

欢迎按以下格式引用:陈群辉,刘聪.法律大模型赋能数字法院建设的价值边界[J].长江大学学报(社会科学版),2025,48(2): 101-107

法律大模型赋能数字法院建设的价值边界

陈群辉 刘聪

(长江大学 湖北省域基层治理研究中心,湖北 荆州 434023)

摘要:法律大模型作为新兴技术,通过“大模型+生成式”架构革新了法律人工智能范式。凭借语义理解、非结构化数据处理及多场景嵌入能力,法律大模型为数字法院建设提供了统一、高效的技术基座。然而,技术赋能有其边界,可分解为技术可行性、法律正当性、社会接受性三大维度的限制。对此,应以有限辅助为原则,确保法官主导裁判权;保障程序供给,防范技术约化倾向;坚持以实质正义为核心,平衡技术工具性与经验法则,确立法律大模型赋能的价值边界。

关键词:法律大模型;数字法院;价值边界;技术赋能

分类号:D926.2 **文献标识码:**A **文章编号:**1673-1395(2025)02-0101-07

一、问题的提出

人们创造实用性人工智能的早期尝试,就是通过规则或事实的集合,以明确方式将人类的专业知识编码到计算机系统中。但是,世界上的许多事物并不是以离散方式组织起来的,也并不容易被还原为简单规则或符号表征。^{[1](P413)}正因如此,符号主义法律人工智能虽在司法数字化的浪潮下大规模嵌入法院建设,但并未带来预期赋能效果,实践与话语冷热差异凸显。^[2]2022年,ChatGPT横空出世,凭借Transformer架构、端到端学习等技术,这一联结主义人工智能里程碑式的产物迅速成为科技热点,也标志着全球科技竞争进入“大模型”(亦称“大语言模型”)时代。面对这一新质生产力的时代风口与前沿趋势,司法界也以高度开放和进取的姿态展开布局。经过短暂的技术积累与沉淀,最高人民法院于2024年11月15日发布国家级AI基础设施——“法信法律基座大模型”,为国内法治领域奠定了“行业基座”。^[3]紧接着,最高人民法院于2024年12月26日

发布《人民法院第六个五年改革纲要(2024—2028年)》,明确提出“加强数字法院基础设施建设”“强化智能算法、大语言模型、数字模型等技术应用,充分发挥人工智能技术辅助办案、集约事务、防控风险、方便诉讼等作用”等改革任务。^[4]由此,法律大模型进一步嵌入数字法院建设已是无法阻挡的潮流。

人民法院的数字化发展呈现“积木式创新”模式,大致可分为法院信息化建设、智慧法院建设、数字法院建设三个演进阶段。^[5]当前的数字法院建设战略是对大数据时代的系统性回应,以对数据积极拓展应用为底层逻辑,重视制度框架与科技要素的并驱效应。法律大模型的嵌入不仅体现于内部性、场域性、功能性的工作范式变革,还将为法院的工作流程、方式、机制带来重塑化革新。应对这一创新驱动的变局,必须多角度、全方位加强理论供给。此前,学界对于法律大模型的研究大多使用其上位概念“生成式人工智能”,近期的研究成果则逐渐开始转向“法律大模型”本身。在此概念下,李鑫提出了法律大模型的构建机理与实践路径^[6],邹劭坤等探

收稿日期:2025-02-20

基金项目:湖北省域基层治理研究中心课题“数字技术赋能地方网络暴力信息治理路径研究”(20241217);长江大学国家级大学生创新训练计划项目“生成式人工智能在司法实务中的应用研究”(202410489006)

第一作者简介:陈群辉(1969—),男,湖北石首人,副教授,主要从事民商法研究。

讨了法律大模型在数字法院内的可能应用场景与挑战^[7],魏斌总结了法律大模型技术的优缺点与应用挑战^[8]。整体而言,既有研究多聚焦于法律大模型在数字法院内的应用场景与风险探讨,而对于技术赋能的边界及其衡量尺度的研究则呈现碎片化、附带化的特点,缺乏系统性聚焦。有鉴于此,本文将梳理法律大模型赋能数字法院建设的当代图景,分解技术赋能的限制要素,并据此划定法律大模型的赋能边界,以期发挥理论研究对数字法院建设的指导与推动作用。

