

# 寻求技术范式与业务体系的深度融合

□彭瑞峰 褚尔康

在数字时代,以“数字革命”驱动新时代法律监督提质增效,进一步探索构建数字检察的数据结构体系,设计数字检察的模型算法,开发数字检察的要素系统,进而逐步构建起数字检察运行体系,对推动检察工作高质量发展具有重要意义。

## 数字检察的逻辑关联

数字检察的逻辑关联是以数据为对象实现监督工具体系的建构,即以数字化为手段、以模型为方式实现法律监督体系更好地融入国家治理体系,进而释放更大治理效能。

理论体系的逻辑关联。从解构学角度看,在数字检察理论体系的构建过程中,通过对原始数据进行深度挖掘获取有效信息,不断刻画出清晰的数据结构,进而形成模型化的检察工作思路,进一步提升法律监督工作的科学性和有效性,这对夯实数字检察底层逻辑关系具有重要意义。

要素体系的逻辑关联。从法理学角度看,数字检察工作中,法律监督的数字化过程并非简单地从形式上转译

法律代码。数字时代的法律监督思维模式首先要摆脱抽样数据思维模式的束缚,采取整体、全方位的数据采集方式,对监督对象行为进行全景式扫描,从而实现更为精准和有效的监督。因此,数字检察平台建设的基础环节是对相关法律法规规定的主体要素、客体要素、程序要素进行全面数字化转换,组成网状结构的监督要素模型和监督要素图谱,这是数字检察的核心要素系统。

对象体系的逻辑关联。在数字检察建设过程中,如何将文本性的数据量化为可供计算机系统识别和计算的数据种类和结构,成为实现监督模型数字化的重要途径。对数字检察及其运行机制的模型化体系构建是一个系统工程,每一个环节都紧密相连,串联起上述系统运行的纽带和桥梁是数据,即以数字化形态存在的监督对象。因此,如何将现实法律监督过程进行模型化处理,进而形成能够被算法识别和处理的数据体系,成为数字检察建设和运行的核心环节。

## 数字检察的体系衔接

以数据要素为核心构建数字检察

信息处理平台成为大数据时代的必然选择。在数字检察体系架构转换建设过程中,必须高度重视检察权运行机制的模式体系转换问题。

监督流程的体系衔接。在法律监督模型设计和实现过程中,往往采取指标化的量化模式,即通过分解检察业务流程相关指标体系,以此为基础对真实的监督对象进行数字化流程转换,从而达到对监督活动数据进行自动化处理的目的。这种量化结构的最大特征是指标体系的离散性,很有可能因指标设置“粗粒化”而出现法律监督工具智能化不足的情形。这就需要数字检察系统流程设计过程中,对实体原型进行必要的抽象化处理,即通过相应的数字化形态转换,将监督事项主要特征标注出来,通过监督要素和行为的对象之间量化数据关系的关联性进行相应的匹配。

证据审查的体系衔接。目前各地检察机关法律监督平台建设不断加强机制创新,坚持监督数据与办案平台的整合力度,建设贯通监督和案件审查的一体化资源共享、数据互联互通机制,为案件办理提供全方位查询、分析及辅助研判服务,逐步构建起具

有一体化融合能力和规范智能辅助工具的智能化平台。基于人工智能本身任务繁重、效率和准确率偏低的突出问题,针对现实电子数据体系庞大、关联性高度复杂等特点,在业务流程实现过程中需开发一系列符合检察业务办理需要的自动化证据审查工具。

处理工具的体系衔接。在大数据时代,实现对检察工作多变量、多因素复杂系统模型的构建,必须要有更为先进的算法做支撑。只有通过行之有效的算法进行拟合化处理,才能将上述复杂系统还原为较为简单的“元素”关系进行处理,从而避免数据类型巨量化导致的系统开发风险。这就需要法律监督平台模型算法的构建要符合全程性、主动性、实时性的功能特征,能够实现事前监督指标对应、事中监督监测规范、事后监督严密准确的全过程分析自动化运行的目标。

## 数字检察的运行转换

数字检察的运行过程不仅涉及监督对象体系的结构转换,还涉及监督工具体系的系统重构。这一过程不仅体现为制度与技术之间的高度耦合

性,更为重要的是还体现了检察业务与数据处理的过程衔接性,后者表现为连续性的社会现实活动与离散性的计算机系统之间的运行模式转换。因此,数字检察建设的关键落脚点在于如何妥善实现上述运行模式的体系性转换。

