点所在,势必无法在短时期内获得当事人和社会公众的广泛认同。(3)司法主体地位的割据。司法人员依法主导量刑活动是现代刑事司法的基本特征。在智能量刑系统中,基于大数据的统计分析并由智能系统自动生成的意见、自主矫正量刑偏差等做法,导致审判主体的结构变成二元化,甚至多元化。智能量刑系统的程序研发与设计人员、司法数据的处理商等,也在实质上与法官共同作出决定。一旦出现司法主体的割据状态,审判主体和决定者难以特定

化,司法权力的边界也就变得模糊不清。当出现分歧或需要作出最终决定时,法官的独占性地位不复存在,司法责任制也恐容易流产。(4)司法责任制的流产危机。智能技术的司法应用可能会导致司法人员将责任转嫁给人工智能辅助量刑系统。法官不能将案件责任转嫁给智能系统,以规避司法责任制。同时,必须正面解决智能量刑系统因技术缺陷等而引发的司法责任归属问题。这些问题的交错出现,也使说服法官相信并主动使用智能量刑系统处于“前后矛盾”的尴尬状态,智能技术伦理的审慎介入显得尤为迫切。

2. 司法规则的受挫。主要为:(1)司法裁判规律的逆反。司法规律是司法权运行的科学之道。为了确保量刑的公正,与之相配套的司法裁判准则是前提,主要包括量刑主体的专属性、量刑依据的法定性、量刑方法的规范化、量刑结果的正当性等。然而,智能量刑系统导致司法权的行使、司法主体的取舍等均发生变化,使传统司法规律受到直接影响。为了确保量刑活动的稳定运行,现阶段仍应坚持人工智能辅助刑事裁判系统的工具性价值,不可断然抛下传统司法裁判的本质规律与法官的职业裁断。新旧司法规律的有序衔接,也可以更好地发挥智能量刑系统的“辅助”作用,但对智能量刑探索的深入挖掘也形成一定的外部拘束力。(2)司法量刑技艺的丧失风险。司法人员的裁判活动具有专业性。长期积累和总结,汇集成特定的司法技艺。量刑活动有其专属的技艺体系,有自成一体的司法知识系统,具有高度的专属性、专业性等特征。智能量刑系统不仅打破司法技艺的形成、传承以及发展模式,也影响传统司法技艺的生存境遇。例如,智能系统虽然可以自主生成参考性文本,但是法官的审阅、矫正等工作未变,在案多人少、审理期限短等办案压力下,智能量刑系统的辅助功能及其优化,可能助长司法人员的惰性,甚至异化为过度依赖智能系统办案的“去人中心化”倾向。

3. 传统定罪活动也会有所“殃及”。传统刑事司法中的定罪与量刑高度结合,甚至融为一

体。从司法的时间逻辑看,先有定罪,后才有量刑。当前,人工智能辅助定罪的探索需突破重大的理论禁区与认识难题,更面临司法公正沦陷的公众隐忧。当前,智能量刑系统的探索走在智能定罪系统的前面。但是,智能量刑系统的深度发展与应用,会对传统定罪体系及理论学说产生“逆向”的渗透效应。

## (二)人工智能辅助量刑的理论赋新

人工智能辅助量刑对传统量刑理论形成了不同程度的冲击,也开始形成相对独立的知识逻辑体系。同时,这场渐进的司法渗透亦预示着无限的理论发展可能。

1. 推动量刑理论发展的特定意义。具体来说:(1)局部消解“人的智力”在量刑环节的短板效应。人类具有“智力”这一尤为特殊的生理结构因素,使其拥有主体性地位。司法人员的“大脑”经过“专业化”锤炼,可以担当并处理纷繁复杂的司法活动与量刑任务。但“人的智力”仍存在局限性,包括智力水平的相对发展性、不均衡性等问题,使人类不足以充分与彻底认识和解决量刑规范化的本质与形式等问题。而人工智能技术及其应用被赋予诸多的厚望,智能优势及其对刑事司法活动的积极意义被不断放大。特别是在量刑规范化领域,智能技术比“人的智力”更具有优势,可以填补“人的智力”在刑事司法领域的短板。当前,人工智能技术及其应用对传统司法的积极助推作用,促成了各国法院司法体系从专家审判司法体系向智能人工司法体系的“大转型”趋势。<sup>[32]</sup>(2)有效降低司法人员统领量刑活动所形成与依赖的司法经验之非理性成分。司法经验是非常重要的司法“资产”,是司法活动的重要前提和基础。司法经验的形成高度依附于司法人员,甚至可以认为是“人的智力”在司法领域的结晶。但是,司法经验的非理性因素及其所暴露的问题常被诟病。从智能技术的特征看,智能技术在量刑领域的辅助应用,可以降低司法经验累积的非理性因素,增加更加客观中立的理性因素。(3)将传统量刑说理切

换为“算法运行”及其智能化优势。量刑的科学化与规范化,主要是通过量刑说理予以实现和论证的。从传统的量刑说理方式看,主要存在说理方式较为单一、说理依据不充分、说理的精细化不足、说理的比较性欠缺等问题。在智能技术的支持下,基于司法大数据及其“活着的”量刑经验与规律,可以为量刑提供更便捷、高效且强力的说理论据、方式等。智能量刑系统依托司法大数据,可以充分评估已决案件的情况,从纵横两个维度进行动态对比,为实现“同案同判、类案类判”,提供相对更可靠、全面、系统以及权威的说理支撑。(4)减轻司法人员的非核心量刑工作并提高效率。当前,司法改革的聚焦点仍是“人”。但是,司法人员承担大量的非审判性等事务,不利于专注于规则适用等核心审判工作。在传统司法环境下,司法人员的量刑活动异常复杂,不仅要应对外部因素,还要应对内部因素。内部因素不仅限于量刑本身,还包括大量的事务性工作、程式性要求、内部考核指标等。为了充分解放司法人员的生产力,使其重心落在量刑等司法主体活动,智能技术的导入便具有非常积极的“解压”作用,促成科技理性与法律理性的深度融合,可显著提高审判效率。

2. 从量刑技术与方法上加速量刑规范化改革的理论升级。量刑规范化改革旨在消除传统量刑中的外部干扰大、量刑说理不够精细、量刑情节的规范评价体系不科学等弊端。从方法论看,在反思经验主义等做法之际,当前仍未实现“质变”的效果。通过逐步导入智能技术后,智能量刑系统的探索与推行,从方法论上实现了质的飞跃。以“高度智能化”为逻辑起点,非常贴切地回答了量刑规范化的目标与要求,也从方法论上恰如其分地予以贯彻和落实。智能量刑系统被赋予使量刑活动具有更高“科技含量”的预期功能优势,量刑结论也被认为更具有个别公正性与可接受性。在司法功利主义的鉴别下,人工智能辅助量刑系统是可能超越传统量刑活动的弊端的,故应主动回应与释放外界期待的司法正能量。

