

探寻数智赋能监督的新路径

编者按 作为使用大规模数据和强大计算能力训练出来的人工智能大模型,具有高度的通用性和泛化能力,随着数字技术的深入发展,检察机关在研用大数据法律监督模型的同时,也在积极学习、探索大模型的应用路径,以期进一步提升数字赋能监督质效。大模型为检察工作带来新机遇,同时也对数据安全、技术应用等提出更高要求。本刊以此为切入,呈现专家学者理论洞察、省院实践经验与基层院的一线反馈,为检察机关智能化转型提供有益思考和借鉴。

法律与AI交融的试炼之路

□沈阳

当今时代,数字化进程正快速拓展至各行各业,司法领域亦置身其中。检察机关作为法律监督机关,正面临信息洪流和复杂案件的双重挑战。此时,AI大模型技术的悄然浮现,既为检察办案带来科技助力,也为整个法律生态注入新鲜活力。检察机关如何与人工智能技术实现深度融合?这场变革不仅是技术与规则的碰撞,更是认知方式和思维范式的重构。

大模型为检察办案提供破局之路

传统的法律实践,常依赖人工在浩如烟海的法规、判例和证据材料中寻觅和研判相关线索。这一模式在信息时代面临严峻考验:案件卷宗日益庞大,检察官难以在有限的时间内精读所有材料;法律文书与司法解释迭代频繁,知识储备更新成为一项长期任务。数字化社会瞬息万变,违法犯罪手段不断翻新,而大模型的出现,为司法实践提供了一条破局之路,借助算法与语料库,将海量文本资源整合提炼,让法律人在信息洪流中能够顺势而动。

过去,检察官办案需先确立检索维度,再辗转于法规库、案例库与学术研究中,一旦遇到生僻条文或跨领域案件,耗时费力难以避免。而大模型的逻辑在于“协同”,检察官可直接提出复杂问题,让模型从多源信息中自动抽取匹配内容,并给予条理化的解

释。这种协同不仅提高了工作效率,还催生一种新型办案范式——人机互补,在机器智能的辅助下,检察官法律智慧的深度与广度将得以同步拓展。

大模型的意义绝非简单的文本整合,它更像是一个精密的知识引擎。在面对棘手案件时,检察官可通过自然语言指令,让模型即时检索数以千计的案卷与条文,并呈现其中的关键点。如此一来,检察官就可不再陷于手动翻阅大量卷宗的复杂劳动之中,能在短时间内获得案件要点与法律逻辑。这不仅缩短了案件审查周期,更让检察官得以将更多精力投入到对案件本质问题的深度思考上。

大模型融入司法实践的良好循环

在实践层面,大模型的应用涵盖多种情境。对重大、疑难或新型案件,大模型可先期梳理证据链条,区分轻重缓急,提示潜在法律监督点。检察官可在此基础上对比分析,无须在冗余材料中反复消耗精力。更有意义的是,透过对海量既往案例的结构化分析,大模型还能对特定类型犯罪的量刑趋势、社会影响进行数据化呈现。这种预测可以为检察官办案提供参考,为提高司法决策的科学性与前瞻性提供助力。

引入大模型不仅限于处理当前案件,更可对司法知识库进行持续迭代。当大模型在大量案例中捕捉到相似争议点和司法难题时,检察机关可据此优化内部培训与办案指引,大模

型也可在更新的数据和规范中持续优化其输出精度,从而形成良性循环。

放眼国际,欧美等国同样尝试将人工智能技术运用于辅助司法办案领域。在跨国司法协作与国际法治对话中,大模型可提供快捷的信息检索和跨境法规比对,畅通国际司法合作路径。在全球化进程中,司法信息协同与经验分享将更为便捷,从而推动各国在共性问题上达成更深层次的共识。由此,法律与智能技术的互动将不再局限于一域之内,而是以普适经验促成更广泛的国际共鸣。

然而,前行的道路并非坦途。一方面,大模型的可信度与权威性取决于数据来源的严谨性和算法的中立性,若输入数据存在偏颇,输出结果便可能引发误导。在法律领域,这种误导的代价尤为严峻,可能涉及侵犯公民权益或影响司法公正。为此,检察机关必须建立严格的数据合规审查机制,对大模型的训练语料进行筛选与校验,并在算法中内置防范歧视与偏见的逻辑规则。另一方面,法律工作者需明确自身主体地位,不能让技术取代理性判断。技术应当是辅助,只有将技术红利与伦理底线有机统一,才能确保大模型成为法治秩序的驱动者,而非潜在风险的引信。

