ICS 35.240.99
CCS L 67

团　　体　　标　　准

T/ISC XXX—XXXX

法律从业人员人工智能使用能力评估规范

Specification for Assessment of Artificial Intelligence Usage Competency of Legal Practitioners

（征求意见稿）

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

中 国 互 联 网 协 会　　发布

目　录

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

以大语言模型为代表的生成式人工智能技术正在深刻重塑法律服务的生产方式，人工智能已从辅助工具逐步深度融入判例检索、案情分析、合同审查、文书起草等法律业务核心流程。

当前，法律从业人员的人工智能使用能力呈现明显的“两极分化”：部分人员仅停留于简单问答式使用，普遍存在幻觉识别能力弱、敏感数据泄露风险高、提示词工程能力不足等问题；同时，现有的人工智能能力评价多面向通用信息技术或开发群体，缺乏针对法律行业“高合规、高严谨性、强伦理要求”特点的专项评估标准。法律业务容错率极低，人工智能使用不当可能直接导致法律风险、执业风险与客户权益受损。

制定本文件，旨在建立法律从业人员人工智能使用能力的基本评价框架，提升其工具判断能力、业务应用能力和风险控制能力，保障法律服务质量、数据安全和客户权益，为律师事务所、企业法务部门、司法辅助机构及高等院校相关培训活动提供可量化、可实施的评价参照。

本文件确立了“基础认知与工具判断＋业务场景应用”的双维能力评估模型，采用分级递进的能力等级设置，并将数据安全合规与人工智能输出校验作为贯穿评估全过程的核心约束指标，突出“场景驱动、实操为主、风险兜底”的评估特色。

法律从业人员人工智能使用能力评估规范

1 范围

本文件规定了法律从业人员人工智能使用能力评估的原则、能力评估模型与维度、能力等级要求、评估方法、评分与合格判定及评估实施要求。

本文件适用于对下列人员开展人工智能使用能力评估：

a）律师事务所执业律师、律师助理、实习律师；

b）企事业单位、金融机构及社会团体的法务人员与合规人员；

c）司法机关、仲裁机构、公证机构中从事法律辅助工作的人员；

d）其他从事法律服务相关工作的人员。

高等院校法学专业学生及相关培训机构可参照本文件开展人工智能使用能力评价或培训，但本文件不规定培训课程设置、培训机构资质及认证活动要求。

本文件不适用于法律职业资格、执业资格、独立办案资格或其他法定从业资格的认定。

2 规范性引用文件

GB/T 41867—2022《信息技术 人工智能 术语》

GB/T 46800—2025《生成式人工智能技术应用社会影响评估指南》

3 术语和定义

GB/T 41867—2022和GB/T 46800—2025界定的术语，以及下列术语和定义适用于本文件。

有关法律业务的专门用语按本文件所界定的评估场景使用。

3.1 法律从业人员

从事律师、企业法务、合规、司法辅助、仲裁、公证及其他法律服务相关工作的人员。

3.2 人工智能系统 artificial intelligence system

针对人类定义的给定目标,产生诸如内容、预测、推荐或决策等输出的一类工程系统。

3.3 人工智能 artificial intelligence; AI

<学科>人工智能系统相关机制和应用的研究和开发。

3.4 生成式人工智能 generative artificial intelligence;GenAI

是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术。

3.5 人工智能使用能力

法律从业人员在法律业务活动中，选择、配置、操作、评估和校验人工智能工具，并对相关安全、合规和专业责任风险进行控制的综合能力。

3.6 法律人工智能工具

面向法律咨询、法律检索、文书起草、合同审查、案情分析等法律业务场景开发或适配的人工智能软件、平台或服务，包括通用型人工智能工具与法律垂直领域人工智能工具。

3.7 提示词工程 prompt engineering

为获得预期输出，对输入给人工智能模型的指令、角色设定、上下文、示例、约束条件等内容进行设计、组织与迭代优化的方法与过程。

3.8 人工智能幻觉

人工智能模型生成表面上合理，但与客观事实、权威法源或任务条件不一致，或者缺乏真实依据的内容的现象。

注：法律场景中的典型表现包括虚构案例、编造或错误援引法条、引用已失效或不适用的法律文件、错误记载裁判文书名称、案号、文号、法院名称或裁判观点，以及将不确定推测表述为确定性法律结论等。

3.9 检索增强生成 retrieval-augmented generation

通过检索外部知识库并将检索结果引入内容生成过程，以提高输出准确性与时效性的技术方法。

3.10 人机协同

法律从业人员与人工智能工具在任务分工、交互反馈与结果校验中协同完成法律业务的工作方式。

3.11 输出校验

法律从业人员对人工智能生成内容进行事实核对、法源核验、时效性核查、适用范围审查、逻辑审查、格式审查和专业判断的过程。

注：输出校验的深度与任务风险相适应，其结果用于支持法律从业人员的独立专业判断。

3.12 法律业务敏感信息

在法律服务、法务、合规、司法辅助、仲裁、公证及其他相关业务活动中，因其泄露、篡改、丢失或被不当使用，可能对个人、组织、案件当事人、客户或机构权益造成不利影响的信息，包括个人信息、敏感个人信息、商业秘密、未公开案件材料、客户依法或依约要求保密的信息，以及其他依法或依约应当保密的信息。

