

以高质量数据集筑牢人工智能发展基石

□吴耀明

数据是新的生产要素,是基础性资源和战略性资源,也是重要生产力。2026年6月,国家数据局印发《关于推进行业高质量数据集建设行动的实施方案》(以下简称《方案》),这是国家层面首次对数据赋能人工智能发展作出的系统性部署。这不仅是“人工智能+”行动落地的关键一步,更标志着我国人工智能发展范式正从“以模型为中心”向“以数据为中心”深刻转型。

从“原油”到“汽油”:准确理解高质量数据集的核心内涵

高质量数据集并非简单的“数据大”,而是“数据好”。它具有三大核心特征:一是高知识密度,蕴含深厚的行业机理与专家经验,而非表面的统计信息;二是高技术含量,经过智能标注、质量检验等多道工序,满足人工智能就绪要求;三是高应用价值,能够显著降低模型训练推理成本,提升在特定场景下的适配性与准确性。

数据之于大模型,就像石油之于汽车。未经开采和炼化的原油,无法直接驱动引擎;同样,海量原始数据只有经过采集、清洗、标注、治理等系统化“炼化”,形成高价值、高密度、标准化的数据集,才能真正助力大模型精准学习规律、提升推理能力。

当前,我国高质量数据集建设已取得积极进展。截至今年一季度,全国已建成高质量数据集超过11.6万个,总体量超过960PB,日均Token调用量突破140万亿。然而,供给的结构性矛盾依然突出:公共数据开放不充分,行业标注标准不统一,企业“不知需要什么数据、不知如何建设数据、不知怎样评估

数据”的三重困境仍普遍存在。《方案》的出台,正是为了破解这些瓶颈,推动我国数据资源从“分散探索”迈向“集约化、标准化、产业化”的新阶段。

从“燃料”到“引擎”:深刻把握高质量数据集对新质生产力的战略意义

高质量数据集的战略价值,不仅体现在技术层面,更在于其对产业生态和经济形态的重塑能力。可以说,高质量数据集不仅是人工智能的“燃料”,更是驱动产业深度转型升级的“引擎”。

高质量数据集是落实“人工智能+”行动的基础支撑。“人工智能+”不是简单的“加法”,而是深度的“化合反应”。没有高质量的行业数据,通用大模型在工业、医疗、农业等垂直领域就会“水土不服”,出现“AI幻觉”问题。《方案》聚焦科学研究、工业制造、低空经济、具身智能等19个重点领域以及低空经济、具身智能、智能驾驶、智慧海洋、生物制造等创新领域,明确提出加快推进行业高质量数据集建设。只有让模型“吃”进足够多、足够好的行业专识数据,它才能真正理解行业经验和专业知识的关系,进而从通用的“聊天助手”进化为专业的“行业专家”。

高质量数据集是催生数据要素价值化的核心载体。数据要素要成为资产、进入市场、创造价值,离不开标准化的产品形态。高质量数据集正是数据从资源到资产再到资本的关键一跃。《方案》提出探索以词元为基础的价值体系,推动商业模式从基础数据包销售向API调用、全栈服务梯次跃升,并鼓励数据集质押融资、作价入股等资产化创新模式。这意味着,未来企业拥有的高质量数据集,可以像专利、商标一样

成为可估值、可交易的资产,真正激活数据要素市场。

高质量数据集是构建区域竞争新优势的战略抓手。在算法趋同、算力普惠的背景下,数据正在成为差异化竞争的核心“护城河”。谁掌握了某一领域最深、最全、最准的高质量数据集,谁就能训练出性能最强的行业模型,进而在该领域形成代际优势。河北作为工业大省,拥有先进钢铁、装备制造、生物医药等优势产业,海量工业场景每天都在产生蕴含丰富机理的“数据洪流”。将这些“沉睡的资源”转化为“醒来的资产”,建设具有本地特色的行业高质量数据集,是人工智能时代河北在京津冀协同发展找准产业链和创新链定位的关键。

从“分散”到“集约”:系统探索高质量数据集赋能发展的实践路径

建设高质量数据集是一项系统工程,必须坚持“需求牵引、急用先行、应用验证、安全保障”的原则,落实《方案》部署的强基扩容、标注攻坚、提质增效、应用赋能、管理服务、价值释放六大专项行动,加快形成“场景牵引数据、数据驱动模型、模型赋能应用、应用创造价值”的“数据飞轮”。

