

团 体 标 准

T/ISC XXX—2026

人工智能应用场景下的法律知识库建设规范

Specification for the construction of legal knowledge bases in artificial intelligence
application scenarios

在提交反馈意见时，请将您知道的相关专利与支持性文件一并附上。

（征求意见稿）

2026 - XX - XX 发布

2026 - XX - XX 实施

中 国 互 联 网 协 会 发 布

目 次

前 言 IV

引 言 VI

人工智能应用场景下的法律知识库建设规范 1

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

3.1 法律知识库 legal knowledge base 1

3.2 法律数据 1

3.3 权威来源 1

3.4 知识条目 1

3.5 元数据 metadata 2

3.6 时效性 2

3.7 效力状态 2

3.8 版本管理 2

3.9 检索增强生成 retrieval-augmented generation; RAG 2

3.10 数据脱敏 data masking 2

3.11 法律大模型 legal large language model 2

3.12 标注 2

3.13 语料 2

4 符号和缩略语 2

5 总体要求 2

5.1 建设原则 2

5.2 总体架构 3

5.3 建设流程 3

5.4 组织与人员 3

5.5 收录范围管理 3

6 数据采集要求 3

6.1 来源管理 3

6.2 采集合法性 4

6.3 权威性核验 4

6.4 采集技术要求 4

7 分类与编目规范 4

7.1 分类体系 4

7.2 编目要求 5

7.3 元数据 5

7.4 标签与知识关联 5

8 存储与结构化要求 5

8.1	数据加工与清洗	5
8.2	结构化要求	5
8.3	存储要求	6
8.4	入库质量管理	6
9	更新与维护机制	6
9.1	时效性监测	6
9.2	更新时限	6
9.3	版本管理	7
9.4	废止与下线管理	7
9.5	维护保障	7
10	数据安全与合规管理	7
10.1	数据分类分级	7
10.2	个人信息保护	7
10.3	著作权与授权合规	7
10.4	访问控制与审计	8
10.5	数据出境管理	8
10.6	安全事件应急	8
10.7	第三方合作与委托处理	8
10.8	系统安全与攻击防护	8
11	应用管理要求	8
11.1	服务接口规范	8
11.2	面向人工智能应用的要求	9
11.3	用户反馈与持续改进	9
12	质量评价	9
12.1	建立知识库质量评价指标体系	9
12.2	质量评价的开展	9
附 录 A	（规范性） 法律知识库核心元数据规范	10
附 录 B	（资料性） 法律知识时效性监测要点清单	11
附 录 C	（资料性） 法律知识库数据结构示例	12
参 考 文 献		14

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国互联网协会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

引 言

近年来，法律科技与生成式人工智能快速发展，法律数据库、智能检索系统、法律大模型等产品在司法办案、律师执业、企业合规、公共法律服务等场景中得到广泛应用。法律知识库作为上述产品的核心数据底座，其内容质量、结构化程度与管理水平直接决定法律信息服务的准确性、权威性与公信力。

从行业现状看，法律知识库建设尚缺乏统一标准：数据来源认定、采集加工规则、分类编目方法、更新维护机制等方面各机构做法不一，部分产品存在法规版本陈旧、失效文件未标注、案例数据引用不规范等问题。法律知识具有强时效性与强权威性，法规的制定、修改、废止以及司法解释的更替若不能及时同步至知识库，将直接导致产品输出错误结论，引发执业风险与合规风险。例如，援引已废止的条文出具法律意见、依据失效的监管规定开展合规审查，均可能对使用者造成实质性误导，甚至引发赔偿责任。同时，法律大模型训练与检索增强生成（RAG）等应用对高质量结构化语料的需求日益迫切，低质量语料会直接传导为模型幻觉与错误输出，知识库建设的标准化有助于从源头降低行业重复投入、提升语料供给质量。此外，裁判文书、案例数据的采集与使用涉及个人信息保护、著作权、数据安全等多重合规要求，司法公开数据的商业化利用边界亦需厘清，亟需明确数据合规边界与安全管理要求。

制定本文件的必要性主要体现在三个方面：一是质量保障需求，通过统一来源核验与更新时限要求，确保法律知识产品输出的准确性与权威性；二是产业发展需求，通过统一分类编目与结构化标准，形成行业可复用的高质量语料供给，支撑法律人工智能产业的集约化发展；三是安全合规需求，通过明确数据分类分级、个人信息保护与授权合规要求，守住法律知识服务的数据安全底线。

