

基于大数据技术支持的企业法律风险识别和防控研究

文 | 翟昌彬

数字经济不断深化发展让企业所处法律环境越发复杂。企业在合同履行、知识产权维护和数据治理等方面面临多重法律风险，传统依靠人工经验的方法效率不够且预警较为迟缓。借助大数据海量数据处理、智能研判和动态预警能力，为企业法律风控提供了技术支持，促使管控模式从被动处置转变为主动预防。



(配图由 AI 生成)

大数据驱动企业法律风险管理的理论逻辑

企业法律风险的类型划分与演化趋势

企业在经营活动过程中始终存在着法律风险，具体涵盖合同、知识产权、数据合规以及劳动用工这四个主要方面。合同风险主要是源于主体资质审核不足、条款表述不清楚以及履约监督不严格等问题，可能会引发合同失效或者违约争议方面的状况；知识产权风险表现为商标被侵权、专利保护存在缺陷以及商业秘密发生外泄等情况，严重削弱企业在市场中的竞争优势；数据合规风险因《数据安全法》《个人信息保护法》的颁布而变得日益突出，数据从采集开始到跨境传输的全链条都面临着合规挑战；劳动用工风险是由劳动合同管理混乱、社保缴纳缺失以及离职程序不规范等原因造成的。在数字经济背景下，企业法律风险的跨领域关联性变得更强、隐蔽性变得更高，且爆发速度变得更快；原有单一风险已经转变为多主体、多环节相互交织的复合型风险，这对企业的风险识别与防控体系提出了更高标准要求。

大数据技术的核心能力与技术架构

大数据技术能处理海量信息，还可实时高速分析，并且具备整合多样化异构数据的能力，其技术体系由数据获取、存储、处理和应用四个层级构成。数据获取层通过爬虫程序、API 接口和物联网设备等方式，全面收集企业运营数据以及

外部司法、监管与市场信息。数据存储层采用分布式文件系统和 NoSQL 数据库等工具，实现大规模数据的高效保存与维护工作。数据处理层借助 MapReduce、Spark 等分布式计算平台，执行数据的净化、转换和深度挖掘相关工作，还融合自然语言处理技术，解析合同和法律文件的语义内容，利用机器学习算法建立风险预测模型。数据应用层把分析成果转化为直观的风险图谱、智能预警提示和决策参考文件。这一技术体系突破传统数据处理在容量和速度方面的瓶颈，为企业法律风险的精准识别、持续监控和智能预警提供核心技术保障。

大数据技术重构企业法律风险管理范式

大数据技术推动企业法律风险管理实现模式变革，从经验主导变为数据主导、从被动应对转向主动预防、从静态管控走向动态监测。以往主要依靠法务人员经验判断的方式，存在风险识别受主观影响、覆盖范围有限且反应滞后等问题；大数据技术能汇聚历史合同、司法判例、监管政策等多元信息，借助关联分析、聚类分析等手段揭示潜在风险规律，实现风险全面筛查与精准定位。传统事后补救方式难以有效规避损失，大数据预测模型通过跟踪合作方信用状况、市场环境变化及政策法规调整等先导指标，可提前预警风险信号并启动处置流程。传统静态防控依赖人工定期检查，大数据技术构建实时数据监测系统，持续监控合同履行情况、知识产

权使用状态及数据处理合规性，形成“识别-预警-响应-优化”闭环管理，大幅增强企业法律风险防控准确性与实效性。

大数据技术在企业法律风险智能识别中的应用机制

合同法律风险的智能识别系统

合同法律风险智能识别系统依靠自然语言处理和机器学习算法，对合同文本进行深度剖析以及风险筛查工作。系统通过构建合同要素提取模型，自动提取合同主体、标的物、金额、履行周期、违约责任等关键条款，运用命名实体识别技术明确权利与义务的具体承担方。风险识别模块把过往合同纠纷案例、司法判例以及法律法规整合为风险知识图谱，借助文本匹配和语义相似度分析，发现条款中的疏漏、矛盾表述及歧义问题。系统还接入企业信用信息公示系统、中国裁判文书网等外部数据平台，建立信用评估体系来评估合作方的履约能力。以思通数科的应用实践作为例子，大型金融机构部署智能合同审查系统后，合同审核耗时缩短了40%，合规审查精准度提高了20%。科大讯飞的合同数字员工产品将法务审核效率提升到每份10分钟，合同比对时间压缩到每份5分钟。

