



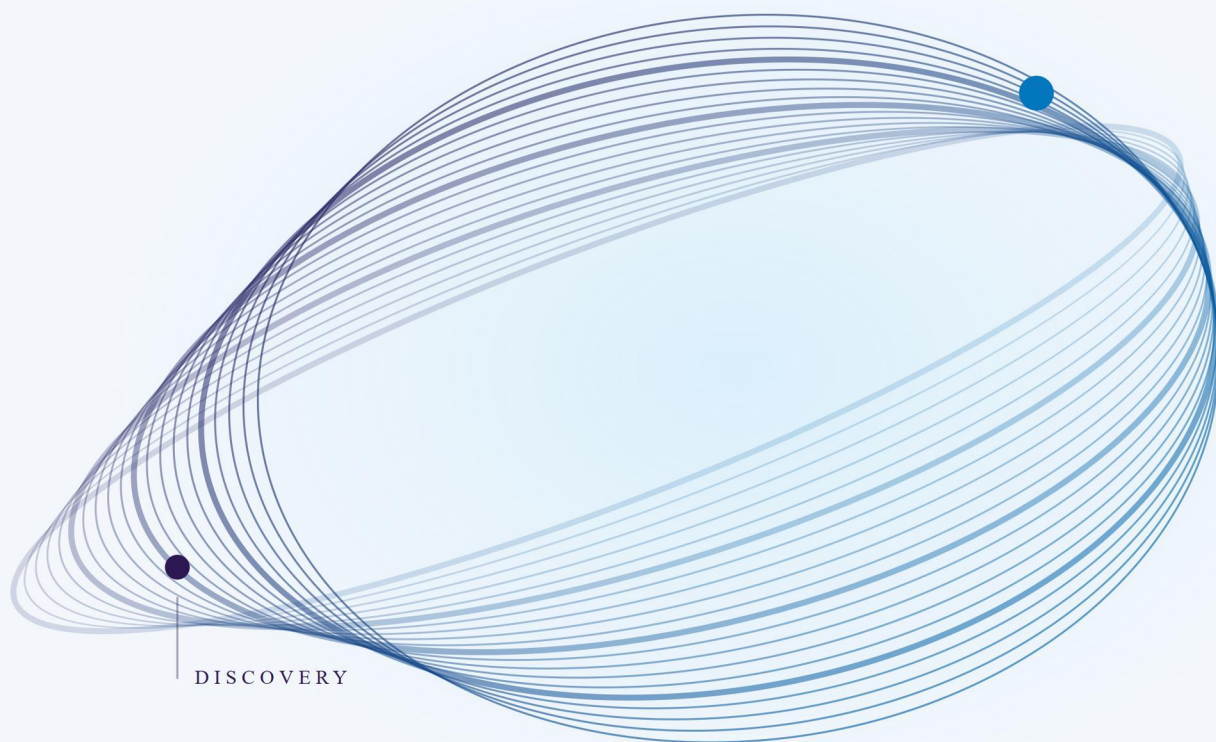
瀚思科技创新与发展协会

HATID

WHITE PAPER · HATID

智能法律服务的可信化 与规模化应用

法律知识库、智能合同与场景化合规系统的方法与实践



2024年8月

thehatid.com

出版说明

智能法律服务的可信化与规模化应用

法律知识库、智能合同与场景化合规系统的方法与实践

作者 衡森飏

出版日期 2024 年 8 月

支持单位 瀚思科技创新与发展协会

研究主题 法律治理

研究定位

本书面向律师事务所管理者、企业法务及合规负责人、法律科技产品团队和相关研究人员，讨论如何使法律知识库、智能合同管理与场景化合规工具形成可信、可持续扩展的服务能力。

研究方法

全书结合公开规范、原始技术研究、流程分析和行业典型场景推演，按照研究范围、问题诊断、对应机制、实施路径与效果评价展开。所提出的架构、操作方法及评价口径属于管理建议，其效果需要在具体场景中验证。

阅读与适用

以 2024 年 8 月为研究时点，主要讨论中国内地一般民商事法律服务与企业合规支持。后续应用应重新核对届时有效的规范，并结合事项事实、委托范围与组织条件判断。

文中方法不构成针对具体事项的法律意见，典型场景不代表任何特定机构的部署或经营成果。读者可据此确定试点边界与验收条件，再依据实际评价决定是否扩大使用。

关键词

法律知识库 · 智能合同 · 场景化合规 · 专业复核 · 可信治理

目录

第 1 章 执行摘要	4
第 2 章 研究范围与概念基础	5
第 3 章 法律服务数字化的参与者与价值链	6
第 4 章 制度环境与技术选择	7
第 5 章 从工具应用走向可持续服务	8
第 6 章 法律知识运营中的质量断点	10
第 7 章 事项流程与专业协作的损耗	11
第 8 章 资源投入与经营判断的偏差	12
第 9 章 数据协作中的权限与语义风险	14
第 10 章 合同交付与自动执行的断层	15
第 11 章 用户价值与信任的形成障碍	16
第 12 章 以可追溯知识为基础的总体架构	18
第 13 章 事项流程与责任分配机制	19
第 14 章 数据质量与经营评价机制	21
第 15 章 跨系统协作与数据边界机制	23
第 16 章 合同管理与受控执行机制	24
第 17 章 分层服务与持续运营机制	26
第 18 章 分阶段实施与典型场景	27
第 19 章 效果评价与风险复评	29
第 20 章 适用边界与发展结论	31

第 1 章 执行摘要

1.1 研究问题与主要结论

法律服务智能化正在将信息检索、文本分析和流程处理连接起来，但处理速度并不能单独证明服务质量。一个系统能够快速生成意见，并不意味着它已经掌握充分事实、使用正确依据或取得外部执行授权。本书研究如何使法律知识库、智能合同管理和场景化合规工具形成可信、可持续扩展的服务能力，面向律师事务所管理者、企业法务及合规负责人、法律科技产品团队和相关研究人员。

核心结论是：规模化必须建立在可追溯的知识、明确的事项流程、受控的数据与执行权限以及有效纠错之上。专业人员不需要重复机器完成的全部工作，但必须在足以改变结论的位置参与判断，并对交付范围和重要后果保持清楚认识。可信系统应能够说明依据、承认事实不足、接受复核并支持修正，而不是依靠流畅语言制造确定性。

1.2 六类障碍与对应机制

本书将实践中需要检查的问题归纳为六类主要障碍。知识运营中的版本和上下文缺失，要求建立分层来源与规则管理；流程交接中的事实和理由损失，要求建立事项卡、状态条件及实质复核；资源配置中的功能导向，要求围绕合格交付计算成本和价值；数据协作中的对象、语义与权限错位，要求在检索和接口阶段落实边界；合同执行中的条件误判，要求保持文本、证据和授权对应；用户理解中的过度信任，要求分层交付并提供有效异议渠道。

这些机制并非独立功能清单。规则修正需要定位历史输出，历史输出需要关联事项，事项需要明确责任人，责任人又需要能够暂停执行和联系用户。只建设其中一个环节，往往只能获得局部效率，无法承担整体服务扩展。全书因此按环境、诊断、机制、实施与评价的顺序展开，让每项建议都能追溯到具体问题。

1.3 建议的实施次序

机构宜从边界清楚的高频任务开始，先完成资料和版本治理，再验证专项辅助，按需要连接审批、交付和履约。受控执行是需要单独验收的可选能力，不是规模化的必经阶段。试点前建立人工基线和测试样本，试点中保持专业复核，扩展前核查维护资源与异常接管能力。案例检索、采购合同和合规事件三个行业典型场景展示了这一方法，均用于解释流程，不代表特定机构的既有部署成绩。

评价应同时观察专业质量、过程完整性、权限安全、执行控制、成本和用户行动。关键风险不能用总体平均分抵消，无法确定答案时的合理追问应受到认可。发生规则、模型或业务变化后，需要针对性复评；条件不再满足时，应能缩小范围、回到人工流程或退出。持续运营能力比一次成功演示更能决定规模化是否可靠。

1.4 读者可以作出的决定

读者可据此确定适合自身机构的首个场景、需要优先补齐的基础能力和扩大使用的验收条件。对法律科技采购，应要求供应商展示来源到交付再到纠错的完整路径；对内部建设，应同时安排知识维护、专业复核和数据管理资源；对服务使用，应让客户知道输出用途、关键条件和后续行动，避免把专业责任隐蔽地转移给用户。

本书以 2024 年 8 月为研究时点，主要讨论中国内地一般民商事法律服务与企业合规支持。公开规范与原始研究用于建立基础，具体架构和流程属于可验证的管理建议，不构成个案法律意见，也不承诺统一收益。研究的最终价值在于帮助机构把技术能力转化为能够解释、能够接续并对实际结果负责的服务。

第 2 章 研究范围与概念基础

2.1 从技术能力转向服务结果

法律服务的数字化可以覆盖资料收集、法律检索、文本拟订、合同审查、履约监测和争议处理，但这些环节并不具有相同的风险结构。检索工具提供一份相关资料，与系统直接发出解除合同通知，对当事人产生的影响明显不同。本书研究的核心问题是：法律服务机构如何把知识处理能力转化为可以持续交付、能够解释并接受复核的服务能力。研究对象因此包括技术系统，也包括专业人员、业务流程、数据权限和纠错渠道。

可信化指一项输出能够说明依据、适用条件、处理过程与责任归属，并在条件变化时得到修正。规模化指在业务量增加、人员轮换、客户类型扩展后，机构仍能维持明确的交付标准。前者不等于宣称系统永不出错，后者也不等于把同一答案发送给更多客户。两者的交点是可重复执行的专业控制：哪些工作允许机器完成，哪些判断必须由人作出，哪些情形应暂停并补充事实。

2.2 法律知识库与规则模型

法律知识库是对规范、案例、业务文书及相关解释进行组织和维护的信息体系。它应保存材料来源、适用范围、效力状态及相互关系，使使用者能够回到原文。知识库不以文档数量作为唯一价值尺度；缺少版本与地域标记的大型资料集合，可能把不适用的规则更快地推送给使用者。将一个条文切成若干检索片段时，定义、例外和过渡安排仍需与主规则关联。

规则模型则把某项业务判断表达为条件、动作和例外之间的关系。法律规则中的开放性概念未必能转换为明确的真假值，因此模型需要保留“事实不足”“存在争议”“需要解释”等状态。知识库回答依据在哪里，规则模型回答在什么条件下触发何种处理。二者可以相互连接，却不能相互替代；没有规范依据的流程规则，也不能仅因写入系统而成为法律要求。

2.3 智能合同的两类含义

本书区分智能合同管理与链上智能合约。前者包括条款识别、版本比较、风险提示和履约提醒，可以完全不依赖区块链。后者通常指按预设条件运行的程序，其执行结果可能连接支付、登