二、法律大模型赋能数字法院建设的当代图景

数字法院建设缘起于司法资源与司法需求的差异化增长。司法领域长期受制于“案多人少”的社会结构性矛盾,审判模式也亟须突破固有的制度性约束,数字法院以数字化为改革目标,由此产生了很多革命性后果,形成了双层空间-虚实同构、人机共处-智慧互动、算法主导-数字生态的时代特征。^[9]法律大模型则作为司法新质生产力,承接了来源广泛、样式众多的数据处理需求。同时,凭借高通量、强交互、易嵌入的技术特征,法律大模型将为审判具体业务场景提供人工智能解决方案,有望实现审判工作与大数据、人工智能联结方式从低频率、低黏度向高频率、高黏度的转变,法律大模型赋能数字法院的当代图景就此展开。

(一)法律大模型的技术革新

在法院信息化、智能化建设浪潮中,人工智能的嵌入可大致分为三个阶段。一是专家系统路线,尝试通过人工预设规则和推理路径来打造机器法律专家;二是知识图谱路线,尝试通过将法律规定、法律文书、证据材料及其他法律资料中的法律知识点以一定的法律逻辑联系在一起,形成概念框架;三是深度学习路线,通过循环神经网络、卷积神经网络、深度神经网络等技术的综合运用,实现智能化的要素识别,但目前国内这方面的研究和应用尚处于初始阶段。^{[10](P337)}总体来看,既往法律人工智能大都受限于自然语义处理与知识生成的技术瓶颈,在可持续发展的制度化逻辑方面与既有实践成果之间尚未实现科学衔接,功能完善、普适性、覆盖面等指标均有待进一步发展完善^[11],因而未曾出现能够让法院系统普遍实现质效跃迁的技术。

法律大模型(如法信法律基座大模型)是在通用

大模型基座上,投入海量的法律数据语料,并经过预训练、优化训练、监督微调等方式构建而成。相较于传统法律人工智能,其技术革新体现在以下三个方面。其一,模型范式的改变。传统范式采用“小模型+判别式”的技术架构,事实上难以胜任法律领域自然语言处理(NLP)技术的需求;法律大模型则凭借“大模型+生成式”的模型构造,将有效破除法律自然语言处理能力的发展瓶颈。其二,交互方式的变革。传统法律人工智能的交互方式仅限于界面任务模块的选择与响应,缺乏高频信息交互;法律大模型并未预设任何任务单元,所有交互均在界面以自然语言完成,这种“以言行事”的方式极大地增加了交互效率与体验。其三,嵌入能力的增进。以往法律人工智能需要面向应用场景进行专门化开发,导致技术路线过于封闭僵化,赋能路径严重受限;法律大模型可以基于不同提示词完成不同任务,使其嵌入任何应用场景都能保持高响应能力,可嵌入性得到增进。

(二)法律大模型赋能数字法院建设的实践逻辑

数字法院的底层需求为数据处理,即通过数据筛选、对比、碰撞产生信息聚类价值,由此可引申出数字法院对于技术工具的核心要求——大数据处理能力,但法院的绝大部分数据都以非结构化的文本形式呈现,机器直接处理效果不甚理想,所以,实务中用于机器学习的数据需要依靠人工筛选、清洗与分类。在这种情况下,各类人工智能工具的“智能”本身都严重依赖人工构造,更遑论承接形式多样、结构复杂的数据处理工作。法律大模型在非结构化数据处理上具有良好表现,凭借其在语义理解、命名实体识别(NER)与关系抽取(RE)等基本任务单元上的出色表现,法律大模型可以对非结构化法律文本进行自动化解析与信息抽取。不仅如此,法律大模型可以联结、转化大量法律文本模式与结构,使其具备强大的上下文理解能力,能够在高信息熵的中文法律语言中理解上下文之间的语义关联,从而提升信息处理的精确度。

法律大模型赋能数字法院建设还体现在其对于多方需求的整合能力。法院信息化建设计划实施以来,全国各级法院根据自身需求,先后开发出众多的人工智能工具。这些智能工具应用大都存在局限性,且不同地区法院的技术路线基本不能兼容,使得“试点-推广”的愿景面临南橘北枳之忧。大模型技术则通过统一的大规模语言模型,改变了传统人工智能分散化、碎片化的建设思路。^[12]凭借通用化的

模型架构,法律大模型可以嵌入全国任何法院,服务任何场景。这与数字法院建设的整体性、系统性战略需求相符,使得全国法院建设可以基于同一个技术底座,不断横向延伸,发展新应用,同时保障纵向监管深度,充分回应系统性响应、集约化管理的现实需求。