基于上述背景，制定本文件的目的在于：确立人工智能应用场景下法律知识库全生命周期建设的基本规范，统一数据采集、分类编目、加工存储、更新维护、安全合规与应用管理的技术要求和管理要求，引导行业建立权责清晰、质量可靠、安全可控的法律知识库体系，为法律信息化与法律人工智能的健康发展提供标准支撑。

本文件的适用对象为从事法律知识库建设、运营与应用的组织，包括法律人工智能产品研发企业、律师事务所、司法行政机关、企业法务及合规部门等。

本文件首次系统提出法律知识库全生命周期建设规范，围绕采集、编目、加工、更新、安全、应用各环节形成闭环要求；面向法律大模型训练与检索增强生成等新型应用场景，提出适配人工智能应用的知识结构化与质量要求；针对法律行业特有的权威性、时效性痛点，建立时效性监测与版本管理机制，以期填补行业标准空白，为各方主体提供可执行的建设依据。

人工智能应用场景下的法律知识库建设规范

1 范围

本文件规定了人工智能应用场景下法律知识库建设的总体要求、数据采集、分类与编目、存储与结构化、更新与维护、数据安全与合规管理以及应用管理要求。

本文件适用于法律人工智能产品研发企业、律师事务所、司法机关、企业法务及合规部门等组织开展的法律知识库规划、建设、运营与应用活动。高等院校、科研机构、行业协会等建设专业法律知识库可参照使用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 22239—2019 信息安全技术 网络安全等级保护基本要求
GB/T 35273—2020 信息安全技术 个人信息安全规范
GB/T 37988—2019 信息安全技术 数据安全能力成熟度模型
GB/T 41867—2022 信息技术 人工智能 术语
GB/T 43697—2024 数据安全技术 数据分类分级规则
GB 45438—2025 网络安全技术 人工智能生成合成内容标识方法

