



刑事审判人工智能的权力逻辑

卫晨曙

(中国政法大学 证据科学研究院, 北京 100088)

[摘要] 刑事智能办案系统是人工智能技术进入刑事司法场域中的主要表现形式。当前,刑事智能办案系统主要集中在证据审查判断、辅助精准量刑、类案推送和案件偏离预警等领域。随着人工智能技术与刑事司法的深度融合,人的“代具性”原理、自动化偏差认知机制、国家权力的背书与支持、技术外包的研发模式以及系统架构的运行机制等因素推动着智能办案系统从“工具”向“权力”的演化。同时,刑事审判中人工智能技术的权力化可能会引发规训风险、排斥风险与误判风险。为了有效规避刑事审判场域中人工智能技术引发的正当性风险,应当树立权力规制理念,从智能办案系统的适用机制、参与机制以及研发机制三方面进行规制。

[关键词] 刑事审判; 人工智能; 技术权力; 智能办案; 正当程序; 司法裁判权

[中图分类号] TP18;D925.2 **[文献标识码]** A **[文章编号]** 1008-245X(2021)03-0141-08

2017年初,一则南京中院决定引入智能机器人以提高审判质量与审判效率的新闻刷爆朋友圈,一时间关于“法官是否会被机器人取代”“机器人审理案件”“案件审理的自动售卖机”等话语成为网络中的高频词汇。这一话题的舆论关注度从侧面显示出了人工智能在刑事审判中应处于何种地位、发挥何种功能的问题,牵动着整个社会的神经。关于这一问题,无论是司法决策层还是学界,都将人工智能技术定位为辅助司法的工具。具体而言,这一观点的论证思路,即司法案件的裁判并非简单的三段论式推理,其中仍然蕴含着案件审理法官对案件事实的感官直觉、对法律条文的理解、逻辑思维的应用以及联想、顿悟等多种心理活动,而当前的人工智能技术尚不具备感性判断的能力,因而刑事审判中人工智能、大数据技术的应用应当定位于辅助地位。然而,这种论证思路尚处于应然层面的逻辑推演,忽视了办案法官与智能系统交互过程中面

对智能系统自动化决策的心理认知等实然因素,面对智能办案系统的自主化、智能化不断提升、人与智能办案系统联系愈加紧密的趋势,智能办案系统是否依然发挥的是以“工具主义”为核心的辅助功能,以及我们该如何审视与规制刑事审判中的智能办案系统是本文所要讨论的主要内容。

一、刑事审判中人工智能的应用现状

当前,司法领域中的人工智能、大数据技术具体表现为智能办案系统,例如上海“刑事案件智能辅助办案系统”、贵州“政法大数据系统”、北京“睿法官”系统等等。根据不同的功能指向,我国当前刑事审判中的人工智能主要应用于以下三个领域:

(一)辅助证据审查判断

上海“刑事案件智能辅助办案系统”(206系统)是这一功能的典型代表。这一系统主要包括证据标准、

[收稿日期] 2020-08-24。

[基金项目] 教育部规划基金项目(19YJA820004)。

[作者简介] 卫晨曙(1995—),男,中国政法大学证据科学研究院博士研究生。

证据规则指引,单一证据校验以及证据链和全案证据审查判断三项功能。首先,就证据指引而言,主要分为证据标准的指引和证据规则的指引,其运行原理是根据不同案件类型,分门别类地对各类案件所应收集的证据形式、证据内容以及收集程序等注意事项进行归纳和总结,形成诸如《上海刑事案件证据收集、固定、审查判断规则》等文本规则,将其嵌入到智能系统中,在办案人员的办案过程中对其进行指引。其次,就证据校验而言,主要是针对单一证据的审查,主要集中在书证领域,其运行原理是依托人工智能深度学习、智能识别技术,对过往案件中被扫描进系统的相关证据种类进行自主学习,进而能够定位、识别并提取出各种证据的签名、印章、手印等,然后再根据相关的证据标准或证据规则校验其证据的形式、收集是否规范。最后,就证据审查判断而言,主要是综合全案证据的审查判断,其运行原理是智能识别技术定位并抓取证据中曾经出现的相关人物、物品、时间以及地点等要素,在此基础上分析这些要素之间的相互关系,理清人物关系、运动轨迹以及物品流动轨迹,从而形成关于某一具体案件的具体故事模型,凭此对各项证据之间是否有内在联系、是否相互印证以及是否有矛盾或疑问进行研判,最终将分析结果提供给司法人员。