大模型引领法律职业思维深刻变革

值得深思的是,当大模型持续为

□鲁建武

2024年12月,最高人民法院下发《关于全面深化检察改革,进一步加强新时代检察工作的意见》,明确指出要“以科技赋能法律监督”积极探索大数据、区块链、人工智能等技术辅助司法办案。有效推进了高新技术与检察工作的有机结合,也为各地检察机关设置了一道如何因地制宜实现创新发展的“必答题”。

安徽省检察院在《2024年全省数字化检察工作方案》(以下简称《方案》)中明确提出“要探索开展人工智能认知大模型在检察办案的落地应用”以来,积极利用当地人工智能产业发展态势,以“人工智能+”行动为基点,稳步构建生成式大模型应用系统,打造服务于全省检察机关的生成式大模型应用系统,逐步探索场景试点等工作,用大模型的“小练兵”“微创新”摸索数字化发展路径,以数字赋能驱动检察工作提质增效。

一、大模型是数字检察的“枝叶”。大模型是具有大规模参数和复杂计算结构的机器学习模型,具备提供自然语言理解、文本生成、知识问答、多模态等能力,为千行百业带来触类旁通、自适应的智慧涌现,也为检察机关大模型的培育和探索提供了智慧化路径。

在数字检察战略中,大模型能充分发挥数据效能,有效“合并同类项”。数字检察战略以个案办理为起点,以系统治理为终点。如果说数据

生成式人工智能的实战应用

是数字检察的“根系”,大模型则是数字检察的“枝叶”。基于数据进行训练和逻辑推理,大模型在辅助检察办案实际运用中,能够实现人机交互方式的自然化、信息承载形式的模型化、案件审查的精细化,有效补足检察人员技能欠缺、补足卷宗材料信息维度,补强法律监督能力,帮助检察人员突破人机交互的语言障碍、信息传递的时空限制和对客观事物的认知局限,能更好地处理“同案同判”和“个案差异”之间的矛盾,有效提升案件规范化程度和办案效率。

二、从“一省之计”变“长治之策”。安徽省检察院充分认识到人工智能大模型的应用优势,围绕《方案》制定目标,推动全省数字检察顶层设计的完善,组织人工智能生成式大模型专班,以基层工作重点、痛点、难点为出发点,明确研发场景和应用需求,梳理实际办案思维逻辑,推动大模型在办案过程中的应用,为检察机关大模型创建从“一省之计”变为“长治之策”打下良好开端。

在安徽,检察机关大模型创建拥有天然的培育土壤。一是需求牵引,随着经济社会的发展,检察机关受理案件量逐年增长,据统计,近5年来安徽省检察机关共受理各类案件85万余件,安徽省检察机关积极参与省委政法委牵头的政法跨部门一体化办案平台的应用,在数据积累和政法一体化机制方面,具有相应的实践积淀。二是技术支撑,依托安徽高科技企业国产化算力模型训练后发布的法律大

模型,实现了多模态交互、语音识别通用能力和制卷、阅卷、文书等工具的自动输出功能,可全面提升法律专业任务上的准确性,为检察机关大模型培育夯实了能力底座。

三、场景驱动应用先行,实现“点穴式”赋能。有了深耕的土壤,便有破土的藤蔓。通过大模型学习证据分类规则,可将检察机关受理后案件的卷宗整合成符合检察官阅卷习惯的电子卷宗,让证据一目了然;通过大模型深度解析卷宗材料,分析笔录规则,可助力检察官自动填写立案信息并回传至检察业务应用系统;通过大模型理解卷宗内容,可自动拆解破案经过、犯罪事实等核心要素提取展示,并定位相应证据材料,帮助检察官快速梳理案情。

检察机关大模型的应用是“点穴式”的赋能。在安徽省检察院的推动下,安徽省检察机关大模型的应用提升了大模型的实战应用价值,推动检察官办案提速增效。目前,安徽省检察院正在探索文书生成等六大场景的大模型应用。