### (三)司法辅助角色的理论定位

传统司法与量刑活动的惯习,以及人工智能辅助量刑系统在专属性与独立性上的薄弱一面,必然决定了人工智能辅助量刑在现阶段扮演辅助角色。这一司法功能的定位,有助于合理释放人工智能辅助量刑与传统量刑理论的可调和性。

1. “辅助”功能的理论适宜性。《意见》坚持“辅助审判原则”。人工智能是审判工作的辅助,不应削弱用户自主决策权。主要的内因包括:(1)竭力吻合刑事司法智能化的公众认同度。刑事司法智能化的进程与效果,应与现阶段的司法认同水平保持一致。人工智能尚未在法律核心领域显示全面超越人类的应用能力。人工智能适宜作为法官裁判的助手而非完全替代法官。现阶段的刑事司法智能化,应定位于辅助法律人决策的助手与“参谋”角色,而非直接取代。这既由人类(法律人与公众)对法律职业的认知决定,也由人工智能技术的不足决定。当下,应当坚持法官的主体性,确认人工智能辅助量刑的工具性,旨在防止出现过度依赖智能工具后的“锚定”效应。智能量刑系统可以与法官量刑的主导性、法定性功能相辅相成,但也不应低估人工智能辅助量刑系统潜在的重大司法意义和正义价值。(2)激活辅助量刑意义应循序渐进。人工智能技术及其应用的深度演变,可以建立高度“类人”的决策机制,<sup>[33]</sup>可以“和人一样”,对复杂场合作出合理决策。在量刑领域,基于“算法”的强大功能,可以高度“模拟”司法人员,依法预测量刑,通过技术优势,使量刑结论更具有合理性、科学性以及参考价值。进言之,智能技术通过大数据分析、数字运算及系统化分析等优势,可以建立更可靠的预测性能与指引效果,为法官持续提供相对“客观”的比照标准与预测结果,对疑难复杂刑事案件的量刑具有积极的意义。人工智能辅助量刑系统在预测上具有非常突出的精准优势,但是,并不能据此直接否认传统量刑规范化体系的合理性及其优势。从功能预设看,智能量刑系统暂时尚不具备直接取

代法官地位的能力,而是努力“思考得像一个法官”,并作出分析、决策、判断。<sup>[34]</sup>法官也不能完全依赖人工智能技术,沦为智能办案辅助系统的附属。此外,司法大数据等外部支撑要素的不足,使刑事司法智能化无法充分展开,它主要扮演辅助性角色而非司法实践的主体角色,完全取代法官不切实际。(3)全流程促进司法改革效能应当被有序伸张。目前,各地探索与研发的人工智能辅助量刑办案系统,既有从全流程推进智能司法办案的角度进行定位,也有专门从量刑预测系统的个别性需求出发。量刑是刑事诉讼活动的终端与归属,人工智能辅助量刑可以注入精准预测的能量。作为新型智能化办案平台,不仅可以释放技术推动量刑规范化改革的潜能,还可以结合司法大数据的丰富内涵,显著提高量刑的精准性与公正性。人工智能辅助量刑是推动以审判为中心的刑事诉讼制度改革落地的有益尝试与重要辅助措施,是刑事司法智能化改革进程的开拓者。不能片面地认为,人工智能裁判的运行逻辑和审判中心主义的内在机理存在显著冲突,从而限制人工智能裁判及其在量刑领域的适用。审判中心主义及其诉讼制度改革是全方面的,并积极追求司法公正与诉讼效率的融合。人工智能辅助量刑可很好地提升量刑规范化质量并提高量刑精准性,其关键是完善配套措施。司法大数据的不断增量与积累,人工智能辅助量刑的实证依据、规律等支撑要素变得更强大,预测的量刑结果将更精准,对刑事司法智能化的贡献将不断释放。智能技术及其应用是进步的,人工智能辅助量刑系统的预测功能将不断升级和完善,辅助的角色及其内涵也是发展的。

2. 适时有序扩充理论演绎的场域。公权力的行使遵循公开性与透明性,而算法具有相当的“黑箱性”。算法的形成、运行规则、运行过程与内容等,目前不便于公开或公示,不容易被外界认知。这就天然地与公权力行使的透明性、刑事司法及其规则的公开性相悖,决定了刑事领域的审慎应用与推广之理性选择。例如,公诉办案辅

助系统和量刑建议分析系统主要适用于案情简单、证据标准较明晰的轻刑案件,如危险驾驶、“两抢一盗”、赌博等。<sup>[35]</sup>又如,对于人身危险性的量刑意义之识别、判断以及智能量化等,仍有很大的进步空间。尽管智能量刑系统适用的罪名仍未全覆盖,但对占比极高的认罪认罚案件,可以优先探索并实现智能量刑系统的一般性应用。《刑事诉讼法》第15条实质上已经将认罪认罚案件与不认罪认罚案件的程序繁简分流规范化、正式化。在全面实施阶段,认罪认罚案件的数量占比不断攀升,成为司法机关最主要的办案对象。在自愿认罪认罚的前提下,定罪基本无争议或已解决,并主要通过程序予以确认“有罪”,使量刑从宽协商、量刑建议的提出、量刑建议的精准等成为重中之重。科学开展量刑从宽协商、精准提出量刑建议并使其具有正当性,成为认罪认罚案件质量考核的核心标准。人工智能辅助量刑的探索与适用,成为认罪认罚办案机制的新生力量。<sup>[36]</sup>相比于不认罪认罚案件定罪量刑的“坚硬性”等法定特征,研发与适用针对认罪认罚案件的人工智能辅助量刑已成为可以优先探索的具体应用问题。

### 三、人工智能辅助量刑的规范理路

规范量刑几乎成为我国近些年量刑改革的核心旨趣。量刑改革的规范化导向,不仅成为量刑科学化的动力,也为智能量刑系统的探索孕育了土壤。人工智能量刑系统应当以量刑规范化的基本原理为前提和基础,充分发挥司法大数据的量刑预测意义,提高量刑预测的精准性,在功能与制度上实现深度升级与优化。