知识产权风险的实时监测平台

知识产权风险实时监测平台依靠网络爬虫与图像识别技术，针对电商、社交平台以及企业官网等场景开展持续性监控工作。该平台运用分布式爬虫来采集商品数据，利用OCR技术识别文字信息，并且借助图像相似度来比对企业注册商标，从而有效发现近似或仿冒商标的使用情况。专利侵权监测部分会抓取竞品技术参数，将其和企业专利权利要求进行特征匹配。商业秘密保护功能能够追踪员工异常操作，像非工作时间登录系统或者批量下载核心文件等情况，还会综合离职时间与竞业条款判断泄密风险。依据广州知识产权法院的判决，珍某公司的“公众号助手”软件非法获取微信公众号用户账号密码，该行为被认定为不正当竞争，最终判决其赔偿300万元。

数据合规风险的全生命周期管控

数据合规风险全生命周期管控体系会贯穿数据采集、存储、处理、共享、跨境传输及销毁这六大流程。在数据采集阶段依托合规审查引擎自动监测采集行为是否遵循最小必要原则，系统会与《个人信息保护法》法规库保持实时联动并对违规采集行为实施阻断控制。数据存储环节采用分级分类管理方法，按照敏感度将数据划分为公开、内部、敏感、核心四个级别，并配置相应的加密保护措施。数据处理环节部署数据血缘追踪功能，完整记录数据流转轨迹、操作行为及访问主体，以确保数据使用过程全程可查，跨境传输环节构建合规评估模型，可根据目的地国家数据保护法规自动生

成风险分析报告。2022年国家网信办对滴滴公司处以80.26亿元罚款，对知网处以5000万元罚款。2021年至2023年间数据安全相关行政处罚案件累计达888件，呈现逐年增长趋势。

大数据技术在企业法律风险动态防控中的协同体系

风险数据的采集整合与质量管控

风险数据采集整合体系把企业内部业务系统、外部公共数据平台以及第三方专业数据服务这三大渠道联通起来了。企业内部通过集成合同管理系统、知识产权管理系统、人力资源系统和财务系统，来获取合同履行信息、知识产权申请信息、员工劳动关系信息以及财务往来信息。外部数据采集对接国家企业信用信息公示系统、中国裁判文书网、知识产权局专利检索系统和海关总署跨境数据监管平台，实时抓取合作方工商变更动态、涉诉裁判文书、专利授权公告和跨境数据传输备案信息。第三方数据服务引入商业征信机构、法律数据库、行业监测平台所提供的信用评级报告、法规政策动态和行业风险指数。数据质量管控通过去重、纠错、补全、标准化处理等操作来提升数据准确性，采用数据一致性校验算法识别矛盾数据。北京知产宝网络科技有限公司通过API接口，以许可使用方式提供数据服务，2023年实现收益大约1800万元。

风险分析的算法模型与预警机制

风险分析算法模型把逻辑回归、随机森林以及神经网络等多种机器学习技术整合起来，以此构建起全方位的风险评估框架。在合同违约风险预测方面，模型会将合作方的财务状况、过往履约记录、行业发展态势以及涉诉情况当作关键输入参数，最终输出违约概率评分与风险等级评定结果。知识产权侵权风险评估模型依托市场监测数据、竞品技术特点以及专利分布密度等要素，来测算侵权行为发生的可能性指数。数据合规风险模型依据数据处理操作记录、法规更新动态以及监管处罚实例等资料，对合规性缺失程度进行量化评估。预警系统设定了红色、橙色、黄色三个级别的警戒线，当风险指标达到相应阈值时，预警信息会自动发送至相关负责人。预警信息内容涵盖风险类别、风险级别、影响区域、证据关联链条以及处理措施建议五个核心要素。杭州互联网法院的判例显示，某直播平台运用数据监测分析系统成功识别出异常刷奖行为，法院最终依据相关法律作出惩罚性赔偿判决，判令赔偿金额为300万元。

风险处置的智能响应与联动机制

风险处置智能响应系统会依据预警等级自动启动对应的应急程序，面对红色预警时，系统激活最高级别应对措施，马上锁定相关业务操作权限，同时向法务、业务及风控部门

发出紧急任务指令，要求法务部门在规定时间内完成风险评估并形成解决方案；当橙色预警出现时，系统自动制作风险评估报告并分发给各部门主管，责令其在限定时间内拿出具体防控策略；黄色预警会被纳入常规监测体系，定期跟踪风险状况的演变。为保障高效协作，系统构建了跨部门的联动工作模式，其中法务团队承担法律解析及诉讼处理职责，业务部门负责提供相关背景资料与证据，风控小组评估可能造成的经济损失，信息技术部门协助进行数据调用和权限管控。系统完整记录所有风险处理的响应时长、具体方法及最终成效，逐步积累形成一套完善的案例知识库，比如最高检曾披露的案例中，某企业因网络安全防护体系存在缺陷，最终导致前员工删除核心数据，造成超过 12 万元的直接经济损失。