记或其他业务动作。自然语言合同承载的权利义务，与程序能够读取的变量并不天然一致。使用智能合约技术，需要另外处理外部事实输入、授权、错误纠正及争议解决。

为避免术语扩大化，后文以“合同管理系统”指文书与业务流程，以“自动执行模块”指能够触发业务后果的程序，以“链上存证”指对记录完整性的技术辅助。电子签名、电子数据证明力和程序执行各有不同判断对象。电子签名法为相关电子形式提供法律制度基础，但可靠电子签名的效力不能直接推导为某段程序生成的全部结果有效。[3]

2.4 研究方法 with 适用范围

研究采用规范分析、流程拆解和行业典型场景推演。公开法律与技术文献用于确定制度要求和技术原理；本书提出的操作表、分层架构及评价方法属于管理建议。文中场景用于展示方法的使用条件，不代表任何特定机构已经完成部署。讨论以中国内地一般民商事法律服务和企业合规支持为主，不把司法审判系统的规则直接移用于律师业务。

全书以 2024 年 8 月为研究时点，使用此前已经公开的规范与研究成果。读者在后续实际部署时，应重新核对当时有效的法律、行业要求及供应商条件。规范的发布日期、施行日期与适用于具体事实的时间可能不同，知识结构应分别记录，不以本书研究时点替代个案的时间判断。对刑事辩护、跨境争议、重大资产处分等高后果事项，本书只讨论辅助工作的治理原则，不提供自动决策的通用许可。明确这一范围后，才能分析各参与方提供什么能力、承担什么义务，以及为何单纯采购软件不足以完成转型。

第 3 章 法律服务数字化的参与者与价值链

3.1 需求由业务事件形成

企业和个人通常不是为了获得一段法律文字而寻求服务，而是为了完成某项交易、降低不确定性或处理已经发生的争议。服务起点应当是业务事件：谁需要作出什么决定，决定的期限是什么，错误的后果由谁承受。把需求直接理解为“生成合同”，容易遗漏交易结构、授权链条和履行能力；把需求理解为“回答法律问题”，也可能遗漏需要立即保存的材料。

需求方内部同样存在分工。采购人员关注交期和付款，财务人员关注结算与资金，法务人员关注权利义务及救济路径。数字工具只有把不同关注点带入同一事项，才可能减少往返沟通。若系统只优化法务部门接收文档后的审查速度，前端事实收集和后端执行中的等待仍会保留，整体服务周期未必缩短。

3.2 专业服务方提供判断与组织

律师事务所和企业法务部门在价值链中的核心作用，是将事实、规范与客户目标结合，形成能够承担专业责任的建议。系统可以帮助定位文本和提示遗漏，但哪些事实值得继续核查、多个风险之间如何取舍，需要结合委托范围和业务环境判断。专业服务方应当决定交付标准，并对机器输出进入正式意见的条件作出安排。

在组织层面，知识维护者、事项承办人和质量复核人可能由不同人员担任。知识维护者负责规则来源及更新，承办人负责事实与业务背景，复核人负责结论能否支持客户决策。小型团队可以兼任岗位，但不宜省略关键检查。把所有义务集中在一名“系统管理员”名下，会让技术维护承担其无法完成的专业判断，也会使承办人员误以为资料已经被全面审核。

3.3 技术供应链中的多重依赖

一个法律服务产品可能同时依赖文档解析、检索组件、基础模型、身份认证、电子签署和云存储。界面上的单一品牌并不意味着后台只有一个处理主体。采购时需要了解哪些环节由供应商自行提供，哪些交给其他服务商，数据是否被用于模型优化，故障时能否切换，以及终止服务后资料如何返还。依赖关系越隐蔽，出现问题时越难定位责任。

专业机构不必掌握每个底层算法的全部细节，但应能识别关键控制点。例如，扫描合同先经文字识别再进入条款模型，错漏可能来自识别阶段，而不是规则本身；知识检索经过第三方服务时，权限边界可能在传输中改变。将供应链画成数据流和动作流，比仅列软件名称更有助于发现风险，也便于把验收要求落实到具体接口。

3.4 公共基础设施与市场分工

公开法规、司法信息和研究成果构成行业知识基础，但公开可见不等于可以不受限制地重新处理其中全部信息。机构仍需识别材料中的个人信息、授权条款及使用条件。数据安全法提出分类分级保护等制度要求，提示行业在开发利用数据时同时安排保护责任。[2]在这一基础上，机构可以形成自己的业务分类、条款库和复核经验，而不必重复建设所有底层设施。

合理分工应当让标准化组件承担重复性工作，让专业人员处理事实差异和利益冲突。行业平台可以促进术语、接口及评价方法交流，但平台展示不能代替产品验收，协会名称也不构成具体法律结论的有效性证明。判断一项合作是否有价值，应观察它能否改善来源获取、交付衔接和责任识别，而不是合作方数量。价值链中这些关系，决定了下一章讨论的制度与技术环境如何进入日常经营。

第 4 章 制度环境与技术选择

4.1 适用规则取决于处理活动

智能法律服务需要按实际活动识别制度要求，不能只根据产品名称判断合规范围。同一软件在机构内部辅助起草，与向公众开放法律问答，服务对象和数据流向可能不同。生成式人工智能服务管理暂行办法明确了向境内公众提供相关服务的适用范围，并对未向境内公众提供服务的研发应用作出区分。[4]机构应先画清服务边界，再判断需要履行的义务。

内部使用不意味着脱离其他法律要求。涉及客户文件、员工资料或商业秘密时，仍需审查处理目的、授权来源、访问范围和保存安排。采购合同中的“合规产品”表述不能替代使用方对具体场景的判断。对外服务上线后，如果用户群、生成能力或外部接口发生变化，原有判断也应重新评估，避免用试点时期的结论覆盖新的业务。

4.2 专业责任限制自动化范围

法律服务的交付往往影响他人对权利和义务的判断，因此准确性要求与一般文本生产不同。最高人民法院关于规范和加强人工智能司法应用的意见强调辅助定位等原则，其直接对象是司法应用。[6]本书借鉴其中对技术与专业判断关系的关注，提出法律服务机构应保留清晰的人工作业位置；这一管理建议不等同于把司法文件的适用范围扩大到所有企业软件。

对于能否自动发出某项意见，应结合后果、可逆性和事实完备程度判断。内部提示遗漏的邮件附件通常可以由系统生成，涉及终止交易、承认责任或放弃权利的文本则需要更严格的授权。系统输出越接近外部行为，控制要求越高。若同一按钮同时生成文本、完成审批并向客户发送，使用者将难以发现中间环节的判断缺口。

4.3 技术进步与经济约束并存

检索增强生成把外部资料检索与语言生成连接起来，为知识密集型任务提供了一条技术路径。相关原始研究展示了将参数化与非参数化记忆结合的方法及其在研究任务中的表现。[8]法律场景可以借鉴这一思路，但研究基准上的结果不能直接作为合同审查准确率。来源过滤、引用对应关系及专业复核仍需针对业务另行验证。

技术采购的经济性也不能只比较单次调用价格。低价组件如果要求大量人工纠错，可能提高每项合格交付的总成本；自建系统若缺乏维护人员，也可能在规则变化后快速失效。机构应同时估算资料整理、权限治理、测试、培训和退出迁移成本。预算应围绕具体事项的全流程配置，而不是只为模型能力或界面功能支付费用。

4.4 信任来自可见的控制

用户可能把流畅语言、完整格式和即时响应视为专业程度的信号，但这些外观与结论是否可靠并无必然对应。产品设计应让关键事实缺口、资料版本和人工复核状态容易被看见。一个明确说明“尚需核对签署权限”的结果，往往比一段没有来源的肯定结论更能支持下一步行动。可信体验应当让用户知道现在能做什么、还不能做什么。

在服务竞争中，长期价值可以体现在问题定位更准确、交付更稳定、纠错更迅速，而不必依赖绝对化宣传。机构应谨慎使用“完全替代”“零风险”等表述，因为这类承诺掩盖了法律解释和事实认定的开放性。制度边界、成本约束与用户理解共同决定技术选择。成熟度评价因此应以可验证的控制能力为基础，不能按使用了多少新技术简单排序。

第 5 章 从工具应用走向可持续服务

5.1 资料电子化阶段

初始阶段的重点是让资料能够被找到、阅读和关联。机构可以先统一事项编号、文件命名、合同版本和归档位置，明确哪些扫描件需要回查原件。此时未必需要生成式能力，简单的检索与权限管理就可能解决大量重复查找问题。评价标准应是资料是否完整、是否能定位来源，以及离开原承办人后其他授权人员能否继续处理。

电子化不等于结构化。将大量文档搬入共享目录，如果没有区分已签署版本、谈判稿和客户提供的参考文本，仍会产生使用错误。进一步建设知识库之前，应当先解决文件状态和权属问题，否则后续系统可能把不同性质的内容一起学习和调用。基础治理看似不直接产生新功能，却决定智能化能力是否建立在可靠输入之上。

5.2 专项辅助阶段

第二阶段围绕边界明确的任务提供辅助，例如抽取当事人、比较条款变化、定位法律依据或形成待核事项清单。机构需要为每项功能建立可接受的输入格式、输出要求和人工检查位置。任务越具体，越容易构造有效测试；“完成所有法律研究”则难以界定成功与失败，也难以对结果负责。

这一阶段应观察使用者是否真正改变了工作方法。如果承办人仍需逐字重做系统的全部工作，工具可能只增加了额外步骤；如果承办人因为输出流畅而跳过核查，则风险可能被扩大。合适的状态是系统减少重复查找，专业人员把时间投入事实差异和关键判断。功能保留与否，应由实际工作记录和错误分析决定，而不是演示效果决定。

5.3 流程协同与受控执行阶段

进一步发展时，工具可以与事项受理、审批、签署和履约管理连接。协同的关键在于每次交接都说明对象、状态和接收责任，避免前一系统显示完成而后一系统尚未接收。某些低后果动作可以在满足条件后自动执行，但应独立设置执行权限、触发证据和停止开关，不能因为检索功能通过测试，就默认开放外部执行。

机构可能在合同归档上达到较高成熟度，在复杂案件分析上仍处于初级阶段。这种不均衡是正常现象，不能用一张整体等级证书掩盖。建议按任务分别记录资料治理、规则维护、人工复核、故障恢复和评价能力。记录可以使用“尚未建立”“已建立但未验证”“验证后受限使用”“持续监测运行”等状态，并附可查看的工作记录；这些状态是机构自评工具，不是外部认证或行业排名。只有相关维度都能支撑业务后果时，才进入下一阶段；任何关键依赖失效，都应允许退回人工流程。