3 术语和定义

GB/T 41867—2022界定的以及下列术语和定义适用于本文件。

3.1 法律知识库 legal knowledge base

以法律领域知识为内容，按照一定的知识体系、数据模型和技术规范组织、存储，可支持检索、查询、分析与智能化应用的数据集合及其管理系统。

3.2 法律数据

法律知识库所收录的各类数据资源，包括法律法规、司法解释、司法案例、行政执法文书、合同范本、学术观点、实务指引以及其他与法律相关的信息。

3.3 权威来源

依法有权发布法律数据的国家机关官方网站、官方出版物，或者经合法授权、能够证明其内容与官方发布一致的传播渠道。

3.4 知识条目

法律知识库中具有独立标识、完整语义和明确来源的最小可管理单元，如一部法规、一个条文、一篇裁判文书、一条观点文献。

3.5 元数据 metadata

描述知识条目的属性、来源、效力状态、版本、关联关系等信息的数据。

3.6 时效性

法律知识库中的知识条目与其权威来源最新有效状态保持一致的程度。

3.7 效力状态

表明规范性法律文件现行有效、已被修改、已废止、尚未生效等法律效力的标识信息。

3.8 版本管理

对知识条目的历次修订版本进行记录、标识、保存和追溯的管理活动。

3.9 检索增强生成 retrieval-augmented generation; RAG

大语言模型在生成内容前，先从知识库中检索相关知识并将检索结果作为生成依据的技术方法。

3.10 数据脱敏 data masking

通过去标识化、匿名化等技术手段，对数据中的个人信息及敏感信息进行处理，使其在不借助额外信息的情况下无法识别特定自然人的过程。

3.11 法律大模型 legal large language model

以法律语料为主要训练或微调数据，面向法律检索、问答、文书生成、合同审查等专业任务的大语言模型。

3.12 标注

按照预设规则对法律数据中的实体、关系、属性、效力等要素进行识别并添加结构化标记的加工活动。

3.13 语料

用于大语言模型训练、微调或检索增强的、经加工整理的文本数据集合。

4 符号和缩略语

下列符号和缩略语适用于本文件。

API：应用程序编程接口（Application Programming Interface）

JSON：JavaScript对象表示法（JavaScript Object Notation）

OCR：光学字符识别（Optical Character Recognition）

RAG：检索增强生成（Retrieval-Augmented Generation）

UTF-8：Unicode转换格式8位编码（8-bit Unicode Transformation Format）

XML：可扩展标记语言（Extensible Markup Language）

5 总体要求

5.1 建设原则

- 1）合法合规。法律知识库的采集、加工、存储、传播与应用应符合国家法律法规和监管要求，尊重知识产权，保护个人信息和数据安全，不得损害国家利益、社会公共利益和他人合法权益。
- 2）权威准确。知识条目应以权威来源为准，内容应与来源保持一致，关键数据应经过核验，确保知识库内容的准确性与权威性。
- 3）完整系统。应按照明确的收录范围与分类体系建设知识库，保障核心法律数据门类的完整性，知识条目之间的关联关系应清晰、可追踪。
- 4）动态时效。应建立与法律数据立、改、废节奏相适应的更新机制，及时反映规范性文件的效力变化，确保知识库内容持续有效。

5) 安全可控。应落实数据分类分级保护要求，建立访问控制、安全审计与应急处置机制，保障知识库全生命周期的数据安全。

6) 开放兼容。数据格式、接口设计宜采用通用技术标准，支持与检索系统、人工智能应用等外部系统的对接与协同。

5.2 总体架构

法律知识库宜按以下层次构建：

- a) 数据源层：权威发布渠道、合法授权数据库、自建实务资料等数据来源；
- b) 采集层：承担数据获取、格式转换、完整性校验与采集日志记录等功能；
- c) 加工层：承担清洗、脱敏、结构化拆分、标引、质量校验等功能；
- d) 存储层：承担结构化数据、非结构化原文、元数据与版本数据的存储管理；
- e) 服务层：提供检索、查询、接口调用、订阅推送等服务能力；
- f) 应用层：面向法律检索、智能问答、大模型训练、合规审查等业务场景。

与上述层次并行，应建立贯穿全程的标准规范体系与安全质量管理体系，覆盖制度、流程、人员与技术保障措施。

5.3 建设流程

法律知识库建设宜遵循规划立项、需求分析、数据盘点、采集加工、入库审核、发布运行、运营维护、评价改进的基本流程。各阶段应形成相应文档记录，包括但不限于建设方案、数据来源清单、加工规则说明、质量检测报告与运维记录，并妥善保存以备查验。

5.4 组织与人员

- 1) 建设主体应设立或明确负责知识库建设运营的管理岗位，配备具有法律专业知识与数据管理能力的内容审核、数据加工、技术运维人员。
- 2) 应建立岗位职责分工与审核授权机制，数据采集、内容审核、发布上线等关键环节应职责分离，重要内容的发布应经双人复核。
- 3) 应对相关人员开展定期培训，内容涵盖本文件要求、法律数据更新动态、数据安全与个人信息保护要求等。

5.5 收录范围管理

- 1) 建设主体应制定成文的收录范围说明，明确收录的数据类别、地域范围、时间跨度、效力层级与不收录事项，并对外公开或向用户提供，避免用户对产品覆盖范围产生误解。
- 2) 收录范围应保障核心数据的完整性：收录法律法规的，应覆盖现行有效的法律、行政法规与司法解释；收录裁判文书的，应说明文书来源平台、年份与地域覆盖情况。
- 3) 收录范围调整的，应留存调整记录；缩小收录范围或停止更新既有类目的，应提前向用户作出提示。

6 数据采集要求

6.1 来源管理

1) 应建立数据来源清单管理制度，明确各类法律数据的权威来源。规范性法律文件应以发布机关的官方网站、官方公报为准；司法案例应以最高人民法院、最高人民检察院等司法机关官方发布平台为准；无法从官方渠道获取的数据，应使用经合法授权的商业数据库或出版机构数据，并留存授权证明。

- 2) 应建立来源准入与评估机制，对数据来源的合法性、权威性、稳定性进行定期评估，来源清单发生变更的应留存记录。
- 3) 不得将来源不明、无法核验真伪的数据作为知识库的正式收录内容。用户上传、网络爬取等非权威渠道获取的数据，经与权威来源核验一致后方可入库。