应用的探索展示了检察机关大模型的优势。具体体现在提升案件分析效率,通过文本数据的深度学习,精确高效提炼事实描述、证据链构建等关键要素,为检察官提供全面而深入的案件分析;高效优化文书撰写流程,自动生成或辅助生成格式化规范的审查报告、起诉意见书等法律文书;决策支持更加全面,通过对以往案例的学习和对现有法律法规的理解,可对复杂案件办理提供科学合理的预测和建议;

检察官提供快速而精炼的分析时,法律人的思维方式也将悄然转型。过往,法律职业建立在经验积累与直觉判断之上,许多法律工作者在长期实践中形成对法律条文本意与案例规律的自身认识与理解。如今,大模型通过数据挖掘与逻辑推演,勾勒出更为宏观与多元的知识图谱。这种外脑式的补充有助于法律人摆脱惯性思维,从自身经验提升至宏观审视,在面对新型案件时更具底气。换言之,大模型的加入是一种智力扩容,将法律工作者推向更为专业、理性、明晰的决策轨道。

面对科技极速演进,法律领域必须以更为灵活和开放的姿态应对挑战。大模型为检察工作注入的不仅是信息处理的提速,还有思维格局的放大与价值理念的升华。在技术、伦理与制度的多维交织中,法律职业共同体有望在大模型的助力下形成更为高效、透明、理性和公正的生态结构。在这一过程中,法律不再是冰冷条文的静态重述,而是与社会现实的动态脉动相契合,在数智时代的律动中砥砺前行。

在未来重新审视当下,这一场法律与AI的交融试炼必将留下一道深刻的印记,它见证了新时代的思维裂变与认知跃迁,以规则与理性化为指引,智能与伦理交相呼应,为法治事业注入更稳固的底色,为社会的发展注入更强大的动能。

(作者为清华大学新闻学院、人工智能学院双聘教授)

□钱飞 王琦

当前,检察机关数字化检察工作正在深入推进,尤其是大数据法律监督模型运行,有效提升了法律监督工作质效。但与此同时,运用大数据法律监督模型时的制约因素也不可忽视,如模型对数据的针对性、适应性要求较高,一旦脱离预先设定的数据格式和要素体系,将无法实现有效监督。而与数据要素密切相关的突出问题就是如何做好文本解构,即如何将非结构化的文本数据利用技术手段有效解析要素信息,形成结构化的数据表。

文本解构之所以成为工作中的难点,主要原因在于法律文本的变化非千篇一律、法律事实错综复杂、法律逻辑精深交错。实践中,即使总结或提炼了某一类法律文书的规范化表述或模板范例,亦不能苛求所有的文书都能按照这样的“标准”来形成。这就决定了不能用固定的句式、结构、连接词来分析法律文书。在大语言模型产生之前,检察人员往往需要通过人工审阅的方式,对文本模型进行理解与判断后提炼出相关内容,形成所需要的数据表、数据库,再进行后续分析。

众所周知,大语言模型已在很大程度上实现了对语义、文本的理解与分析,如果能将其有效运用于法律文本的解析,势必会带来突飞猛进的跨越。

大语言模型运用于法律文本解构的现实困境

利用大语言模型的第一步,是在确保数据安全和隐私保护的前提下,将大语言模型部署在检察机关的专用局域网内,以隔绝与外部网络的直接连接,避免数据泄露的风险。互联网上开放的大语言模型虽如雨后春笋,但受制于司法办案工作的涉密性,不允许检察人员将相关法律文书直接上传到互联网。因此,应运而生的是在政法智能化建设中崭露头角的可在内部网络中单独部署实施的大语言模型。然而,部署这样的模型并达到成熟的运用,离不开顶层设计、资金投入、人员支持、科学管理等多种因素。客观来看,特别是对于基层检察机关而言,想要理顺关系、走上正轨绝非一蹴而就,需要各方面的大量投入。