#### (一) 与量刑规范化改革的深度契合

《人民法院量刑指导意见(试行)》《人民法院量刑程序指导意见(试行)》《关于规范量刑程序若干问题的意见(试行)》以及《关于规范量刑程序若干问题的意见》(法发〔2020〕38号)相继出台,标志着我国量刑规范化改革进入新时期。持续修订《关于常见犯罪的量刑指导意见》



(2021年)使规范化改革持续推进。《最高人民法院司法责任制实施意见(试行)》(2017年)、最高人民法院《进一步加强最高人民法院审判监督管理工作的意见(试行)》(2019年)等相继出台,既明确要求法官审理案件应全面检索类案和关联案并制作检索报告,加强统一司法裁判及量刑标准的工作,也优化了量刑的程序。这些都有助于纠正量刑不均衡与规范量刑活动,然而量刑规范化改革的现实阻力却不减。与大数据技术的对接虽然提升了动能,但是实践中的问题仍然不少。

目前,处于“深水区”的量刑规范化改革,其理论与司法痛点集中表现为:(1)量刑可预测性的司法改革目标不达标。回顾迅猛推进的量刑规范化这项重大司法改革,虽然取得了有效转变“重定罪而轻量刑”观念等成绩,但个案裁量的规范化、标准统一化等僵局性问题依旧存在。特别是传统“估堆”量刑等方法仍然根深蒂固,经验量刑的不确定性与不统一性、量刑标准的随意性与不统一性、基准刑的认定过于僵化等问题依旧存在。然而,公认的是,提高量刑可预测性是量刑规范化改革的目标,促使公民自愿遵循和认同法律。<sup>[37]</sup>所谓“量刑的可预测性”,是指对诉讼各方及社会公众而言,经过量刑规范化及其程序得出的最终宣告刑,与各方对刑罚的比较与估量具有“同一性”或“相似性”,属于一般可预测或可接受的范围。实践证明,司法结果的可预测性,是司法正义的重要内容。量刑规范化以提高量刑的可预测性作为目标,旨在改变量刑的标准不统一,破解“同案不同判、类案不类判”等问题。目前,量刑规范化改革对量刑的可预测性问题着力不够,尚未进入“深水区”。最高司法机关之所以力推强制类案检索报告制度,就是为了提高量刑的可预测性。(2)司法大数据与类案推送制度的粗糙化。大数据技术“进入”量刑领域,主要是借助司法大数据蕴含的量刑经验,最大程度减少个体司法人员决策所产生的量刑误差以提高量刑质量。立足司法大数据,对裁判文

书的海量汇集与类型化分析,通过“类案推送”等工作方式,抽取实践中的量刑规律与规则,可以消解量刑经验主义的样本不足、类型不充分、规律不完整等问题,并稀释以司法人员为核心的经验主义模式及其不足。但是,“类案推送”等类案类判系统相对粗糙和简单,主要是建立更全面的司法数据系统与案件数据库,是对司法数据进行更大规模的类型化整合。虽然优化了司法数据分析能力,但是预测量刑建议的精准性相对有限,对“同案同判、类案类判”的促进效果尚不明显。此外,检索推送案例的精准性不足、案例的关联性不高、案例的来源及层级不充分、类案类判的实践差异显著等问题相互叠加。相比之下,对司法大数据的全面分析、量刑结果的智能预测等做法,能够达到更直接的效果,超越了案例大数据库建设、案件检索及类案推送等粗糙的数据分析与加工阶段。

#### (二)“精准预测”量刑功能的整体融入

传统量刑与智能量刑两种模式,在定罪与量刑的关系与安排、刑法规范的适用模式等核心内容上有所差别。但是,从量刑逻辑看,都遵循演绎与归纳的方法论。通过“不断往返一般经验与个案需要之间”的方式,尽最大可能针对个案预测量刑建议,最终实现公正的量刑。传统司法人员预测量刑与智能预测量刑虽存在运行模式的差异,但并未彻底变成完全对立的两种模式,而是相互依存并嵌入使用的。这是现阶段的实际样态。预测是一切量刑活动的本质逻辑与属性,精准预测是智能量刑系统的核心司法价值。这是对量刑规范化改革的最大积极推力。

1. 个体经验性预测向算法指引智能预测的转向。尽管传统量刑模式与人工智能量刑模式存在激烈的“对冲”现象,但二者都是为了实现量刑公正。在核心的实施要素上,都是根据量刑原理与量刑经验,作出最符合个案的“预测”量刑意见,实现量刑规范化的基本目标。在量刑活动的“预测”之基本特征下,人工智能辅助“预测”的功能效果更值得期待。其根据为:(1)“预

测”量刑是基本原点。从量刑的过程看,个案的量刑相互独立,均需作出个别化的结论。量刑公正正是司法责任制度的核心内容,迫使量刑主体必须穷尽所有可能的方法,将既往的量刑经验等信息,作为具体某一次量刑的“参考”。这些量刑参考发挥(个别性)“量刑预测”功能。目前,对刑事司法智能化的担忧仍未停止。2019年3月,法国颁布了“第2019-222号”法律。<sup>[38]</sup>其中,第33条规定,法官和书记官的个人数据不能被用于评估、分析、比较或预测他们的实际作出或将要作出的专业行为,任何违反该条规定的人都可能被判入狱5年。这对基于大数据、人工智能技术及其应用所形成的“司法预测功能”持消极态度。但是,人工智能“指引(预测)”司法裁判的标准化、集约化乃大势所趋。预测量刑是规范量刑活动与实现量刑公正的重要问题。量刑规范化旨在实现量刑公正,在方法论上必然自成体系。可以概括为更好地根据经验与法则“预测”量刑结果,在不断新出现的个案中,实现量刑结果的公正与有效。(2)两种预测模式的比较。无论是经验量刑方法,还是常见的规范量刑方法,都是以人的智力为前提的。实践证明,人的智力具有一定的局限性,特别是面对大规模案件、大数据分析等问题时,存在“能力不济”的结构性短板。在面对海量的司法数据与复杂案件时,传统的“人”预测之经验与方法,难以和智能技术的“计算”能力与预测效果相比。事实上,智能技术辅助的量刑,在“智能程度”的技术上更胜一筹,也因此受到各界的推崇。智能预测量刑与传统量刑活动的“不适”,并非实质的功能对立,只是方法上的差异。前者依赖智能技术,后者以“人的智力”为基础。这是技术与人的“较量”。究竟是“人定胜天”还是“技术更优”的争论虽难有定论,但不变的是,任何量刑系统实质上都是为了实现量刑过程与结果能够无限接近科学、精准,最终实现罪责刑相适应原则。(3)精准预测是智能技术与量刑的制度性耦合产物。通过深度融合人工智能与现代法律,可以