5.4 成熟度是持续状态

成熟度不是项目结束时的永久标签。规则更新、人员流动、供应商变更和业务扩张都会改变系统表现。组织应为每一类场景设置维护责任和复评条件，尤其关注长期未触发的规则、持续被人工覆盖的提示以及无人认领的异常。这些现象通常比功能数量更能说明系统是否仍适合使用。

本书将成熟度理解为能否在变化中保持可解释、可纠错和可接续的服务。管理者可据此选择建设次序：先确保资料可用，再验证任务辅助，之后按需要连接业务流程；是否增加受控执行能力，应另行论证和验收。后续六章将分别分析知识运营、流程交接、资源投入、数据协作、交付执行及用户理解中的障碍，说明为何成熟度提升必须解决具体问题，而不能依靠一次性技术升级。

第 6 章 法律知识运营中的质量断点

6.1 检索相关与法律适用的距离

检索系统通常从文字相近、语义相似或历史点击等信号中寻找相关材料，但法律适用还依赖事项性质、地域、时间和主体身份。同样包含“解除”的材料，可能分别涉及劳动合同、买卖合同或委托关系。若界面只展示相关度排序，使用者容易把排在前面的结果理解为最适用的依据。检索效率的提高，反而可能加快错误材料进入分析的速度。

更隐蔽的问题是材料之间的关系被切断。例如，某条规则的通常条件被抽取出来，而同一文件中的例外没有进入检索结果；某裁判观点被概括为一般原则，而支持该观点的事实组合没有保留。问题并不一定出在语言模型生成虚假内容，也可能出在真实材料被不完整地组织。质量诊断应当同时查看原文、片段和最终回答，找出失真发生的位置。尤其需要检查否定词、条件连接词和例外引用是否完整；文字识别少了一个“不”字，即可能反转条款含义，后续模型再准确也无法自动弥补输入失真。

6.2 版本混用改变结论

法律资料具有时间属性。同一名称的文件可能经历修改、废止或实施范围调整，历史事项又可能需要考察当时有效的规则。若系统只保存一个“最新版本”，可能无法处理历史争议；若保留全部版本却没有状态标记，则可能把旧规则用于新事项。因此，版本治理不是简单删除旧文件，而是建立事项时间与规则时间之间的对应关系。

业务文书也存在相似风险。客户发送的第二稿未必是最后稿，文件名中的“最终版”不等于已签署版本，附件更新还可能改变正文的含义。当知识库将未经确认的谈判文本沉淀为标准条款时，个案妥协可能被误当作机构推荐。需要追问的不是文件有没有保存，而是这份内容在什么状态下、基于何种理由被允许再次使用。

6.3 维护责任缺位使知识失效

知识库上线通常有明确的建设团队，长期维护却容易成为附带工作。新规则由谁发现、谁判断影响、谁修改解释、谁通知使用者，如果没有形成责任链，数据库即使持续增加文档，也可能没有提升可用性。资料数量增长与质量改善之间不存在自动关系。不断新增而不清理重复、冲突和失效内容，还会增加检索噪声。

维护负担也不能完全交给资深律师逐条阅读。机构应识别高频、高后果和近期变化的知识主题，按风险安排更新顺序；对暂时无法完成复核的内容，应限制推荐或明确状态。若系统允许未经维护的规则继续自动触发外部行为，维护延迟就会从资料问题变为业务风险。知识运营需要被纳入工作计划，而不是依赖个别人员的额外投入。

6.4 缺少负面样本掩盖风险

演示资料常选择容易识别的条款、完整的问题和已知答案，但真实使用中存在缺页、模糊扫描、相互矛盾的陈述以及没有明确结论的法律问题。只用标准样本检验系统，会得到过于乐观的

印象。诊断时应加入表面相近但规则不同的材料，并观察系统是否会追问或拒绝给出确定答案，而不是只统计能够完成多少次输出。

知识质量还应接受反向检查：如果结论错误，能否追溯到具体材料和版本；如果来源被撤回，能否找出受影响的历史输出。缺少这种可追溯关系，纠错只能依赖逐个通知使用者，规模越大成本越高。本章的问题在第十二章通过分层知识架构回应，并在第十九章以引用核验、时效测试和回溯能力进行验证。知识可靠之后，服务仍需解决跨岗位交接中的事实损失。

6.5 知识抽样应覆盖材料关系

一次知识抽样可以从“交付验收”主题开始，检查主合同、技术附件和补充协议是否被共同识别，再选择一项存在例外的规则查看上下文。随后分别使用当前事项与历史事项提问，观察系统是否选择对应版本。抽样记录应写清输入、预期关系和实际缺口，不能只记“资料已入库”。这种检查可以在小型团队中人工完成，目的在于发现关系缺失，而非增加复杂评分。

第 7 章 事项流程与专业协作的损耗

7.1 受理环节的事实缺口

法律问题的初始表达往往只是业务人员对风险的直觉描述。“这份合同能不能签”没有说明交易背景、议价空间和履行安排，“对方违约怎么办”也未必说明已经采取的措施。系统如果立即把问题转成法律结论，会把缺少事实的输入包装成完整意见。受理环节应首先识别需要解决的业务决定，并明确结论依赖哪些事实。

事实缺口还可能来自上传材料之间的不一致。合同正文、报价单、邮件和口头说明可能使用不同金额或交付条件。若每份材料分别处理，系统不会自然发现这些冲突。承办人必须知道哪些事实由客户确认，哪些只是文件陈述，哪些尚有争议。缺少这一层区分，后续专业复核只能检查文字是否通顺，无法判断建议是否建立在正确基础上。

7.2 交接丢失判断理由

在多人协作中，结果通常比理由更容易传递。业务人员收到“建议修改”，却不知道为什么要修改；复核人员看到修改后的条款，却不知道原始商业目标。结果是每次交接都需要重新询问，或者在信息不足时依赖个人经验补齐。自动摘要若只保留结论，可能进一步压缩掉判断条件，使协作表面更快、实质更脆弱。

交接记录至少应保留事项目标、材料范围、重要分歧、已作决定和下一步责任人。不是所有讨论都需要原样保存，但影响后续判断的理由不能仅留在即时聊天中。当客户后来改变条件时，团队才能区分是原意见错误，还是新的事实导致结论改变。这种区分关系到服务质量评价，也关系到责任边界是否清楚。

7.3 复核容易成为形式动作

如果系统已经生成完整报告，而复核人员只被要求点击通过，人工参与未必能够降低风险。高工作负荷、时间压力和格式完整性会强化对系统的依赖。复核流程需要让人员看到关键来源、变化条款与尚未解决的问题，并赋予其退回、补充材料和停止交付的权限。只记录复核人的名字，不足以证明完成了实质检查。

还应区分事实复核、规则复核和商业授权。法务确认条款可以接受，并不当然意味着财务同意付款安排，也不意味着业务负责人接受交付风险。将不同判断合并为一个总审批，容易让各方误解他人的确认范围。流程设计应使每次确认对应明确对象，并在条件发生实质变化后触发相关人员重新审查，而不是沿用旧的通过状态。

7.4 例外处理脱离主流程

标准流程之外的事项并不罕见，例如对方坚持使用自身模板、紧急签约、关键附件无法及时取得，或客户决定接受已提示的风险。若系统不容纳这些情形，人员会转向邮件和线下沟通，最终形成一套“系统内合规、系统外实际办理”的双重流程。例外越多，系统记录越不能代表真实工作。

例外机制应要求说明偏离什么规则、由谁决定、依据是什么以及是否需要后续补救，同时保留拒绝办理的可能。记录例外不是为不当行为提供形式正当性，而是让能够被合理处理的特殊情况进入可监督流程。第十三章将用事项状态、交接对象和独立复核点回应这些问题。其前提是机构愿意配置持续维护与专业复核资源，这也是下一章的重点。

7.5 交接记录保留业务语境

对合同谈判的交接，可以采用“当前版本、已接受条件、保留意见、下一轮重点”四个说明项。若客户接受较短索赔期限，记录应说明其接受决定所依赖的交易安排，而不是仅保存修改痕迹。后续条件改变时，接续人才能判断原决定是否仍适用。机构也应记录谁拥有最终对外回复权限，防止多人同时发送不同版本造成新的分歧。

第 8 章 资源投入与经营判断的偏差

8.1 功能采购与业务价值错位

技术方案常以模型规模、功能清单和响应速度呈现优势，业务价值却产生于某个具体流程的改善。一个能生成长篇报告的产品，未必比准确定位关键条款的工具更适合小型团队。管理者若先购买功能、再寻找用途，容易形成高投入低使用的系统。合理起点应是高频工作中的主要耗时和主要错误，而不是市场宣传中的最新概念。

采购比较应使用同一组实际任务、相同的材料范围和一致的评价方法。对无法处理的输入、需要人工补充的步骤和失败后的恢复成本，也应纳入比较。供应商演示通常在理想环境中进行，不能直接代表机构自身的数据质量和人员能力。只有把产品放回真实业务约束，才可以判断价格与价值之间的关系。

8.2 被忽略的全周期成本

系统上线后仍需支付资料维护、测试、权限管理、人员培训和故障处理成本。规则库由谁更新、例外由谁审批、客户投诉由谁分析，都需要可支配的工时。若预算只覆盖初始建设，机构可能在试点阶段表现良好，却无法维持长期质量。运营成本应与功能范围一同确定，不能留待使用中临时解决。

评价成本时还要避免只计算机器替代的显性工时。系统增加的复核时间、重复输入和错误追查可能抵消部分节省。建议把成本分为直接使用、专业处理、维护治理和退出迁移四类，分别记录；返工成本按所属活动归入前三类，并另加返工标记，避免重复计入。没有实测资料时，可以建立估算方法和敏感性分析，但不能把假设结果写成已经实现的收益，更不能据此承诺所有机构都会获得同样回报。

8.3 专业人员与技术人员之间的配置

知识结构化需要法律理解，也需要可执行的数据规范。如果仅由技术人员设计字段，可能遗漏适用条件与法律概念差异；如果全部依赖律师手工整理，则可能难以形成稳定的数据结构。双方应围绕具体样本共同定义输出，并通过争议样本发现理解差异。工作界面越明确，沟通成本越可控。

复核资源不能平均分配给所有输出。高后果事项、来源冲突和首次出现的业务应获得更多检查，低风险且规则稳定的重复任务可以采用适当抽查，但是否减少复核必须有评价结果支持。节约资源不是简单减少人工，而是让专业注意力集中在最能改变结论的位置。否则，规模化将变成错误处理能力的扩张。

8.4 供应商锁定与退出成本

机构对某一供应商的依赖，不仅来自技术接口，还可能来自无法导出的标签、规则、历史意见和审批记录。即使原始文档可以下载，若关系数据无法迁移，新系统仍需重新整理。采购时应核查可导出的对象、格式、完整性和时间安排，并通过小范围导出验证，而不是仅依赖合同中的概括承诺。