三、法律大模型赋能数字法院建设边界的多维分解

法律大模型对司法实践产生了系统性影响,使数智化司法审判从憧憬性的模糊概念逐渐变为可触碰的现实。但在对其附魅的同时,应当认识到“倘若有人只想使直接成本最小化,则错误成本可能升得很高”^{[13](P23)}。审判不能因片面追求效率而悬置公平正义,应用任何技术手段都不能超出其应有范围。法律大模型的嵌入是全方位、系统性的,在对其赋能边界进行考察时,有必要依据其技术表现,阐明各维度的限制条件,为确立价值边界设定基本框架。

(一)技术可行性边界

从技术原理上来看,以大模型、大算力、大语料为基础的法律大模型完成了法律领域自然语言处理任务的性能跃迁,可以在准确理解用户意图的基础上进行上下文贯通的多轮对话。维特根斯坦认为,语言的边界就是思想与世界的边界。^[14]司法场景中的一切事务都被囊括于“语言之网”中,法律大模型在自然语言理解(NLU)与自然语言生成(NLG)上的优越性能将被直接转化为对于司法工作的赋能。尽管如此,法律大模型与真正意义上的“AI法官”仍相去甚远,司法实践中依然有大片区域是其无法真正涉足的。从司法审判的现实需求衡量,这主要源于现行技术路线在以下两方面的短板。

其一是逻辑推理能力的短板。司法实践中,法官以形式逻辑作为裁判基础,遵循“三段论”演绎推理方式,其中往往涉及长链逻辑和多步骤推理,需要一系列法律规范的连贯适用。法律大模型的推理是基于海量数据的模仿,追求对于既往模式的复现,所展现的往往是统计关联而非严密的逻辑推导。例如,有观点认为,大模型难以理解裁判任务的本质,可能永远无法进行真正的推理、想象或超越瞬间的思考。^[15]不仅如此,法律大模型还在“大前提”(即能引起法律效果的法律规范)的选择和验证上存在缺陷。法律规范分为法律规则和法律原则,法律规则的适用采取“涵摄”即全有或全无的方式;法律原则的适用则需要穷尽规则、更强理由或个案正义,适用

时可以叠加。在规则层面上,法律大模型可以通过统计复现来进行正确选择,但原则选择往往是默会的、经验性的,需要人类的理性与法治精神,因而也是法律大模型难以触及的。加之大模型缺乏反思与纠错能力,无法将“目光反复流转于规范和事实之间”,进一步造就了其推理能力的不足。

其二是准确输出能力的短板。自大模型诞生以来,围绕其输出准确性问题的诟病便从未止歇,技术术语中,大模型不准确或无意义的输出被称为“幻觉”(Hallucination)。“一本正经地胡说八道”,便是对其幻觉问题的最好概括。应当承认,法律大模型在通用大模型基座上应用数据标注工程、知识体系工程、检索增强生成技术(RAG)等手段后,幻觉问题的确可以得到很大程度的改善。但作为千亿参数级别的大模型,幻觉问题可能由诸多因素叠加形成,难以根除。例如,数据标注工程可能因标注人员的主观主义或者疏忽大意而造成机器识别位点的不确定性,知识供给领域最为前沿的 RAG 技术也可能因知识库本身存在错误内容或者索引机制不够完善而导致输出精度下降。所以,法律大模型的应用理应以“可能形成不准确输出”为前提性认识。

(二)法律正当性边界

法律大模型作为一种能动要素嵌入实务工作能够有效提升司法生产力,但这种有效性主要是工具性的,而合法性是评价性的。^{[16](P47)}质言之,法律大模型作为效率工具,不能突破司法审判在实体和程序上的正当性要求。司法系统的正当性主要源于其公开性、公正性与程序严谨性,而法律大模型的技术特性与司法价值的张力在此领域尤为凸显。

首先,算法黑箱可能有悖于司法公开理念。司法公开不仅要求裁判结果的公开,更要求裁判逻辑的可追溯与说理过程的明晰化。传统司法模式中,法官需通过判决书阐明事实认定、证据采信及法律适用的完整逻辑链,使当事人得以理解裁判依据并行使救济权。然而,算法“输入-输出”的决策机制往往隐藏于技术黑箱之中,其内在推理路径、权重分配及数据关联性仍难以被法官乃至开发者完全理解。即便在行为外观上,大模型能够生成看似完备的裁判文书,也无法达到法官裁判说理的水准,因为大语言模型的底层运行逻辑并非推论,而是基于大数据的符号操纵。它从一开始就不要求不同事态之间的统一性,也没有预先确立某种形式系统,所以无法满足“因果推理”的要求。^[17]从这一角度来看,大模型看似“滴水不漏”的文本输出并不能有效解码其决策