夯实底层要素体系。使用数学语言可以精确表达监督对象的内在特征,通过数学上的演绎推理和分析求解,检察人员能够深化对实际问题的认识和判断。要实现上述目标,首先需要在法律监督行为要素定义和计算逻辑体系数字化基础之上,进一步拓展和深化两者之间体系的耦合,设计开发出基于数据搜集、管理、监测和评估为一体的法律监督平台。只有研发出符合检察业务需要的底层监督要素系统,形成数据挖掘依据的实体要素系统,达到数据比依据的规范要素体系,才能实现以数据辅助办案工作流程要素的目的,从而能够以更加科学的手段对搜集到的信息资源进行汇总、分析和反馈。

建构普适数据模型。目前法律监督模型的构建往往按照不同的业务领域特点进行构建,在一定程度上能够发挥不同模型赋能法律监督的作用。

但从全局角度看,这些模型大多应用在个别领域,没有达到全领域和全流程数据模型的效果,构建具有普适性的法律监督业务模型需求较为迫切。特别是在模型体系构建过程中,各类案件情况各异,涉及的法律规范也多种多样,在模型设计过程中,做好核心算法体系的设计与开发成为重中之重。

畅通数据信息渠道。在数字检察建设过程中,各级检察院要抓住运用模型方法推进法律监督全面深化变革的契机,从关键业务环节监督的多跨场景出发,既要善于从政务数据共享交换平台、政法跨部门大数据办案平台、行政执法与刑事司法信息共享平台及多跨场景中连接数据,又要充分挖掘检察机关内部数据,积极借鉴先行地区的好经验。要强化检察一体化履职理念,上下级检察院要统筹数据资源运用,形成监督合力;各业务部门在运用大数据中,加强协作配合,敏于发现问题线索,通过模型化体系的不断完善促进各项检察业务的实质性融合。

(作者为山西省人民检察院检察技术处副处长,天津师范大学政治与行政学院副教授)

## 数字漫谈

## 走进科技

# 数据共享如何实现安全可控

□冯煜清

数字检察战略下,检察机关推动数字检察改革,通过对大数据、人工智能等科技创新成果的深度融合,建立法律监督模型及配套系统,发现与推动破解司法活动及社会治理中的深层次问题。实践中,监督模型、算法、系统所依托的数据往往超出检察办案内部数据的范畴,高效、安全地实现数据共享是做好检察办案工作的一环。

从政策视角看,中央对于司法数据共享有明确要求。中央政法委2021年印发了《关于充分运用智能化手段推进政法系统顽瘴痼疾常治长效的指导意见》,明确要求“完善执法司法数据跨部门共享机制”,提出建设“跨部门大数据办案平台”等打破数据壁垒的创新举措。

然而,数据共享问题并没有得到充分解决,不愿分享、不敢分享,甚至不能分享数据的情况时有发生。从共享主体看,现有的多数大数据法律监督平台应用于检察机关内部,跨单位、跨系统的平台较少,即便有外部单位参与,往往也限于司法执法机关,行政

机关积极性不高;从平台接入的数据类型看,绝大多数来自检察业务应用系统的内部数据,主要服务于提升检察办案效率、加强内部监督管理等目标。外部数据相对匮乏,数字检察的协同社会治理功能难以充分发挥;从数据安全看,数据安全、隐私保护等问题不仅是抽象的技术问题,更是切实的主体责任。当数据汇总到大数据法律监督平台后,数据使用主体发生变化,但责任主体并未完全转移。

造成上述困局的根源主要在于数据共享机制的局限性。现有的大数据法律监督平台多采用以平台为中心的数据共享机制,各方数据汇聚至检察业务应用系统,检察机关依托平台开展数字检察工作。这种平台中心模式在数据集中管理方面有一定优势,但容易引发相关方对数据所有权、使用权及相关隐私问题的担忧。同时,由于数据权利主体和责任主体错位,平台中心模式很难由数据共享主体自发形成,而是高度依赖于自上而下的行政驱动。要进一步完善数据共享机制,需在共享机制的底层逻辑上解决痛点、消除顾虑。