退出计划还应覆盖服务中断期间的人工接续、密钥与权限撤销、备份资料处理和历史记录读取。依赖越关键，越应提前演练。第十四章将把投入选择转化为数据与指标机制，包括合格交付成本、风险加权评价和供应商比较，帮助管理者判断哪些能力值得建设、哪些应保留为外部服务，以及何时应停止无效投入。

8.5 采购问答应转化为验收证据

供应商声称支持完整导出时，可以要求现场导出一项测试事项，核对文档、标签、关系和操作记录是否齐全；声称支持权限隔离时，可以用两个测试客户验证交叉访问；声称能够更新法律知识时，可以检查一项变更从发现到上线的记录。这样的证据比概括承诺更便于比较，也能揭示后续仍需使用方自行承担的工作。

第 9 章 数据协作中的权限与语义风险

9.1 同一事项在不同系统中失去一致性

合同、客户和项目在不同系统中可能使用不同编号和名称。销售系统中的客户简称，不能直接替代合同中的法律主体；财务系统中的付款对象，也可能与实际履约方不同。若通过模糊匹配自动合并记录，系统可能把关联公司、历史名称或同名主体混为一谈。数据协同首先需要确认对象身份，之后才是传输效率。

字段一致也不等于含义一致。“完成”可能表示审查结束、签署结束或履约结束，“金额”可能指含税总价、阶段付款或预算额度。接口如果只校验数据格式，不解释业务语义，错误会在多个系统间被稳定复制。机构应对关键字段说明定义、来源、允许状态和变更责任，尤其注意会触发审批或执行的字段。

9.2 访问便利与最小范围之间的张力

为了提高回答效果，团队可能倾向于让检索系统访问更多资料，但不同客户、事项和岗位之间存在保密边界。用户无权查看的文档，也不应通过摘要、相似问题推荐或缓存结果间接暴露。权限检查如果只发生在界面打开文件时，而检索阶段已经把内容送入生成模块，控制就发生得太晚。

个人信息保护法提出明确合理目的及最小必要等基本要求。[1]在工程上，可以据此对输入字段、检索候选、日志和导出分别设计范围控制。对合同审查并非必要的身份信息，不应为了“资料完整”而默认传入外部组件；对确实必要的信息，则应说明处理依据和保护措施。具体法律适用仍需结合实际处理活动判断，不能只靠字段脱敏完成全部审查。

9.3 去标识处理并非风险终点

删除姓名并不一定使资料无法关联到个人。职位、交易时间、项目地点和特殊事件的组合仍可能具有识别性；结构化标签、向量索引和日志也可能保留敏感内容。机构应沿着数据复制路径检查风险，而不是只查看上传界面是否遮挡电话号码。测试环境尤其容易被忽视，因为真实资料常被复制用于问题排查。排查人员应先判断是否可以用最小片段复现问题，并对复现资料设置独立权限与清理安排，避免把一次技术排查变成新的资料扩散路径。

需要区分去标识化与匿名化。去标识化使信息在不借助额外信息时不能识别特定个人，但并不因此排除个人信息保护要求；匿名化则要求处理后的信息不能识别特定个人且不能复原。[1]不能仅凭删除姓名或更换字段名称判断已经实现匿名化。对模型服务商、运维人员和分包方的访问也应分别控制，并核查数据留存、再次使用和删除安排。任何超出原目的的新用途，都应重新判断其依据。

9.4 外部内容可能改变系统行为

客户合同、网页和附件应被当作待分析资料，而不是系统操作指令。若文档中出现要求忽略规则、泄露其他事项或调用外部服务的文字，系统不能因其被检索到就执行。这种风险说明内容

处理与权限执行必须隔离。即使资料来自熟悉客户，也应按内容来源处理，不能赋予其修改系统规则的权力。

在设计上，应限制生成模块能够使用的工具和数据范围，对外发、下载、删除等动作另设校验。测试中可使用包含无关指令的合同附件，观察系统是否继续完成原定分析、是否泄露信息以及是否触发额外动作。第十五章将通过统一语义、事项隔离和接口校验回应上述问题，并将异常内容处理纳入协作流程，而不是仅依靠使用者自行辨认。

9.5 接口失败也需要业务定义

接口超时并不等于业务没有发生。发送合同后未收到回执，可能是发送失败，也可能是接收成功但回执丢失。系统应保留中间状态，并通过查询或人工确认消除不确定性，不能直接重复操作。数据协作因此还包含失败含义的统一：哪些错误可以重试，哪些必须核对，哪些需要暂停。若双方没有约定，技术上的自动重试可能变成业务上的重复行为。

第 10 章 合同交付与自动执行的断层

10.1 文本审查与交易安排脱节

合同风险不仅位于某一句条款，也存在于正文、附件和履约能力之间。系统可能正确识别责任限制条款，却未发现交付标准只写在未附上的技术文件中；也可能建议增加违约责任，却忽略对方没有可执行资产。文本分析能够提供重要线索，但交易判断必须结合实际安排。把风险清单直接当作完整法律意见，会夸大工具的覆盖范围。

交付物应明确审查对象和未覆盖的事项。例如，条款一致性检查不包含商业可行性评价，签署形式检查不包含代理权限的全部事实调查。范围描述需要与实际工作一致，并让客户能够据此安排补充核查。过度宽泛的免责声明同样没有帮助，因为它无法告诉客户哪项风险已经处理、哪项仍需采取行动。

10.2 程序条件难以覆盖法律情境

自动执行模块依赖明确输入，而合同中的“合理”“重大”“及时”等概念通常需要结合事实解释。将这些概念直接映射为一个数字阈值，可能改变当事人原本的安排。程序还可能无法感知双方后续协商、争议或履行障碍。只要外部事实没有进入系统，程序就会在错误或不完整的条件下持续运行。

以采购付款的行业典型场景为例，物流系统显示签收，并不必然等于质量验收完成。如果程序收到签收状态就自动付款，可能绕过合同约定的验收环节。技术设计应把“收到货物”“完成验收”“付款获批”区分为不同事件，并确认由谁提供证据。自动化范围应从这些可验证事件推导，而不是从程序能够调用支付接口倒推。

10.3 存证能力不能替代事实调查

链上存证有助于核对记录是否变化，但不能仅凭技术记录证明原始陈述真实。人民法院在线诉讼规则对上链后未篡改的认定及相关异议审查作出安排。[5]这一制度区分提醒产品设计者把记录完整性、形成过程和内容真实性分别处理。扫描件被保存后没有改变，不等于扫描件内容与实际交易一致。

当机构选择存证方案时，应明确要证明什么问题，保留必要的原始文件、采集过程和身份关联信息。不能把所有合同原文直接公开上链，也不应把“已存证”作为审查通过的替代标记。对于敏感资料，应评估保存摘要或验证信息的必要性以及与原文的对应关系，并安排长期可读取、可解释的证据路径。

10.4 错误执行缺少停止与恢复

文本中的错误可以修改，外部行为的后果却未必能够直接撤回。系统重复发送通知、重复发起付款或错误变更状态时，简单重新运行可能进一步扩大影响。设计中需要考虑重复请求识别、执行前校验、结果确认和异常接管，并明确哪些动作必须暂停等待人工。系统提示成功，也应有外部接收或处理结果支持。

恢复机制应区分技术恢复与法律处置。修复数据库不能自动消除已经发出的意思表示，也不能抹去对方已经产生的合理信赖。出现重大异常后，机构应保存过程、确定影响范围，由适当人员决定补救沟通和后续法律行动。第十六章将据此安排合同交付、受控执行与供应商责任接口，使执行能力始终处于明确授权和可接管的范围内。

10.5 验收对象不能停留在界面

供应商展示“合同风险已识别”时，验收人员应追问识别范围是否包括附件、表格和跨条款引用，以及提示能否定位具体版本。展示“履约自动化”时，应核对触发条件来自哪里、谁能修改和谁能停止。功能按钮存在并不说明控制有效。把验收对象拆到实际输入与后果，可以避免将产品展示中的能力名称直接当作可交付承诺。

第 11 章 用户价值与信任的形成障碍

11.1 用户需要决策支持而非风险堆积

一份列出大量风险的报告并不必然帮助客户决定下一步。如果没有区分严重程度、发生条件和可采取的措施，客户可能只感到所有选项都不可行。系统还可能反复提示形式问题，掩盖真正影响交易的核心条款。评价交付质量应观察客户能否理解主要选择，以及是否知道需要补充什么事实、向谁提出何种要求。

风险排序应结合后果和行动窗口。即将发生的付款或通知期限，可能比可在谈判后期修订的格式问题更需要关注。系统可以协助组织信息，但排序理由必须可解释，不能仅依赖一个不透明分数。对高不确定问题，应说明不同事实成立时结论如何变化，而不是给出看似精确却缺少依据的风险百分比。

11.2 流畅表达造成过度信任

自然语言输出具有交流便利，也容易让使用者把系统理解为已经掌握完整案情。尤其当回答使用肯定语气、引用法律名称并生成正式格式时，客户可能忽略事实不足。产品需要让输出状态与实际完成的工作一致：资料提示、分析草案和经专业复核的正式意见，应具有不同的呈现和流转规则。

用户教育不应仅放在服务协议中。关键操作前的简短说明、来源展开功能和可见的复核状态，更容易影响实际使用。对外输出应使客户能够辨认哪些结论依赖其陈述，哪些来自书面材料，哪些需要进一步确认。一个无法解释来源的结果，不应因为语言更清晰就获得更高的信任等级。

11.3 申诉与纠错渠道不完整

法律服务中的异议可能来自事实补充、对规则理解不同或对服务范围的误解。若客户只能给出满意或不满意，机构无法判断问题来源。反馈机制需要接收具体争议点，允许上传相关材料，并将其分配给有权处理的人员。对已经影响外部行为的错误，应优先确认是否需要暂停后续流程，而不是只把反馈作为产品改进建议。

纠错还应追踪到知识和流程。若客户指出某条款已被修改，仅修正当前报告可能不足以防止再次发生；若多名客户反复误解同一提示，问题可能在界面和措辞。机构应保存处理结论、解释方式和影响范围，并判断是否需要通知其他受影响事项。规模化服务必须具备规模化纠错能力，否则早期的效率收益可能被后续补救消耗。

11.4 不同用户的使用条件存在差异

专业人员可以理解法律术语和条件限定，普通用户却可能把“可能”“原则上”理解为无法行动。不同规模企业的数据基础、内部审批能力和预算也不同。统一产品不应要求所有用户自行完成相同复杂度的判断。对缺少专业支持的使用者，应提供更清楚的事实提问和转人工入口，避免把专业审查负担隐蔽地转移给客户。

用户价值最终体现为行动质量和权利保障，而不是咨询次数或页面停留时长。第十七章将围绕分层交付、解释、异议和持续运营提出机制。至此，六类问题构成完整诊断：知识来源决定分析基础，协作决定事实传递，资源决定持续能力，数据决定边界，执行决定后果，用户理解决定结果如何被使用。后续方案必须逐项回应，不能只在其中某一环节提高自动化程度。