(二)辅助精准量刑

辅助精准量刑的原理一般是通过“正则表达式、上下文无关文法、命名实体识别等方式提取法律文书的情节信息,之后对提取的信息进行结构化,用数学模型来描述数据的特点和规律,算出与其最相符的数学模型或算法,人工智能通过模拟算法,根据情节推导量刑结果”^[1]。以法律智能公司华宇元典公司研发的“元典量刑辅助系统”为例,通过整合人民法院的公报案例、指导案例以及裁判文书网收录的刑事裁判文书,根据案由和影响量刑的案件要素,对当地法院的生效判决进行过滤筛选,最终由系统给出所选取时期内当地法院就该类型案件和类案的量刑区间分布、集中量刑值、平均量刑值等,如类似情节案件的刑期的“众数”、最低刑期、最高刑期及平均刑期。让办案人员根据具体案件通过界面勾选与自定义检索相组合的方式进行操作,辅助量刑参考结果更为精确。

(三)类案推送与案件偏离预警

类案推荐、偏离预警是实现同案同判的重要保障。一方面,当前各地法院的智能系统中,类似案件推荐功能是最为常见的功能之一,北京、上海、江苏等地法院均有使用。类案推荐功能实现的前提是构建同类案件

的知识库(数据库),一般通过搜集同类案件海量的裁判文书,提取其中最为相似的案件情节来完成。在此基础上,类案数据库根据法官所办的具体案件,为其推送与之最为相似的历史案件以供法官参考。另一方面,偏离预警是指人工智能依据算法推测的结果与法官裁判的结果进行比较,在上海、江苏、浙江、贵州、云南等地的法院智能辅助系统中,都包含案件偏离预警的功能模块。主要分为实体偏离预警与程序偏离预警。前者是智能系统对案件实体裁判,例如罪名认定是否正确、证据形式是否合法、证据链条是否完整以及量刑幅度是否合理等与案件实体判决相关的问题进行比对的偏离预警。后者是智能辅助系统对法官办理案件的时限、步骤等程序性规定,与系统相似案件的程序性规定比对,对案件办理的全流程进行预警。

总而言之,目前智能办案系统是技术与司法深度融合的产物。一方面,智能办案系统与系统操作者之间的关系更为紧密,二者呈现出相互依存、彼此塑造的交互关系。智能办案系统的运行离不开系统操作者与系统之间的双向问答,智能办案系统根据操作者的勾选指令进行即时反馈,以满足系统操作者的预期需求。另一方面,目前的刑事司法人工智能已经扩展到证据审查判断、类似案件推送、社会危险性审查以及案件办理纠偏预警等领域,逐步深入到法官的心证及自由裁量的过程。面对人工智能在刑事司法领域高歌猛进的态势,难免引发人们对智能办案系统分割甚至架空司法裁判权的隐忧,“法庭正暗地里将做出裁判的权力托付给没有政治合法性的外人。难道法庭名义上的助手成了它背后的主宰者吗?”^{[2]210}在权力理论的视角里,这一担忧似乎正在成为可能。

二、刑事审判人工智能技术的权力演化

进入数据社会之后,数据逐渐成为资源、财产甚至社会的建构力量,而处理海量数据所需要的计算能力已经远超人工计算水平,因而人类对这些数据资源的利用与分配逐渐让位于智能算法,在这一过程中算法呈现出具有支配性力量的权力化趋势。这种权力更倾向于福柯眼中的权力范式,身处权力场域中的人对这种权力难以察觉,这种网络化、弥散化的权力犹如毛细血管一样渗入到刑事审判场域的各个角落,通过对人进行监控、规训以及控制的方式实现支配控制。

(一)人的“代具性”:人工智能技术权力演化的哲学根基

从技术哲学的角度而言,人对技术的依赖从哲学

层面解释了人工智能技术可能对人产生的支配关系。法国当代技术哲学家贝尔纳·斯蒂格勒在论述人与技术的关系时提出了人的“代具性”概念,代具(prothèse)原意是指用于替代人肢体的器具(例如肢体残疾的人所使用的假肢),斯蒂格勒借此概念来泛指用于补遗的、人身体之外的一切技术物。斯蒂格勒认为没有属性就是人的第一属性,这种天生的、原始性的缺陷是人之所以为人的本质属性,这就要求人不得不发明和利用技术来补救、超越自身的天生缺陷,所以说超越这种缺陷是人之为人的首要条件^{[3]312},进而人也成为技术性的存在。因而,人与技术则形成一种既相连又相离的关联,离开技术,人就不存在,因为人的生命延续需要以生命之外的技术为前提。反之,离开人,技术也不存在,因为技术的存在,需要人的能力作为超前动力^[4]。“人的代具性”为审视裁判者与刑事审判智能系统的关系提供了新的视角,人对技术天生的依赖性使得裁判者在应用智能系统时,它们之间所形成的并非是一种“法官—辅助工具”的二元关系,而是法官与刑事审判辅助系统协作互补的一体关系,智能系统所拥有的文书自动生成,类似案例推荐、证据形式审查、证据标准、规则指引等功能补救了法官作为个人在面对繁琐海量复杂的审理业务时体力与脑力的不足与缺陷,同时这种智能办案系统的功能在潜移默化中成为如同办案法官手足的一部分。

