

在多维数据与多种方法间寻求突破

业务数据资源化

大家谈

□柴慈永 王斌

仅仅收集数据并不是数字检察,将数据变成实实在在的办案资源,是每一个站在数字检察前沿的检察官必须面对的问题。

随着数字检察工作的深化和工作思路的拓宽,更多的检察官跳出“就案办案”的思维定式,努力去发掘案件背后的共性特征,探讨共性特征与数字技术的结合方式。在深入推进数字检察工作时,检察人员还应不断提高数据分析能力,寻求更高效多元的数据利用方式,真正发挥数据的长效价值。

数据思维的培养

检察业务应用系统的应用涉及各个业务部门、案件管理部门、信息技术部门等,每个检察人员都是数据的生产者和使用者。数据思维是检察官能否对数据进行有效分析并加以利用的前提,如果检察官没有数据思维,就感知不到数据对业务的价值,只会看到一系列的标准、一堆堆的表格和凭空增加的工作量。对此,应从以下三方面强化数据思维的培养。

一是解开思想桎梏。检察官开展数字检察工作的核心价值,就是提供办案思路,这个过程要尽可能避免因资源的局限所造成的影响。因此,要让检察官破除对技术和数据的思维限制,在寻找思路、设计模型时,树立“所有数据都可获取”“所有技术都可实现”的前置思维。

二是正视数据价值。数据之“大”并不一定是“体量大”,而是“价值大”。或许一份数据并不全面,无法覆盖所有的细节和要素,但只要能够反映足够的特征,在数据维度上足够细致,也能够产生较大的数据价值。检察官应反复审视自己能够访问的数据,从中寻找到更多的工作切入点。

三是重视数据价值。数据思维的本质是发掘数据潜藏的价值,而不是关注数据本身。案件管理部门的检察人员在开展统计工作时,具有最大的数据访问权限,但如果开展的业务分析仅仅是罗列数据的升降、比率的高低,就没有真正发挥数据价值。只有将业务逻辑投射到数据中,找到问题症结,才能有效指

导相关业务开展。

数据资源价值的发掘

一是内部数据的价值亟待挖掘。检察业务应用系统中的数据全量且精准,其中的案卡项目是典型的结构化数据,法律文书则需要按照一定规则书写,具有半结构化的特征,也易于拆分整理,信息抽取分析成本较低,这些数据都可进行深度分析的优质对象。但是,从传统业务来看,这些数据除了被用于案件管理部门的业务分析研判外,更多时候仍在“沉睡”,其最常被使用的场合和方式,就是与外来数据进行数据碰撞,但这仅是数据分析的方法之一,其他诸如关联规则挖掘、回归分析、聚类分析、时序分析、异常检测等方法,以及对法律文书进行自然语言处理、文本挖掘等应用还未能充分得以运用。这种情况的出现,一方面,是由于检察官对技术不了解。在众多数据分析工具中,数据碰撞的概念最易于理解,运用起来也不需要技术知识背景,通过 Excel 表格即可完成,而更为复杂的分析方式则需要更多的技术知识,这对检察官而言是个较难跨越的门槛。另一方面,新的方法试错成本高,结果难以预测。在借助数据碰撞这一思路构建模型时,具备共性的数据项可以直接观察到,分析的结果也可以预知,而其他众多分析方法则需要采用专业的分析工具,或需借助编程技术才可实现,在构思阶段,无法对结果做出预期,很有可能在投入时间和精力后,结果并不如人意,这无形中给这些分析方法的应用带来了阻力。加之通过数据碰撞等简单分析方法取得成功的案例有示范效应,使得检察官更加信任经过验证的思路,本能地会减少对新思路、新方法的尝试。

二是外来数据的治理要“以终为始”。随着各种模型的不断构建涌现,检察官逐渐意识到,多维的数据可以互相支持、互相印证。在这个过程中,检察机关汇入了大量的外来数据,这些数据形

式多样、标准不一。要想让这些外来数据与内部数据实现融合,便于进一步进行数据检索和分析,就需要进行相应的数据治理。将相关信息按照一定的规则提取出来,经过清洗、转换、集成,最后成为适于检察机关利用的数据形态并汇入数据仓库。数据治理的最终目的是服务于检察业务,从这个“终点”出发,在数据治理过程中,技术分析必须与业务应用紧密协作,让业务部门感受到数据的显性价值。对于数据治理工作,业务部门是最终的使用者,对数据的准确理解、调度至关重要,整个数据治理过程也都需要将重心前移到业务一线。唯有如此,才能让业务部门认可数据治理这一工作的价值,更为积极地投入这项工作中。对于业务部门而言,也不是“要我做数据治理”,而是“我要做数据治理”,不再是“遵从你的标准”,而是“一起制定标准”。

