

窦立博:

# 借一“数”光,让办案有了“新模样”

·以概念解析入手·  
案件里抓“准”规律

“大数据法律监督模型,是检察官在个案办理中或者在数据异常中,发现规律性、共性、普遍性的问题,归纳特征和业务逻辑,应用数据关联碰撞,从海量数据中获取线索的一种思维载体。”每次以实训教师身份在北京市检察院分院举行的数字检察实训提升研修班授课交流时,窦立博总是先会先解释“是什么”切入。看似普通的开场语,却也是他自己最初认识“数字检察”、走进“数字检察”的敲门砖。

窦立博告诉记者,最初的他,面对检察官应该如何参与到数字检察战略中并没有明晰的思路。“有点迷雾中看不到点的感觉,好像找不到抓手在哪儿,不知道该如何发挥业务主导作用。”既然脑中一片混沌,不如就从源头切入——他从基本概念着手研究,随后慢慢悟出来,虽然大数据法律监督模型看似是个新事物,但它依托于数据却回归于监督,不断加深对案件的研究就不会错。于是,作为知识产权检察条线的检察官,他以创建一个符合涉知识产权案件的数字检察模型为目标,将近三年全市检察机关办理的侵犯商标权类的案件仔细梳理,着重关注一些规律性、共性的

问题并总结特点,最后这些一步步的实践摸索都成为创设销售伪劣类投诉涉刑线索移送大数据法律监督模型业务规则的基础。

截至目前,销售伪劣类投诉涉刑线索移送大数据法律监督模型已经在全市范围内推动刑事立案38起,公益诉讼立案5件,打掉制假售假窝点百余处;推动行政执法机关对26个重点区域开展专项检查,查处68起行政违法案件。据悉,2023年度北京市检察机关十个知识产权保护典型案例,有4个直接来源于该模型筛查出的线索。

·以调查研究深入·  
经验里找“难”问题

据了解,除了销售伪劣类投诉涉刑线索移送大数据法律监督模型外,窦立博还先后参与了经济犯罪检察领域等多个重点模型的算法规则研判工作。基于此,他针对模型创建初期怎么找异常监督点、异常监督点确定后如何结合异常监督点构建算法规则以及模型运转成熟后如何进行推广等问题,分别撰写了多篇业务文章进行深入论述。如在《创建模型需夯实“三个基础”》一文中,窦立博总结了刑事检察官如何在大数据法律监督工作中发挥业务主导作用的心得,并在系

统分析的基础上提出,刑事检察条线大模型的三种类型分别是追诉漏罪漏犯类、侦查违法监督类和行刑衔接类。

“数字检察历经电子检务、智慧检务再到数智赋能的发展,其间必然伴随着一些配套问题亟待厘清思路并完善解决。有些问题需要理论思考,有些问题需要实践积累。”采访中,窦立博告诉记者,检察官在办案之外,也有义务基于履职实践去探究和总结数字检察的应用路径和场景。“这本身也是数字检察工作的一部分。”窦立博给出了自己笔耕不辍的答案。

通过不断地总结梳理,目前销售伪劣类投诉涉刑线索移送大数据法律监督模型的算法规则以知识产权领域行刑衔接为履职点,已经不断升级完善,其中,他参与创设的“检察院12345智能解析程序”,可在30分钟内一键完成对36万余条投诉数据的数据清洗,并动态、精准、高效识别包含知识产权行刑衔接在内的多领域重点监督点和重点线索,实现监督场景应用的可视化。“每个阶段都有不同的问题出现,都需要不断地梳理、总结,而在此过程中,又会不断地发现新的问题,这本身就是数字检察作为新兴领域的必然特点。”在他看来,对数字检察的思考与研究来源于办案实践,也必将反哺于办案实践。