大数据支持下企业法律风险智能防控的保障机制

一体化数据治理体系的构建

以数据中台作为核心架构来构建一体化数据治理体系，以此对企业法律风险相关数据开展统一管理工作。数据中台借助元数据管理系统，去制定统一的数据标准、编码规则及命名规范，从而实现合同、知识产权、合规、人力资源等异构数据的标准化整合。数据资产目录明确各类数据的归属部门、访问权限、共享范围及更新频率，能够有效打破部门间的数据壁垒。数据质量管理模块设置完整性、准确性、一致性、及时性四维评估指标，针对入库数据实施自动化质量检测。数据安全采用分级加密技术，核心敏感数据运用 AES - 256 加密算法，并且设置多因素身份认证与操作审计日志。数据生命周期管理规范了数据保存期限与销毁流程。国网北京海淀供电公司已经完成 2.5 亿条数据登记，还获得北京市首批 15 项数据知识产权登记，“城市大脑园区碳监测分析”数据资产已完成入表并计列存货。

智能化分析与预警能力的提升

智能化分析能力的强化，需依靠算法模型的持续精进以及技术工具的深度整合。企业要建立算法模型的迭代流程，定期把新增风险案例、司法裁判数据和监管处罚信息当作训练素材，对模型参数与特征工程进行优化。采用深度学习技术处理非结构化数据，运用 BERT 模型解析合同条款语义，借助卷积神经网络识别商标图形相似度，利用循环神经网络预测数据合规风险走向。预警能力提升需要构建动态阈值调整机制，依据行业风险状况、企业风险容限以及历史预警准确率自动调整阈值。开发可视化风险监控驾驶舱，通过热力图呈现各业务板块风险分布，以趋势图展示风险指标变化，运用关系图谱揭示风险传导链条。培养法律科技复合型人才，为法务人员提供 Python 编程与数据分析培训，对技术人员开展合同法、知识产权法及数据合规法律培训。艾瑞

咨询数据显示，2020 年至 2025 年中国法律科技年复合增长率预计为 28.6%，2025 年合同智能审查工具覆盖率有望超过 45%。

数据驱动的合规管理制度完善

数据驱动的合规管理体系借助大数据分析结果，为制度优化和决策制定提供有力支持。企业构建合规风险定期分析报告机制，法务部门通过大数据分析生成法律风险专项报告，明确高频风险类型、高风险业务环节以及高风险合作对象。合同管理制度依据智能审查系统识别的高频风险条款，完善合同模板库和必备条款清单，规范不同金额和类型合同的审批权限与审查流程。知识产权管理制度根据监测平台发现的侵权高发领域，强化商标注册布局、专利申请时机以及商业秘密保护措施。数据合规管理制度结合合规风险评估结果，细化数据分级分类标准、优化数据授权审批流程并更新数据安全保护措施。建立风险防控绩效考核机制，将各部门风险识别数量、预警响应速度以及风险处置效果纳入年度考核指标。知产宝案例库显示，驰名商标侵权案件中 80.99% 的判赔金额低于 10 万元，1 亿元以上赔偿案件仅占 0.14%。

结束语

大数据技术给企业法律风险的识别与防控提供技术支持，还革新了原有的管理方式。借助智能识别系统能精准锁定合同、知识产权、数据合规等领域风险，依托协同防控体系可完成数据采集、算法分析及智能响应全链条管理。通过建立保障机制能强化数据治理根基、提升技术水平、完善合规体系。大数据技术成功解决传统模式中效率不高、排查不彻底等问题，促使企业法律风险管理从依赖经验转向依赖数据。未来随着人工智能、区块链等技术的深度融合，企业法律风险防控将朝着智能化、自动化方向发展，这需要企业、监管机构及学术界共同协作。在技术创新与制度建设方面齐头并进，从而构建起支撑高质量发展的法律保障框架。

作者简介：翟昌彬 同方威视技术股份有限公司

责任编辑：杨佳宇 **投稿邮箱：**zhouhl@staff.ccidnet.com