以场景为牵引,破解“建什么”的难题。数据集建设不能脱离实际应用、“为数据而数据”。要坚持从行业痛点出发、“以用促建”,梳理数据资源清单和数据集需求清单。对河北而言,应优先围绕算力、钢铁、医药、农业、交通等优势产业,建设一批具有区域特色的行业数据集和专识数据集。同时,积极对接国家数据集管理服务系统,争取纳入国家“物理分散、逻辑集中”的资源目录体

系,实现“一次测评、全国互认”。

以标注为抓手,解决“怎么建”的痛点。数据标注是将知识和经验注入训练数据的关键环节,也是当前制约数据集质量的突出短板。要推动数据标注从“以人为主”向“人机协同、专家深度参与”的多层次标注模式转变。河北要引导有条件地区建设数据标注创新试验区,发展智能化标注工具,建立行业专家认证机制,让各行各业专家深度参与专业知识标注,提升数据集的知识密度。

以流通为纽带,打通“怎么用”的堵点。高质量数据集的价值在于流通与应用。要落实数据持有、使用权、经营权“三权分置”制度,在保护各方权益的前提下,鼓励“以数换数”“数模互换”等多元模式。要培育“为高质量数据付费”的市场共识,推动政府部门、国有企业率先将数据采买纳入预算,发挥示范引领作用。此外,要积极参与开源社区建设,推动基础性、公益性数据集向社会开放,激发全社会的用数活力。

以安全为底线,守护“可持续”的根基。高质量数据集建设决不能以牺牲安全为代价。要落实数据安全相关法律法规,建立覆盖采集、加工、标注、流通全流程的安全治理机制,防范数据投毒、隐私泄露等风险。同时,坚持伦理先行、坚守“科技向善”理念、筑牢科技安全底线,严禁非法收集或使用敏感数据,防范数据集产生偏见与歧视。

总之,要加快构建具有河北特色、服务国家大局的高质量数据集体系,为人工智能筑牢数据基石,以高质量数据集建设为支点,撬动千行百业的智能化改造和数字化转型。

(作者单位:中共河北省委党校)

以共生之思 绘绿色之路

——评《共生视角下重污染企业环境绩效与财务绩效协同发展研究》

□梁英华

在“双碳”战略深入推进、实体经济加速绿色转型的背景下,如何实现生态保护与经营效益双向共赢,是当前重污染企业绿色转型升级面临的核心难题。史宝娟、周思彤、郑亚男共同撰写的《共生视角下重污染企业环境绩效与财务绩效协同发展研究》一书,基于国家自然科学基金项目的研究成果,紧扣现实痛点、回应时代需求,既拓展了企业绿色治理领域的理论研究边界,又构建了绩效协同演化的分析范式,研究论证严谨扎实,为企业绿色转型和环境政策优化提供了重要理论支撑与实践借鉴。

本书引入“共生”这一源自生态学领域的概念,用以解析企业环境系统与财务系统二者之间的关联。作者未将两者视为对立或权衡的关系,而是将其视为两个相互依存、相互影响的复杂系统,专注于探寻它们“耦合共生”以及“演化共生”的内在运作机制。该视角突破了传统环境经济学中“合规成本”的静态思路,契合了“波特假说”中严格环境规划可能激发创新、提高竞争力的动态论断,为理解企业绿色转型提供了更为深入且系统的理论框架。

该书整体架构清晰且明确,逻辑关系严谨且紧密相连。遵循现状研判→经验借鉴→机理剖析→综合测度→实

证检验→博弈施策,构建起一个从理论到实证、从问题到对策的完整研究闭环。书中综合运用耦合协调度模型、Dagum基尼系数模型、核密度估计法、Lotka-Volterra模型以及博弈论等多种研究方法,兼具定量测算与理论推演、动态演化与博弈分析,范式规范严谨,多维研究视角相互支撑,有力保障了研究结论的科学性与可靠性。

该研究以重污染企业为研究对象,此类主体是产业绿色转型的关键载体,直面重污染企业环保与效益难以兼顾的突出矛盾,聚焦现实发展中的突出瓶颈,精准契合了当前生态文明建设与高质量发展的核心研究议题。全书系统阐释了我国重污染企业环境绩效与财