6.2 采集合法性

- 1) 数据采集应遵守《中华人民共和国著作权法》《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律法规，不得侵犯他人的著作权、商业秘密及其他合法权益。
- 2) 通过网络自动化方式采集数据的，应遵守目标网站的 robots 协议、服务条款及访问频率限制，不得采取破坏、绕过技术保护措施的方式获取数据，不得对目标网站的正常运行造成影响。
- 3) 采集涉及个人信息的裁判文书、案例数据时，应核验来源平台是否已依法进行脱敏处理；对未脱敏的原始数据，应在入库前完成脱敏加工。
- 4) 通过采购、合作等方式获取数据的，应签订书面协议，明确数据内容、授权范围、使用方式、转授权限制、保密义务与违约责任，并核查数据来源方的合法授权链条。

6.3 权威性核验

- 1) 入库前应对知识条目进行权威性核验。法律法规、司法解释等规范性文件应与权威来源全文比对，核验内容包括标题、发文字号、发布机关、发布日期、施行日期、效力状态及正文内容。
- 2) 重要数据宜采用两个以上独立来源交叉校验；不同来源内容不一致的，应以权威来源为准，并记录差异处理过程。
- 3) 对官方发布的修订、修正、废止信息应及时核验并同步更新，确保效力状态标识准确。

6.4 采集技术要求

- 1) 应根据数据更新频率制定采集计划，对核心法律数据源实行常态化监测采集。法律法规、司法解释等时效性要求高的数据，宜每日监测；案例、文献类数据可按周或按月采集。
- 2) 采集过程应记录采集日志，包括采集时间、来源地址、数据量、校验结果与操作人员，日志保存期限不少于一年。
- 3) 应建立采集完整性校验机制，对漏采、重复采集、格式损坏等情况进行自动检测与告警，并形成处理记录。
- 4) 采集工具与采集脚本应实行版本管理与变更审批，采集规则调整可能影响已采数据的，应评估影响范围并留存评估记录；不得使用来源不明的第三方采集工具处理包含敏感信息的数据。

7 分类与编目规范

7.1 分类体系

- 1) 应按照法律知识的内容属性建立基本分类体系，至少包括以下类目：
 - a) 法律法规：法律、行政法规、监察法规、司法解释、部门规章、地方性法规、地方政府规章、自治条例和单行条例等规范性文件；
 - b) 司法案例：指导性案例、公报案例、典型案例、裁判文书及其他案例资源；
 - c) 学术与实务观点：学术论文、专著、实务文章、立法释义与官方解读；
 - d) 合同与文书范本：合同示范文本、诉讼文书样式、法律文书模板；
- 可基于上述基础类目，按行业、业务领域搭建专题知识库，如金融、数据合规、知识产权专题库。

- 2) 法律法规类目应按照效力层级细分，并标注制定机关与施行地域；案例类目宜按照审判机关层级、案件类型、案由细分。同一文件涉及多个类目的，应确定主类目并建立交叉索引。
- 3) 应在基本分类之上建立多维分类标签，包括但不限于：效力层级、制定机关、地域、行业领域、案由、业务场景。多维标签规则应成文发布并保持稳定，调整时应有版本记录。
- 4) 分类体系应兼顾法律专业习惯与计算机处理要求，类目划分应互斥清晰、层次分明，避免出现条目无法归类或多重归类无规则可依的情形。

7.2 编目要求

- 1) 应为每个知识条目分配唯一标识符，标识符应稳定、不可复用，条目内容修订或版本更替时标识符保持不变。
- 2) 编目信息应完整著录知识条目的来源、责任主体、时间、效力等要素，著录规则应统一、成文，并保持前后一致。
- 3) 同一文件的不同版本、同一事项的不同表现形式（如原文与修订对照文本）应建立关联编目，便于检索与引用。

7.3 元数据

- 1) 应为知识条目建立元数据记录，核心元数据字段及填写要求见附录 A。附录 A 所列字段为规范性要求，建设主体可在此基础上扩展，但不得删减或变更其语义。
- 2) 元数据应与知识条目同步创建、同步更新，元数据变更应留痕可查。
- 3) 效力状态、施行日期等关键元数据缺失或存疑的知识条目，不得标注为“现行有效”对外提供服务。

7.4 标签与知识关联

- 1) 宜为知识条目设置主题标签、关键词，标签词表宜统一管理，避免同义标签混用。
- 2) 宜建立知识条目之间的关联关系，包括法规与条文、法规与案例、案例与案例（如援引、类案）、法规与观点之间的引用与关联，关联类型应有明确定义。
- 3) 采用自动化方式进行标签标注与关联抽取的，应建立人工抽检复核机制，抽检比例不应低于 5%，标注准确率合格标准不应低于 95%；抽检比例与合格标准应成文规定。