·以技术赋能升级·  
成效里寻“新”增长点

“大数据、人工智能发展对检察履职变革产生的规定性影响之一,就是驱动法律监督提质增效‘反哺’保障相关科技创新和产业发展,要推动检察监督模型从数据碰撞比到初级形态向人工智能的高级形态演进。”在谈及从事数字检察工作两年后的自己,与最初有何不同时,窦立博思考片刻后,引用该院检察长朱雅频在2024全球数字经济大会“人工智能产业发展与知识产权保护”专题论坛上的讲话来回答。在他看来,新技术的发展也为自己拓展了数字检察素能提供了新的思考和借鉴。

记者了解到,在北京市检察院的指导下地处“京西智谷”的北京市门头沟区检察院的支持配合下,销售伪劣类投诉涉刑线索移送大数据法律监督模型已经插上了“人工智能”的翅膀,借助通用大模型的底层架构,实现了数据清洗、数据筛查的全流程自动化。“每个月定期灌入数据后,不足5分钟就能筛查出本月可能存在重点监督线索。”窦立博告诉记者,前期的“小米加步枪”是必要的,因为要通过实践去验证算法规则的科学性、精准性,后期人工智能的加持更是必需的,

因为一旦确定了方向是正确的,就要借助技术“全力奔跑”,真正发挥出数字检察战略的叠加倍增效应。

提及该模型未来的走向,他表示履职路径仍是“个案办理—类案监督—系统治理”。因此,模型未来的发展应是融入社会治理,以机制方式把制假售劣涉及的问题进行系统性解决。而其中,人工智能则是助推这一愿景实现不可或缺的力量。

“我接下来还有‘一大一小’两个计划。大的方面,已经在着手开展搭建知识产权全领域监督场景,探索将模型监督点关联、算法规则同类、数据源趋同的模型,利用相关技术实现大模型应用场景的构建,提升成效的规模化。小的方面,更加专注模型规则要素的梳理和研究。”采访的最后,窦立博告诉记者,他又在“小米加步枪”的类案中,发现因不同机关存在数据壁垒,导致虚假财务报告审计难以被识别的深层次法律监督问题,目前已梳理出十余条规则。对他而言,两年前那种与数字检察相遇时“初始创业”的激情,一直都在。

数字先锋

□本报记者 简洁

“这个模型算法规则层次清晰,和我们实际办案的思路也是一脉相承,数据来源也打通了,这就是为什么它可用、好用……”每每介绍起数字模型的要素特征和建用心得时,北京市检察院经济犯罪检察部检察官窦立博总是滔滔不绝。从最初的数字检察“小白”到如今的数字检察素能实训教师、北京市检察机关数字检察专家,对窦立博而言,除多年深耕的知识产权检察工作领域外,大数据法律监督模型的建与用,无疑是他近年来最充满干劲的尝试,更是甘之如饴的自我赋能。

## 一场思维方式的悄然重构

数字漫谈

□卞叶

在万物互联的智能时代,数据已成为驱动社会发展的新型生产要素。当传统司法经验遭遇数字洪流,检察机关的思维正在经历转换,这场思维变革同时也在重塑法律监督的认知框架与实践路径。

首先,从“概念涵摄”到“类型归属”的思维重构。与“概念”相比,“类型”具有开放性、层级性、流动性、整体性等特点。在数字检察背景下,可以突破成文法“概念涵摄”的思维定式,在传统法律适用的“三段论”演绎之外构建线索发现的“类型归属”思维。类型思维的本质是建立动态开放的认知框架,放弃“因果关系”桎梏而投向对“相关关系”的挖掘。如江苏省苏州市检察机关构建的破产领域虚假破产债权监督模型,将破产领域虚假破产债权行为分解为涉诉案件集中性、债权金额异常度、当事人关系网络、诉讼时间序列等多个分析维度,通过机器学习对66万份裁判文书进行特征提取,以类型化建模形成虚假破产的识别图谱。这种“分子式”的认知解构,使法律规范与复杂事实得以在数字空间实现映射。