突破法律文本解构困境的路径

在实际工作中,将大语言模型运用于法律文本解构,立足基层院现有的软硬件条件,可从以下路径进行探索。

一是对已公开的相关法律文书,可运用互联网上的大语言模型进行分析。

已公开的起诉书等法律文书,已经按照有关规定,隐去自然人姓名、身份证号等关键信息,转以“某某”或“××”代替。这些文书均可在12309中国检察网上通过公开渠道获取。在需要提取的内容不涉及被隐去或代替的内容时,这些被处理的文书因内容完整、句子通顺,完全可以被大语言模型用来分析和提取相关要素。如当需要提取某交通肇事案的案发路段、地点时,可以将公开法律文书上的相关内容作为待分析对象,将其输入到大语言模型,即可准确地将无关的语句内容剔除,给出较为准确的案发路段、地点等信息。

这种方法的局限性在于,仅能对在公开法律文书中完整展示的信息进行分析和提取,一旦所需的信息在法律文书公开时被全部或部分隐去,就无法通过此种途径来实现。

二是对可能涉密的法律文书,可运用文本处理技术剔除敏感内容,保留非涉密内容后运用互联网大语言模型进行分析。

要分析在公开法律文书中已被处理的信息,检察人员必须立足于原始文书来想办法。笔者发现,利用简单的文本处理技术进行脱敏或脱密处理,提取出包含待分析文本内容的碎片文字或语句,再将这些并不包含任何敏感信息,甚至无实际含义的文字,上传至互联网的大语言模型进行分析,即可得到所需要的要素结果。如在运用监督模型提取起诉书中相关的公司、企业名称,一般来说,公开的文书中对具体的名称已经隐去。此时,检察人员可以用“公司”“企业”等关键字对原始文书进行提取,求得原始文书中的一句话或语句。再对这些表述“不完整”“无意义”的文字进行脱敏处理,比如将“犯罪”“涉嫌”等可归纳的关键字词删除,即可得到更加无意义、无逻辑的只言片语,这些文字几乎没有任何含义,但却包含了完整的公司、企业名称,而这些名称恰恰是需要提取的信息。

检察人员将这些已经涉密的碎片文本上传到大语言模型进行分析,可便捷有效地提取相关的公司、企业名称,为下一步比对工作提供了充足的基础数据。这种方法的局限性在于,如果要对更加复杂的案情细节进行提取和文本预处理,就很可能“误伤”一些关键字,导致信息碎片化程度太高,难以实现有效提取。

三是利用检察网上的普通服务器、电脑,单独部署简单实用的大语言模型。当前,最新开源的一些国产大语言模型,已经可以依托本地网搭建在普通电脑上,即使没有高配置的硬件,也能实现运行。但需注意的是,利用低配置电脑运行大语言模型,其分析能力可能较弱。因此,在应用这类大语言模型时,不妨缩小切口,找到最适合的应用场景,针对某些或某类问题具体予以解决,避免一网打尽式的阐释与分析。例如,对上述举例中提及的地点路段分析、公司名称提取,利用内部网络大模型也可实现有效分析。同时,还可以利用大模型核对、纠正法律文书中的错别字,按照文书公开的要求规范身份信息号码信息、人名等,都能得到较为理想的结果,同时,在规范检察文书制作、辅助业务数据管理等方面,也能够解决很多具体问题。

(作者分别为江苏省泰州市海陵区人民检察院检察长,海陵区人民检察院副检察长)

□强峰 张宏

类案监督是司法实践中探索并形成的法律监督路径,并在司法实践中不断发展,是检察机关强化法律监督的着力点和突破口。数字检察战略的实施为检察机关高质量开展类案监督提供了新的动力引擎,基层院在推进类案监督数字化应用过程中,需明确方向指引,细化实践举措,立足区域特点有效提升数字赋能类案监督质效。

数字化类案监督的探索与成效

一是用好数字化平台。最高检部署的全国检察机关大数据法律监督模型管理平台(以下简称“模型管理平台”)为基层院数字模型的复制推广应用和自主研发提供了运行环境和实践平台,全面推动了基层院数字检察工作跨越式发展。模型管理平台上架了全国各地检察机关优秀成熟的模型,不仅提供了可复制推广的法律监督范式,也

集中展示了各级检察机关在司法实践中总结提炼出的数字检察应用案例和办案实践经验,是蕴含法治思维与数字思维融合的“富矿”,为基层院开展数字检察学习交流提供了新阵地。作为数字检察工作发展较慢的西北地区基层院,甘肃省张掖市甘州区检察院紧紧依托模型管理平台,组织检察干警学习优秀模型的办案指引和数据研判方法,在实战应用中培养干警的数字思维和数字能力,促使将应用模型变为办案习惯。积极应用涵盖“四大检察”业务的数字模型171个,获得初筛线索1300余条,取得类案监督成效的数字模型79个,办理各类监督案件100余件,模型管理平台内模型的本地化推广应用为该院提升法律监督质效提供了强有力支撑。