11.5 用户理解需要任务验证

理解测试可以让使用者阅读一份经过适当处理的分析，随后说明自己准备采取什么行动、哪些条件仍未确认，以及遇到异议应联系谁。若其行动与报告意图不一致，应检查结构和措辞，而不简单归因于用户专业能力不足。测试不追求用户记住全部法律术语，重点是其能否辨认关键限制和下一步选择。

第 12 章 以可追溯知识为基础的总体架构

12.1 将依据与生成分开管理

针对第六章的知识质量问题，本书提出五层架构：来源层保存原始材料，知识层组织概念与适用关系，规则层表达条件与例外，任务层完成检索和分析，交付层控制复核与使用。各层可以由不同组件实现，但必须保留可回溯的关联。最终输出不仅要能找到一个链接，还应能定位到支持相应命题的具体内容。

来源层不接受无法说明来历的材料直接进入正式知识集合。对公开规范，应记录发布机构、文件名称、效力状态和适用范围；对内部经验，应说明其属于机构业务建议，不能冒充法律规定。来源发生变化时，系统应能够标记依赖该来源的知识和规则，等待复核后更新。这样可以把一次资料修正传递到实际使用位置，而不是只替换存储目录中的文件。

12.2 建立可维护的知识单元

知识单元的粒度应围绕可独立判断的命题，而非机械按字数切分。一项合同审查规则可以关联通常条件、例外、所需事实和相关原文；一个案例知识单元可以关联争议焦点、关键事实及裁判理由。若某段材料依赖上下文才有完整含义，应同时保留上下文入口。机器处理需要简化结构，但简化不能破坏法律关系。

可使用统一的知识卡记录主题、来源位置、适用时间、地域、规则摘要、例外、维护责任人与状态。摘要是帮助检索的解释，不取代原文。维护者应对新增、修订、暂停使用和恢复使用分别留下原因。使用者发现问题时，可以直接反馈到对应知识卡，避免问题只停留在某次聊天记录。稳定的标识使知识变更可以被统计和追踪。

表 12-1 五层架构的控制对象

层级	主要对象	验收问题
来源层	原始材料与版本	能否回到对应原文
知识层	概念、关系与例外	上下文是否完整
规则层	条件、动作与限制	何时必须停止判断
任务层	检索、比较与分析	输出是否受来源约束
交付层	复核、授权与纠错	能否定位责任和影响

本表为本书提出的管理框架，不含实测效果数据。

12.3 检索与回答实行分步验证

任务开始时先确定事项类型、时间范围和允许访问的资料，再进行关键词与语义检索。候选结果需要检查来源可靠性、适用性和冲突情况。只有在材料足以支持分析时，才进入回答组织。若缺少关键事实，应形成补充问题；若规则冲突或资料不足，应呈现不同路径及其条件，而不是让生成模块自行选择看起来最肯定的结论。

生成后还应核验引用是否真正支持对应语句，是否存在把资料中没有的内容归于来源的情况。引用数量不是质量指标，重复列出无关法律名称反而妨碍复核。对于影响结论的命题，应尽量使用直接依据；对于基于来源进一步提出的管理建议，应明确推导条件和建议性质。技术架构的目标是把事实、来源和分析分离存储，再在交付时有序连接。

12.4 规则发布与回溯修正

新规则不能从编辑界面直接进入全部业务。建议经过专业审核、样本验证和受限使用后再扩大范围，并保存旧版本以处理历史事项。发布记录应说明改变了什么、哪些场景受影响、如何恢复到原状态。所谓回滚仅指系统配置恢复，不能自动撤销已经交付的法律意见；受影响事项仍需由承办人评估和处理。

当发现依据不再适用于相应事项或解释存在实质错误时，应限制受影响规则的继续使用，再定位历史输出，按后果确定通知和重审顺序。若仅为来源链接迁移，应先核对原文和有效版本，修复引用入口；新规则施行也不意味着旧规则对所有历史事项失去适用性。轻微文字问题可以在下次使用时修正，影响客户行动的结论则应主动进入纠错流程。知识架构由此不只是一个查询入口，也成为质量维护的基础设施。其有效性将在第十九章通过规则替换、历史检索和影响追踪测试验证。

12.5 以最小完整能力开始建设

机构可以先选择一类高频合同建立小型知识集合，确保每条知识都具备来源、适用条件和维护责任，再逐步扩展。选择范围时应避开事实高度开放、专业分歧难以收敛的复杂事项。小范围不意味着降低严谨程度，而是让团队有能力完成一次完整的知识生命周期，并看清维护成本。

架构验收不以是否采用特定数据库或模型为标准，而以是否能够演示一项结论从原文到交付、再到纠错的全过程为标准。若无法定位来源或无法处理失效规则，就不宜扩大使用。完成知识基础后，第十三章把这些能力嵌入事项流程，使每个环节都知道接收什么、核查什么及向谁交付。

12.6 知识卡的最小信息结构

以验收条款知识卡为例，卡片应把“是否约定验收方式”作为问题，把合同正文与附件作为检查范围，把需要业务确认的标准单独列出。规则摘要说明审查理由，适用条件限制使用场景，例外字段容纳不同交易安排。发布时同时保存反例：附件缺失、双方另有补充约定和验收存在争议。这样的结构使规则不仅告诉系统何时提示，也告诉系统何时不能直接下结论。

第 13 章 事项流程与责任分配机制

13.1 用事项卡明确工作对象

针对第七章的受理与交接问题，建议建立统一事项卡，记录客户目标、业务类型、决定期限、材料清单、授权范围和承办责任。事项卡不是要求客户一次性填写所有细节，而是随着调查

逐步完善。关键事实应区分客户陈述、书面材料支持和仍有争议的内容，并允许承办人说明为何某项缺口不影响当前阶段的工作。

系统可协助检查必需字段，但不能把字段非空视为事实可靠。例如，签署人姓名已填写，不等于代理权限已经核实；金额已录入，不等于正文与附件一致。受理完成的标准应是当前任务具备足够输入，未解决问题已明确归属。对于紧急事项，应记录时间压力和先行处理的范围，避免临时判断在后续被当作全面结论。

13.2 状态转换对应真实条件

事项状态可以设置为已受理、材料整理、专业分析、复核、交付及后续跟踪，但每次转换必须对应具体条件。进入复核意味着材料范围和分析版本已经固定；进入交付意味着相关复核完成且接收对象明确。状态名称应简洁，转换条件却需要清楚，防止人员以“系统显示通过”为由忽略实际缺项。

若客户补充重要材料或改变交易条件，事项应返回受影响环节。系统不必每次重做全部工作，但应提示哪些结论依赖改变的事实。这里需要使用第十二章建立的关联关系：事实影响规则，规则影响意见，意见影响后续动作。工作流因此能够识别实质变化，而不是只跟踪文件上传次数。

13.3 复核按对象而非职级划分

事实复核检查材料与陈述是否一致，专业复核检查依据和推理，业务授权确认是否接受商业安排，交付检查确认发送对象与版本。这些工作可以由不同岗位承担，也可以在小团队中合理兼任。对高后果事项，应尽量避免同一人员完成全部关键判断；具体安排应与组织规模和业务风险相匹配。

复核界面应突出结论改变、关键引用、例外和未解决问题，允许直接回到原文。复核人员填写必要理由，而不是重复承办人的完整报告。若选择接受某项风险，应写明接受的对象与条件，不能使用笼统的“同意全部风险”。人员完成复核后，如对应文本又被修改，应重新判断原确认是否仍然有效。

13.4 例外与升级进入同一流程

例外申请应说明标准路径无法满足什么业务需要、偏离可能产生什么后果，以及是否存在替代措施。批准人只在自身权限内决定，超出权限的事项继续升级。系统应保留拒绝和暂停选项，不能通过默认勾选或隐藏提示引导批准。例外到期后需要复查，防止一次临时安排长期成为事实上的新规则。

对于明显超出服务范围、关键事实无法核实或客户要求不当行为的事项，应停止相关任务并转交适当人员处理。此处的停止是专业控制，不应被评价体系一概视为效率下降。管理者应单独观察合理退回和无效往返，前者保护质量，后者才是需要减少的流程损耗。二者不区分，会鼓励承办人把不完整事项尽快推向下一环节。

13.5 交付包与接续责任

正式交付应包含结论、重要依据、适用条件、建议行动和未覆盖范围，并明确后续由客户或服务方完成的工作。内部还应保留定稿版本、复核记录和交付凭证，便于争议时重建当时的判断。对需要持续跟踪的事项，例如付款、通知和整改期限，应形成独立任务，不能只写在报告末尾等待客户自行发现。

离职交接、休假和项目转组也应沿用同样的事项记录。接续人能够看到已作决定及其理由，才不必重新询问全部背景。流程机制是否有效，可通过一次人员替换演练检验：新承办人能否定位材料、说明当前状态并继续下一步工作。第十八章将把这种演练安排进试点，第十九章再评价交接完整性和复核的实际作用。

13.6 紧急事项仍保留必要记录

紧急办理可以缩短沟通路径，但不应删除关键授权与版本记录。机构可以先形成简短的决定说明，列明当前可确认的事实、尚未覆盖的部分及后续补充责任，再按风险完成必要复核。紧急状态解除后，应将零散材料归入同一事项并检查是否需要修订原意见。对无法在期限内可靠完成的工作，应说明能力边界，而不是以系统生成速度替代专业准备。

第 14 章 数据质量与经营评价机制

14.1 先建立可比较的基线

针对第八章的投入偏差，机构应在引入工具前记录现有工作方式的基本情况，包括任务类型、复杂程度、材料完整性、处理时间和返工原因。基线不是为了证明旧流程低效，而是为改进提供比较对象。不同复杂度事项应分组，不能把简单合同的工具处理时间与复杂交易的人工处理时间直接相比。

比较期间应说明哪些工作由系统完成、哪些由人员完成，以及复核是否达到相同标准。若新流程减少了审查范围，时间变短不能视为同等服务下的效率提升。若人员因熟悉测试材料而处理更快，也应识别学习效应。没有足够样本时，可以先报告观察和不确定性，不必急于得出一个看似明确的收益百分比。

14.2 指标围绕合格交付组织

建议以合格交付作为基本评价单位。合格意味着在约定范围内完成必要工作，通过相应复核，并向正确对象交付正确版本。成本沿用第八章的直接使用、专业处理、维护治理和退出迁移四类口径，返工按所属活动归类，不另行重复相加。比较当期交付成本时，将约定观察期内的运行成本除以同期合格交付量，失败任务消耗也计入分子；评价采购方案时，另列一次性建设与退出迁移成本，并说明分摊期限和预计业务量。工具调用次数和生成字数只能解释工作量，不能独立证明价值。