从平台中心模式转型为去中心化的数据共享模式,有助于从源头破解数据共享面临的困局。在架构设计上,可以在保障数据安全和数据权益的前提下,实现多源异构数据的协同处理。具体来说,去中心化的数据共享模式,其基本框架由平台层、数据层、应用层组成。

平台层由数据站和不特定数量的数据节点组成,通过网络连接。在数字检察的数据共享体系中,大数据法律监督平台是数据站,每个参与数据共享的单位(部门)可以看作是一个数据节点,数据站和各数据节点之间采用分布式架构。

数据层解决的是数据安全和权益保障问题。数字检察涉及的相关数据,可转化视角细分为以下三类:一是开放数据,即依法应当向社会公开的数据,对此类数据应尽可能接入数据站并公开公示。二是限制数据,即可以在特定范围内共享但不得公开的数据,对其采用两权分离的方式(数据所有权归本地节点所有,数据站享有使用权)。为保障数据安全,不能直接对限制数据进行复制或转移。数据站在

使用数据时,通过隐私计算、多方安全计算等加密机制对数据进行计算和分析,使数据得到最大程度的利用。三是保密数据,即依法不能共享流通的数据。此类数据一律不出本地,一般也不能使用。开展数字检察工作确有需求时,可以将数据分析需求提交至保密数据的本地节点,按相关规定一事一议。

应用层解决的是共享数据的应用领域和场景问题,主要围绕数字检察法律监督的具体工作展开。各数据节点与数据站处于平等地位,相关应用工作既可以依托平台开展,也可以在某个节点实施。上述安排有助于提升数据共享的交互性,激发不同部门参与数据共享的积极性,强化数字检察工作的部门协同。

数据共享问题的解决是一个贯穿政策、管理、技术、思想观念的系统性问题,是数字中国建设面临的一个基础性问题。要建立完善的数据共享机制,尚需更加深入、细致的理论研究和实证检验持续跟进。

(作者为东南大学法学院副教授、博士生导师)

# 加速实现智能计算“中国定义”

算力资源是数字经济发展的底座。随着数字经济蓬勃发展,数字化新事物、新业态、新模式推动应用场景向多元化发展,算力规模也在不断扩大,算力需求持续攀升。工信部日前发布的数据显示,2022年,全国在用数据中心机架总规模超过650万标准机架;近5年,算力总规模年均增速超过25%。

智能计算将人工智能技术与算力相结合,能够更好地满足实际应用场景的复杂计算需求。如何推动我国智能计算产业发展,已成为政产学研各界高度关注的问题。其中,标准制定工作尤为重要,它对智能计算及相关产业的发展具有基础性、支撑性、引领性作用,是推动智能计算产业创新发展的关键抓手。

智能计算属于新兴领域,产业发展方兴未艾。在国际标准领域,已有的国际标准化组织和国际先进的标准化机构中还没有智能计算领域的专业技术委员会和工作组开展标准制订工作,国内智能计算尚未形成完善、统一的标准体系。因此,迫切需要把握机遇,加快对智能计算产业发展进行研判,系统梳理,加快建立智能计算领域的标准体系,明确标准之间的依存性与制约关系,建立统一完善的标准体系,以标准化的手段促进我国智能计算技术产业蓬勃发展。

据悉,刚成立的全国智能计算标准化工作组将全面梳理智能计算产业标准化需求,制订并持续优化智能计算标准体系。围绕“基础通用”“计算技术”“计算架构”“应用”“计算安全”5个方面开展标准制修订工作,联动产业平台和创新载体,结合智能计算发展趋势和行业应用需求,在图计算、类脑计算、超算互联网、科学计算等关键技术领域加快标准研制工作,持续开展智能计算标准需求征集,激发智能计算行业创新活力。

——《科技日报》

## 数字视角

# 大模型发展的“四问四答”

□陈静 王霞 徐颖

随着人工智能技术的不断发展,大语言模型技术日趋成熟和统一,大模型应用已成产业趋势。在检察领域,基于大模型的应用也会不断扩容。大模型究竟是什么?大模型在应用检察工作多种场景时如何实现科技赋能的效果?笔者对此逐一展开介绍。

## 何谓大模型

大模型是指参数量巨大的深度学习模型,其中包含大量的人工神经网络和

连接,可以处理大量的数据和任务。简而言之,大模型就是机器通过学习大量的数据后,不断提高自己的智能水平和性能。大模型的优势在于机器可以学习到数据的深层特征,从而具有更好的泛化能力和处理复杂问题的能力。