(二)自动化偏差:人工智能技术权力演化的认知基础

计算机技术、人工智能等一批具有自主性技术的普及与出现,推动了传统的认知科学、计算机科学以及工效学的交叉融合,以应对自动化时代人机关系研究的需要,从而催生出了“认知工程学(Cognitive Engineering)”。认知工程学中的“自动化偏差”现象揭示了人对智能机器的偏信心理。具体而言,自动化偏差指的是人们一般倾向于将计算机生成(computer generated)的结论认定是正确的,并且不进一步探究或者直接忽视可能与计算机系统生成结论相异的其他信息^[5]。在刑事司法领域,美国学者的一项实验^[6]也验证了智能系统的“自动化偏差”心理对裁判者决策的影响^①。

(三)国家权力:人工智能技术权力演化的权力基础

一方面,国家权力为刑事审判人工智能的正当性提供话语保障。2016年7月,在由中共中央办公厅、国务院办公厅印发的《国家信息化发展战略纲要》中,“智慧法院”成为国家信息化发展战略的重要组成部分,“建设智慧法院,提高案件受理、审判、执行、监督等

各环节信息化水平,推动执法司法信息公开,促进司法公平正义”。2017年5月,周强在全国法院第四次信息化工作会议上指出:“建设智慧法院,是顺应新一轮科技革命浪潮的必然选择,是提升司法公信力的重大举措,是提升人民群众获得感的有效手段,是深化人民法院司法改革的重要支撑。”从政治系统到司法高层都为刑事审判中智能办案系统的应用提供了充分的话语准备,为其运行的合正当性与合法性提供了背书。

另一方面,国家权力为算法权力提供司法数据保障。算法的运行离不开数据的“喂养”。在万物皆可数据化的时代,刑事司法大数据成为司法机关握有的特殊权力资源。司法领域内形成的数据资源不同于以社会为背景的大数据资源。一是司法大数据的占有主体具有单一性,司法大数据的占有主体仅为司法机关,辩护律师、诉讼当事人以及其他社会组织能够通过裁判文书公开网等司法公开途径获取司法数据,但是这样获取的司法数据是十分有限的,例如裁判文书公开网公开的仅仅是判决书与裁定书,而与案件相关的其他诉讼文书,如审判委员会的会议纪要、合议庭会议记录以及法院的案件审结报告等文件则未予公开。二是司法数据与社会数据之间数据流动的单向性。为保障案件认定的准确性,司法大数据能够主动吸纳与犯罪嫌疑人、被告人有关的社会数据,而司法数据因其特殊性

① “再犯风险预测工具”(Recidivism Risk Prediction Instrument, RRPDI)是美国刑事司法为保障量刑的准确性而采取的、对被告人的再犯风险进行预测的智能化工具。该工具根据被告人的职业背景、婚姻状况以及教育程度等因素将被告人的再犯风险量化为低中高三种档次,法官在量刑时根据具体的量化风险作为其量刑决策的参考。美国密歇根大学法学院斯塔尔教授为了考察 RRPDI 对司法决策的影响,将 83 位法学院学生分为 A、B 两组,其中 A 组 43 人, B 组 40 人,并将智能系统与案件背景等相关情况对两组学生进行了告知,唯一不同的是: A 组在进行量刑决定时不会获取 RRPDI 的计算结果, B 组在进行量刑决定时会参考 RRPDI 的计算结果。斯塔尔教授提供了两则案例,罗伯特与威廉分别是实验中两则案件的被告人,他们所犯刑事案件的案由相同,即盗窃价值十万美元的珠宝。他们都被判处相同的罪名(重大盗窃罪),两人皆无犯罪前科(只不过其中一人之前因未到法定饮酒年龄饮酒被判处有期徒刑),两人都是男性并且二者的种族并未告知被测试者。但是罗伯特与威廉两人的个人情况差别极大:罗伯特,中年人,已婚,大学学历,珠宝连锁店主管,犯罪动机是为自己的女儿筹集大学学费,其犯罪经过精心策划,将珠宝首饰盗取后,又用赝品予以代替并售卖,掩饰其犯罪行为;威廉,21 岁,单身未婚,无业,酗酒,最近被父母赶出家门,在商场寻找零售工作时看到珠宝展示柜台被打开,遂生歹意,因而当即决定盗取大量珠宝首饰。经过 RRPDI 的计算,罗伯特的再犯风险等级处于“低等”,威廉的再犯风险等级处于“中高等”。最终,在 A 组中,17 人决定判处罗伯特更长的刑期,13 人决定给予判处二人相同的刑期,13 人决定判处威廉更长的刑期;在 B 组中,仅有 8 人决定判处罗伯特更长的刑期,9 人决定判处二人相同的刑期,23 人决定判处威廉更长的刑期。