质量指标需要与任务对应。知识检索可检查来源适用性和引用支持程度，合同审查可检查关键风险遗漏与误报，流程协作可检查交接缺项和等待，用户服务可检查解释是否被理解。指标应

保留分母定义和样本范围。例如，遗漏率中的“应发现风险”需要由专业人员事先标注，不能只把系统已经提示的风险作为评价范围。

表 14-1 服务评价的指标口径

维度	观察指标	解释限制
知识质量	适用依据遗漏、引用支持	按预设任务和来源范围评价
专业质量	关键风险遗漏与误报	区分后果，不用平均分抵消严重错误
协作效率	有效工时、等待与返工	保证前后服务范围一致
经营价值	每项合格交付的总成本	含维护、复核和迁移投入
用户价值	行动理解与纠错完成	满意度仅作为补充

本表为本书提出的管理框架，不含实测效果数据。

14.3 对错误后果进行分层观察

一个错别字与一个导致错误付款的判断，不应在评价中被等量计算。机构可以按后果、可逆性和发现时点划分错误等级，并分别记录。高后果错误即使数量较少，也可能决定某功能不能进入自动执行。加权得分有助于比较，但不得让大量低风险正确结果抵消不可接受的严重缺陷。本文的“风险分层”指对任务和错误后果作管理分类，并非声称已经通过统计模型计算出真实发生概率；在缺少有效数据时，不应把分层标签包装成精确预测。

不确定事项应单独处理。如果专业复核者本身存在合理分歧，宜记录分歧理由和可接受答案范围，而非强行设定唯一标签。评分规则应在测试前确定，避免看到结果后修改标准。通过这类安排，评价既能识别确定性错误，也能保留法律分析的开放性，不把复杂判断简化为机械匹配。

14.4 投入组合与敏感性分析

经营判断可以比较不同方案在相同业务量和质量要求下的成本结构。例如，自建方案增加固定维护投入，外部服务增加使用费用和供应商依赖，混合方案则需要额外接口管理。分析应说明业务量变化、复核工时和资料维护频率如何影响结果。参数来自实测时注明范围，参数属于估计时则用多种情形观察结论是否稳健。

当某方案只有在极低错误率或极高业务量下才具经济性，管理者应看到这一条件，而不是只看到理想收益。试点可以设置停止条件，如关键质量问题无法解决、维护成本持续超出预算或人员无法完成必要复核。退出不是项目失败的唯一含义，也可能是有效评价后作出的合理选择。

14.5 指标治理与复评责任

指标需要指定维护人和解释人，定期检查采集是否完整。系统自动记录的时间可能包含等待，也可能遗漏线下工作，应通过抽样核对确认口径。对人员绩效使用指标时，应避免鼓励减少合理追问、隐瞒错误或过快结束事项。质量评价应支持改进，而不是把所有复杂情况转化为个人扣分。

经营复评可与功能升级和业务扩张同步进行。每次扩大使用范围，应确认原基线是否仍适用、测试样本是否覆盖新场景以及成本结构是否改变。NIST 人工智能风险管理框架提供了治理、情境识别、测量和管理相互衔接的参考思路。[7]本书据此强调持续评价，但具体指标和门槛应由机构根据服务风险确定，不作为通用行业认证标准。

14.6 用成对任务检验实际改善

小范围评价可以选择结构相近的任务，分别记录人工方式和辅助方式的处理过程，并交由独立复核人员按相同标准检查。应避免同一人员连续处理同一材料而产生明显记忆优势，也应记录其经验差异。若只能采用前后比较，则报告期间业务组合与人员变化。评价设计不必追求复杂统计形式，但必须足以识别最明显的混杂因素。

第 15 章 跨系统协作与数据边界机制

15.1 用统一对象承载业务关系

针对第九章的语义和权限问题，机构应为客户主体、事项、合同及文档建立稳定标识，并保存它们之间的关系。主体信息变更时，不应简单覆盖历史记录，而应保留对应时期的名称与身份。对于相似主体，系统可以提出匹配建议，最终合并应由有权人员确认，尤其不能仅凭名称接近自动合并客户资料。

关键字段需要形成数据字典，明确含义、格式、来源、允许值和更新责任。合同总价、已付款和应付款应分别记录，审查完成、签署完成和履约完成也应分别定义。接口双方应使用具体样本验证理解是否一致，并测试字段缺失、顺序错误和重复提交的情况。语义清楚后，自动流转才有可靠基础。针对第九章的超时风险，接口还应区分未提交、处理中、已完成和结果待确认。具有外部后果的请求应携带稳定的业务标识，重复提交时先查询原结果；接收方不能保证同一请求只执行一次时，应先核对或转人工，不能仅凭超时自动重试。

15.2 权限在检索前生效

访问控制应在材料进入检索候选时执行，并在内容生成、缓存和导出时再次保持一致。用户只能得到其授权事项范围内的内容；跨客户相似条款推荐应使用已经适当处理且允许复用的知识，而不能直接引用另一客户的原始合同。权限模型应能表达岗位、事项关系、资料类别和有效期限，避免仅依赖宽泛的部门共享。

测试权限不能只使用正常账号。应构造离职人员、临时协作人员、权限刚被撤销的账号，以及访问其他事项编号的请求，检查正文、摘要、搜索提示和历史缓存是否同时受限。若权限变化后旧缓存仍可显示内容，撤销就没有完整生效。测试记录应保留预期行为和实际结果，修复后重新验证受影响路径。

15.3 最小数据流与受控日志

每项任务应定义需要的最小输入集合。例如，初步比较条款差异通常不需要当事人的全部证件资料；解释通用法律概念也不需要上传客户合同。系统可先在本地完成字段筛选和必要处理，再把任务所需内容交给后续组件。减少无关数据既有助于控制风险，也能降低检索噪声和处理成本。

日志应保存足以定位问题的信息，同时避免无差别复制全部敏感正文。可以记录事项标识、处理版本、操作类型和必要摘要，原文访问则通过受控路径完成。保存期限应根据目的与适用要求确定，不能统一设为永久保留。备份、测试副本和供应商留存也应纳入同一管理视野，防止主系统删除后其他副本长期存在。

15.4 将内容与动作权限隔离

外部文档中的任何操作要求都应保持资料属性，不能改变任务权限。模型可以分析合同中“删除”“发送”等文字的法律含义，但不能据此执行删除或外发。对具有外部后果的工具，系统应使用明确的参数校验、对象校验和授权条件，并让用户能够看到将要发生的动作。文本分析结果只能成为后续判断的输入。

安全测试应覆盖恶意附件、隐藏文字、引用网页中的无关指令以及试图获取其他客户内容的提问。测试的重点是控制是否生效，而不是系统能否识别某个固定关键词。遇到异常内容时，可以继续分析合法部分，也可以暂停并交由人员处理；无论选择哪种方式，都不应扩大数据访问或执行权限。

15.5 供应商协作与变更管理

接入外部服务前，应了解数据接收主体、分包关系、存储位置、再次使用安排及故障处置方式，并结合实际活动审查合同与保护措施。对跨境处理不能仅凭服务器宣传位置作出结论，还需核查运维访问、备份和其他传输路径。具体适用程序应按当时规则 and 实际数据情况判断，本书不提供一套适用于所有业务的固定申报结论。

供应商升级模型、调整数据政策或改变接口时，应触发影响评估。机构可先在隔离环境中验证，再决定是否切换，并保留恢复路径。协作机制的验收应包括一次权限撤销测试、一次数据导出核对和一次故障切换演练。只有这些基础能力可靠，第十六章讨论的合同交付与受控执行才具备可接受的运行条件。

第 16 章 合同管理与受控执行机制

16.1 把合同拆为关联对象

针对第十章的交付断层，合同管理应同时保存文本、主体、条款、附件、审批和履约事件。条款识别不能只提取正文段落，还应识别其引用的附件和定义。每项风险提示需要定位原文，说明影响、所需事实和可选处理，便于承办人判断。对扫描不清、附件缺失或版本不明的情况，应先标记输入问题，避免形成伪完整报告。

修订建议应与客户目标相联系。增加责任上限、调整付款条件或改变验收程序，都可能影响交易可接受性。系统可以生成选项，但不宜将某种条款风格视为唯一正确答案。承办人应说明建议改善什么风险，同时可能增加什么谈判成本。交付的重点是支持取舍，而不是让所有合同机械地变成同一模板。

16.2 自然语言与程序保持对应

对准备自动执行的合同事项，先建立自然语言条款与程序条件的对应表，逐项说明触发事实、数据来源、授权人、执行动作和例外。无法可靠获取的外部事实，不宜直接作为自动执行依据；存在解释空间的条件，可以保留人工确认。程序说明应让非技术承办人能够理解，不能只有开发人员才知道实际触发逻辑。

当自然语言条款修改时，系统应提醒检查对应程序，而不是默认新合同继续使用旧配置。反过来，程序升级也需要判断是否改变业务效果。合同和程序可以有不同版本号，但必须明确哪一组经过联合确认。未经确认的组合不得进入实际执行。通过双向关联，可以防止文本与代码各自更新后产生隐蔽偏差。

16.3 执行前校验与停止条件

每次外部执行前，应检查事项状态、授权有效性、输入证据和动作对象，并识别重复请求。付款、通知和权限变更等动作应分别设置控制，不能因为同属一个合同就共享全部权限。对于已经出现争议、材料变化或外部系统异常的事项，应暂停自动执行并进入人工处理。暂停条件应事先定义并可以测试。

行业典型场景中，验收资料缺失时，系统可自动提醒承办人补充，而不直接推定验收通过。若经授权人员确认验收并批准付款，执行模块再核对金额和账户。完成后保存结果与接收反馈。这里的自动化只覆盖已经明确的动作链，不替代对验收事实和付款授权的专业、业务判断。

16.4 证据记录与异常补救

记录体系应能够说明谁在何种权限下，根据什么材料批准了哪项动作，以及外部系统返回了什么结果。需要采用电子签名或存证时，应根据证明目的选择方案，核查身份关联、完整性验证和长期读取能力。技术记录应服务于真实业务过程，不能事后补造未发生的审批或签署步骤。

发生异常后，先阻止重复或继续执行，再判断已经产生的后果。技术人员负责定位故障，事项负责人负责业务影响，专业人员负责相关法律处置。补救过程应保留原始记录和修正记录，不通过覆盖历史掩盖错误。对外沟通应准确说明已知情况与处理措施，避免在事实尚未查清时作出不必要的责任承诺。

16.5 交付验收覆盖完整链条

合同系统验收至少应覆盖正常签署、附件缺失、签署权限变化、付款条件未满足、重复请求和外部服务中断。测试不是要求所有异常都能自动解决，而是确认系统能够识别不具备执行条件的情况，并交给适当人员。对无法保证安全接管的功能，应维持辅助模式或限制使用范围。