数据资源与业务需求的衔接方式

数据资源是决策过程的“外脑”,只有有效地管理数据,高效地利用数据,探寻其与业务需求的结合点,才能真正发挥其价值。

一是数据的资产化运营。对数据的利用,应以“让数据产生价值”为目标,从单纯的储存、调用、分析,走向主动式的运营管理。首先,数据资源是无限的。数据的流通和使用并不会消耗数据本身,反而会产生更多的数据。其次,数据的“折旧率”低,复用率高。以往的数据可以借助新的分析方法得出新的结论,也可以与当前数据结合,产生新的数据。再次,数据使用的边际成本低。数据治理仅需要投入必要的采集、清洗、整理成本,但该投入往往是一次性的,数据的可复制性使其能够应用于更多的场景中,数据价值的发掘成本相对较低。随着数字化转型的推进,数据生产要素的特征愈加明显,将成为提升检察机关社会治理能力的重要组成部分,检察机关对数据的运用也愈加成

数字赋能需在融合上下功夫

□魏家齐 刘绍宇

数字时代的发展,为加强检察机关信息化、智能化建设指明了方向。实践证明,大数据在纵深推进数字检察工作中发挥着突出作用,当前,还有待进一步在融合上下功夫,不断深化大数据法律监督工作,更好地融入国家治理大局。

一、促进多源数据融合贯通

在数字检察战略下,面对全面深化司法体制改革的新形势新需要,推动释放以融合为特征的数据内生价值成为当务之急。检察机关应打破数据壁垒,做好数据归集,整合形成数据资源集,提炼法律监督逻辑要素作为法律监督决策的参考依据,不断推进监督模式从案卷审查、个案监督转化为数据赋能、类案监督,从中找准数字监督的热点重点、执法司法的难点堵点、社会治理的焦点弱点,以点带面激发大数据赋能监督办案的辐射效应。

如何促进法律监督业务多源数据的融合,实现数据共享应用与信息互联互通,是数字检察亟须破解的难题。笔者认为,应加强统筹协调,从更高层次

出发,横向推动贯通民政、卫健、公安、政法、城管、交通等多部门的数据资源,实现必要的深度共享。如在执法司法数据和“两法衔接”平台,围绕从案件受理到执行终结,打通公检法司业务系统,以业务一网办理、流程一网流转实现数据共享的全覆盖与规范化;纵向注重检察系统内部数据的上下贯通、汇聚整合,以“四大检察”数据融合为导向,建立内部数据跨区域、跨业务条线、跨部门间的共享、开放、会商和研判机制,加快推进检察机关上下数据一体统筹,避免资源浪费。此外,在保证数据合法性与正当性的前提下,融合跨不同行业领域的社会来源数据,比如,挖掘关于企业治理、消费者投诉等数据,为检察机关发现数据背后的逻辑与治理价值提供广泛的基础素材。

二、推动技术与业务深度融合

数字技术与检察业务的融合,是构建大数据法律监督模型的核心,检察人员通过对海量数据进行快速处理与精准分析,在类案中发现异常、总结规律、归纳要素。检察机关可搭建数字办案治

理单元,分级分类审批使用数据,增设数据分析模块,重塑条线分工与业务流程,促进信息上下互联、线索互送共用、监督事项数字化办理。

从实践看,大数据法律监督模型作为各地检察机关有效开展法律监督工作的数字化工具,是基于对数据资源集的归纳、总结和分析,根据业务类别选择或关键词搜索提取不同类型数据,以此为依据设计模型架构,从而找出异常类案监督线索,解决类案发现难、取证难以及跨行政层级、跨行政区域、跨部门职能等带来的履职难题。如可建设统一标准的云平台,整合各类相关信息化数字化业务系统数据,打通检察机关纵向与横向上的“数据壁垒”和“信息孤岛”,推动监督办案模式从“数量驱动、个案为主、案卷审查”的个案办理向“质效导向、类案为主、数据赋能”的类案治理式转型,不断提高资源配置和利用效率,实现精准、高效与协同办案。