发现法律运行发展的深层规律,从而促进司法预测等在内的计算法学的发展。智能预测量刑的优势是可以借助量刑原理与司法大数据的经验,进一步提高量刑预测结论的有效性及其程度。这正是智能量刑与传统量刑可以自然融汇的结合点。目前,人工智能辅助预测量刑系统的主要工作原理存在高度相似性。通常是基于算法及司法大数据等,结合算法的函数、逻辑等运算规则,建立独特的量刑算法系统、量刑预测系统及模型,自动对海量案件量刑情节进行提取、分析与组合等;通过海量的精准算法与匹配、对比,自动根据司法大数据中的量刑规律,为具体个案输出精准的量刑预测建议。而且,还可以强化量刑过程的动态说理性与量刑结果的有效性。这都是为法官提供量刑参考的技术平台,凸显智能预测量刑的精准并提供“量刑建议(参考)”的辅助功能。

2. 精准预测功能的实现路径。以量刑规范化原理为前提和基础,遵循传统量刑理论及其方法的有益部分,通过深度激活、挖掘、总结与提炼司法大数据中“活着的”量刑规律与经验,可以自主“发现”与“固定”司法大数据中的“实然”“中立”之量刑逻辑与经验,对个案展开智能预测,从而超越司法人员的个体性或“经验式”等传统量刑预测的技术壁垒与不足。据此,其运行逻辑和实现机理包括:(1)预测结果的实体裁判属性。在量刑领域,通过深度整合司法大数据与人工智能技术,可以实现量刑建议的智能精准预测功能。进言之,通过算法及其规则的同比模拟与计算,根据已经设置的量刑裁判模型及其参数等,可以自动抓取个案的量刑事实与情节,通过类案推送与匹配的方式,对比司法大数据中的规律与经验,预测正在审理案件的实体性量刑裁判并提出参考意见等。因此,人工智能辅助预测量刑的功能,是以实体法为前提,以定罪活动为基础。基于量刑的基本原理及规定,智能输出的量刑建议具有实体裁判属性,决定了刑罚处罚的结果与效力。当然,实体裁判中的量刑预测,主要

是为法官审理具体个案提供量刑参考的智能化工具,直接由人工智能作出裁判,不符合当前司法运行的基本伦理。(2)算法规则实现量刑情节的规范化、智能化裁量。量刑情节是刑罚裁量活动中的“重头戏”。司法人员对量刑情节的认定,在方法论上可以概括为“经验+规则”。前者是指司法人员既有的量刑经验,后者是指刑事司法广泛存在的立法解释、司法解释以及细则等。它们都体现了司法人员的主体性、主动性与主导性,司法机关的集体讨论等特殊情形亦不除外。司法人员完全统辖下的个体性、经验式量刑预测模式客观上存在不足,是我国量刑规范化改革进程中的“老大难”问题。在人工智能辅助预测量刑中,对个案量刑情节的认定,是通过事前的数据输入、分析等,与设定的计算规则、分析规则进行最佳匹配后,智能输出经过理论与实际相互验证后的预测性参考意见。这使“个体经验性”“内心独白式”“封闭性”的传统量刑做法,变成了以知识谱系、系统设置、数据输入、算法规则及其运算体系等智能运要素为核心的说理过程。通过算法规则的科学设置与深度学习,更能确保裁量过程及结果的合理与公正,排除外部因素的不当干扰。(3)个案与类案的智能精准匹配与量刑结论校正。在传统量刑模式中,司法人员的精力与能力受限,主要是通过一般理论的演绎、经验的积累、典型案例或指导性案例的提示等方式进行量刑。原则上无法通过穷尽已有的案件,全面掌握或精准类型化采用“活着的”量刑经验,存在司法经验的不完整性、不全面性,甚至极端化等问题,容易影响量刑公正的结果实现。相反地,算法规则依托一般量刑原理、司法大数据与智能分析等优势,可以对海量的同案、类案等进行比对和分析,推送最具系统性、综合性的“预测性”参考意见。这直接填补了个体经验性量刑的“比对基数不足”问题。通过最大限度的“类案的海量实质对比与校正”,强化量刑预见的可预测性与有效性。(4)实体裁判预测的自主监督与自觉纠错。基于司法大数据、深度学习等条

件,智能量刑系统不仅可以自动抓取“同判或类判”的类案或个案,也具备强大的审查与纠错能力。对基于量刑原理得出的量刑建议,经过智能对比发现有重大偏离的,可以自动向办案人员发出预警,防止裁判尺度出现重大偏离。例如,苏州法院的智能系统不仅可以统计类案的裁判模式与结果,还能根据历史裁判模型,对具体待决案件进行模拟裁判。法官作出的判决结果,一旦发生重大的量刑失衡等偏离问题,系统会自动预警。<sup>[39]</sup>这确保了量刑预测建议的准确性,有助于确保刑罚裁判的公正质量,同时也不会削弱审判监督的应有力度与效果。

### (三)人工智能辅助量刑规范的保障措施

量刑规范化改革旨在更好地实现量刑公正。智能量刑系统的尝试与推进也是在此背景下进行的。从技术支撑与方法论的角度看,提高量刑预测的精准度与质量,特别需要“发现”并“运用”过往司法经验中的量刑逻辑,以“活着的”量刑经验与规律“校正”量刑规范化的理论偏差,合理限制自由裁量权的边界。智能辅助精准预测量刑的核心功能与量刑规范化完全契合。在深度衔接的过程中,为了改革和完善量刑规范化,可以借助智能量刑系统的正向应用优势予以巩固。为了推进二者的深入衔接与发展,也应从应用基础、技术支持等方面进行完善。

1. 同案同判、类案类判的宗旨设定。刑事司法的智能化可以降低司法人员的主观性中的不确定性,减少或消除裁判的模糊性、标准的不确定性,并增加裁判的统一性和预见性。在量刑问题上,人工智能技术的精确计算功能,与司法大数据的深入融合,很好地消除了同案不同判的量刑失衡问题,更好地做到了同案同判。人工智能辅助量刑系统是以量刑原理与司法大数据库为基础的,通过算法的强大优势,对司法大数据进行智能分析,深度挖掘类案的具体结构与影响要素。以数据化与智能化的方式,自动且个别地、具体地输出量刑预测建议,辅助法官对类案的裁判,保障司法裁判标准的统一,促进同案同判、类