交付完成后还需要明确维护与退出安排：谁保管签署文件，谁更新提醒，供应商服务终止后如何读取历史记录，争议发生时如何导出必要资料。这些事项决定合同系统能否支持长期服务。技术能力由此转化为一套可以解释、能够停止并允许人工接续的交付机制，而不是单纯追求更多自动动作。

第 17 章 分层服务与持续运营机制

17.1 按决策用途设计交付

针对第十一章的用户理解问题，建议把交付分为资料定位、分析支持和正式专业意见等不同用途，并在产品流程中明确区分。资料定位展示来源与相关内容，分析支持呈现条件和选项，正式意见则按照机构约定完成专业工作与复核。分类不是通过名称免除责任，而是让实际服务范围、人员参与和用户预期保持一致。

客户收到结果时，应首先看到需要决定什么、主要依据是什么和下一步如何行动。详细分析可以随后展开，避免核心事项被长篇文本淹没。对存在多种路径的问题，可以并列说明适用条件和代价。系统不应为了显得果断而消除合理不确定性，也不应以大量抽象提示让客户无法辨认任何可执行建议。

17.2 解释与事实确认相结合

解释应围绕客户能够核对的事实展开。例如，付款建议依赖验收完成，就应明确询问何种文件证明验收，而不是只列出一段法律原则。对专业术语，可以在首次使用时给出简短定义，并说明其对当前选择的影响。来源入口应易于打开，重要结论与依据之间的对应关系也应清楚，避免客户面对一串参考文献却无法理解。

用户确认应针对具体事实或行动，不能使用笼统勾选替代知情理解。对于高后果操作，可以要求复述关键条件或再次核对对象；对于一般资料查询，则不宜增加不必要的负担。交互设计应与风险匹配，使用户在真正需要注意的地方获得足够提示，而不是因过多重复警告产生忽视。

17.3 异议处理形成闭环

反馈入口应支持事实补充、结论异议、技术故障和服务投诉等不同类型，并提供适当的处理路径。接收反馈后，先判断是否存在正在发生的风险，再分配给有权人员处理。客户应能了解事项是否已接收、由谁负责以及下一步需要什么材料。流程不必承诺无法实现的固定时限，但应有明确的内部响应安排。适用法律、监管要求或服务约定已有时限的，应优先落实；内部安排不能降低这些要求。对临近权利行使期限的反馈，应按影响优先处理并明确转人工路径。

处理结果应说明是否修改结论、为什么修改，以及是否影响其他交付。对知识性错误，应回到第十二章的维护流程；对权限或接口问题，应进入第十五章的变更机制；对交付理解问题，则调整表达和提示。反馈闭环的价值在于减少同类问题反复发生，不能只以关闭工单作为完成标准。

17.4 持续服务的组织保障

机构应设立能够跨专业、业务和技术协调的运营责任，定期审查未解决异常、过期规则和高频投诉。小团队可以用固定会议和事项清单完成，大型机构可以设置专门岗位，但都需要把维护工作纳入计划。系统上线后若没有人有权决定暂停功能，治理机制就不完整。

培训应使用真实工作类型和经过适当处理的材料，让人员练习识别输入缺口、核查引用、拒绝越权和接管异常。培训完成不以观看课程时长为唯一依据，而应观察人员是否能够完成关键操作。对新入职和岗位变动人员，应重新确认权限与职责，防止历史授权长期保留。

17.5 扩展服务前验证用户价值

扩大场景之前，应观察现有用户是否能够理解输出、正确完成后续行动，以及在遇到问题时找到有效帮助。可以通过任务完成访谈、错误分析和实际交付记录进行判断，而不只依赖满意度。用户满意可能来自快速响应，也可能来自系统迎合其预期，因此需要与专业质量结果一起解释。

规模化的条件是知识维护、事项流程、数据边界和纠错能力能够同步扩展。若只有生成能力扩展，复核和投诉处理却没有相应资源，服务质量可能下降。下一章将把前述六类机制组合为分阶段实施路径，并用三个行业典型场景展示它们如何共同运行；随后再用独立评价确认这些机制是否真正有效。

第 18 章 分阶段实施与典型场景

18.1 准备阶段形成最小工作单元

实施首先选择一个边界清楚、资料相对稳定且有专业负责人承担维护的任务。建议以一类合同审查或一个知识检索主题为最小工作单元，列明当前流程、输入材料、常见错误、质量要求和停止条件。业务负责人确认需求，专业负责人确定判断标准，技术人员梳理系统依赖，数据管理人员核查访问和处理范围。角色可以兼任，但每项决定必须有明确承担者。

准备阶段的里程碑是形成可测试的任务说明，而不是完成产品采购。说明中应包含任务可以处理的输入、必须补充的事实、输出用途以及不能自动完成的行为。与此同时建立人工基线和测试样本，记录复杂度与材料完整性。尚未具备合法使用条件的客户资料，不应因为试点规模小就直接投入测试；可以采用适当处理的材料或明确标识的构造场景验证流程。

18.2 受限试点与扩大使用

试点先在受控范围内运行，由承办人员保留完整复核，比较工具结果与既有工作方式。每次错误应定位到资料、规则、模型、接口或人员操作，不把所有问题归为“模型不准”。修正之后使用未参与调试的样本再次验证，避免系统只对熟悉材料表现良好。试点期间发生关键权限泄露或错误外部动作，应暂停相关功能并查明原因。

扩大使用需要满足三个方面的条件：质量达到预设要求，维护资源能够持续投入，异常可以由人员接管。这里没有适用于所有机构的统一分数，门槛应在试点前结合风险设定。受限试点通

过也不自动授权外部执行；检索、起草、交付和执行各有验收对象，后一环节须验证自身控制后方可开放。扩展可以按合同类型、使用团队或处理环节逐步进行，每次只增加可管理的范围。若试点成功仅依赖个别专家持续兜底，应先解决知识和流程沉淀，再推广到其他人员。

表 18-1 分阶段实施的进入条件

阶段	应形成的记录	下一阶段条件
准备	任务说明、范围与人工基线	输入合法可用，责任明确
受限试点	错误分析、修复及复测	达到预设质量要求
扩大使用	新范围测试、维护配置	资源与异常接管能力充分
持续运行	变更、反馈及复评记录	条件变化时重新判断

本表为本书提出的管理框架，不含实测效果数据。

18.3 案例知识检索的行业典型场景

假设承办人需要研究一项买卖合同争议。任务开始时先填写争议事项、关键时间、交易结构和需要回答的问题，再上传经过授权的材料。系统据此检索相关规范与案例，展示关键事实差异和来源位置。对没有明确交易时间的输入，先追问；对看似相近但合同类型不同的案例，说明其关联限制，不能直接合并为同一规则。

承办人检查候选来源后，系统可以组织研究提纲和待核事项，专业人员再决定哪些观点进入正式分析。交付时保留实际使用的来源版本和重要事实假设。如果后来新增材料改变交易性质，事项返回分析环节，重新检查受影响的依据。该场景验证的是检索、事实补充与回溯修改之间的衔接，不以生成一份长报告作为成功标准，也不推定检索结果能够预测裁判结论。

18.4 采购合同审查的行业典型场景

假设企业提交一份设备采购合同及技术附件。系统先检查正文引用的附件是否齐全，抽取主体、价款、交付、验收和责任条款，并比较不同版本。若正文写明一次性验收，而附件写明分阶段验收，系统将其作为冲突提示，要求业务人员确认真实安排。承办人随后结合采购目标提出修订选项，财务与业务负责人分别确认相关条件。

签署前固定最终文件与附件组合，检查授权及签署对象。履约阶段将到货、验收和付款作为不同事件，提醒功能可以自动运行，付款动作则依据明确授权和证据另行控制。若验收存在异议，系统暂停自动付款并分配核查任务；是否暂缓履行付款义务，仍须由有权人员结合合同条件和实际事实判断。系统应继续提示到期时间，避免将技术暂停误当作合同义务已经中止。此场景的价值在于让文本、业务状态和执行权限保持一致，不把条款识别准确直接等同于交易风险已经解决。

18.5 合规事件处理的行业典型场景

假设企业收到涉及客户资料使用的投诉。受理人员记录投诉内容、相关业务和现有证据，系统协助定位处理流程及责任岗位。最初不急于生成是否违法的确定结论，而是梳理收集、使用、传输和留存环节，标记需要补充的信息。对可能仍在持续的异常处理，可先由有权人员决定采取必要的限制措施，同时保存调查所需记录。

专业人员根据查明事实分析适用要求，提出整改、解释和后续检查安排。技术团队负责验证配置是否修正，业务团队负责调整实际操作，客户服务人员负责适当反馈。事项关闭前应确认问题原因已经处理，并检查相似流程是否存在同类风险。系统可以组织任务和资料，但不应代替有权人员作出对外责任认定或监管报送决定。

18.6 交接演练与正式运行

正式运行前至少进行一次人员交接、一次供应商中断和一次规则错误修正演练。交接演练观察新承办人能否继续事项；中断演练观察能否读取必要资料并转入人工流程；修正演练观察能否找到受影响输出及责任人。演练使用清楚标识的测试对象，避免对真实客户产生外部后果。

运行移交应包含知识维护清单、权限管理安排、异常联系方式、评价口径和退出路径。完成移交并不意味着项目不再需要投入，而是从建设责任转为运营责任。实施阶段主要确认机制能否运行，第十九章则从相对独立的评价职责检验服务结果。评价口径和验收样本应在试点前准备，试点期间即执行评价，不是等实施结束后才开始检查。职责可以分开，工作仍应衔接，防止实施团队只用成功演示证明产品有效。

第 19 章 效果评价与风险复评

19.1 评价样本覆盖真实差异

评价应在预先确定的范围内选择代表性任务，覆盖不同文档质量、业务复杂度和事实完整性。除正常样本外，还应纳入条款冲突、失效依据、缺失附件、无权访问、重复请求和无法确定答案的情况。不能只从系统已经成功处理的记录中抽样，否则失败输入会被排除在评价之外，结果难以支持推广决策。

样本应区分用于建设调试的集合与用于验收的集合。专业人员根据来源和任务范围确定可接受输出，并记录分歧。对于开放性法律分析，评价可检查必要问题是否识别、关键依据是否适用和结论条件是否清楚，而不是要求与某段标准答案逐字一致。样本数量和分布应如实报告，没有充分覆盖的场景不得借用其他场景的结果。

19.2 分别评价知识与专业输出

知识评价检查来源可达、版本正确、引用支持和检索遗漏。可以让评价人员从最终结论反向定位依据，再从一项已知重要规则正向检查系统是否能够找到。两个方向结合，能发现“有链接但不支持结论”和“关键依据根本未检索到”的不同问题。检索结果越多并不必然越好，相关性、适用性和完整性需要共同判断。