黑箱。算法层面的不可解释性实质架空了对于裁判结果充分说理之要求,可能使当事人陷入“知其然,而不知其所以然”的困境。当算法决策过程成为各方都无法审视的“黑箱”时,当事人对错误裁判的异议权与救济权便可能失去着力点,进而影响司法制度的正常运行。

其次,算法歧视可能违反司法公正原则。公正是法治的生命线,公正理念下的人不是数据,更不是电子痕迹的汇总,因而有权利获得基于独立人格的“个殊化判决”,但算法的数据依赖性与模式化推理可能将历史裁判中的隐性偏见固化为系统性歧视。法律大模型的训练数据往往源自既往司法数据,而历史上因法官认知局限或社会环境因素导致的裁判偏差——如劣迹人员犯罪可能性更大,经济犯罪更容易发生在发达地区——可能被算法识别为“客观规律”并加以强化,形成“偏见进,偏见出”(Bias in, Bias out)的情况。美国 COMPAS 系统曾被证实对于黑人的系统性歧视,其本质上也是数据偏见的外在反映。^[18]这种技术性偏差与司法公正原则的冲突具有双重性:其一,算法通过复杂参数构建的“歧视性关联”往往超越人类直觉可察觉的范围,导致歧视更具隐蔽性;其二,算法决策常以“技术中立”为免责抗辩理由,将责任转嫁给数据质量问题或模型复杂性,这与司法机关的主动纠偏义务相悖。

最后,算法驱动可能消解程序严谨性。程序作为“社会制度化最重要的基石”,在司法制度中有着独立的不可取代的价值与地位。相较于实体正义因评判标准的不确定性而难以衡量,以利害关系者的实质参与和程序保障为核心要义的程序正义获致了前所未有的珍视与推崇。^{[19](P22)}从技术角度来讲,法律大模型仅凭案卷材料就可以为法官提供案情概要、证据审查乃至最终判决,司法审判并不需要像以往那样强调法官的亲历性。正因如此,这一强力的智能工具可能引起人为压缩甚至剪裁程序空间,仅借助各方参与达成裁决外观,最终演变为“事实和法律经法官剪裁和取舍用以证成数据博弈的最终结论”的荒唐情形。^{[20](P203)}算法的无序驱动最终可能指向“技术司法”的境地——当事人真正面对的不再是能感知情感、沟通理解的法官,而是无法质疑、不可协商的算法程序,严重突破法律正当性底线。

(三)社会接受性边界

司法作为社会纠纷的断后性解决机制,其权威性不仅源于法律体系的自我证成,更依赖于公众对司法裁决的普遍接受与自愿遵从。对于法律大模型

等智能体嵌入法院系统,乐观者普遍认为,法官的判决和决策模式将从“单一人脑决策转向聚合智脑决策”,进而减少法官的任意性,提升司法裁判的公信力。^[21]诚然,对于知识和科技的信任可以在一定程度上巩固和维系数字法院的司法权威,但也需考虑技术本身带来的风险在自反性意义上对公众信任的冲击。正因如此,技术理性与公众认知之间的鸿沟仍会形塑隐性的社会接受性边界,这主要体现在以下三方面。

第一,法律大模型难以满足社会公众对于实质正义的需求。博登海默认为,正义有着一张普洛透斯般的面孔,变幻无常,随时可呈现不同形状,并具有极不相同的面貌。^{[22](P238)}实质正义与千篇一律、讲求平均主义的形式正义不同,对于当事人来说,实质正义不仅停留在法律规范形式适用层面,更强调对个案特殊性的充分考量以及对人性尊严的终极关怀。这种正义的实现有赖于法官在制度框架内进行创造性的价值判断——当法律条文存在模糊地带时,需要结合立法目的与客观情势进行自由裁量;当法律规则适用导致明显不公时,应当通过法律原则的衡平适用实现矫正正义。法律大模型的司法参与建立在“过去决定未来”的统计正义之上,不涉及微观的个案正义与社会关系的修复,更不具备基于实践的经验法则,因而难以满足社会公众对于独特性、差异性等实质正义的现实需求。例如,于欢辱母杀人案,一审法院以故意伤害罪判处被告人于欢无期徒刑,而二审法院最终认定其行为构成防卫过当,以故意伤害罪改判被告人于欢五年有期徒刑。这一裁判结果的历史性跨越体现了法院对于弱势群体的关注与保护,是司法俯身倾听民意,弥合分歧,最终实现个案实质正义的结果,因而也是法律大模型依靠自然语言处理技术所无法实现的。