务绩效耦合的行业特征及区域分异规律,在此基础上提出的协同发展路径与优化对策具有针对性和现实可操作性。研究结论可为企业绿色转型、政府环境规制及产业政策制定提供扎实理论参考与实践借鉴。

总而言之,《共生视角下重污染企业环境绩效与财务绩效协同发展研究》是一部紧跟时代步伐、视角独特、方法严谨规范、内容充实可靠的优秀学术著作,既丰富了企业环境绩效与财务绩效共生领域的理论体系,又为企业绿色经营、政府环境规制及产业政策优化提供了重要参考,兼具较高学术价值与现实指导意义。

(作者单位:华北理工大学)

推动党建与业务同频共振互促共进

□刘海超

党建是引领各项事业发展的根本保证,是提升治理效能、优化服务供给、强化业务支撑的核心引擎。新时代新征程,高质量发展对各领域运行提出更高标准,人民群众对服务品质寄予更高期待。面对个别党员协同性不足、创新性不够、执行力不强、融合度不深等现实课题,必须坚持以习近平新时代中国特色社会主义思想为指导,把政治优势、组织优势、作风优势深度融入业务全流程、各环节,以党建强基、赋能、聚力、增效,推动党建与业务同频共振、互促共进,为现代化建设提供坚强政治保证与组织支撑。

坚守政治引领 把稳事业发展前进航向

政治建设统领事业发展全局,是各项工作行稳致远的根本遵循与首要准则。各项业务绝非简单的事务性运转,而是践行初心使命和服务发展大局的重要抓手。持续深化党的创新理论武装,健全常态化、制度化学习机制,推动理论学习与岗位职责相融、与业务流程贯通、与民生服务对接,彻底破除重业务、轻党建的固化思维,牢固树立以党建领航发展、以党建赋能履职的工作理念,真正将理论素养转化为干事创业、履职尽责的实际成效。坚持从政治高度审视工作大局,研判发展形势,落实重点任务,以高标准政治建设引领高质量业务推进,为各项事业高质量发展筑牢政治根基,保

驾护航。

深化组织赋能 构建协同高效运行体系

基层党组织是凝聚人心、统筹资源、攻坚破难的坚强战斗堡垒。针对当下业务链条繁杂、环节衔接紧密、协同推进难度偏大、岗位沟通不畅、资源整合不足、整体联动合力欠缺等短板,以组织能力提质增效带动业务水平全面提升。坚持党建与中心工作一体谋划、一体推进、一体提升,做到发展目标同向、工作节奏同步、攻坚发力同频。依托党组织统筹协调职能,把组织建设、队伍管理、作风培育、廉洁建设贯穿业务运行全周期,围绕重点攻坚任务、常态化治理事务、民生保障事项,完善统筹协调、分工协作、闭环落实的运行机制,打破岗位壁垒,整合各类资源,畅通协作沟通渠道,形成党组织统筹引领、各岗位协同发力、全体人员履职尽责的一体化工作格局。依托组织机制优化、人力物力整合、办事流程完善,推动工作模式从分散运行转向系统治理,从被动处置转向主动谋划,从常规推进转向精准提质。

创新实践载体 激活干事创业内生动力

党建工作的生命力扎根实践,履职成效依靠实干落地。主动突破传统党建工作固有模式,推动党建阵地从会议室课堂延伸至工作一线。在业务前沿设立党员示范岗位,关键领域划定党员责

任区域,攻坚任务组建党员先锋队,督促党员主动亮明身份、发挥表率作用,立足本职岗位争先创优。丰富组织生活形式内涵,把主题研学、实践锻炼、难题化解融入服务现场、群众身边与工作一线,切实增强党建工作吸引力与实际价值。紧扣事业发展堵点、日常运营难点、群众服务痛点,针对性设计党建实践载体,让党建工作贴合业务实际、贴近群众需求。常态化挖掘宣传岗位先锋,以先进榜样带动作风转变,助推业务精进升级。