并不会流向社会为公众所知,例如刑事被追诉人的轨迹数据信息、账单数据信息以及音视频数据信息等会因刑事诉讼的进行而进入司法场域,成为案件办理的重要数据资源,在此基础上形成的数据信息则在圈存在司法场域内。

(四)技术外包:人工智能技术权力演化的知识基础

技术“外包”(outsource)是目前刑事司法领域智能系统所采取的普遍的研发模式。所谓技术“外包”是司法机关将审判智能系统研发所依赖的专业技术交由司法机关以外的数据服务商或者科技公司等主体,司法机关主要提供司法活动中产生的大数据,同时向技术公司准确描述智能系统所应具备的目标需求。当前各地法院智能办案系统合作研发情况如表1所示。

表1 法院智能办案系统合作研发情况

| 智能办案系统名称 | 研发法院 | 合作主体 |
|------------|------|--------------|
| 上海刑事审判辅助系统 | 上海高院 | 科大讯飞 |
| 北京“睿法官” | 北京高院 | 华宇元典 |
| 重庆“易审”平台 | 重庆高院 | 百度智能云 |
| 浙江“智慧法院” | 浙江高院 | 阿里巴巴集团 |
| 安徽“类案指引项目” | 安徽高院 | 安徽富驰信息技术有限公司 |

技术“外包”的研发模式是智能系统研发中公私主体的一次合作,但更为重要的是,这一过程中存在着法学专业知识与智能技术知识两套知识体系的分野,这种专业精细的学科知识为刑事审判场域算法权力的产生提供了知识基础。法国哲学家福柯认为权力与知识二者密不可分,相互勾连,知识造就权力,权力生产知识^{[7]20-25},可以说掌握专门知识的人就是掌握权力的人。

从目前刑事审判智能系统的研发实践来看,科学技术知识与法学专业知识之间在一定程度上处于隔绝的状态^[8],法学专业知识与智能算法的契合度较低,这两套知识体系的差异形成了智能算法技术对司法人员的支配和控制。办案法官并不了解智能系统的运行原理,智能办案系统在收集、筛选以及处理数据时因“黑箱原理”难以完整呈现其具体运算过程,加之前文提及的“自动化偏差”认知机制的影响,法官无法对智能系统生成的结论进行深刻的批判与质疑,最终使得法官臣服于智能系统算法决策的结果。这种技术“外包”研发模式所导致的法学知识与智能算法知识之间的不平等性,推动了刑事审判场域中智能系统的权力化。

(五)系统架构:人工智能权力演化的现实动力

架构,又称软件架构(software architecture),是计算机科学中的专业术语,架构并非软件,而是软件设

计的宏观策略,为特定问题提供抽象解决方案。架构的本质是其所包含的解决特定问题的既定路径或预设步骤所产生的刚性效力,核心在于对架构内行为主体的规制作用。人工智能时代,算法搭建了诸如淘宝、亚马逊以及脸书等一系列电商平台、社交媒体等系统架构,这些系统架构规范着系统内用户的具体行为。以亚马逊公司的“混沌存储算法管理仓库”为例,在此智能仓库内,货物先由算法根据其体积、种类以及货架空间进行摆放、存储,人工通过算法定位所需货物,根据架构指引进行取货调货,这一过程可以说是人的行为受到了系统架构的支配^[9]。

具体到刑事审判领域,以“上海刑事案件智能辅助系统”为例,这一系统也可以看作是一个系统架构,该系统为实现刑事证据标准的指引功能,预先设定了实现该功能的路径与规则,例如先根据地方司法机关制定的类案证据标准构建不同案件的证据模型,根据不同模型形成刑事案件证据指引的知识库,在此基础上根据刑事诉讼活动不同阶段的不同证据标准,将知识库嵌入系统,对办案人员的证据收集行为进行引导。在这一过程中,智能办案系统产生了对办案人员的约束力,这一系统会对办案人员收集的证据进行校验和审查,如果证据出现瑕疵,除非司法办案人员能够对瑕疵证据进行补正或说明,那么该系统就会自动阻断证据进入下一个阶段。在司法实践中可能出现办案人员为了使证据标准能够符合系统要求,不得不按照系统的要求再次进行证据收集与补正,在这一过程中,刑事审判智能系统凭借架构优势所形成的规范效力,支配着办案人员现实的证据审查行为。

三、刑事审判人工智能权力化的正当性风险

人工智能技术作为不同于刑事司法的社会系统,其并不具备传统刑事审判的中立、公开、准确等正当性品格,反而可能会对其造成侵蚀或不当干预。因而,有人将数据时代的人工智能算法称为“数字杀伤性武器”,随着人工智能技术与刑事司法的深度融合,其引发的正当性风险也将逐渐显现。