司法实践中的类型化探索正在催生新的方法论,具体呈现为三个转

变:通过数据挖掘发现类型化监督线索,从个案办理转向类案治理;以类型化数据作为决策支撑,从经验洞察转向趋势评估;运用类型化分析发现隐藏规律预测社会治理风险,从事后处置转向事前防范。

其次,从“经验判断”到“数据决策”的思维重构。大数据法律监督模型的核心就是数据思维,模型构建即将海量数据转化为有价值的决策依据的过程。回顾数字检察监督建模实践,笔者总结出七种数据思维,即数据筛选思维,是最基础的基于数据本身承载的信息功能开展的数据挖掘,通过各类数据要素的组合分析发现疑点、筛选线索。数据比对思维,或称之为数据碰撞思维,是突破单一数据源局限,通过将不同来源、类型或维度的数据进行交叉对比,从而挖掘疑点发现线索的思维。数据聚类思维,是将海量数据按某一特定数据要素特征或相似性特征进行自动归类,从而发现疑点线索的思维。数据穿透思维,是以特定数据为“线头”(或称之为“引子”)逐层溯源、反向追踪,通过“数据钻探”穿透表面数据信息直抵问题根源的思维。数据穿透可以帮助我们从一个微小的异常点出发,深入挖掘隐藏在背后的系统性问题。数据画像思维,是指通过多维度、多层次数据标签体系构建对象特征的思维方式。具体来说,可以通过收集、整理和分析大量数据,对某个群体的特征、行为、偏好等进行标签化、类型化描述,它类似于用数据“画”出一个对

象的“肖像”,帮助人们更直观地理解其特点。数据分析思维,其本质是对数据进行态势分析,通过分析数据的分布、趋势和异常点,发现隐藏在数据中的问题和线索。数据串联思维,是通过构建数据要素之间的动态关系网络,将孤立信息节点转化为有机风险体系,整合多维度数据挖掘潜在关联关系从而发现疑点线索的思维。数据决策的核心是以数据为基本单元,通过系统解析和模型建构整合数据碎片,将海量信息转化为可操作的监督智慧。数据决策不是对经验的否定,而是对经验迷雾中认知盲区

的照亮。数据要素的深度应用正在重构检察工作场景。如在公益诉讼检察领域,卫星遥感数据与行政执法数据的碰撞,让生态环境损害无所遁形;在未成年人检察方面,教育、医疗、民政等多源数据的融合分析,构建起未成年人立体化保护网络等等。“数据感知—模型推演—决策输出”的新认知链条,使法律监督从被动响应转向主动预见。

再次,从“线性逻辑”到“融合认知”的思维重构。数字检察是多元协同的,是数据、技术与算法的深度融合。要实现检察工作向数字检察的代际转型,除了需要具备数字化技术能力之外,数字资源和要素同样必要。数据提供基础底座,若仅有数据而无其他要素,数据可能成为“沉睡的矿藏”。技术扩展监督能力,是数据决策、治理、应用的基础设施。算法提炼法律经验,是法律规则的

量化表达。三者之间的融合是提升法律监督效能的关键,这一过程不仅是工具革新,更是检察履职方式的变革。

数字检察并非单纯的技术应用,而是检察业务与技术的深度融合。数字检察整合数据、技术、算法,以监督模型为载体,实现“业务需求—数据汇集—技术攻关—场景验证”的闭环创新机制,催生具有检察特色的数字化解决方案。融合认知的本质是以业务为主导,建立“业务+技术”的双重解释框架,核心要义是实现法律语言到数据语言再到技术语言的转化。检察业务与数字技术不是简单的叠加关系,而是形成了相互滋养的共生系统。

数字检察可以打破刑事、民事、行政、公益诉讼“四大检察”的职能壁垒,实现数据共享与监督协同。通过打通各业务条线数据壁垒构建融合性监督模型发现跨领域监督线索,传统检察业务条块分割的格局正在被数据要素重构。