案类判。智能技术导入量刑环节后,类案类判不再是纯粹的大数据司法分析技术,而是集中到提升量刑的智能预测及其精准度。通过充分发挥“算法”及其深度学习的能力,提高智能运算、分析、决策等能力,可以超越一般性的经验积累与人的理性在量刑活动中的“极限”,最终在方法论上实现“高度智能”预测量刑的绝对优势。这拓宽了传统裁判思路与思维,甚至会发展为统一司法适用的强化手段,并上升为新型司法判决的质量控制机制。为此,应继续推动量刑规范化改革,制定出台精细的量刑指导意见;加快建立健全司法大数据库,优化类案类判系统的司法管理体制、制定标准流程与规范管理制度,配备类案检索、类案推送、类案类判的复合功能。

2. 量刑规范化的智能标准化体系。统一司法裁判尺度,是对刑事司法智能系统功能的基本定位。这一核心功能被通俗地概括为“标准化”。以标准化为核心功能导向,催生了新的司法运行机制。司法判断和决策模式在算法的推动下,呈现出智能化与标准化、集约化、模块化等特征。竭力追求量刑预测的“标准化”,会要求法官专享审判权,以确保类似案件的裁判尺度尽量相同。这也会生成更加开放与包容的第三方评估标准,强化量刑预测的监督属性。然而,“标准化”在契合智能技术的“流水”作业属性时,也要坚持适度的灵活性。刑事司法智能系统的“标准化”应用不能走向极端,个案公正是不能突破的底线。应正确对待“标准化”的可靠性问题,重大、疑难、新型案件更复杂,对智能量刑系统的标准化要求更高,以此确保预测结论的可靠性,真正促成量刑标准的统一化。在量刑预测过程与结论的“标准化”上,应留足必要的“差异性”空间与干预范围。

3. 人工智能辅助量刑的专业化。只有将算法、司法大数据与量刑基本原理进行高度的黏合,才能达致精准预测。在定罪是量刑的前提、个案的量刑要素分析、同案或类案的分析与对比、量刑预测与量刑建议等核心环节,注入算法

的计算能力与运行因子,提升量刑过程的科学性与其结果的有效性。目前,人工智能辅助量刑系统仍隶属于智能审判系统或智能检察办案系统,没有完全独立的测算模块与运行系统,缺乏专属性与独立性。为了破解专门化、专业化、专属性相对不足的问题,应加快应用层面的专业化建设,实现原理、数据以及智能预测的三维合一。因此,一是应当完善司法大数据库的建设及分析能力,<sup>[40]</sup>优化对司法数据的智能挖掘能力,改进算法的规则与运行。二是在系统应用的便捷性上,优化量刑建议的推送方式,附上相应的“算法说理”内容。三是在系统结论的正当程序性上,增加量刑建议的透明性与验证性,使其更符合刑事司法规律、量刑活动的需求。

4. 优化人工智能辅助量刑的说理性能。相比之下,数据支持、说理的表述方式、论证方法、说理的过程公开等不足,削弱了传统量刑说理及量刑结论的可接受性。以算法为核心的智能量刑系统,构建新的量刑说理机制,有助于消除个体性、经验性量刑预测模式的封闭性与独断性。通过统合量刑规范化原理与司法大数据中的量刑规律,可以提供更中立、直接以及具有验证性的说理机制,以更立体、可视化的方式,展示同案同判、类案类判的实现过程,强化预测结论的正当性与可接受性。当前,应当继续以算法为核心,不断优化算法规则、运行方式以及深度学习条件等条件,完善人工智能辅助量刑的说理水平。初步看:(1)提高自然语言的“解码”能力,与人类语言实现无缝对接,避免出现语用逻辑等信息偏差、数据输入错误等深度学习的误区。(2)要尽快实现算法规则及其运行的透明性,加大智能量刑过程的公开性,强化智能量刑说理机制的外部可识别性、可解读性与公众可参与性等。(3)具体操作与实施环节不仅应以推送确定或具体的量刑建议为常态做法,更应附上算法层面的说理依据,实现说理与结论的相互验证。(4)建立智能量刑系统说理的外部监督机制。应当允许司法人员或利害关系人员提出意见和

监督,对于明显错误或不当的,设置纠正或不予以采纳的补救通道。

#### 四、人工智能辅助量刑的公正达致

人类的智力水平及其经验所形成的刑事法治知识体系,是司法活动的重要基础。得益于智能技术与司法大数据的融合技术优势正在壮大以及集体释放的积极效果,人工智能辅助量刑快速发展。但是否更科学、可靠及其有效性等深层次“追问”也随之出现。尤其在“算法”主导下,智能量刑尚存在可视化不足等问题。“算法”合乎正当程序运行是人工智能辅助量刑系统达致司法公正的保障。

##### (一)警惕“算法”司法化的隐忧

人工智能技术高度依赖“算法”及其运行规则。对于这一新事物,实践中不乏各种担忧的态度。主要是因为“算法”的自主运行很可能超出人的控制,继而导致人工智能辅助量刑系统的运行及其结论可能脱离司法正义的底线之设定。

##### 1. 算法的司法应用之技术风险。主要为:

(1) 算法系统的非透明性。数据和算法是算法治理的核心,智能系统高度依赖算法与自主的深度学习能力。但是,算法是“黑盒子”,“黑箱性”特征是其天性。数据在完整性和真实性上的瑕疵以及算法的不可解释性,使算法及其运行具有不透明性等问题,天然地与法律决策和司法裁判的“透明性”“公开性”等要求相冲突。算法的隐蔽性和当代刑事案件的裁判过程透明性之间的不适乃至冲突问题由此而生。在此基础上,人工智能及“算法”有可能成为内在的技术障碍。在现阶段,智能量刑系统无法真正全面实现透明化、可视化。通过正当程序的方式,合理稀释非透明性并增加公开性、可视化,是一项需要长期努力才能实现的任務。(2) 算法规则的权力垄断倾向。智能量刑系统的自动生成判决、根据大数据矫正法律决定的偏差等功能,都是对目前量刑活动的有益修正与补充。然而,自动生成判决的普及适用,也使法官的地位和作用被极端弱