(一)规训风险

人工智能技术权力作为一种弥散于刑事审判场域的权力范式,其与作用主体之间形成了“泛在”的权力关系,表现为对人或主体无处不在的引导或操纵,这种“无意识”的权力治理术规训着法官在裁判过程中的自主性。

一是智能办案系统对法官裁判行为的引导作用。

智能办案系统的证据指引、量刑辅助以及类案推送功能通过构建数据模型实现对法官办案行为的导引,这种数据模型形塑了法官办案的行为。法官在智能办案系统中无外乎一个助推系统运行的“零件”,无论办案法官自身有着怎样的办案惯习,在智能办案系统整齐划一的运行指令下,他们认知案件事实的步骤必须满足智能办案系统的要求,司法办案人员的行为受到智能办案系统的调控与约束,法官的主体性在这一过程中可能会遭受消解,导致机械司法的出现。

二是智能办案系统对法官裁判行为的监督作用。大数据技术能把法官在办案过程中所生成的全部数据悉数保存,形式不仅局限于以往审结率、上诉率等以数字呈现的数据,同时也包括如法官的庭审视频、语音记录以及电子卷宗等等,法官无异于置身巨量数据所形成的“数据牢笼”之中。仍以上海 206 系统的偏离预警功能为例,“在法官决定判决时,写好判决书放入系统里,系统会自动提示跟本院以及上级法院有将近 85% 的类似情况下的判决是不一样的,问法官是否坚持这样做,若坚持该判决,那么系统将自动把判决推送给庭长以供讨论。”^①这种智能办案系统的监测预警功能与办案法官之间形成了微妙的控制关系,智能办案系统通过对法官判决的监测,可能成为规训法官裁判符合其预警标准的调控技术。

(二) 排斥风险

随着智能办案系统逐渐深入刑事审判程序的关键领域,诸如刑事证据审查、辅助量刑等关切刑事被追诉人程序性权利与实体性权利的决策由智能系统作出,然而智能办案系统排斥刑事被追诉人参与到司法决策的程序中。

一是算法黑箱限缩了刑事被追诉人行使程序性权利。智能办案“算法黑箱”的封闭性加剧了刑事审判中控辩双方信息不对称的状况,使本就处于弱势地位的刑事被追诉人难以行使程序性权利。以质证为例,最高人民法院《刑事诉讼法解释》第七十一条规定:“证据未经当庭出示、辨认、质证等法庭调查程序查证属实,不得作为定案的根据。”证据质证是被追诉人程序参与,实现庭审实质化的重要方式,被告人行使质证权的前提是能够提前获知控方证据的具体信息,作为质证意见的依据。但是,面对控方借由智能系统生成的指控证据,由于算法黑箱的存在,被追诉人难以获知智能系统生成结论依赖的海量数据及智能算法,即便能够获知这些信息,被追诉人也缺乏相关知识进行计算和理解。例如在一起涉税案件中,整个案件涉及的资金

流水数据达到上亿条,账户数据上万个,凭借人力难以进行梳理与分析,检察机关通过搭建数据模型,利用数据挖掘等技术刻画了资金来源和去向的案件网络^[10],面对这一案件中数据模型生成的关于资金网络的证据材料,数据量巨大,数据关系复杂,被告人及辩护律师难以对其有效进行质证。

二是智能系统的商业秘密特免权排斥刑事被追诉人获知智能办案系统的代码信息,被追诉人难以参与到事关自身权益的决策程序中。这一现象在国内并未引发争议,在国外已经成为智能办案系统妨碍被追诉人正当程序权利的争论焦点^[11]。当受智能办案系统决策影响的刑事被追诉人提出公开智能办案系统源代码(source code)的请求时,这些商业科技公司以这些代码属于本公司的商业秘密,应享有商业秘密特免权(trade secret privilege)保护为由拒绝公开智能办案系统的代码信息,并且这一主张得到了法院的支持。今后,商业秘密特免权在我国很可能成为横亘在智能办案系统与刑事被追诉人之间的鸿沟。2014 年以来,我国人工智能专利授权量增速较快,从人工智能专利的申请主体而言,2017 年排名前 20 的中国申请人仅中国科学院为科研单位,其余均为企业申请人^②。同时,我国刑事诉讼法第五十四条第三款规定:“对涉及国家秘密、商业秘密个人隐私的证据,应当保密。”这种商业秘密特免权在今后可能成为被追诉人参与智慧司法决策的阻碍。

(三) 误判风险

科技系统与司法系统分属两个不同的社会子系统,二者在目标利益与运行逻辑上差异较大,例如科技注重对效率的追求,司法更强调公平而兼顾效率,这在刑事司法领域可能导致误判的风险。

一是智能办案系统的转译模式导致的误判风险。在刑事司法场域,为了保障智能算法能够识别法律文本,需要对其进行代码化转译。一方面,从法律文本的转译主体而言,智能系统的设计工程师和程序员将法律规范及相关司法解释转译为能够被算法识别运算的程序代码。但是,程序员并非法定的立法主体和法律解释主体,同时其也缺乏相关的法律知识,在转译的过

① 最高法院改办何帆:中国法院正努力把人工智能引入办案系统[EB/OL]. (2017-07-29)[2021-01-24]. <http://finance.sina.com.cn/roll/2017-07-29/doc-ifyinryq7012174.shtml>.