当数据要素深度融入检察工作,这场思维重构既是时代发展的必然选择,也是司法认知的自觉回应。当类型思维打破概念枷锁,当数据理性照亮经验盲区,当融合认知消融专业壁垒,我们看到的不仅是工作方法的革新,更是司法认知论层面的范式变革。唯有主动拥抱数字文明,方能在时代浪潮中勇立潮头,让公平正义以更智慧的方式抵达人心。

(作者为江苏省苏州市检察院检察信息技术部副主任、数字办副主任)

智能体的进化之路

智能体是一个能够感知环境、作出决策并执行任务的“智能个体”,核心能力是自感知、自决策、自执行。它可以是一个物理机器人,比如仓储中的分拣机器人,也可以是虚拟程序,比如家里的智能音箱。

随着信息技术的发展,AI智能体涌现。它在智能体的基础上融合了人工智能技术,能够进行更复杂的决策并适应动态环境。它可以开展机器学习和深度学习,从数据中学习并优化行为,具有根据新输入数据调整决策和行为的自适应能力,可在复杂场景中进行分析并作出个性化的复杂决策。一般应用于智能客服、语音助手、汽车自动驾驶等。

AI智能体的发展与大模型息息相关,大模型通过增强智能体的自然交互能力、环境感知精度以及学习优化能力,使其从简单的自动化工具转变为能够理解、决策和自我改进的智能系统。这种多维度的提升,让AI智能体在语言理解、环境适应和任务执行上能够更接近人类智能的表现,真正实现智能化的飞跃。

未来,每个人可能在生活和工作中用到多个AI智能体。用更长远的眼光来看,未来的AI智能体将不再只是单一的个体,而是一个复杂生态系统的核心。它的角色是协调者,将人类、机器和环境整合到一个动态协作的网络中,能在复杂的生态系统中解决跨领域、跨场景的问题。

——摘自《环球时报》

走进科技



## 司法救助实现“致广大”

针对司法救助线索发现渠道单一等问题,宁夏回族自治区银川市检察院在办案过程中,深入挖掘检察业务应用系统内部数据资源,与从民政、妇联、残联等单位获取的困难群众数据进行比对分析,创建多元化司法救助监督模型。模型内设性侵害、人身伤害、侵犯财产权3项类案线索司法救助规则,可有效助力检察机关及时发现司法救助线索。2024年以来,银川市检察机关利用该模型办结司法救助案185件。目前,该模型已上架全国检察机关大数据法律监督模型管理平台,向全国检察机关进行推广。图为该院检察官利用模型研判案件线索。

(本报记者单曦莹 通讯员郝雅婷/文 马瑞/图)

## 民事案件公告送达准时“从这里出发”

□郑燕 余传备

法律文书的送达既是法院审理案件的重要程序,也是维护当事人合法权益的重要保障,在诉讼过程中发挥着重要作用。新疆生产建设兵团石河子市检察院在办理民事监督案件时发现,由于法律和司法解释对公告送达的规定较为分散,个别法官对公告送达的条件、公告内容、公告方式的运用存在一定的认识偏差,出现被诉当事人因未参加诉讼而不服裁判,造成“案结事未了”的情形。据此,该院研用民事案件公告送达违法大数据法律监督模型,精准开展法律监督,促使法院规范公告送达管理,严格落实法律和司法解释有关公告送达的规定,进一步保障当事人的诉讼权利。

【数据采集与转换】

数据来源:该模型所需数据为诉状数据和裁判文书数据。诉状数据来源为法院公告网发布的起诉状副本及开庭传票公告、上诉状副本公告;裁判文书数据来源为法院公告网发布的裁判文书公告。其中,需从起诉状副本及开庭传票公告中获取答辩期限、举证期限、开庭时间、起诉状要点、公告刊登日期等信息;从上诉状副本公告中获取上诉状要点、公告刊登日期等信息;从裁判文书公告中获取裁判的主要内容、公告刊登日期等信息。