三、以数据思维推动检察人员数字化办案方式融合

大数据法律监督是一项系统工程,要求办案人员不仅要具备高质量的

事,会主动寻求与其他部门进行数据共享和业务联动。

二是构建数据资源目录。想要充分利用数据,需要让检察官知道哪些数据可以利用,这些数据在哪里、质量如何。没有对数据状况的充分了解,就无法将模型构想具体落地。编制本单位的数据资源目录是高效利用数据的前提,它是开展数据管理的抓手,又是利用数据资源的引擎。一方面,从需求出发,检察官在提出一个模型构思后,可以参考这一目录,根据当前数据资产的具体情况,制定每一步的具体路径。对已经在库的数据,不需要消耗人力、技术去收集整理;对当前缺少或者不完备的数据,可以尝试借助其他已有数据进行推理,或者明确规划下一步获取数据的工作。另一方面,从数据出发,检察官获知了本单位的数据资源目录,可以看到哪些数据是可利用的,由此可激发更多新的设想。数据资源目录集中、便捷、标准化地管理了所有的数据资产,使检察官可以通过跨越时间、空间、业务领域,以及价值、场景等多维度进行数据的查询、获取和加工,能够显著提高数据资源的利用效率。

三是实现数据支撑的平台化、接口化、服务化。技术只是工具,开展业务才是目标,要避免数据资源的收集治理与业务应用相分离的情形。首先,数据支撑需要平台化。平台化的管理模式可以将检察机关获取的内外部数据资源进行整合,按照统一标准和口径加工成数据资产。同时,借助多渠道汇总的数据进行交叉验证,可确保数据的准确性,将差错排除在应用之前。平台化的数据支撑使数据资产具备高度的可复用性,在业务应用中,检察人员可忽略收集、治理、存储的成本,不需要担心数据质量,直接拿来构建模型。其次,数据支撑需要接口化。“万变”的是数据,“不变”的是方法,接口化的数据支撑可以提供通用型的技术模型,构建面向共性的分析平台,将分析方法封装成拿来即用的“黑箱”,检察官不需要参与技术实现,在一端输入数据,就可以在另一端直接获取结论。再次,数据支撑需要服务化。在业务构想与数据结合的过程中,数据及其分析方法是检察官放在手边随时取用的资源。服务化的数据支撑致力于提升数据及工具的易用性和可获得性,让数据资源变得人人可用,有效降低业务与技术的沟通成本以及数据资源的应用门槛,进而改善资源利用效率,促进检察业务的发展。

(作者单位:天津市人民检察院)

办案能力,还要有高水平的数据思维,其中主要包含数据逻辑思维、数据融合思维、数据分析思维与数据决策思维,依托思维的转换提升,推动数据管理,共享路径优化,最终实现数据的归集融合、共享协作和应用管理。由此,只有业务和技术双精通的融合型检察人才,才能够最大限度地激发数据治理价值,做好业务数据汇集、统计、研判、碰撞、比对工作,逐步构建起数字检察监督办案的新形态。

对此,要全面提升检察人员的数据思维,提高其运用数据进行决策、发现、治理的能力。强化培训,推动法律监督业务与信息化专业知识的融会贯通。在提升法律监督质效的目标指引下,不断迭代检察人员的数字赋能思维,与法院检察院联合开展“订单式”培养,打造一批既具有数字化思维又熟悉检察业务的专业技术人员,加强信息化专业人才培养,推动其深度参与一线检察办案实践,让技术真正发挥支撑作用。

(作者单位:北京市科学技术研究院数字经济创新研究所,浙江大学光华法学院)

□杨军 严俊

当前,生成式人工智能技术的迅猛发展,加快了传统生产力的转换迭代。检察机关也在积极探索生成式人工智能技术在检察实践中的应用,但如果不聚焦当前检察实践中的具体应用需求,不从应用场景和用户思维出发,难免会陷入“一哄而上、一哄而散”的情形。因此,笔者认为,要真正让生成式人工智能技术赋能检察工作,应坚持打造“两个模型”、实现“四个蝶变”。

打造垂直领域大模型和自主学习的内部语料模型

要运用生成式人工智能技术赋能检察工作,前提是要具有经过充分训练的大模型。然而,大模型的搭建成本较高,基层检察院暂不具备独立建设通用大模型的能力,对此,可通过构建垂直领域大模型的方式来实现相关功能。

垂直领域大模型以通用大模型为基座,通过对检察工作相关知识标注后进行训练调校,使其能够更好地理解和处理检察办案中的专业术语和法律条文对应关系、总结提炼案件事实证据等,有效减少部分事务性工作量,破解案多人少的困境。