② 国家知识产权局规划发展司. 2017 年我国人工智能领域专利主要统计数据报告[EB/OL]. (2018-10-18)[2021-01-24]. <https://www.cnipa.gov.cn/20181114100805476353.pdf>.

程中程序员可能会曲解、简化法律文本,存在着被程序员简化或误读而改变立法原意的风险。例如,在美国,从2004年9月到2007年4月之间,程序员将900多条理解错误的法律规范编写嵌入到科罗拉多州的公共福利系统(CMBS)中,这一失误波及所有受该系统影响的人群,造成了成千上万受该系统配给社会福利的人丧失了公共福利待遇^[12]。另一方面,法律文本转译的迟滞性问题可能会带来智能办案系统的准确性降低。政策法规向计算机代码的转换之间具有一定的时间差,一些新近修订的政策法规无法及时嵌入自动化决策系统从而导致系统决策准确性降低。

二是智能办案系统的数据处理方式导致的误判风险。一方面,在数据输入层面,刑事司法人工智能离不开司法大数据的喂养,在数据与智能的关系上,即“rubbish in, rubbish out”(垃圾进,垃圾出),司法大数据的质量一定程度上决定了司法人工智能决策的质量。我国目前司法大数据仍然存在着不充分、不真实、不客观以及结构化不足的特点^[13],匮乏且低质的数据输入难以保障智能办案系统的准确性。另一方面,在数据处理层面,如果说数据是人工智能运行的“燃料”,那么处理数据的算法即是人工智能的“引擎”。在智能时代,算法已经逐渐脱离“技术中立”的价值立场,一些对性别、地域、种族等因素的预设观念在进行算法设计的同时也被植入到智能办案系统中,由此导致的“算法偏见”“算法歧视”问题也成为掣肘智能系统办案准确性的重要因素。

四、刑事审判中人工智能技术的规制路径

面对刑事审判场域中人工智能的权力化趋势,人工智能规制更应注重技术的权力倾向以及由此引发的对相关主体的引导与操控。因而,刑事审判场域中人工智能技术的规制应主要从智能系统的操作主体——裁判者、智能系统的作用对象——辩护方以及智能系统自身三方面展开。

(一)裁判自主化的适用机制

首先,根据案件类型合理安排智能系统的适用。目前,我国刑事辅助办案智能系统主要具备证据标准指引查验、量刑辅助以及案件裁判偏离预警等功能,但是目前刑事智能系统的发展尚不完善,仍旧存在着案件情节提取准确度不足、类案识别准确率偏低以及模型训练的样本瑕疵等问题。为了保障准确审判案件,保障被追诉人的合法权利不受智能系统决策的影响。一方面,根据案件的复杂程度以及量刑幅度决定是否

适用智能系统辅助决策。倘若案件情节复杂、量刑可能偏高的案件应当减少智能系统的使用。另一方面,根据案件中被告人及辩护人的认罪态度决定是否适用智能辅助系统。若被告人不认罪或者辩护人做无罪辩护的,法院不宜使用智能系统,防止智能系统产生的决策在认知偏差的作用下使法官产生预判从而导致未判先决的情况发生^[14]。

其次,根据智能办案系统的功能合理选择智能系统的适用。根据目前智能系统的功能,可以将其分为预警性功能与引导性功能,前者如案件审判偏离预警、量刑幅度预警,主要提醒裁判者案件审理的程序或实体性事项,注重防错纠偏,这种功能往往集中于事后纠偏,对裁判自主性影响相对较小;后者如证据审查指引、量刑辅助以及类案指引等,这类功能倾向于事前引导,可能会对裁判者产生心理锚定,影响司法权的自主运行。因此,对于预警性功能而言,裁判者要认真比对、核查人工决策与系统预警之间的差异,依据系统的预警提示作出的更改应当标注留痕。对于引导性功能而言,可以借鉴美国法院使用COMPAS系统的经验,可以对裁判者提供办案系统局限性的书面提示(written advisement of its limitation),告诫智能系统的操作者(法官)该系统的准确性、透明性尚不完备,系统决策仅具有参考意义,在参照决策时应当审慎。