登日期等信息。

数据转换:借助人工智能助手,将相关公告文书转化为结构化数据。具体操作如下:

对于诉状数据而言,需将相关文件导入人工智能助手,按照受送达人姓名、原告姓名、案由、送达文书名称、答辩期限、举证期限、开庭日期、诉状要点、文书刊登日期等数据要素进行提取,对公告中包含“诉讼请求”“上诉请求”“请求”“诉状要点”等关键词的,将诉状要点数据要素标记为“是”,不包含的标记为“否”。

对于裁判文书而言,需将文件导入人工智能助手,按照受送达人姓名、原告姓名、案由、案由、裁判主要内容、文书刊登日期等数据要素进行提取,对公告中包含“判决如下”“裁定如下”“主要内容”等关键词的,将裁判主要内容数据要素标记为“是”,不包含的标记为“否”。

【模型运用】

一是要发现公告内容不规范的监督点。

第一步:数据过滤。分别将诉状要点和裁判主要内容数据要素标记为“否”的数据进行筛选,作为可能未依法说明起诉要点和裁判主要内容的线索。第二步:分析研判。对线索涉及的公告,用受送达人姓名在法院公告网进行检索和分析研判,将起诉状副本和上诉状副本公告中没有写明

诉状要点、裁判文书公告中没有写明裁判主要内容的,判定为有效线索推送成案,以公告内容不规范为监督点制发检察建议。

二是要发现开庭期限、审理期限和移送上诉状副本期限不规范的监督点。

第一步:数据关联和运算。以“诉状数据和裁判数据中受送达人姓名、原告姓名要素相同”为条件,关联同一案件数据,并将该数据中裁判文书公告刊登日期与起诉状公告刊登日期进行日期运算,计算时间差。第二步:数据筛选。以举证期限要素中“15日”或者“30日”为关键字,筛选时间差小于45日(即公告期限30日加举证期限15日)的案件,作为实际开庭日期可能早于公告确定的开庭日期线索;以送达文书名称要素中包含“起诉状”为关键字,筛选时间差大于210日(含30日公告期)的案件,作为审限可能大于法律规定6个月的线索;以送达文书名称要素中包含“上诉状”为关键字,筛选时间差大于55日(即公告期限30日,上诉状期限15日,二审法院收到上诉状,移送原审法院期限5日,原审法院收到上诉状副本,送达当事人期限5日),作为可能超期移送的上诉案件线索。同时,需进一步调阅法院案卷,核查研判异常线索,对有效线索推送成案,并以办案期限不规范为监督点制发检察建议。

【应用成效】  
模型赋能,唤醒“沉睡”数据。该院通

过对法院公告网信息的深度挖掘和智能转换,将PDF格式的文本文档转换成适应模型运行的表格数据,有效解决了大数据法律监督数据获取的难题。该院运用模型对石河子市法院近3年共1500余条公告数据进行筛查比对,发现有效线索30余条,进而跟进开展监督。

精准定位,提升监督质效。公告送达不规范可能导致当事人未收到或者未及收到法律文书,不仅影响当事人及时履行义务,还可能引发新的矛盾纠纷。检察机关开展法律监督,发现法院存在法律文书未依法有效送达的问题,通过制发检察建议等方式督促法院予以纠正,有效维护司法公正。

以点带面,推进系统治理。检察机关在履行法律监督职责中,针对一类案件发现深层次社会治理问题,会通过类案监督促进源头治理。该院充分运用该模型发现法院在公告送达程序中存在的问题,通过制发类案检察建议,促使法院完善制定了对公告送达文书及案件规范管理的机制,统一了公告文书的格式和开庭日期的计算标准,促使问题得到有效解决。该模型上架全国检察机关大数据法律监督模型管理平台后,推广效果良好。

(作者单位:新疆生产建设兵团石河子市检察院)