第二,法律大模型无法回应社会公众对于人文关怀的期待。司法技艺可分为司法技术与司法艺术两大维度,法律大模型是对司法技术的单向强化,无法复刻法官对于社会情理与人文关怀的认知与实践。传统司法模式中,法官除了需要增进在法律知识和方法方面的技术水平之外,也需要以正常人的角色理解其所生活的社会实际,理解其置身的社群现状,深入社会社群之中理解特定时代的中国社会。^[23]法官在家事纠纷中弥合亲情裂痕,在行政协议纠纷中平衡契约精神与公共利益,在刑事量刑中思考被告人的社会危害与改造可能,实质上都是在进行“法理与人情的对话”。正如哈贝马斯所言,任

何一种沟通交往的过程都发生在文化前理解的背景下。^{[24](P132)} 没有对潜藏在交往行为中的情感脉络、社情民意的理解能力，触及矛盾本质的有效沟通就无法形成。而诸多实验证明，大模型在换位思考和混合情感识别任务上的表现较差，在情感困境中选择有效解决方案时，模型更依赖显式模式而非深层推理，并不具备在法律纠纷中处理复杂情感并施以人文关怀的能力。在此情况下，司法系统对司法技术的狂热追求，可能使得法官特有的“情境化考量”被排除，社会公众将随之形成类似韦伯设想的“司法即自动售货机”的认知，司法审判蕴含的教化功能与社会整合价值也将逐渐消失。

第三，法律大模型加剧社会公众的“技术利维坦”焦虑。大模型技术与司法权力的结合正在催生出一种新的权力形态——“技术利维坦”。传统的司法权运行模式下，承办法官需要与当事人进行联系，非涉密案件的庭审允许旁听，裁判文书上网公开接受审视，权力运行始终处于可见的网状结构之中。法律大模型的嵌入则创造了隐性的权力层，从案情概要到智能量刑建议，这一系列过程都在直接或间接地影响着当事人的权利行使，却因保密程序或者算法黑箱而无法接受任何监督。这种监督失效的危机在刑事审判领域尤为尖锐，即便法律大模型仅承担辅助职能，其介入人身自由权、生命权处分的现实，已足以触发公众对“技术利维坦”的本能抗拒。因为机器参与人类命运裁决的事实本身，已然逾越了社会认知中的伦理边界。不仅如此，个体之于“技术利维坦”的绝对弱势可能瓦解人的自主性，面对国家权力系统掌握数字技术的绝对优势，个体几乎很难找到可行的、有效的维权方式，也无力制衡国家权力体系以数字形式对自身的权力支配。^[25] 这种基于优势地位的技术规训可能引发社会公众的系统性焦虑与抗拒。

四、法律大模型赋能数字法院建设的价值边界确立

司法审判不能完全交由法律大模型，数字法院也不能在算法权力的影响下，将社会公众作为可计算、可预测、可控制的客体。法律大模型对于数字法院的赋能必须符合人本主义的价值观念，由此可以根据上文已述的边界要素确立其价值边界。基于数字司法的技术性、创新性、复杂性和系统性，这个边界并不会是清晰的、确定的、图表式的答案，更多的是一些应遵循的基本原则、程序和机制。^[26] 这