结合本地化部署的要求以及检察工作自身的特殊性,仅有调校后的垂直领域大模型还无法真正实现生成式人工智能的自主学习、自我纠偏功能,对此,还需要建立一个涵盖检察工作的法律法规知识、司法政策知识、用户反馈知识的内部语料模型。如检察官发起“请打开王某某危险驾驶案卷宗,结合案情事实证据帮我撰写审查报告”的指令后,人工智能大模型根据指令阅读电子卷宗,根据犯罪嫌疑人基本情况、诉讼经过、案件事实、现有证据等进行分析,访问内部语料模型,然后按照标准化审查报告格式自主撰写报告内容。

接下来,办案检察官还需对人工智能大模型生成的审查报告进行人工校验,认为有部分错漏或不够满意之处,可以继续发出修正优化指令,结合人工智能大模型的多种提示进行结果优化修正,直至符合检察官的预期。同时,大模型系统还可将检察官提出的任务问题和最终结果集结作为训练语料,定期归集到训练集中,进而训练微调人工智能大模型,增量更新人工智能大模型的基座。

实现检察办案、业务管理、队伍管理、智能办公的蝶变

一是检察办案。在初级阶段,人工智能大模型需根据检察官指令对案件卷宗进行识别分析和推理总结,生成阅卷笔录、案件审查报告、起诉书或不起诉决定书、逮捕决定书或不捕决定书等法律文书;通过量化的标准根据个案情节制作量刑建议书;根据犯罪嫌疑人、被告人、犯罪事实等制作公诉意见书、刑诫词等。在进阶阶段,人工智能大模型需根据个案电子卷宗上下文情况发现侦查机关漏罪、遗漏事实、遗漏同案犯或者侦查活动违法等情形,在审查报告中标注或单独提醒检察官;对审判机关的判决书进行解析,自动生成裁判结果审查表,并与起诉内容进行核对,对不一致情形及是否存在审判监督线索进行标注或提醒;根据指令或自动将相关数据回填至案卡。在高级阶段,人工智能大模型需根据指导性案例以及本地区、本部门和检察官自己办理的所有案件,自动对正在办理的个案进行核对,根据同案、类案的不同处理意见进行必要的偏离度提醒。

二是业务管理。生成式人工智能大模型需根据用户指令统计一定区域、一定时期、某一部门或某一个检察官的办案质效,如直接生成本部门所有检察官的平均办案时长、撤回起诉、案件被判无罪或免于刑事处罚等情况,对案件办理流程、办理程度进行自动质量评价。同时,大模型还可根据用户指令查询特定区域、特定业务条线、特定时期的办案态势、案件“结构比”等数据,辅助上级检察机关或本院党组及时调整优化办案力量。此外,大模型可根据用户指令,对本地区或跨区域的特定罪名发案情况进行数据分析,形成针对性的社会治理研判报告。在能够获取相关行政执法机关数据的前提下,可以自主排查和推送行刑正反向衔接线索及检察建议。

三是队伍管理。检察机关组织人事部门可以将所有检察官、检察辅助人员、司法行政人员、聘任制书记员等个人档案信息扫描后上传至生成式人工智能大模型,由大模型自动梳理建成人员信息库,根据人员身份、部门、年龄、学历、工作经历、专业背景、任职经历等进行标注。相关院领导和组织人事部门可以发出指令,大模型会自动给出相应名单。同时可根据个人职级、任职年限等情况梳理计划表,提醒组织人事部门作出是否需要启动相应调整的决定。此外,还可以根据单位、部门、个人的办案情况等开展相关辅助测算和量化工作。

四是智能办公。检察机关综合管理部门的工作人员可以提出需求,生成式人工智能大模型便可将上级或外单位文件自动推送给院领导。院领导阅读后以口头或书面方式进行批示,大模型根据批示情况自动转给相关部门阅处,并列入部门办事议程和领导专班议程进行后续跟踪提醒。综合管理部门工作人员还可依托大模型发布会议通知,通知参会人员并自动统计参会信息,根据报名信息自动生成座位表,向参会者发送提醒。此外,大模型还支持用户指令,根据相关内容撰写材料、报告等,有效提高办公办公文的便捷性。

(作者单位:江苏省南京市人民检察院)

检察长眼中的数字检察

□湖北省咸丰县人民检察院
检察长 谭远磊

数字检察是数字时代推动检察工作高质量发展的新模式。基层检察机关在开展数字检察实践中,需立足区域实际,从制约自身数字检察发展的难题入手,以“微更新”举措夯实数字检察基础建设根基,更好为人民司法、为法治担当贡献检察力量。