二是找准类案监督方向。实践中,模型应用所针对的监督点越明确具体,就越容易直指问题要害,找准核心特征要素,精准筛选出类案监督线索。“小切口”模型具有监督方向具体、数据要素少、初筛线索精准、成案率高、易复制推

广等特点,用好用足“小切口”模型,可在“四大检察”深度融合应用中实现一体化监督。甘州区检察院结合本地办案实际,针对养老金、低保金、退役军人优抚金发放中存在的违法问题,应用模型管理平台内相关模型开展国有财产保护领域专项监督活动,筛查核实服刑人员违规领取低保金、养老金、退役军人优抚金线索22条,制发检察建议3件。甘州区检察院以“小切口”类案监督为模型应用方向,在食品行业终身从业禁止、个人信息保护、游泳场所安全隐患排查、违规销售隐形眼镜等医疗器械、放射诊疗机构安全许可、外卖商家违法经营、医疗美容行业整治等民生领域开展公益诉讼专项监督,办理监督案件26件。

三是有效拓展数据资源。在开展类案监督过程中,类案识别的精准度与案件相关数据开发应用的深度、广度成正比。甘州区检察院高度重视数据资源的获取应用和维护管理,依托检察业务应用系统深度挖掘内部数据,同步积

极拓展合并使用外部数据。坚持“不求所有,但求所用”原则,与公安、法院、司法、市场监督管理、民政、教育等部门建立横向数据共享协作机制,以法律监督办案需要为基准,以模型办案应用数据要素最小化为原则,严格限定数据获取的周期、范围,以共赢理念和方式创新拓宽数据获取渠道。

四是抓实线索核查办理。数字模型初筛线索和研判是监督成案的关键步骤,还需进行人工核查和后续调查,确定线索的真实性、有效性。甘州区检察院探索建立线索移送、核查机制,明确由数字化办案团队负责线索研判、流转、办理,案件管理部门负责线索台账管理、成案情况和办案进度核查,实现从模型管理平台线索分析研判推送至检察业务应用系统成案办理的全流程闭环监控管理。

优化数字化类案监督的方向与举措

一是围绕高层部署,深入实施数

字检察战略。基层院在全面贯彻落实《关于全面深化检察改革,进一步加强新时代检察工作的意见》时,要牢牢把建立横向数据共享协作机制,以法律监督办案需要为基准,以模型办案应用数据要素最小化为原则,严格限定数据获取的周期、范围,以共赢理念和方式创新拓宽数据获取渠道。

二是聚焦“小切口”模型,加大应用的广度和深度。基层院应持续有效发挥成熟模型“一域突破、全域共享”的应用规模效应,充分发挥“小切口”模型精准高效的特点,持续加大应用广度和深度,在服务市场经济、保障民生、保护未成年人、保护生态环境等各类领域中赋能法律监督。同时,积极探索人工智能等新兴数字技术的应用,推进数字模型的系统集成和智能化迭代升级,依法拓展监督应用领域,丰富类案监督应用场景,全面提

升法律监督质效和服务社会治理工作实效。

三是研发内部监督管理模型,赋能检察管理高质量发展。基层院要在如何更好优化“四大检察”履职结构比、一体抓实“三个管理”方面进一步探索数字赋能检察管理。在类案监督中探索研发内部监督模型场景数字模型,开展案件质量评查、业务质效风险预警、一体化监管的实践应用。紧密依托模型管理平台,将业务管理、案件管理中的数据要素、办案流程、纠错监管和质量评查规则等实现数字化、智能系统集成,搭建针对流程监控、案件受理和审结、出庭公诉、裁判结果审查、文书公开等各个环节的监督管理模型和质量评查模型。

【作者分别为甘肃省张掖市甘州区人民检察院检察长,甘州区人民检察院副检察长。本文为2024年甘肃省人民检察院理论研究课题“类案监督数字化应用的实践与思考”(GSJC2024-54-01)阶段性成果】

解构法律文本要素的困境与突破