化,甚至可能出现算法支配或控制审判。一旦变成司法常态,算法的独裁问题随之出现。例如,深度学习可能使既有判决中存在的失误、质量问题以及偏差值在无意间被固定化。“将错就错”的风险,容易压抑通过个案发现并确认合法权利、创新规范、推动制度升级的动态进化机制。(3) 算法预测偏好的极端化。智能量刑系统依靠“算法”进行预测,在一定程度上使传统量刑的程序独立改革化为泡影,转而变成不可视觉化的“算法”之运算过程。这可能导致量刑辩论等核心的程序性活动走向“消亡”。在“一切取决于既定的软件,面对面的对话式论证不需要”的算法运行逻辑下,可能会引发负效应的扩大化,特别是量刑过程中的法律推理、法律议论、法律解释学等本就薄弱的部分变得更弱不禁风。而且,来源多样化、规模不断膨胀的法律数据的可靠性与真实性仍存在问题,难免诱发“算法”的暗箱效应。片面且过度强调智能量刑系统的可靠性,可能加剧原有司法实践的弊端或放大某些缺陷,甚至制造新的弊端和不足。(4) 算法语言切换的隐忧。如何借助智能技术,将人类语言、思维、经验等法律知识,在遵循人类社会信赖的公正标准与正义要求的前提下,通过自动解码,将人工语言转换成人类可以识别的电子信息、数据以及结论。这是两种语言之间的“技术鸿沟”。只有实现信息对等、充分沟通、有效对话,才能保障算法的程序正义与结果公正。“语言鸿沟”使智能量刑系统背后隐藏一定的“技术暗箱”操作与“流水线”生产方式等隐忧。这是不可测且不可控的不确定性司法风险。

##### 2. 算法的功能性与制度性缺陷。主要包括:

(1) 算法的数据制度跟进滞后。司法大数据库的迅猛发展以及人工智能技术的充分发掘,使“活着的”量刑经验与规律日益公开、丰富与可靠,为智能量刑系统的实践提供基本的“类比依据”。但是,我国司法大数据库建设仍存在不少问题,包括体量不足、内容不充分、真实性与客观性有待加强、司法数据的结构分类不到位、一定

范围内的数据缺失与匮乏等问题。司法数据是智能量刑系统中“算法”的运行基础,是精准预测的匹配对象。司法数据体系尚不完善,降低了算法运行的顺畅与结论的科学性。(2)算法逻辑的社会属性匮乏。在美国,不透明的算法规则,部分取代法官的自由裁量活动,已引发批评与担忧,特别是智能技术应用脱离了社会现实。<sup>[41]</sup>社会属性与人文关怀的缺失,既是由于智能技术应用过程中的“去人中心化”效果的必然产物,也使智能应用在量刑领域必须面对“去人性化”等的制度性困难。量刑活动既是法律行为,也是社会问题。量刑的标准及其裁量,必然受到各种社会经济因素的影响,在整体上也必须有效回应社会诉求。在现阶段,智能量刑系统在有效植入并充分评价社会因素上,仍存在一定的技术短板,可能会削弱精准预测的有效性与公众认可度。对算法司法化的疑惑与发酵,折射出人工智能辅助量刑的“正当性”隐忧。如何规制算法并确保人工智能辅助量刑的过程与结果公正才是关键。

## (二) 遏制消损算法公正的制度供给

以司法大数据和算法为基础建立的智能决策模式,可以较大程度地克服人类因认知局限或主观恣意等导致的偏见,使司法结果更客观、准确和公平。人工智能技术与司法的结合,往往被认为更接近公平。但智能技术及其应用的好坏并非绝对。智能司法引发的歧视和偏见并不鲜见,应坚守司法公正的底线。“算法”及其规则体系作为智能量刑应用的关键支点,及其所得出的智能量刑预测结果,应当兼顾司法的一般公正与个别公正。立足当下的技术应用水平,既要对外部控制“算法”的司法化风险外溢,也要从内部强化“算法”逻辑的科学性,为人工智能辅助量刑打造可靠的公正性基础。为此,应建立一套有效的规范化预防措施。

1. 量刑公正的指标设定。为了充分发挥人工智能辅助量刑的技术优势,克服算法本身的技术风险,应设定一套司法公正的指标体系。其指

标要素体系应当包括:(1)司法大数据的公开与统一。除了量刑规范化的基本原理,司法大数据是人工智能辅助量刑系统的基础要素,也是算法运行的主要对象,对算法规则与结果形成具有前端影响力。向智能系统输入、分析的司法数据必须是公开的,这不仅可以从源头确保算法的科学性与公正,也可以回溯算法的源头,增加算法在前端的透明性;而且,为了统一刑事司法智能化的适用标准,应当建立统一的司法大数据库,打破当前司法“数据孤岛”的消极局面。<sup>[42]</sup>确立国家统一的标准化共享共用机制,在更大幅度内通过司法大数据实现预测量刑的规范化、标准化。(2)量刑情节的法定化、类型化与酌定情节的规范化。量刑情节是量刑活动的关键要素,量刑情节的法定化是确保量刑规范化、统一化的重要前提。算法系统对量刑情节的自动识别和分析,也应当以法定化为原则,并增加类型化的分析功能。应当规范酌定量刑情节的适用,强化个案的公正效果。只有对量刑情节的识别、分析与处理,符合法定原则、规范要求,算法运行所得出的量刑建议才具有正当性。(3)算法过程的透明性。在技术应用层面,应当考虑建立一套可以被外部观察或体验的机制,以优化算法运行过程的透明性问题。例如,算法运行的主要步骤、阶段、分析方法等要素,可以通过人类语言的方式对外公布,或有相应的预警、监督措施。此外,需强化量刑建议的说理性。在自动输出量刑建议之际,应当附上相应的说理内容。既要阐明结论的形成过程,也要立体地、可视化地阐明理由,使结论具有辩论性、检验性。在现有技术条件下,司法人员基于公正的需要,可以主动干预,对预测的量刑意见作出合理修正,或直接不予以采纳并说明理由。(4)自然语言处理技术的衔接升级。算法的运行依托自然语言,与人类语言不同。智能量刑系统需要通过自然语言“解码”人类语言表述的法律问题。这引发了不同语言的衔接与冲突问题。以法律文书的读取、分析等为例,电子法律文书大致按照统一的格式编辑,但司法人

员撰写法律文书时,对相同事项的文字表述方式具有多样性,如自首可以表述为自动投案、代为投案、如实供述罪行等形式。在人工智能辅助量刑中,算法构建日益理性与科学的知识图谱后,由于自然语言技术尚不成熟等原因,可能无法通过传统的关键词匹配等方式,全部、完整地提取情节。这无疑弱化了智能技术本应发挥的技术优势。人工智能辅助量刑在自动高效、准确提取案件事实、证据以及情节等数据信息上仍有提升空间。为了消除自然语言处理技术短板,应优化自然语言处理技术的外部便捷性,显著提高算法的沟通、协作能力。