一、基层数字检察建设现状

以我所在的湖北省咸丰县检察院为例,数字检察战略实施以来,我院不断培育数字思维,在数据融合基础上推动监督线索的筛查研判,着力依托大数据赋能法律监督提质增效。

聚焦理念革新。根据“高质量办好每一

个案件”的要求,我院多次开展数字检察“为什么做、做什么、怎么做”的大讨论,督促干警领悟数字检察是培育新质生产力的重要抓手,不断强化大数据战略思维,调整充实数字检察工作领导小组,厘清数字检察短期、中期发展目标,加强数字检察工作的统一领导、部署、协调。

聚焦靶向攻坚。我院注重做好数据归集工作。对内,加强与省、州院联动,用好自主研发的环境保护税领域公益诉讼案件监督模型、非法收购运输林木法律监督模型和20个省级推荐模型。对外,以类案监督为切口,牵头搭建行政检察与行政执法信息衔接平台,联合司法局推动本地区行政执法数据向政务大数据中心归集,实现行政执法数据信息资源跨部门共享。2023年以来,行政机关在“两法衔接”信息平台录入行政处罚案件235件,我院对不起诉案件制发检察意见66件。在数据归集过程中,突出抓好数据安

全管理,建立数据使用、移交等工作机制。

聚焦溯源治理。我院加强数据线索获取、整理和分析,通过“正向检索+反向筛查+内部核查+外部调查”四步工作法,形成数字检察工作闭环,将检察机关内部数据和县域治理平台共享数据进行碰撞对比,进而通过大数据法律监督平台推送相关线索,依法深入实地查看,并到相关单位开展核查。通过大数据赋能推动专项监督,发现社会治理难点堵点和执法司法“病灶”,抓实溯源治理。

二、基层数字检察建设存在的主要问题

虽然数字检察建设取得一定成效,但在深入拓展过程中,我院发现在理念自觉转变、部门信息共享、数据开发应用等方面还存在不足。

当前,还有不少检察人员对数字检察的认识比较片面,将其简单等同于信息化建设,认为数字检察是技术部门和技术人员的

事情,只是在办案的辅助,而忽略了“业务主导”这一基本要求,部分检察人员开展数字检察工作仍存在畏难情绪,对数字检察理念的转变还停留在“要我用”阶段,未能真正切换到“我要用”状态。

在数据获取方面,虽然建立了内外联动获取数据的通道和方式,但因部门间信息共享机制不健全,部分部门出于不同考量而拒绝共享数据,数据获取阻力较大。同时,数据的共享还只是简单地对某类数据实现汇聚,在各部门业务融合过程中,还存在司法办案各流程数据共享的欠缺。

在数据应用方面,因基层院数字模型开发力量不足,既懂技术又懂业务的检察人员相对匮乏,加之上级检察院大数据法律监督平台推送、交办的线索有限,数据赋能法律监督的应用成效还有待提升,数字模型的复用和优化,也亟须在拓展应用中不断迭代升级。

三、加强基层数字检察建设的对策建议

数字检察工作的推进不是“大包大揽”,更不是“一哄而上”。同样,解决基层院在开展数字检察过程中存在的问题,也不是完全“推卸重来”,而是要因地制宜、有的放矢地针对关键问题,以细微创新举措,推动解决难题卡点。我认为可以从以下三方面展开探索。

一是建立“小而精”的工作推进机制。探索建立数字检察办案指挥中心,对线索办理进行全流程督办。条件成熟时,可探索建立数字专班直接办案、跨部门一体化指挥机制。强化协作配合,各业务部门积极落实办案主体责任,分条线分类型主动利用模型推线索、办案等。同时,可尝试构建点对点、院对院的珠网式协作体系,积极开展县(市)内外、省内外区域协作。

二是打造数据共享新通道。深化数字平台建设,充分整合裁判文书网、检察业务应

用系统、公益损害与诉讼违法举报中心等平台,加强数据开发、整合和运用,以“一月一调度、一月一通报”工作模式整体推进全院数字检察工作,及时发现数据共享中存在的问题。同时,建立数据共享机制,推进跨部门大数据协同办案,积极争取被监督部门的支持和配合,促进信息交换共享。

三是更新模型建用思路和人才培养方式。在建模过程中,我们需着眼本地化需求,立足“四大检察”职能,构建符合区域特色和案件特点的本地化数字模型,打造“人在干、数在转、云在算”的数字化检察监督模式。同时,在利用上级检察院和先发地区检察院研发的成熟数字模型时,应注意结合本地区实际,优化改造后“为我所用”。在数字人才培养方面,强化以模型研用为实训课堂,根据案件类型打造灵活机动的数字办案组,可拆分可整合,通过各条线业务人才实训,夯实本地区数字检察发展的基础底座。

生成式人工智能嵌入检察工作的「图景」