

缪恬辞：所有的摸索，每一步都作数

缪恬辞和同事们从该院办理的一起刑事诈骗案中，被告人通过虚构招用的370余名就业人员与多家企业之间存在劳动关系的事实，从人社局骗取稳就业补贴111万元。在排查同类案件时还发现，有的被告人专门成立200余家空壳公司，骗取创业带动就业补贴达1000余万元。

闯新路，树立数据思维

公益诉讼检察在“四大检察”中具有主动监督的属性，是大数据智能化应用的有效“试验田”，检察官如何破解公益诉讼“数字密码”，激发数据对法律监督工作的放大、叠加、倍增作用，拥有数据思维很重要。

缪恬辞和同事们从该院办理的一起刑事诈骗案中，被告人通过虚构招用的370余名就业人员与多家企业之间存在劳动关系的事实，从人社局骗取稳就业补贴111万元。在排查同类案件时还发现，有的被告人专门成立200余家空壳公司，骗取创业带动就业补贴达1000余万元。

就业补贴属于国有财产，但当前国有财产流失等案件线索来源面窄，案件呈现涉及监管主体多、流通环节广、专业性强的特征，检察机关难以在履职过程中有效发现国有财产流失线索，线索来源也多为接受移送。

“在一些新型侵害国有财产的违法犯罪活动中，网络化、科技化、智能化趋势更加明显，如果我们公益诉讼检察监督还停留在传统方式，仅凭检察官‘眼睛’发现‘脚步’丈量，是不能满足办案需求的。”在缪恬辞看来，数字检察通过搭建大数据法律监督模型、自动发掘违法犯罪监督线索的方式，恰恰能有效突破当前“案多人少”“监督线索发现获取难”的实践困境。

为了让公益诉讼检察开启“智慧眼”，缪恬辞和同事们着手研发公益诉讼检察全流程一体化监督模型。

“如果让检察官从海量数据中去

寻找线索，无疑是在增加工作量。所以，必须对基础数据池进行数据清洗，保证推送到检察官面前的数据，都是公益诉讼立案指数较高的线索。”建模之初，缪恬辞就坚定了这样的想法。她在梳理模型设计思路时，一方面对检察业务应用系统中与公益诉讼案件相关联的刑事案件信息进行筛选，依据关键词、案源类型、案发区域等设计立案指数并形成得分值。另一方面对12345政务服务热线举报数据、行政机关行政执法数据、互联网抓取信息等外部数据进行筛选，设计二次及以上投诉举报、办理结果不满意等规则，实现智能筛选。

“打开模型，案件线索就根据立案得分值从高到低显示在检察官面前，点开一条线索，模型便可自动按领域、类型推送适用的法律依据、指导性案例、典型案例等，还能自动生成法律文书参考模板，为检察官办案减负增效提供便捷服务。”说起这个历时两年研发的模型，缪恬辞如数家珍。

该院第六检察部副主任余斌娜谈起使用该模型的感受，直言“模型既能帮助检察官完成线索发现、知识服务获取、智慧辅助办案、调查取证资源协同、监督成效跟进、典型案例评选，又能全面对接检察业务应用系统，实现一系统多服务功能。”

至今，该模型仍是广州市检察机关公益诉讼检察干警使用频次最多、监督办案最有成效的模型。在模型应用的有效支撑下，2023年，全市检察机关有关国有财产保护领域的立案数量同比上升42.86%。

精钻研，破解研发难点

“本质上，监督模型建构的底层逻辑是业务经验，关键是要从监督办案实务出发考虑数据需求和模型建构方向。”如何发现模型监督点，从规律中提炼规则，摸清监督模型研发的方法和路径，是模型研发中需要直面的难点问题，也是缪恬辞一直探索的方向。

以她和同事们研发的失业保险稳岗返还国有财产保护公益诉讼监督模型为例，失业保险稳岗返还还是政府对未裁员或裁员率低的参保企业，按其上年度实际缴纳失业保险费的特定比例返还相应金额。“经过整理，它的申领要求企业生产经营符合产业结构调整和环保政策，不属于‘僵尸企业’、关停或失信企业；依法在本市参加失业保险并足额缴纳保费一年以上；企业上年度未裁员或裁员率不高于控制目标这三个条件。”