2. 遵循正当程序的算法逻辑。算法的运行方式具有隐秘性与非透明性,目前主要由研发者、设计者及其相关公司负责管理与控制,在算法的运行规则、过程以及结果等方面,容易形成“黑箱”效应。程序公正是保护被告人基本诉讼权利的前提。智能量刑系统的不公开问题,可能会压缩公正性的应有含量。在美国艾瑞克·鲁米斯(Eric Loomis)一案中,人工智能辅助量刑引发的主要争议是人工智能辅助刑事裁判的不确定性、可解释性风险。克里斯托弗·斯洛博金教授指出,如果认为有必要,法院应该有权强制披露相关智能分析工具算法的代码。<sup>[43]</sup> 遵循正当程序原理,司法裁判活动必须具有公开性,利益各方可以充分参与并辩论。不过,算法作为智能量刑系统的核心要素,在应用方式和效果上,与正当程序及其要求之间仍有差距。在艾瑞克·鲁米斯一案中,威斯康星州最高法院支持初审法院的判决,否定鲁米斯提出的正当程序和平等权主张。被告在算法评估结果前,对 COMPAS 分析的记录在案的公共数据(犯罪记录)和基于被告对问题的回答而形成的数据,有机会否认或解释相关信息以及验证相关信息的准确性,其事后质疑信息准确性缺乏合理依据。<sup>[44]</sup> 这在一定程度上肯定了算法的程序正义性。算法的程序正义性,与司法说理下的程序正义,在价值取向与功能配置上相似,在程序正义的实现上可以相互

借鉴。对于算法的运行而言,建立健全法律议论机制有积极意义。<sup>[45]</sup> 法律议论机制可以充分关切利益攸关方,通过对话等程序达成共识。相应地,应当接纳司法人员的外部监督与介入,容忍辩论等交互、兼容的形式。在刑事司法智能化的背景下,智能量刑系统发生在封闭的裁判场合,算法要处理个人与国家之间的复杂法律关系,在证据、事实、法律以及价值中进行审查与综合判断,应充分考虑社会因素等案外要素,而不单纯分析“法律规范”。智能量刑系统的公正性,取决于算法的外部公开性与可视性。应当进一步优化算法及其运行过程与程序正义的契合性。其中,算法规则与法律的正当程序必须深度嵌合,消除算法的不透明问题,使算法运行的规则、方法、逻辑、过程以及结果生成等要素符合正当程序,消除算法的垄断、偏好以及极端等问题。

3. 算法输出量刑说理机制。“算法”是智能技术的生存“灵魂”,“算法”规则是智能技术走向应用层面的道德伦理规则。智能量刑系统所依赖的“算法”,其意义与功能大体相当于传统量刑说理制度。从可视化的角度看,算法运行与传统量刑说理具有本质上的相似性,肉眼无法直观。但在思维的层面上却是可以被“看见”或“被发现”,也是可以被讨论、检验、修正以及改良的对象。智能系统所依赖的“算法”,在发挥精准预测等量刑功能时,正好扮演提供新型“说理系统”的角色。只是说理的主体、原理、对象、目的、表现形式等,与司法人员的量刑说理机制差异甚大。但并不能阻隔智能系统所依赖的“算法”运行规则,承担起输出人工智能辅助量刑的说理任务。基于此,算法逻辑支撑下的人工智能辅助量刑,实际上正在生成一种新的量刑说理模式,其最大的特点是通过“算法”而非“人脑”,智能预测个案的最佳量刑意见,并无限接近合理性与可接受性。这与传统量刑的说理机制在本质与功能上是一致的,也是智能量刑系统嵌入量刑规范化改革的最强内生动力。而且,通过算法及其规则,不仅可以恰如其分地褪去“非可视化”

“透明性不足”等技术缺陷所引发的外部纷扰,还可以更专注地提升人工智能辅助量刑预测的科学性与精准性。更重要的是,算法逻辑与量刑说理的搭配与组合,也孕育了人工智能辅助量刑系统的规范性基础与教义学属性,使其不再单纯停留于运用司法大数据获得“数据对比优势”等智能技术应用优势的单一层面。这更充分、具体彰显了人工智能辅助量刑系统拥有可靠的规范基础、法理依据、功能体系。人工智能辅助量刑系统变成了事实分析、规范判断、价值取舍的综合体。

4. 算法风险的自控规则。量刑建议的适当与否,直接依赖于算法规则、算法运行的精确性。目前,基于深度学习而自主升级的算法,仍依赖于人类在前期的海量数据输入。海量数据输入的优化学习方式,导致早期的学习积累过程会受人的主体性地位及其能力的影响。在升级与完善算法及其运行上,既要强化改进司法数据等前端要素,也要通过增加自主学习和优化积累能力等,提高算法的法律理性要素。对于算法与司法的结合,要在准确性与科学性、透明性等方面寻找最佳结合点。这赋予人类一定的干预必要性,对算法及其运行摄入正当性的要素。具体而言:

(1)应当披露智能裁判辅助系统的技术参数、算法逻辑等核心司法要素,使外部利益攸关者可以充分知情、有效参与。从现有的主客观因素看,智能量刑系统的数据公开、算法公开、过程公开等一系列问题,并不单纯是法律适用与理解的问题,首先是智能技术攻关问题。推进此项正当程序工作,仍需要很长一段时间。这是智能量刑系统无法与法官量刑“平起平坐”的技术内因。应当率先从智能技术升级的层面解决算法公开的难题,实现智能量刑的预测过程与结论之程序公开。(2)可以建立和完善算法审查制度。司法智能辅助系统在运行中可能会出现不确定性风险,影响司法决策的科学性。可以建立算法的合法性、正当性审查机制,通过人类的外部干预予以消解。例如,可以考虑成立由司法人员和智能