8 存储与结构化要求

8.1 数据加工与清洗

- 1) 入库数据应经过去重、纠错、格式规范化处理。经 OCR 识别的文本应进行校对，识别错误率应控制在成文规定的阈值以内，关键字段（如文号、日期、条文序号）应逐一核对。
- 2) 应保留知识条目的原始文本或原始文件，加工过程不得破坏原始数据的可恢复性；加工规则、加工工具版本应记录备案。
- 3) 清洗过程中发现数据内容矛盾、缺漏或明显错误的，应记录并回溯来源核验，不得在无法核实的情况下擅自修改实质性内容。

8.2 结构化要求

- 1) 法律法规类数据应结构化至条文层级，编、章、节、条、款、项层级关系应完整保留，条文内容不得截断或缺漏。

- 2) 结构化数据应采用 JSON、XML 等通用格式组织，字符编码应统一采用 UTF-8，数据模型应文档化并对外保持一致。
- 3) 面向大模型训练与 RAG 应用的知识内容，应在保持语义完整的前提下进行合理分块，分块应携带所属条目标识、条文定位与版本信息，确保检索结果可溯源至原始出处。
- 4) 语义标注（如实体识别、法律关系标注）应有明确的标注规范与质量验收标准；自动标注结果用于对外服务的，应经抽检合格。

8.3 存储要求

- 1) 应根据数据类型与访问需求选择适当的存储方式，结构化数据、非结构化原文与索引数据可分开存储，但应保持关联一致。
- 2) 应建立数据备份机制，重要数据实行定期备份与异地容灾备份，备份数据应定期验证可恢复性；备份周期与保留策略应成文规定。
- 3) 存储系统应按照 GB/T 22239—2019 相应等级的要求落实安全防护措施，知识库对外提供商业化服务的，安全保护等级不宜低于第三级。

8.4 入库质量管理

- 1) 应建立入库质量校验规则，对完整性（字段齐全、内容无缺漏）、准确性（与权威来源一致）、一致性（格式与关联关系符合规范）进行校验，合格后方可入库发布。
- 2) 应建立质量抽检制度，定期对在库知识条目抽样核验，抽检结果与整改情况应形成记录。
- 3) 发现已入库内容存在错误的，应按更新流程及时更正，并评估错误影响范围，必要时向受影响用户进行提示。
- 4) 宜建立关键数据的度量指标并持续监控，包括在库条目总量、各类目覆盖情况、平均更新耗时、错误率与返工率等，度量结果作为质量评价（见第 12 章）的输入。

9 更新与维护机制

9.1 时效性监测

- 1) 应建立法律数据时效性监测机制，对权威来源的发布动态进行持续跟踪，监测范围应覆盖知识库收录的全部规范性文件类别。时效性监测要点清单见附录 B。
- 2) 应实行分级监测：法律、行政法规、司法解释及收录范围内的核心规章，宜每日监测；地方性法规、其他规范性文件可结合更新频率确定监测周期，但不得超过一周。
- 3) 监测手段可结合官方渠道订阅、自动化采集与人工巡查，监测结果应形成记录。

9.2 更新时限

- 1) 新发布的法律、行政法规、司法解释，应自权威来源发布之日起 2 个工作日内完成入库；部门规章、地方性法规宜在 5 个工作日内完成入库。
- 2) 规范性文件被修改或废止的，应自权威信息发布之日起 2 个工作日内完成效力状态更新与标注。
- 3) 因特殊原因无法在规定时间内完成更新的，应记录原因并采取临时提示措施，避免用户误用失效内容。
- 4) 对法律体系影响重大的立法事件（如重要法律颁布、司法解释集中清理），应制定专项更新预案，明确责任人员、完成时限与质量复核要求，必要时暂停相关自动化推送服务，待人工复核完成后统一发布。