以上述条件为基础，需要获取经营异常企业目录和失信企业目录数据、企业失业保险参保缴费记录数据、企业裁员数据等。此外，缪恬辞和同事们发现，失业保险稳岗返还的概念中提示这项补贴是按“特定比例返还相应金额”，这也意味着还需要调取比例返还的依据和相关数据等。

拿到数据后，缪恬辞和同事们并没有急于比对，“要先分析数据特点，做实数据探查，比如该模型围绕企业主体领取补贴，调取的数据中，也可以围绕企业主体找到共同的数据项，这个数据‘主键’就是统一社会信用代码，围绕它可以将所有数据串联起来。”缪恬辞一针见血地指出问题关键。

以统一社会信用代码为不变项，该模型将异常经营企业、缴保不足期企业、裁员率高的企业设为否定条件，即失业保险稳岗返还发放记录中，企业的统一社会信用代码不得出现在异常企业、缴纳失业保险费不足期企业和裁员率高于20%的企业中。

“做完第一次比对后，我们再分析企业是否按照大型、中小微型分类标准返还特定比例。通过设计分年度返还比例公式，计算出补贴应发放金额，再将该数额同补贴实际发放金额对比，不一致的数据即为比例返还错误企业。”完成了全链条的比对分析后，该院向相关行政机关发函移送模型筛出的线索，行政机关收到复函称，确有企业错领补贴。其中，70余家企业属于企业划型不准确，200余家企业属于比例返还计算错误，还有的企业属于经营异常。模型推送的精准度得到实践验证。鉴于仍有部分补贴未追回，该院立案督促相关行政机关继续履职。以此案为契机，5家行政机关也强化衔接机制，出台企业划型规范意见。

以此模型规则为蓝本，缪恬辞和同事们还在以工代训补贴、创业带动就业补贴、海外留学人才补贴等国有财产保护领域构建起大数据法律监督模型的应用矩阵。

专推广，发挥数据成效

缪恬辞深知大数据法律监督模型的价值不仅在于研发，更在于推广与应用。

为深入推进全市公益诉讼条线大数据模型应用工作，她和同事们列好模型研用“任务表”，先后组织开展广州市检察机关公益诉讼大数据法律监督模型竞赛等活动，制发《关于贯彻落实全国检察机关公益诉讼工作推进会会议精神的工作方案》《关于加强数字赋能公益诉讼检察监督的工作方案》，对已获省级荣誉类、重点研发类、完善培育类模型“量身”制定推进计划。

为挖掘模型应用潜力，她所在的公益诉讼检察部门创新模型推广复用机制，绘好模型研用“路线图”。建立模型“原创单位、应用单位”对口互帮机制，对全市重点模型，属于两级院联合研发的，由中院专业化办案组负责推广；属于区院单独研发的，由部分区院交叉应用，视应用情况统筹安排全市推广。

为进一步运用数字化思维、数字化认知推动实现监督方法的迭代，缪恬辞先后撰写多篇调研文章，提出信息技术与检察监督深度融合思路，分析实现数字赋能公益诉讼检察监督的最优路径。

今年，在广州市检察机关报送参加全省检察机关大数据法律监督模型推广应用活动中，全市公益诉讼业务条线模型上架数量占全省三分之一，缪恬辞和同事们研发的漏缴防空地下室易地建设费公益诉讼监督模型也在全国检察机关大数据法律监督模型管理平台上参加模型应用评选。

“路虽远行则将至。近年来，以缪恬辞为代表的公益诉讼数字检察工作专班团队，做到了‘积力之所举，众智之所为’，以数据为支撑，找准了服务大局着力点、监督视角创新点、模型应用关键点。”在该院副检察长曾晖看来，正是有了从一个人再到一群人的不断发力，破解社会治理难题的“高含金量”检察方案，才会源源不断地在数字赋能下迸发创新。

模型推广的“三大法宝”

数字漫谈

□杨文

今年4月，最高检组织开展全国大数据法律监督模型推广活动，各地检察机关踊跃参加，活动热度持续升温。如何结合本地实际选准模型推广和应用赛道、如何研发易学好用的推广模型、如何应用好上架模型快速成案？面对这三个问题，笔者认为，应当把握好检察履职的“三大法宝”——严格依法、实事求是、遵循规律，持续激励各级检察机关运用大数据法律监督模型赋能高质量办案，实现推广活动的最大成效。