## 大模型的“大”体现在哪里

首先是参数量大。大模型的参数量足够多,能够实现更快的训练和推理,具备实现更准确的预测结果及处理更复杂任务的能力。时间向前回溯,最早的模型是线性参数,是以有限的高级

参数来参与分析。如今,大模型在参数量方面增强了1200多倍,如此庞大的参数迅速形成神经网络,可以实现大体量的知识涌现。

其次是算力量大。大模型需要大算力来提供强大的计算能力和存储容量,实现对海量数据的处理和分析,并支持模型训练和推理的复杂计算。因需要大量的算力资源投入,大模型的训练门槛普遍较高。

最后是数据量大。大模型需要海量的数据参与训练,如网页数据、搜索数据、视频数据、语音数据、图片数据等,这些数据量可达十亿级、百亿级、万

亿级甚至更多,这都是大模型可以实现深度学习和训练的基础。在参数和数据量足够多的情况下,训练出来的大模型才会形成思维链并具备知识涌现能力。大模型的成长类似人的学习,先有0-18岁的通识教育,再有18岁以后的专业教育。在通识教育阶段,需要学习世界范围内大量的基础知识;在专业教育阶段,则需要对某一领域进行深入学习,成为该领域的专家。

## 大模型如何应用在检察领域

一是构建检察知识库。大模型的知识增强技术可以实现构建检察知识库的需求。检察知识库涵盖法律法规、司法解释、指导性案例、典型案例等各类内容。通过大模型的文本分类、实体识别等功能,将实现对海量法律知识的自动化标注、分类和检索,帮助检察官快速定位所需法律知识和参考案例,提高办案质效。

二是生成法律文书。大模型可以通过对大量法律文书的学习和训练,增强专业知识,自动生成符合法律规范的法律文书,如起诉书、判决书等。通过输入案件的基本信息和要求,大模型能够快速生成符合规范的法律文书,大大减轻检察官的文书写作负担,提高工作效率。

三是提供检察智能“客服”。基于大模型的对话增强技术,可以构建检察智能“客服”系统,通过自然语言处理技术实现与用户的智能交互和问题解答。通过大模型的知识系统和深度学习,检察智能“客服”系统能够针对用户的问题进行精准回答和处理,为律师或当事人提供更加智能化、个性化的服务。

## 大模型的“成长”需要哪些训练

当前,大模型的训练分为基础预训练、通用有监督精调训练和场景化有监督精调训练三个阶段。

基础预训练阶段:基于检察行业数据,对预置的大模型进行训练,提升模型对检察专属知识的理解度和丰富度。通用知识数据包括各业务及关联学科的相关文献、标准、规范、学术论文和实践经验等数据资料,在基础预训练阶段,可对这些包含检察领域的基本理论、原理、技术、方法、案例等方面的知识(包括文本、图像、视频等不同形式)展开基础学习。

通用有监督精调训练阶段:根据检察应用场景需要,进行针对性数据标注,形成通用精标注数据,对基础预训练后的模型进行精调训练,满足不同检察场景需求,提高大模型的理解和生成能力,不断提升指令的泛化能力。比如,科技人员可以准备一个包含各种法律文书和相应标注信息的标注数据集,然后使用这个数据集对预训练好的大模型进行训练,让大模型逐渐学会如何根据法律文书的特征进行分类和预测。在这个过程中,技术人员还可以通过调整模型学习率等参数来优化模型性能。

场景化有监督精调训练阶段:需准备场景化标注数据,采用特定方式对模型进行场景化精调训练,可以有效提升检察大模型对应用场景支持的满足度,持续提升检察大模型应用的专业度和准确度。

科技的一小步,注定是人类进步的一大步。大模型的产生、发展和应用,将会“反哺”到人类生活、工作的方方面面。

(作者单位:北京百度网讯科技有限公司政务创新业务部)

## 科技守护江河源头

在信息技术日新月异的今天,科技手段无疑成为检察办案提质增效的“发动机”和“助推器”。青海省检察院检察官不断提升运用科技赋能办案的能力,通过无人机等技术支撑,获取法律监督线索,立足办案,用心用情守护长江、黄河源头。图为检察官在检测无人机设备。

(本报记者马会平 通讯员王玉婷/文图)