些原则性的条件限制可以防止法律大模型应用的“滑坡”与异化，使其成为兼具能动性、稳定性与谦抑性的技术要素。

（一）以有限范围的辅助性应用为基本原则

在推进法律大模型深度嵌入数字法院建设的过程中，“有限范围”与“辅助性”两大限定条件共同构成技术赋能司法审判的基本原则。所谓“有限范围”，是指通过规范性法律文件或司法政策明确划定法律大模型参与的禁入领域，尤其针对涉及重大公共利益、伦理争议或高度依赖司法自由裁量的案件类型。举例说明，对于刑事诉讼中可能涉及生命权剥夺以及有重大社会影响的案件、民事诉讼中关乎身份关系或未成年人权益保障的案件、行政诉讼中涉及重大社会公共利益的案件等，均需严格排除法律大模型介入裁判环节。之所以只在裁判环节排除，是出于司法效率的考量，文书校对、案卷移送、法条检索等非裁判环节的应用不会对最终结果产生实质性影响，故可不予排除。综合来看，这种限制既源于司法权作为国家公权力的本质要求——其判断过程必须由具备自由意志与责任能力的人类主导，也基于法律大模型在技术路线、人文关怀及社会效果预测等方面的固有局限，不应当在特定领域的裁判环节发挥任何价值。目前，对于“有限范围”的规定较为缺乏，还没有形成体系化的禁入制度。未来，应当发挥最高人民法院在数字法院建设中的引领作用，补全相关规定，推动法律大模型应用的安全、合理展开。

辅助性原则确立了法律大模型在司法审判中的工具属性。正如威廉所说，机器智能更多的应是与人类理性的相互支持，而不是取代与被取代的零和关系。^{[27](P233)} 对此，最高人民法院在 2022 年发布的《关于规范和加强人工智能司法应用的意见》（以下简称《意见》）中就明确了人工智能的辅助审判原则，并强调“无论技术发展到何种水平，人工智能都不得替代法官裁判，人工智能辅助结果仅可作为审判工作或审判监督管理的参考”。这一概括性的规定主要针对的是裁判环节，厘清了法官裁判与机器裁判之间主次关系的同时，也从侧面肯定了机器辅助结果的参考价值。与此同时，根据权责统一原则，责任归属的清晰界定是辅助性原则落地的制度基础。《意见》指出“司法责任最终由裁判者承担”，在此基础上还需确立技术与司法二分的责任框架，若因大模型技术缺陷导致裁判错误，应由外部技术提供方承担产品责任；若法官未履行审查义务而直接采纳

错误建议,则需追究司法主体责任。其中,法官的审查义务不应过于苛刻,因“一般人难以察觉的问题”而造成的过失可予以免责。通过限定责任促进法官学习和使用法律大模型,防止其因过重的责任负担而排斥创新。

(二)以保障程序供给为必要条件

在数字法院建设语境下,原本“重实体,轻程序”的司法价值取向必然得到加强。“正义不仅要实现,更要以人们看得见的方式实现。”以制度化的纠纷解决过程保障行为正当,是实现司法公正、累积司法公信力不可或缺的制度安排或技术装置。^{[28](P152)}保障程序供给的核心在于确保法律大模型的嵌入始终以维护诉讼程序法定性、正当性为基本遵循,并通过程序规范的体系化构建,防止技术工具对审判权力运行模式的异化。从程序法定原则出发,需在立法层面确立技术介入规则,明确其功能实现必须严格遵循的步骤、形式与期限要求。例如,在智能分案、电子送达、线上庭审等数字化司法场景中,需通过司法解释对技术应用的开展形式、效力认定及异议处理机制作出专门规定,确保技术赋能始终运行在程序法预设的轨道之上。与此同时,程序供给还需维护当事人的诉讼权利平衡,在提升司法效率的同时,应当防范技术优势为当事人打上“诉讼客体”标签。这要求数字法院为当事人提供必要的法律大模型服务,既赋予法官对法律大模型的控制权,也保障当事人对大模型介入环节的知情权、异议权及人工复核请求权,避免数字法院落入“程序黑箱”的窠臼。

司法程序的数字化转型还需补足人机耦合的程序规范体系。若没有人机耦合的程序规范,法律大模型就有可能在“滑坡效应”下成为实质性审判者。因为单纯的错案责任承担是结果导向的,无法必然对过程形成规制。由此,应当通过程序规范来确保法官的定位不发生偏离。举例说明,在起诉受理阶段,法律大模型对于虚假诉讼案件的自动化审查只能作为初步程序,虚假诉讼的确定需要法官进行人工复查,建立“机器筛查-人工确认”的耦合式审查机制,防止技术直接剥夺当事人诉权。在证据审查环节,法律大模型的证据分析功能必须服从于证据规则,其生成的证据关联分析或证据链条评估仅可作为法官自由心证形成的参考,且应当通过质证程序接受各方当事人的辩论检验,防止出现“数字心证”。法律大模型应用于裁判文书生成与校验需恪守程序公开原则,直接智能生成的裁判文书必须经过法官实质性修正与理由补强,并在文书中释明法律大模