专业输出评价应检查事实假设、推理链条、重要风险和建议可执行性。对每项遗漏或错误说明可能后果，并区分系统错误与输入缺陷。输入不足时正确追问，应计为符合要求的处理，而不应被视为回答失败。相反，在没有足够事实时给出肯定结论，即使偶然与实际结果一致，也不应被评价为可靠能力。

表 19-1 问题与验证的对应关系

问题章节	机制章节	关键验证
第 6 章 知识	第 12 章 架构	版本替换、引用回溯
第 7 章 流程	第 13 章 责任	人员交接、实质复核
第 8 章 投入	第 14 章 指标	同口径比较、全成本
第 9 章 数据	第 15 章 协作	越权检索、权限撤销
第 10 章 执行	第 16 章 合同	重复请求、停止接管
第 11 章 用户	第 17 章 运营	行动理解、异议闭环

本表为本书提出的管理框架，不含实测效果数据。

19.3 检查人工复核是否实际有效

复核效果不能只统计通过率。可以比较复核前后发现的实质问题、修改理由及仍然遗漏的风险，并检查复核人员是否查看了关键来源。若长期没有任何退回，也没有记录合理分歧，应调查是否存在形式复核，而不能直接认定系统质量完美。对高负荷时段和新入职人员，还应观察复核能力是否下降。

评价应避免把复核人员置于无法完成的任务量下。若系统每次输出过长、来源过多，人员可能无法在合理时间内完成检查。此时应调整输出结构和任务范围，而不是一味增加签字层级。复核的价值在于发现足以改变结论的问题，合理设计应帮助人员集中注意力，并保留补充调查和暂停交付的权力。

19.4 权限与执行测试采用否定路径

权限测试应在明确的账号、数据和攻击路径范围内，验证未授权内容能否被拦截；执行测试应验证预设的条件缺失时动作是否被阻止。通过测试说明已测路径符合要求，不能据此推断所有未知路径均无风险，仍需持续监测和复评。可以检查越权检索、过期授权、撤销权限后的缓存、异常附件指令和错误对象参数。对自动执行模块，还要测试重复提交、外部超时、部分成功和回执缺失，观察系统是否保持可解释状态并进入人工接管。

测试结果需要保存预期、实际行为和必要日志，不能只记录“安全通过”。发现缺陷后，应明确影响范围，修复并复测相关路径。关键隔离或执行控制存在未解决问题时，不应通过其他指标的高分抵消。对于无法在现有条件下消除的风险，可以缩小功能范围或保持人工模式，但应让使用团队了解实际限制。

19.5 价值结果与不确定性共同呈现

效率评价应比较同等范围和质量要求下的处理时间、等待、返工与成本，并报告样本分组。平均时间可能掩盖少数复杂事项的大幅延迟，因此可以同时观察中位数、较慢任务和异常原因。用户评价则关注能否理解结论、完成下一步行动和获得有效纠错。满意度、使用次数和系统在线率可作为补充，不能代替专业质量。

在没有随机对照或充分控制条件时，应谨慎解释因果关系。人员熟练度、业务组合和流程调整都可能影响结果，不能把所有变化都归于智能工具。报告可以说明观察到的变化、可能原因和后续验证计划。明确不确定性不会降低评价价值，反而帮助管理者判断结果能够支持多大范围的推广，以及哪些投入还需要继续观察。

19.6 将复评嵌入变更与退出

规则、模型、数据来源和业务范围发生实质变化时，应进行针对性复评。复评不必每次重复全部测试，但应覆盖变化影响的链条，并保留关键回归样本。若长期出现同类错误、复核负担不断增加或供应商无法提供必要保障，应重新评估继续使用的条件。停止某项功能可以是保护服务质量的主动选择。

退出评价应确认历史资料可以读取、关键关系可以迁移、未完成事项有人接续，以及不再需要的权限和数据得到适当处理。对外部执行产生的历史后果，仍应保留必要的业务记录与责任安排。由此，评价贯穿采用、运行、变更和退出全过程，使规模化始终受到证据和资源约束，而不是依靠一次验收长期放行。

19.7 验收记录支持可复查决定

每项验收记录应包含场景、输入版本、预期要求、实际输出、问题后果、处理人和复测结论。对通过的任务也保留必要记录，便于后续版本比较。验收负责人应形成明确决定：开放哪些能力、限制哪些条件、由谁负责监测。若结论是继续受限使用，应写明下一次复评的触发条件，避免试点状态长期悬置却实际承担正式业务。

第 20 章 适用边界与发展结论

20.1 可信与规模化的共同基础

智能法律服务的持续价值来自专业控制的可重复执行。知识来源能够追溯，事项事实能够传递，执行动作受到授权，错误能够定位和修正，这些能力共同支撑服务扩大。单一模型的语言表现只能解决其中部分工作，不能替代组织对来源、责任和客户行动的管理。机构应把建设目标从“能够生成什么”转向“能够稳定交付什么”。

全书提出的五层架构与六类机制相互依赖。知识治理解决依据失真，事项流程减少事实与理由丢失，评价机制支持资源配置，数据协作保护权限和语义，受控执行限制外部后果，分层服务改善用户理解与纠错。任何一项明显不足，都可能成为规模化的限制。改善顺序应根据本机构的主要障碍决定，而不是机械地一次性完成全部技术功能。

20.2 需要保留专业判断的领域

当关键事实无法核实、法律适用存在重大争议、业务后果难以逆转或委托关系存在冲突时，应保留充分的专业介入。工具可以帮助整理材料和比较路径，但不宜独立形成最终处置。对刑事、重大争议和复杂跨境事项，尤其需要依据具体情境配置人员和程序，不能用一般合同试点的结果推定能够处理同等复杂任务。

区块链、电子签名和语言模型各自解决不同问题。存证辅助核对记录完整性，签署机制连接身份与意思表示，模型协助处理语言与信息。它们可以组合，却不会相互传递无条件的法律可靠性。机构应持续检查技术能力与法律判断之间的空隙，并用事实调查、授权和复核填补，而不是通过概念叠加形成过度承诺。

20.3 推广取决于组织准备

具备稳定业务分类、明确资料权限和可持续维护资源的机构，更容易把方法转化为日常工作。基础薄弱的团队可以从资料管理、版本控制和专项检索开始，不必追求完整平台。跨团队推广之前，应确认识识结构、交接规则和评价口径能够被接收方理解，必要时根据业务差异重新验证，而不是直接复制配置。

推广也需要保留退出与调整空间。当业务量不足以支撑维护、供应商依赖无法管理或用户无法正确理解输出时，缩小范围可能比继续扩展更合理。专业服务的衡量尺度应包括纠错和接续能力。能够及时识别不适用场景并停止自动化，是成熟治理的组成部分，而非对技术价值的否定。

20.4 面向持续改进的研究方向

后续研究可进一步关注法律概念的结构化表达、引用与命题之间的自动核验、复杂事实的交互收集及跨系统责任记录。这些方向需要法律专业人员与技术团队共同定义问题，并使用真实任务检验。研究成果转化为产品时，应清楚区分实验表现、受限试点和稳定服务，避免将不同层次的证据混为一谈。

管理者可以据此作出一项清晰决定：选择边界明确的场景，配置来源与责任，建立独立评价，再依据结果扩大使用。以 2024 年 8 月为研究时点，本书提供的是一套可实施、可检验的法律服务治理方法。其推广价值最终需要由不同机构在各自条件下持续验证，并随着规范与业务变化进行修订。

参考来源

正文以方括号序号标注来源。以下列出所依据的公开规范与原始技术研究；制度文件与技术文献的适用对象各有不同，正文已区分事实依据与管理建议。

[1] 全国人民代表大会常务委员会. 中华人民共和国个人信息保护法.

https://www.cac.gov.cn/2021-08/20/c_1631050028355286.htm

[2] 全国人民代表大会常务委员会. 中华人民共和国数据安全法.

https://www.cac.gov.cn/2021-06/11/c_1624994566919140.htm

[3] 全国人民代表大会常务委员会. 中华人民共和国电子签名法（2019年修正）.

https://www.miit.gov.cn/jgsj/zfs/fl/art/2022/art_e3f623f70c23497e88a941170093446a.html

[4] 国家互联网信息办公室等七部门. 生成式人工智能服务管理暂行办法.

https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

[5] 最高人民法院. 人民法院在线诉讼规则.

<https://www.court.gov.cn/zixun/xiangqing/309551.html>

[6] 最高人民法院. 关于规范和加强人工智能司法应用的意见.

<https://www.court.gov.cn/zixun/xiangqing/382461.html>

[7] 美国国家标准与技术研究院（NIST）. Artificial Intelligence Risk Management Framework (AI RMF 1.0).

<https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10>

[8] Patrick Lewis 等. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.

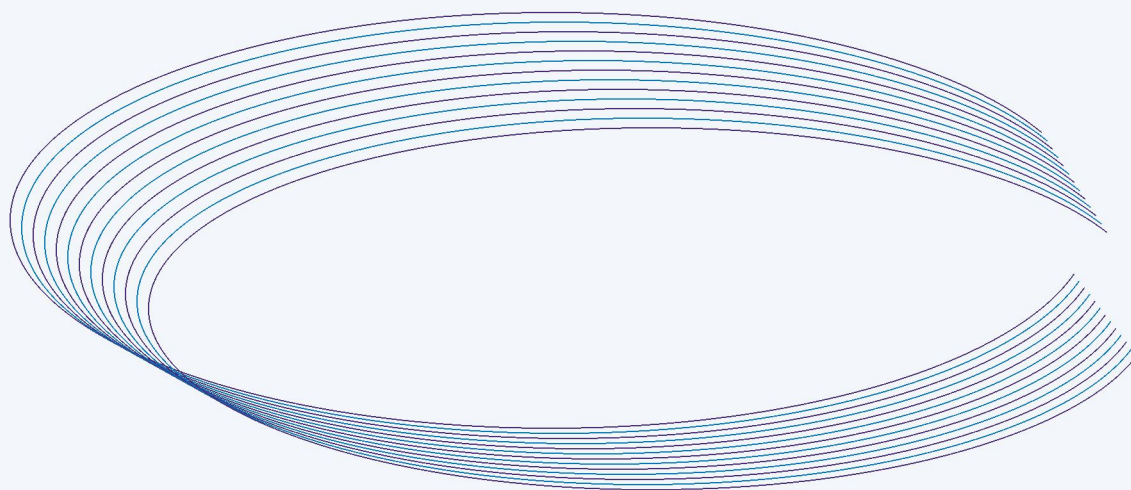
<https://arxiv.org/abs/2005.11401>



瀚思科技创新与发展协会

智能法律服务 可信化与规模化应用

法律知识库 · 智能合同 · 场景化合规



2024年8月

thehatid.com