研发、设计专家等组成的伦理审查委员会,对不透明、反歧视等热点伦理和法律问题进行审查,借助外部力量督促算法的公正和合法性。同时,在研发与设计的阶段,应当预设一定的可回溯性功能,利用算法系统的可回溯性功能,对算法的运行过程及结果进行回查和监督,作出必要的矫正与修正。(3)确立权威的算法价值调和与解释规则。对于算法在刑事司法应用领域引发的风险与价值冲突,不仅要确立基本的调和原则,明确当前智能技术辅助的角色以及人类利益优先原则,用于解决重大疑难问题;也要确立权威的解釋规则,用于个案或类案的止纷,以及界定司法人员的渎职责任。制定解释规则的主体应当是最高司法机关,并且必须确保利益攸关方的充分参与,以对话、协商的方式,妥善解决价值冲突。(4)加强智能司法伦理建设。智能裁判辅助系统在程序层面缺乏可责性与回溯机制,使司法责任制几乎无从谈起,无形中弱化了对智能量刑系统的质量控制、责任监督。人工智能辅助量刑的进一步发展,应始终贯彻合乎技术伦理的设计理念,制定与完善设计者和使用者的司法技术伦理指南,使算法的运行及结果更接近司法正义。

#### 注释:

[1]参见雷磊:《司法人工智能能否实现司法公正?》,《政法论丛》2022年第4期。

[2]参见张凌寒:《智慧司法中技术依赖的隐忧及应对》,《法制与社会发展》2022年第4期。

[3]参见刘艳红:《人工智能技术在智慧法院建设中实践运用与前景展望》,《比较法研究》2022年第1期。

[4]参见孙道萃:《人工智能刑法研究的反思与理论述进》,《学术界》2021年第12期。

[5]参见雷磊:《中国特色社会主义智慧法治建设论纲》,《中共中央党校(国家行政学院)学报》2020年第1期。

[6]参见栗峥:《人工智能与事实认定》,《法学研究》2020年第1期。

[7]参见周佑勇:《智能技术驱动下的诉讼服务问题及其应对之策》,《东方法学》2019年第5期。

[8]参见戴佳:《检察业务应用步入2.0时代》,《检察日报》2020年1月4日。



[9][29]参见陈学勇:《更高水平推进量刑规范化工作》,《人民法院报》2020年11月6日。

[10]参见刘哲玮:《迈入2.0时代的在线诉讼》,《人民法院报》2021年7月3日。

[11]最高人民法院还发布《人民法院在线调解规则》,推动构建中国特色、世界领先的互联网司法模式。

[12]参见李占国:《“全域数字法院”的构建与实现》,《中外法学》2022年第1期。

[13]参见左卫民:《从通用化走向专门化:反思中国司法人工智能的运用》,《法学论坛》2020年第2期。

[14]参见戴佳:《把认罪认罚从宽制度落实到具体案件中》,《检察日报》2019年4月15日。

[15]参见史兆琨:《深入推进量刑建议工作有效开展》,《检察日报》2019年4月29日。

[16]参见苗生明:《认罪认罚量刑建议精准化的理解与把握》,《检察日报》2019年7月29日。

[17]参见罗庆东:《以精准化量刑建议落实认罪认罚从宽》,《检察日报》2020年2月10日。

[18]参见汤瑜:《全国政法机关主动提升司法效能》,《民主与法制时报》2017年7月13日。

[19]参见彭波:《贵州:大数据点亮“智慧检务”》,《人民日报》2017年5月31日。

[20]参见鲍健、王瑛:《智慧公诉建设与未来发展——以浙江省杭州市检察机关的实践探索为例》,《人民检察》2018年第4期。

[21]参见范跃红、张永睿:《杭州西湖:多举措提升办理认罪认罚从宽案件量刑建议精准度》,《检察日报》2019年5月8日。

[22]参见罗书臻:《挖掘“富矿”“反哺”审判——运用裁判文书大数据促进司法公正的地方经验》,《人民法院报》2017年9月1日。

[23]参见方茜:《海南法院大数据人工智能助力司法改革》,《人民法院报》2017年7月27日。

[24]参见叶子:《“智慧司法”开启法律服务新时代》,《人民日报(海外版)》2018年1月26日。

[25]参见辛红:《识别证据瑕疵 推送类案及量刑参考》,《法制日报》2017年9月18日。

[26]参见陈学勇:《最高人民法院立项开发建设量刑智能辅助系统》,《人民法院报》2017年9月27日。

[27]参见罗书臻:《周强:认真学习贯彻党的十九大精神 深入推进智慧法院建设》,《人民法院报》2018年1月5日。

[28]参见杨帆、金鸿浩:《统一业务应用系统2.0版试点倒

计时 助力履职能力现代化》,《检察日报》2019年12月10日。

[30]参见樊崇义:《关于认罪认罚中量刑建议的几个问题》,《检察日报》2018年7月15日。

[31]参见刘卉:《确定刑:认罪认罚从宽制度下量刑建议精准化之方向》,《检察日报》2019年7月29日。

[32]参见程金华:《人工、智能与法院大转型》,《上海交通大学学报(哲学社会科学版)》2019年第6期。

[33]参见孙道萃:《人工智能刑法主体地位的积极论——兼与消极论的答谈》,《重庆大学学报(社会科学版)》2022年第4期。

[34]参见左卫民:《AI法官的时代会到来吗——基于中外司法人工智能的对比与展望》,《政法论坛》2021年第5期。

[35]参见宫鸣:《推动现代科技与公诉工作深度融合》,《人民检察》2017年第20期。

[36]参见孙道萃:《人工智能辅助精准预测量刑的中国境遇——以认罪认罚案件为适用场域》,《暨南学报(哲学社会科学版)》2020年第12期。

[37]参见高憬宏、黄应生:《积极稳妥推进量刑规范化改革》,《法律适用》2009年第8期。

[38]参见施鹏鹏:《法国缘何禁止人工智能指引裁判》,《检察日报》2019年10月30日。

[39]参见丁国锋:《苏州法官判案,8种“机器人”智慧平台当高参》,《法制日报》2017年4月15日。

[40]参见左卫民:《迈向大数据法律研究》,《法学研究》2018年第4期。

[41]参见李本:《美国司法实践中的人工智能:问题与挑战》,《中国法律评论》2018年第2期。

[42]参见赵龙、刘艳红:《司法大数据一体化共享的实践探微——以破除“数据孤岛”为切入点的逻辑证成》,《安徽大学学报(哲学社会科学版)》2019年第6期。

[43]参见朱体正:《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》,《浙江社会科学》2018年第6期。

[44]法院承认COMPAS算法评估结论,可以揭示与鲁米斯相似的一类人的再犯风险,但该评估结论不是法院作出量刑判决的唯一依据,法院拥有同意评估结论的裁量权和相关证据的裁决权。该量刑判决是充分个性化的结论。参见郑戈:《算法的法律与法律的算法》,《中国法律评论》2018年第2期。

[45]参见季卫东:《人工智能时代的法律议论》,《法学研究》2019年第6期。

【责任编辑:邹秋淑】