型的参与环节与程度,通过充分说理要求来保障法官的实质性裁判。通过各类程序规范的叠加应用,确保法官在案件审理过程中的亲历性,有效阻止算法黑箱与算法歧视负面效应传导,并弱化社会公众对于“机器裁决人类”的“技术利维坦”焦虑。

(三)以实现实质正义为核心价值追求

康德认为,法则一般被看作是实践理性产生于意志,准则出现于意志在作出选择过程的活动之中。^{[29](P21)}因而实质正义只能通过人的自由意志来定义与实现。法官对法律大模型的应用需厘清人文性与工具性的主客之序,通过技术工具与经验法则的有机融合,实现裁判效能的跃升。为此,法官需要做如下考虑。首先,法官需要对个案的技术介入程度进行考量,基于案由与复杂程度自主设定技术介入的维度与深度,对事实清晰的简单案件可开放文书生成、类案推送、证据形式审查等全流程辅助功能,而对涉及价值冲突的疑难案件则应当限定于关键信息抽取、法条检索等基础性文本功能。其次,法官需要对辅助结果进行一定的检视。既借助大模型的分析效率快速掌握案件事实与法律关系,又持续保持对其幻觉等输出局限性问题的警觉,自行比对事实概要的上下文关联,评估法律规范适用的社会语境,预防辅助结果的误导化、机械化倾向。最后,法官需要将辅助结果融入裁判逻辑。例如,可以将智能系统推送的类案裁判统计数据转化为论证裁判结果合理性的论据,并在裁判文书中列明。由此看来,实现实质正义,法官需以法律规范为大前提,借助法律大模型减少人类重复性劳动,参考社会政策、人文伦理,最终作出符合人类理性预期的裁判结果。

实现实质正义,还需同步推进法律大模型的正向应用和反向应用,构建技术赋能双向平衡机制。正向应用,是指借助法律大模型实现裁判效率提升。通过智能分案系统优化司法资源配置,利用文书自动生成技术压缩事务性工作耗时,借助关键信息抽取功能、深度挖掘功能提升事实查明效率等,都属于正向应用。此类应用直接面向案件审理,所以受到较多关注。反向应用,是指借助法律大模型实现审判监督管理。基于大模型的类案偏离预警功能,对裁判尺度异常波动进行监测;通过裁判文书质量智能评查系统,自动识别法律适用错误或说理不充分问题等,都属于反向应用。此类应用属于司法管理制度范畴,所以易被人们忽视。《意见》将“审判工作”与“审判监督管理”并列,说明法律大模型的应用并不能因前者而偏废后者。对此,数字法院建设应

当重新审视法律大模型的司法管理功能,通过建立大模型内部应用数据核查与终本案件抽查机制,保证应用过程和结果可预期、可追溯、可信赖。如此一来,将实现技术应用“放权”与“控权”的有机统一,使得数字法院在司法效率与实质正义的张力中达到动态平衡,确保技术工具赋能处于价值边界之内。

五、结语

法律大模型完美契合了数字法院对于统一、集约化技术工具的需求,将在未来法院建设图景中占据重要位置,对此应当以更加开放的姿态去迎接机会、规避风险。^[30]但“科技发展本身就是一种悲喜交加的进步”,在对法律大模型的赋能效果抱有憧憬的同时,应当认识到,它有自身的能力边界,在技术可行性、法律正当性、社会接受性三个维度上都有各自的限制因素,最终形塑其赋能的价值边界。应当以原则性的价值约束来调整与斧正数字法院的新司法生态,确保技术工具性定位与人的自由意志不被减损或异化。

展望未来,应当广泛探索法律大模型在数字法院建设中的应用场景,发掘技术潜力、激活沉睡数据。同时,也要防止技术理性的约化主义倾向,始终确保司法审判遵循“人在回路”原则。还需注意到,司法并非完全存在于宏观叙事,每位当事人在面对时代潮流时的内心境遇也应当成为法律人的关切内容。数字法院应当建立一套面向当事人的诉讼服务体系,让每一位当事人都能享受技术工具带来的红利,以缩小公共法律服务与司法机关之间的“数字鸿沟”,进而让人民群众在每一个司法案件中感受到公平正义。

参考文献：

[1](美)亨利·基辛格,(美)埃里克·施密特,(美)丹尼尔·胡滕洛赫尔. 人工智能时代与人类未来[J]. 胡利平,风君,译. 北京:中信出版社,2023.