9.3 版本管理

- 1) 应对规范性文件的历次版本进行完整保存，形成版本链，明确各版本的施行日期与失效日期，支持按时间点查询历史有效版本。
- 2) 现行版本与历史版本应有明显区分标识；废止文件应保留可查并标注“已废止”，不得简单删除。
- 3) 版本变更应记录变更原因、依据、时间与操作人员，确保变更过程可追溯。
- 4) 历史版本数据应保障可正常查询调用。涉及诉讼追溯、合规审计等需要还原特定时间点有效法规状态的场景，应能够依据版本链准确提供当时有效的文本，避免因版本缺失导致事实认定偏差。

9.4 废止与下线管理

- 1) 知识条目因内容失效、授权终止或质量不合格需要下线的，应履行审核程序，保留下线记录与必要的档案副本。
- 2) 已下线内容不得继续通过服务接口对外提供；已被外部应用引用的，宜通过接口返回状态标识等方式告知调用方。

9.5 维护保障

- 1) 应明确内容维护责任岗位，建立 7×24 小时监测告警与工作日响应机制，重大更新事项（如重要法律颁布）应制定专项保障预案。
- 2) 应定期评估更新维护工作的及时性与准确性，评估结果纳入质量评价（见第 12 章）。

10 数据安全与合规管理

10.1 数据分类分级

- 1) 应参照 GB/T 43697—2024 确立的规则，结合法律数据特点对知识库数据进行分类分级，至少区分：公开发布的法律数据、经脱敏处理的案例数据、涉及个人信息的未脱敏数据、商业秘密及未公开实务资料。
- 2) 不同级别数据应实行差异化管控，明确各级数据的访问权限、使用范围、对外提供条件与存储要求。涉及重要数据的，应依法履行相应保护义务。

10.2 个人信息保护

- 1) 裁判文书、案例数据入库前应完成脱敏处理，隐匿或删除依法不应公开的个人信息，包括但不限于自然人的身份证号码、住址、联系方式、银行账号，以及依法不公开案件中当事人的可识别信息。
- 2) 脱敏规则应成文规定并定期评估有效性，防止脱敏不彻底或可复原的伪脱敏；脱敏处理应保留处理记录。
- 3) 因业务需要处理未脱敏个人信息的，应具备合法基础，遵循最小必要原则，并落实 GB/T 35273—2020 规定的要求。

10.3 著作权与授权合规

- 1) 收录他人享有著作权的作品（如学术文章、实务专著、数据库汇编成果），应取得合法授权，明确使用方式、范围与期限，并按约定支付报酬。
- 2) 对知识库中具有独创性的汇编成果、加工成果，宜通过版权声明、技术措施等方式维护自身合法权益，同时不得阻碍他人对公共领域法律数据的正当使用。
- 3) 对外提供知识库服务时，应通过协议明确用户使用边界，约定禁止大规模抓取、反向工程等限制条

款，并保留必要的维权证据。

10.4 访问控制与审计

- 1) 应建立账号权限管理制度，按照岗位职责与最小授权原则分配访问权限，高权限操作应履行审批程序；人员离职或岗位变动应及时回收权限。
- 2) 管理后台与数据加工环境的访问应实施身份鉴别与访问控制，敏感操作应留存审计日志，日志保存期限不少于三年。
- 3) 应定期开展安全风险评估与漏洞排查，参照 GB/T 37988—2019 持续提升数据安全能力。

10.5 数据出境管理

知识库数据向境外提供，或者境外主体可访问、调用境内知识库数据的，应事先评估数据类型与规模，依法履行数据出境安全评估、个人信息出境标准合同备案或认证等程序；涉及未脱敏案例数据、重要数据的，未经合规审查不得出境。

10.6 安全事件应急

应建立数据安全事件应急预案，发生数据泄露、篡改、丢失、非法访问等事件时，立即采取隔离、止损、取证等措施；涉及个人信息泄露的，依据《中华人民共和国个人信息保护法》第五十七条采取补救措施并履行通知义务。事件处置过程与整改情况应记录归档。

10.7 第三方合作与委托处理

委托第三方开展数据采集、加工、标注或存储服务的，应评估受托方的数据安全能力，签订书面协议明确数据安全责任与保密义务，约定委托事项完成后的数据返还或删除要求，并对受托方的处理活动进行监督。因第三方原因造成数据安全事件的，建设主体不免除对用户的责任。