最后,裁判者应提升自身的综合素养。刑事智能系统对裁判者的控制支配源于裁判者的法学知识与算法的编码知识之间的专业槽,这种专业之间的差距形成了算法对裁判者的相对权威,破解这种权力关系的途径即是裁判者在自身的专业技能之外,能够对大数据、智能算法以及办案过程中应用到的智能系统的工作原理有一定的了解,提高对刑事司法智能系统决策的认知水平,逐渐破除因知识权威而形成的对智能系统的神秘感与依赖感。

(二)辩护平等化的参与机制

首先,拓宽智能办案系统的公开途径。一是制作并公开刑事司法智能系统的操作指南。美国刑事司法领域中COMPAS系统的操作指南(Practitioner's Guide to COMPAS Core)即面向全社会公开,这份指南由Northpointe公司编写,一共分为六部分,对COMPAS系统的历史发展、系统构成、主要功能、运行原理、系统估值含义以及系统决策的正确率与可靠性作了全面详细的介绍与论证。我国的刑事司法智能系统目前尚未有公开的操作指南,可以根据智能系统的具体功能制作并公开完备的系统操作指南。二是设置

智能系统的算法研发听证程序。可以设立刑事司法智能系统算法评估委员会,委员会里应当有技术专家、法学专家、社会公众代表等人员的参与,就数据来源、数据提取和删除标准、数据赋权比重、算法种类选取、算法运行一般原理以及智能系统主要功能指向等问题展开介绍与讨论,这样可以在一定程度上增强智能系统的透明性,同时也使自身利益受智能算法影响的诉讼当事人更易接受智能办案系统的决策结论。三是关于智能办案系统代码商业秘密保护的问题。根据我国《刑事诉讼法司法解释》第八十一条规定:“公开审理案件时,公诉人、诉讼参与人提出涉及国家秘密、商业秘密或者个人隐私的证据的,法庭应当制止;确与本案有关的,可以根据具体情况,决定将案件转为不公开审理,或者对相关证据的法庭调查不公开进行。”这一规定旨在通过不公开审理等程序性技术措施缓解因商业秘密保护带来的诉讼困难,因此当刑事被追诉人提起公开智能办案系统相关涉及商业秘密信息的请求时,可以通过不公开审理、不公开法庭调查等方式将商业秘密规制于庭审范围内,保障智能办案系统的核心算法信息仅在控辩之间公开。

其次,保障刑事被追诉人能够有效参与智能系统的权利。从数据层面而言,被指控人的辩护律师可以请求查看、修改、更正以及解释智能系统中与自身权益相关的数据。我国刑事诉讼法第四十条规定:“辩护律师自人民检察院对案件审查起诉之日起,可以查阅、摘抄、复制本案的案卷材料。其他辩护人经人民法院、人民检察院许可,也可以查阅、摘抄、复制上述材料。”这一规定赋予辩护律师的阅卷权可以作为刑事被追诉人要求获知智能系统内相关数据的依据。例如辩护律师可以对系统中所录入的前科信息、户籍信息、工作信息以及教育水平等信息进行核查与校正,保障数据的真实性与准确性。从算法层面而言,可以赋予被指控人算法解释权。算法解释权是诞生于民商事纠纷领域内的请求性权利,其在自动化决策领域维护着双方地位平等的法律价值,在刑事司法领域,由于智能系统导致控辩审在信息收集、分析能力上的不对称的现象,司法机关应当对智能系统的算法进行解释。具体而言,解释的主体包括智能系统的使用者和研发者;解释的内容应当主要包括输入的数据类型、算法的运行原理、智能办案系统的主要功能、智能办案系统的结论与现实决策之间的关联性等;解释的标准是以受系统决策影响的被指控人及其辩护律师能够理解为限度,鉴于算法黑箱与技术专业性,解释主体不必确切描绘系统运

算的具体过程细节;解释的方式可以通过口头解释或者通过可视化的动画演示。

最后,司法机关应当履行告知义务。刑事被指控人应对智能系统抗辩权的行使离不开司法机关的有效告知。司法机关在应用刑事智能系统之前,尤其是在影响被指控人权利的领域,例如辅助量刑时,应当告知犯罪嫌疑人、被告人其司法决策的决定过程掺入了智能系统的计算结论。例如公诉机关应用智能系统量刑辅助功能制作量刑建议书时应当告知犯罪嫌疑人,法官在应用智能系统进行证据查验、量刑辅助或者被告人的再犯可能性审查时,也应当告知被告人。如果犯罪嫌疑人、被告人及其辩护律师明确表示不愿意适用智能系统的自动决策结论,应当尊重其选择,退出依靠智能系统辅助司法决策的决议过程,进行人工审查决策。