[2]左卫民. 热与冷:中国法律人工智能的再思考[J]. 环球法律评论, 2019(2).

[3]平钰骁. 国家级 AI 基础设施正式发布! 最高法召开“法信法律基座大模型”研发成果新闻发布会[EB/OL]. <https://mp.weixin.qq.com/s/Dya3YvbZ9oKSTqaGoXUQSQ>, 2025-02-03.

[4]最高人民法院新闻局. 人民法院第六个五年改革纲要(2024—2028 年)[EB/OL]. <https://mp.weixin.qq.com/s/kna3nqbVRwAYR1E3ix-MqQ>, 2025-02-03.

[5]贾宇. 论数字法院[J]. 法学研究, 2024(4).

[6]李鑫. 法律大模型构建的模式选择和实践路径[J]. 吉首大学学报(社会科学版), 2025(1).

[7]邹劭坤,刘奕群. 司法人工智能在数字法院生态构建中的应用[J]. 数字法治, 2025(1).

[8]魏斌. 法律大语言模型的司法应用及其规范[J]. 东方法学, 2024(5).

[9]景汉朝. 数字时代在线诉讼模式特有原则与制度构建[J]. 华东政法大学学报, 2025(1).

[10]姜伟. 法律人工智能导论[M]. 北京:北京大学出版社, 2023.

[11]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[J]. 比较法研究, 2022(1).

[12]王禄生. ChatGPT 类技术:法律人工智能的改进者还是颠覆者?[J]. 政法论坛, 2023(4).

[13](美)迈克尔·D·贝勒斯. 法律的原则——一个规范的分析[M]. 张文显,等,译. 北京:中国大百科全书出版社, 1996.

[14]Worthington B. A. Ethics and the Limits of Language in Wittgenstein’s Tractatus[J]. Journal of the History of Philosophy, 1981(4).

[15]Michelle Mohny. How Chatgpt Could Impact Law and Legal Services Delivery [EB/OL]. <https://www.mccormick.northwestern.edu/news/articles/2023/01/how-chatgpt-could-impact-law-and-legal-services-delivery>, 2025-02-05.

[16](美)西摩·马丁·李普塞特. 政治的社会基础[M]. 张绍宗,译. 上海:上海人民出版社, 2011.

[17]郭贵春,梁德柱. 大语言模型应该有“内在的语言生成能力”吗? ——论乔姆斯基批评 ChatGPT 的局限性[J]. 科学技术哲学研究, 2025(1).

[18]陈杰. 宪法平等框架内的算法公平标准与价值权衡问题——兼论 COMPAS 算法公平争议的启示[J]. 浙江社会科学, 2023(7).

[19](日)安平. 程序的正义与诉讼[M]. 王亚新,刘荣军,译. 北京:中国政法大学出版社, 1996.

[20]左卫民. 中国法律实证研究(第 3 卷)[M]. 北京:法律出版社, 2018.

[21]赵杨. 人工智能时代的司法信任及其构建[J]. 华东政法大学学报, 2021(4).

[22](美)埃德加·博登海默. 法理学——法哲学及其方法[M]. 邓正来,姬敬武,译. 北京:华夏出版社, 1987.

[23]王静. ChatGPT 技术嵌入智慧司法的伦理风险及其法治应对[J]. 上海政法学院学报(法治论丛), 2023(4).

[24](德)尤尔根·哈贝马斯. 交往行为理论(第 1 卷)[M]. 曹卫东,译. 上海:上海人民出版社, 2018.

[25]陈兰馨,沈桂龙. 驯服还是被驯服:数字利维坦下人的自主性问题[J]. 学术月刊, 2024(10).

[26]马长山. 数字司法的法治边界[J]. 东方法学, 2024(4).

[27](英)威廉·特文宁. 证据理论:边沁与威格摩尔[M]. 吴洪洪,杜国栋,译. 北京:中国人民大学出版社, 2014.

[28]Coglianese C., Marchant G. E. Shifting Sands: the Limits of Science in Setting Risk Standards[J]. University of Pennsylvania Law Review, 2004(4).

[29](德)伊曼努尔·康德. 法的形而上学原理[M]. 沈叔平,译. 北京:商务印书馆, 1991.

[30]陈欢,刘聪,齐浩宇. AIGC 技术辅助刑事审判的优势、风险与规制——以 ChatGPT 为例[J]. 长江大学学报(社会科学版), 2024(4).