坚持严格依法。严格依法履职，意味着“法定职责必须为，法无授权不可为”。大数据法律监督模型是检察官从个案办理或数据异常中发现规律性、共性问题，总结、归纳特征要素，提炼研判逻辑，从海量数据中分析、挖掘类案线索的一种工具，是数字检察的重要载体和突破口。数字检察坚持“业务主导”，大数据法律监督模型的研发、推广和应用，都是为了服务检察业务、赋能法律监督，属于检察工作的一部分，绝不能偏离检察履职的范畴。模型推广活动评的是“院”，但拼的是模型的硬实力。在监督模型研发过程中要下好“先手棋”，就必须严格按照法律规定，注重检察履职边界，既不能脱离检察职能，更不能超越检察职权。比如，有的模型对法律文书的错字进行校验，仅仅是服务于内部管理优化，虽然对于提高检察工作效率确实有所助力，但不宜作为法律监督模型；有的模型直接获取行政机关的数据后，再对行政机关履职情况开展一般违法监督，甚至代行行政机关、审判机关的职责，显然超越了检察机关法定职权，这类模型也无法上架推广应用。

注重实事求是。模型推广的赛道选择很重要。除了省级院对应的活动组织外，模型推广活动还设置了推广和应用两个优胜项，就是为了兼顾各地检察机关研发应用模型情况的差异化。有的地区自身基础较好，选择推广、应用双赛道，推广、应用双榜排名均在前列；有的辖区院案件较少，在应用上不占优势，便集中精力抓推广，筛选本地重点模型向全国“推销”，甚至上门服务帮助应用落地；有的地区研发能力较弱，便选择

更适宜本土化的应用赛道。无论选择哪个赛道参与该项活动，都要因地制宜，既不能脱离实际盲目作战，也不能为了逞强而分散精力。此外，推广活动是为了促进“更多模型推广、更多地区应用、办理更多高质量案件”，讲究“真学、真用、真有效”，而个别院为了提升推广成效，采取一个案件关联多个模型、重复应用多个同质化模型，甚至将与模型无关的案件强行倒置关联到模型上，这些“变形动作”不仅无助于活动本身，还消耗了大量成本和精力，影响了正常检察履职，与推广活动的初衷背道而驰。

遵循活动规律。如何用好“巧劲”，在有效参与推广活动的同时，真正实现数字模型赋能高质量办案的双赢，根本是要把握好活动规律。一方面，要坚持需求导向、业务主导。监督模型由谁研发、谁来应用？答案不言而喻，是检察人员。因此，无论是为了便于推广，还是更好地落地应用，模型的监督点必须紧贴当前检察业务实际，在务实的基础上，确保好用、管用，比如要具备业务流程图简单易懂、数据源获取简便、监督规则明确易行、应用案例针对性强等特点，有的院致力于构建规则复杂的大模型，或选取的数据难以获取，或业务流程图充满技术术语让人难以理解，这些都会影响模型的推广应用。另一方面，要注重开放共享，合作共赢。有的好模型在本地取得成效后便“善刀而藏”，生怕被别人学了去，而有的院在各地寻求好模型而不得法。模型推广活动开始后，研究院院逐渐成了“推销员”，有的主动上门帮忙排查线索，有的“合纵连横”达成合作联盟，更多好模型在更多地区焕发出新的生命力，也助力更多的高质量履职案件在监督模型的赋能下应运而生。

目前，推广活动已行程过半，各地检察机关依托全国检察机关大数据法律监督模型管理平台上架并在全局推广的模型有450余个，更多模型还在陆续研发、宣传推广和应用成案中，在推广和应用两个优胜项中，更多地区检察机关研发应用模型情况的差异化，用好双赛道，推广、应用双榜排名均在前列；有的辖区院案件较少，在应用上不占优势，便集中精力抓推广，筛选本地重点模型向全国“推销”，甚至上门服务帮助应用落地；有的地区研发能力较弱，便选择

大小模型“齐步走”

将训练好的机器学习或深度学习模型直接部署在智能手机、平板电脑、个人电脑等终端设备，而不依赖云端服务器进行推理计算，是目前人工智能大模型发展的热门方向。但由于所需算力资源过于庞大，大模型几乎无法在手机、人形机器人等小型终端上开展本地部署，限制了大模型的应用场景。于是，更多的人工智能企业及研发机构另辟蹊径，加码小模型赛道。