建强骨干队伍 锻造担当作为中坚力量

干事创业,归根结底依靠人才队伍支撑。恪守党管干部、党管人才基本原则,将高素质队伍建设视作事业长远发展的基础性、战略性工程。围绕综合履职能力提升,常态化开展系统化、专业化技能培训,全方位锤炼人员综合素质,打造政治过硬、本领精湛、作风务实、勇于担当的复合型队伍。完善岗位历练、结对帮带、实战练兵的培育模式,选派优秀人员进驻关键岗位、基层一线,攻坚前沿磨砺成长,激励党员干部在急难任务面前冲锋在先、真抓实干。健全正向激励、人文关怀、思想疏导相关机制,及时舒缓工作压力,解决实际问题,提振团队士气,不断增强全体人员归属感、凝聚力与向心力。发挥党员先锋模范的带动作用,以骨干力量引领全员奋进,凝聚履职尽责、踏实笃行的

磅礴合力。

砥砺过硬作风 涵养清廉实干优良生态

作风建设直接决定事业走向、服务质量与群众口碑。始终恪守从严管控准则,常态化开展纪律规矩宣讲、廉洁警示教育,引导全体人员心存敬畏、严守底线,恪守各项规章制度。紧盯核心岗位、关键流程、业务节点,完善风险排查、全程管控、内部监管体系,坚持抓早抓小、防患未然,从源头规避履职风险,规范日常言行作风。大力弘扬优良风气,坚决抵制形式主义、敷衍懈怠等不良倾向,引导工作人员下沉基层一线,坚守本职岗位、深耕业务本领、用心做好服务。以清正作风守护事业根基,以严实举措保障工作落地,让群众真切体会到党建引领带来的新气象,凭借扎实工作成效稳固发展根基,赢得群众广泛认可。

奋进新发展阶段,立足高质量发展时代坐标,唯有坚持党建统领全局,推动党建业务深度融合,以实干担当开拓发展新局面,方能推动各项事业稳步提升。始终坚守政治铸魂根本,夯实组织建设根基,激发全员干事潜能,锤炼清廉务实作风,持续释放党建引领下的治理效能、服务效能与发展效能,稳步提升工作规范化、精细化、科学化水准,为推进基层治理现代化、增进民生福祉、实现事业高质量发展筑牢坚实支撑。

(作者单位:唐山市数字化城市管理指挥调度中心)

□魏耀涛

随着《网络安全法》《数据安全法》《个人信息保护法》全面实施,税收征管迈入数字化、智能化新阶段。电子发票、大数据风控、智慧稽查等广泛应用,推动征管效能与办税便利度显著提升,也使纳税人身份、财务、经营等敏感信息高度集中流转。同时,网络黑产、不法分子将涉税系统与人员作为重点攻击目标,钓鱼诈骗、木马入侵、内部泄密等事件时有发生,数据泄露易引发电信诈骗、身份冒用、商业窃密等连锁危害。实践表明,涉税领域安全事件多与人的因素直接相关,人既是信息安全的第一道防线,也是最薄弱环节。新形势下,加强涉税人员保密意识,是维护国家经济安全、保护纳税人权益、防范重大安全风险的关键举措,对推动税收治理现代化具有重要意义。

新形势下涉税信息安全的总体态势与核心特征

从总体态势看,数据规模持续扩张、敏感程度不断提升,涉税数据覆盖征管链条,延伸至社保、银行、海关等多领域,单点泄露易引发连锁风险。攻击手段呈现专业化、产业化特征,外部勒索攻击、钓鱼诈骗与内部违规泄密风险交织,安全防护面临多重压力。监管要求日趋严格,相关法律法规大幅提高违法成本,信息安全纳入绩效考核与追责问责体系,泄密倒查、终身追责成为常态。安全事件危害传导快、修复难,泄露信息易在黑产圈快速扩散,不仅损害纳税人权益,更会冲击社会诚信体系、损害税务公信力。

从核心特征看,涉税信息安全具有全域性、高敏性、高频性、链条化、责任重大特征。数据覆盖线上线下、内部外部全场景,是黑产重点目标;办税操作全天候不间断,风险点无处不在;数据贯穿采集、传输、存储、销毁全生命周期,任一环节失守均会引发全链条风险;保密工作要落实党政同责、一岗双责、失职追责的责任机制。