10.8 系统安全与攻击防护

- 1) 应建立网络与系统安全防护体系，部署防火墙、入侵检测与防御、抗分布式拒绝服务攻击等安全防护措施，定期开展漏洞扫描与渗透测试，及时修复安全漏洞。
- 2) 对外服务接口与数据加工环境应实施网络隔离与边界防护，防范恶意爬取、数据批量窃取、暴力破解与接口滥用；发现异常访问行为的，应采取限流、封禁等处置措施并留存记录。
- 3) 面向人工智能应用的知识库服务，应防范提示词注入、数据投毒、对抗样本等针对人工智能系统的恶意攻击，对输入内容与输出结果进行安全检测与过滤。
- 4) 应建立安全监测预警机制，对异常访问、异常查询与异常下载行为进行实时监测、告警与处置，处置过程应记录归档。

11 应用管理要求

11.1 服务接口规范

- 1) 对外提供知识库服务的，应提供标准化 API 或等效服务方式，接口文档应公开数据结构、调用方式、状态码与错误说明，并保持版本兼容管理。
- 2) 接口服务应实施身份鉴权、访问频率限制与调用日志记录，防止未授权访问与数据批量爬取；接口返回的失效或废止数据应带有明确状态标识。
- 3) 应明确服务可用性目标与故障响应机制，服务中断、数据异常等情况应及时告知用户并留存处理记录。

11.2 面向人工智能应用的要求

- 1) 用于大模型训练的语料，应满足本文件关于来源合法性、权威性核验与个人信息保护的要求；训练语料的构成、版本应有记录，确保训练数据可说明、可追溯。
- 2) 用于 RAG 场景的知识内容，应保证分块完整、定位准确、效力状态随检索结果一并返回；知识库宜支持按效力状态、时效版本进行过滤检索，避免失效内容进入生成环节。
- 3) 基于知识库提供智能问答、文书生成等服务的，应建立输出内容的随机复核或抽检机制，对可能存在的错误援引、虚构条文等风险设置提示与拦截措施，不得将未经核验的机器生成内容直接作为确定性法律结论对外提供。面向公众提供生成式法律内容服务的，还应遵守《生成式人工智能服务管理暂行办法》及 GB 45438—2025 规定的人工智能生成合成内容标识要求。
- 4) 面向用户提供检索结果的，应在结果中展示知识条目的来源、效力状态与最近更新时间，保障用户对信息时效性与权威性的知情权；引用知识库内容形成研究成果或对外文书的，宜注明数据来源与版本。

11.3 用户反馈与持续改进

- 1) 应建立用户纠错与反馈渠道，对用户报告的内容错误、时效问题及时核查处理，处理时限与结果应反馈用户。
- 2) 应定期汇总用户反馈、质量评价与运维情况，持续优化分类体系、加工规则与更新机制，相关制度文件的修订应留痕管理。
- 3) 宜建立与行业组织、科研机构、用户单位的交流机制，跟踪法律数据标准与人工智能技术的发展动态，适时评估本文件各项要求的适配性，为知识库的持续升级与标准的修订完善积累实践依据。

12 质量评价

12.1 建立知识库质量评价指标体系

- 1) 内容准确率：抽检条目与权威来源一致的比例，法律法规类核心数据应不低于 99.5%；
- 2) 收录完整率：既定收录范围内实际收录条目的覆盖比例，现行有效的法律、行政法规、司法解释等核心类目应达到 99%，其他类目应不低于 95%；
- 3) 更新及时率：在 9.2 规定时限内完成更新的事项占比，核心法规更新及时率应不低于 98%；
- 4) 溯源完备率：具备完整来源与版本信息的条目占比，应不低于 95%，其中法律法规类核心条目应达到 100%；
- 5) 脱敏合格率：案例数据脱敏抽检合格的比例，应达到 100%。

12.2 质量评价的开展

质量评价应定期开展，每年不少于一次，抽检样本量应满足统计显著性要求，每类目抽检样本不少于 30 条，该类目总量不足 30 条的实行全量检查。评价结果应形成报告并向相关管理部门通报；评价不达标的，应制定整改计划并跟踪落实。