小模型，即参数少、占用算力资源少、反应速度快，可以本地运行的模型。大多数小模型参数量在几百万至数千万，结构也更简单。参数量虽小带来的明显改变，是对功耗以及算力需求的降低，许多智能手机都能轻松“驾驭”小模型。模型虽小，但在部分专业领域，其功能表现并不输大模型。例如，小模型通过专注于学习某

个细分领域的精华数据，可降低不相关、意外或不一致的内容输出风险，更贴近用户端需求，小模型的本地化部署也可以更大程度保障用户的数据控制权和隐私权。

不过，在具体研发层面，大、小模型之间的关系更多是协作而非竞争。当下，许多科技巨头的做法是先训练出通用能力足够强的大模型，再借助大模型对数据进行初步筛选处理，构建站在大模型“肩膀”上的小模型，用质量更高、数量更多的数据完成训练，以更低成本实现不输大模型的应用效果。

——《科技日报》

走进科技

筑牢大型基建工程文物考古的“防护墙”

近看模型

□丁园园

文物是历史的见证。按照文物保护法规定，进行大型基本建设工程，建设单位应当事先报请省、自治区、直辖市人民政府文物行政主管部门从事考古发掘的单位在工程范围内有可能埋藏文物的地方进行考古调查、勘探。

山东省滕州市检察院在开展文物和文化遗产保护专项监督活动中接到群众反映，称某工地未进行考古调查、勘探，擅自开工建设，险些损害一处汉代墓葬。经办案检察官进一步调查发现，由于超过2万平方米的大型基建工程大都是当地的重点项目，施工方为赶工期，往往存在“先上车后补票”的问

题，即先开工建设后补考古勘探的相关手续，大型基建工程未经考古调查、勘探就开工建设的问题普遍存在，造成了“先考古后出让”制度不同程度地存在虚化弱化的情况，给文物保护带来极大风险隐患。2023年，该院构建大型基本建设工程“先考古后出让”法律监督模型，筛查问题线索，通过制发检察建议，督促相关部门形成保护合力，推动大型基建工程“先考古后出让”制度落实，确保文物得到有效保护。

【模型构建】

数据来源：模型所需的数据主要包括自然资源部门出让土地名单及申请文物考古项目名单、文旅部门已完成文物考古名单、行政审批服务部门建筑工程施工许可台账等4类数据。

数据模型：

第一步，将自然资源部门出让土

地名单与已申请文物考古项目名单进行比对碰撞，以项目名称、地址为关键词，经过数据清洗，筛选出“已出让土地但未申请考古勘探”项目名单，以及“已出让土地也已申请考古勘探”项目名单。

第二步，将“已出让土地也已申请考古勘探”项目名单数据与文旅部门已完成考古勘探名单进行碰撞，从中发现“已申请但未完成考古勘探”项目名单。

第三步，将文旅部门已完成考古勘探名单与行政审批部门开工项目名单进行碰撞，剔除已完成考古项目名单，筛选出“已经开工但未申请考古勘探”项目名单。

【应用成效】

在公益诉讼检察方面，该院通过相关数据碰撞，共发现已出让土地但

未申请考古勘探项目名单，已申请但未完成考古勘探项目名单，已经开工但未申请考古勘探项目名单等3类线索200余条，督促文物保护部门对21个项目出具文物影响认定意见函，对15个项目完成了考古调查勘探，实现了当地大型基建工程开工前文物考古调查勘探的全覆盖。

该院推动自然资源部门建立土地出让计划备案程序，与文旅、行政审批等部门共享土地出让信息，建立文物覆盖、勘探、保护三阶段函询机制，文旅部门建立考古调查勘探动态监管台账，行政审批部门将同意供地意见书作为土地出让供应及工程开工的必要条件，实现了监督的闭环。该院还以模型构建为契机，在文旅、自然资源等部门设立检察官联系点，定期召开联席会议、同堂培训等，畅通文物综合保护渠道。

在刑事检察、行政检察方面，该模型的构建和案件办理，实现了监督理念和手段的系统融合。该院检察官在摸排公益诉讼线索的同时，还发现非法占地、扬尘污染等行政监督、刑事案件线索，发现线索后，根据内外部线索移送机制规定，及时将相关线索移送

刑事检察部门和相关行政机关。

目前，该模型已在全国检察机关大数据法律监督模型管理平台上架使用，实现了全国多地检察机关的共享共用。

（作者为山东省滕州市检察院第四检察部三级检察官）