(三)决策可靠化的研发机制

首先,提高刑事司法智能系统的技术研发水平。一是提高司法大数据数量与质量。目前我国刑事司法智能系统的可输入运算的数据量仍较小,上海刑事案件智能辅助办案系统仅学习了1.5万份卷宗材料,这一问题应当给予司法机关和技术部门一定的时间寻求技术上的突破,中国的司法数据是较为丰富的,我国法院系统也正在不断地将过往裁判文件数据化。同时,应当注意输入系统内的数据与案件事实的相关性、地域差异等因素。二是保障法律文本向程序代码转译的准确性。一方面,司法机关要主动对嵌入系统中的法律文本进行筛选,对已经失效或修订过的法律规范进行剔除与校正,对智能系统进行周期性评测,保障系统运算结果的准确性。另一方面,应当促进法学知识与算法知识的融合,技术人员对法律文本进行代码化的过程中,司法人员应当介入并对相关专业术语作出解释,将司法专业知识与司法实践知识结合,将司法理论知识与计算机科学知识结合,确保办案系统研发的专业性与准确性。

其次,对刑事司法领域中的智能系统进行第三方评估监管。第三方监管评测是国外规制刑事司法风险评估工具的常见举措。例如,哈佛大学法学院开展了对美国若干州内审前风险评估工具效能评估的项目,对美国若干州内的公共安全评估(public safety assessment, PSA)系统的功效进行评测,目的是为了测试风险评估工具生成的审前羁押决策(pretrial incarceration decisions)的准确性和实效性。我国可以考虑将科研院所、高校以及非盈利机构作为智能办案系统评估的主体,将智能系统的数据采集、算法模型选取以

及系统决策错误率作为评测的主要内容,对国内刑事智能办案系统的可靠性进行评测。此外,应当注意的是第三方评估的周期性,应当定期开展系统评估,为司法机关与技术部门提供长效稳定的评估报告。

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): 46-53.
- [2] 达马斯卡. 漂移的证据法 [M]. 李学军, 等译. 北京: 中国政法大学出版社, 2003.
- [3] 贝尔纳·斯蒂格勒. 技术与时间: 爱比米修斯的过失 [M]. 裴程, 译. 上海: 译林出版社, 2000.
- [4] 顾世春. 斯蒂格勒人的代具性思想研究 [J]. 创新, 2017(2): 54-58.
- [5] CUMMINGS M L. The social and ethical impact of decision support interface design [C]// WALDEMAR K. International encyclopedia of ergonomics and human factors, London and New York: Taylor and Francis, 2006.
- [6] STARR S B. Evidence-based sentencing and the scientific

rationalization of discrimination [J]. Stanford Law Review, 2014, 66(4): 867-869.

- [7] 福柯. 规训与惩罚 [M]. 刘北成, 杨远婴, 译. 上海: 上海三联出版社, 2012.
- [8] 王禄生. 大数据与人工智能司法应用的话语冲突及其理论解读 [J]. 法学论坛, 2018(5): 137-144.
- [9] 张凌寒. 算法权力的兴起、异化及法律规制 [J]. 法商研究, 2019(4): 63-75.
- [10] 何家弘, 邓昌智, 张桂勇, 等. 大数据侦查给证据法带来的挑战 [J]. 人民检察, 2018(1): 54-57.
- [11] WEXLER R. Life, liberty, and trade secrets: intellectual property in the criminal justice system [J]. Stanford Law Review, 2018, 70(3): 1343-1429.
- [12] CITRON D K. Technological due process [J]. Washington University Law Review, 2008(6): 1249-1313.
- [13] 左卫民. 关于法律人工智能在中国运用前景的若干思考 [J]. 清华法学, 2018(2): 108-124.
- [14] 马靖云. 智慧司法的难题及其破解 [J]. 华东政法大学学报, 2019(4): 110-117.

(责任编辑: 张园)

The Power Logic of Artificial Intelligence in Criminal Trials

WEI Chenshu

(Institute of Evidence Law and Forensic Science, China University of Political Science and Law, Beijing 100088, China)

Abstract Criminal intelligence case handling system is the main manifestation of artificial intelligence technology in the criminal justice field. At present, the criminal intelligence case handling system is mainly concentrated in the fields of evidence review and judgment, auxiliary accurate sentencing, delivery of category cases and warning of case deviation. As the deep integration of artificial intelligence technology and criminal justice, the “prothèse” of technological philosophy, automation bias recognition mechanism, endorsement and support of state power, the model of technology outsourcing and system architecture and so on are driving the transformation of the intelligent case handling system from “tools” to “power”. At the same time, the power of artificial intelligence technology in criminal trials may cause disciplinary risks, rejection risks, and misjudgment risks. In order to effectively avoid the legitimate risks caused by artificial intelligence technology in the criminal trial, the idea of power regulation should be established, and the three aspects of application mechanism of criminal intelligent case handling system, participation mechanism and research and development mechanism should be regulated.

Key words criminal trial; artificial intelligence; technical power; intelligent handling of cases; due process; judicial jurisdiction