附 录 A
(规范性)
法律知识库核心元数据规范

A.1 规定了知识条目应具备的核心元数据字段。建设主体可根据业务需要扩展字段，但不得减少本表所列字段或改变其含义。

表 A.1 法律知识库核心元数据规范

字段	说明	填写要求
唯一标识符	知识条目的全局唯一编号	必填，终身不变，不得复用
标题	条目的正式名称	必填，与权威来源一致
数据类别	所属基本类目（法规/案例/观点/范本）	必填，按 7.1 分类体系填写
来源	权威来源名称及获取渠道	必填，可溯源至原始出处
发文字号/案号	法规的文号或案例的案号	适用时必填
发布机关/审理法院	内容发布或作出的责任主体	必填，观点类可标注作者 / 发布机构
发布日期	权威来源的发布日期	必填，格式 YYYY-MM-DD
施行日期	规范性文件生效日期	规范性文件必填
效力状态	现行有效/已修改/已废止/尚未生效	规范性文件必填，动态维护
效力层级	法律/行政法规/司法解释/部门规章/地方性法规等	规范性文件必填
版本号	条目内容版本标识	必填，随修订递增
入库时间	条目首次入库时间	必填，系统生成
最近更新时间	最近一次内容或元数据更新时间	必填，系统生成
脱敏标识	是否已完成脱敏处理	案例、含个人信息的文书类必填
关联条目	与其他条目的关联关系及类型	宜填，关联关系需标注类型
授权信息	授权来源、范围与期限	涉及授权数据时必填

附 录 B
(资料性)
法律知识时效性监测要点清单

B.1 给出了法律知识库时效性监测的要点，供建设主体建立监测机制时参考。

表 B.1 法律知识时效性监测要点清单

监测对象	监测内容	建议频率	主要权威渠道
法律	新法颁布、修正、修订、废止	每日	全国人大及其常委会官方发布渠道
行政法规	制定、修改、废止	每日	中国政府网、国务院公报
司法解释	发布、修改、废止	每日	最高人民法院、最高人民检察院官方平台
部门规章	制定、修改、废止	每周不少于两次	国务院各部门官方网站
地方性法规	制定、修改、废止	每周	地方人大及其常委会官方网站
指导性案例/公报案例	新增发布、不再参照	每周	最高人民法院官方平台
裁判文书	文书公开与撤回动态	每周	中国裁判文书网等官方文书公开平台
行业监管规定	金融、数据、医疗等领域监管规则更新	每周不少于 1 次	行业监管部门官方网站

附录 C
(资料性)
法律知识库数据结构示例

本附录给出了法律知识库典型条目的数据结构示例，供建设主体设计数据模型时参考。

本附录示例中的法律法规条目取自公开立法信息，司法案例条目及相关案号、当事人信息均为虚构，仅用于说明数据结构；示例中 item_id 采用“类别代码-年份-流水号”的编码方式仅为参考，建设主体可结合自身知识体系自行设计条目标识规则。

C.1 法律法规条目数据结构示例（JSON 格式）

```
{
  "item_id": "FL-2023-0042",
  "title": "中华人民共和国公司法",
  "doc_number": "中华人民共和国主席令第十五号",
  "issuing_authority": "全国人民代表大会常务委员会",
  "publish_date": "2023-12-29",
  "effective_date": "2024-07-01",
  "validity_status": "现行有效",
  "hierarchy": "法律",
  "source": "中国人大网",
  "version": "2023 年修订",
  "articles": [
    {
      "chapter": "第一章 总则",
      "article_no": "第一条",
      "text": "为了规范公司的组织和行为,保护公司、股东、职工和债权人的合法权益……"
    }
  ]
}
```

C.2 司法案例条目数据结构示例（JSON 格式）

```
{
  "item_id": "AL-2023-1580",
  "title": "张某诉某网络科技有限公司网络侵权责任纠纷案",
  "case_number": "(2023)京01民终0000号",
  "court": "北京市第一中级人民法院",
  "trial_level": "二审",
  "case_type": "民事",
  "cause_of_action": "网络侵权责任纠纷",
  "judgment_date": "2023-06-15",
  "source": "中国裁判文书网",
  "desensitized": true,
  "related_statutes": [
    "《中华人民共和国民法典》第一千一百九十四条"
  ]
}
```

参 考 文 献

- [1] GB/T 1.1—2020 标准化工作导则 第1部分：标准化文件的结构和起草规则
- [2] 《中华人民共和国网络安全法》
- [3] 《中华人民共和国数据安全法》
- [4] 《中华人民共和国个人信息保护法》
- [5] 《中华人民共和国著作权法》
- [6] 《生成式人工智能服务管理暂行办法》
- [7] 《数据出境安全评估办法》