涉税信息安全面临的主要风险

外部网络攻击风险。外部威胁以勒索软件、撞库攻击、钓鱼社工攻击为主。攻击者通过加密系统、窃取数据双重勒索牟利,或利用泄露账号密码批量入侵系统,也会伪装成税务工作人员,以退税审核、发票核查为名诱导泄露信息,攻击手段隐蔽性强、迷惑性强。

内部违规泄密风险。内部风险多源于人员不良操作,常见账号混用、密码随意告知、越权查询导出数据、私自留存敏感信息、通过社交软件传输涉密资料、内外网混用、移动介质交叉使用、离职未注销权限等行为,此类问题隐蔽性强、发生频率高,是泄密高发诱因。

第三方与制度执行风险。代账公司、运维外包、跨部门共享单位流动性大、保密管理松散,易成为泄密突破口。同时,部分单位保密制度流于形式,缺乏常态化培训、考核与应急演练,泄密后处置不当、隐瞒不报,进一步扩大风险影响。

提升涉税人员保密意识的总体思路与原则

以总体国家安全观为指导,以数据全生命周期安全为主线,聚焦压实责任、强化认知、规范行为、提升技能、严肃问责,构建“教育常态化、制度刚性化、技术可控化、监督穿透化、考核显性化、应急实战化”六位一体提升体系,推动涉税人员保密意识从“要我安全”向“我要安全、我会安全、我能安全”转变,筑牢涉税信息安全防线。

坚持政治引领、人民至上,统筹国家安全与纳税人权益保护;坚持预防为主、人防为先,以意识培育为基础、习惯养成为核心;坚持权责对等、失责必问,构建明责、履责、督责、问责闭环;坚持全员覆盖、全程管控,实现全岗位、全流程、全场景保密管理;坚持教管结合、刚柔并济,融合教育引导、制度约束、技术管控与文化浸润。

提升涉税人员保密意识的具体路径

强化思想引领,筑牢底线思维。将保密意识教育融入党建、廉政与作风建设,开展专题教育,树立“保密就是保安全、保稳定、保形象”理念;设立“安全警示日”,通报典型案例、观看警示教育片,强化震慑效应。

健全制度体系,压实全链条责任。制定岗位保密责任、敏感信息、操作负面、考核问责四张清单,明确各岗位保密职责与禁忌;全员签订保密承诺书,关键岗位签订专项保密协议,新入职必签、轮岗重签、离岗清签;严格落实账号实名制、权限最小化、定期复核、离岗即销、日志留痕、脱敏共享等制度。

实施精准培训,构建实战化教育体系。推行分层分类培训,针对领导干部、窗口人员、稽查人员、运维人员、外包人员等不同岗位,定制差异化培训课程;聚焦法律法规、制度规范、风险识别、应急处置、报告流程等核心内容,强化实用性。

强化技术防控,以技管人控险。搭建统一身份认证体系,推行强密码、多因素认证、自动锁定、操作审计与异常告警机制;加强终端管控,落实防拷贝、防截屏、防外联、数据脱敏、水印溯源、移动介质管控等措施;完善邮件安全、网关防护、入侵检测、勒索防御、数据防泄露等技术系统,以技术手段筑牢安全屏障,降低人为违规风险,实现“不能违、不敢违”。

严格监督考核,以问责倒逼意识提升。提高保密安全在绩效考核中的权重,实行一票否决、末位约谈制度;开展常态化、穿透式、突击式保密检查,建立问题台账,限期整改并跟踪复查;对失泄密事件从严追责、通报曝光,以案促改,形成强大震慑。

培育安全文化,营造全员守密氛围。打造“保密无小事、岗位有责、操作守规矩、异常即报告”的安全文化;通过宣传栏、短视频、海报、内网专栏等载体,开展常态化浸润式宣传;选树保密标兵、安全模范,宣传先进事迹,以身边人教育身边人,推动保密理念内化于心、外化于行,形成全员重视、全员参与、全员负责的保密工作格局。

面对复杂严峻的安全形势,必须坚持思想铸魂、制度立规、技术赋能、教育强基、问责倒逼、文化浸润,全方位提升涉税人员保密意识与防护能力,坚决守住不发生重大数据泄露、重大安全事件、重大舆情风险的底线,切实维护国家税收安全、纳税人合法权益与社会和谐稳定,为税收现代化建设保驾护航。

(作者单位:河北省税务局干部学校)

投稿邮箱:hbjrbll@163.com