注：法律业务敏感信息是基于法律业务场景确定的风险概念，不等同于《中华人民共和国个人信息保护法》规定的“敏感个人信息”。法律业务敏感信息的范围可能大于敏感个人信息。

3.13 数据脱敏

通过删除、泛化、替换、掩码、扰动或其他技术措施，降低数据中特定对象被直接或间接识别可能性的处理过程。

注：数据脱敏不等同于匿名化，脱敏后的数据仍可能具有可识别性或保密属性。

4 缩略语

下列缩略语适用于本文件。

AI：人工智能（Artificial Intelligence）

AIGC：人工智能生成内容（Artificial Intelligence Generated Content）

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

5 评估原则

5.1 客观公正

评估指标应可量化、可观测，评估过程应可追溯，评估结果应可复核。

5.2 场景驱动

评估应以真实法律业务场景任务为核心载体，重点考察实操能力而非单纯理论知识。

5.3 安全合规优先

数据安全、保密义务、个人信息保护、执业伦理和人工智能输出校验应作为评估的基础要求贯穿评估全过程。评估应在维度一和维度二中分别设置安全合规相关能力要素作为独立评分指标，并纳入相应维度权重；同时设置否决项。否决项不计入总分，但一经确认触发，即直接判定评估不合格。

5.4 分级递进

能力等级由低到高逐级递进，高等级要求涵盖并高于低等级要求。

5.5 动态更新

评估内容、题库、场景库和评分规则应根据人工智能技术发展、法律法规变化、监管要求变化及法律业务实践变化进行动态调整。题库和场景库原则上每年至少全面复核和更新一次；发生重大法律法规变化、重大技术风险变化或评估结果显示题目失效时，应及时组织专项更新。

评估机构应对题库和场景库的版本、更新内容、适用时间及生效范围进行记录和管理。

6 能力评估模型与维度

6.1 模型结构

能力评估模型由两个维度构成：

a）维度一：AI基础认知与工具判断能力——底层素养，包括技术基本原理认知、法律AI工具甄选与评估、数据安全与伦理合规判断三个能力要素；

b）维度二：法律业务场景应用能力——核心实战，包括法律检索与分析、法律文书生成与审查、案情拆解与逻辑推理辅助、业务流程自动化与效能优化、法律业务场景中的安全合规与输出校验五个能力要素。

数据安全、保密义务、执业伦理和输出校验要求贯穿两个维度。维度一重点评估评估对象对安全合规规则的认知和工具风险判断能力;维度二重点评估评估对象在具体业务任务中的合规操作、数据处理、法源核验和人工复核能力。除作为独立能力要素计入相应维度权重外，对严重违反保密义务、数据安全要求或评估纪律的行为设置否决项。

第6章规定评估内容，第7章规定各能力要素的分级要求，第8章规定取得评分证据的方法及合格判定条件。6.2和6.3列示的能力内容应按7.5对应等级的任务复杂度、独立完成程度和风险控制要求实施，不应将高级任务要求直接用于初级评估。

6.1.1 评分项目归属

评估机构应将理论知识测试、场景实操演练、案例分析、成果评审与答辩中的试题、任务或答辩问题设置为评分项目，并在评分表中明确每一评分项目的：

a）评估方式；

b）能力维度；

c）能力要素；

d）对应等级要求、满分及分档评分标准；

e）可核验的评分证据及记录方式。

每一评分项目只归属于一种评估方式和一个能力要素。确需考察多个能力要素的，应拆分为独立评分点，分别明确归属和分值。同一证据可支持不同能力的观察，但同一表现不得重复计分。

评估机构应在评估实施前制定评分表或评分细则，明确评分项目与表2规定的评估方式、附录A规定的能力要素之间的对应关系。评分表或评分细则应作为评估档案保存。

6.2 维度一：AI 基础认知与工具判断能力

6.2.1 技术基本原理认知

评估对象应能够：

a）理解生成式人工智能基于概率统计的内容生成机制，知晓其输出不具备天然的准确性保证（理解模型表现出的推理能力不等同于法律意义上的事实认定、法律适用或专业判断能力）；

b）说明训练数据来源、知识截止时间等因素对输出时效性与适用性的影响；

c）识别人工智能幻觉的成因、表现形式及其在法律检索、文书起草等场景中的危害；

d）了解检索增强生成、模型微调等可能改善特定任务表现的技术路径，并认识到其效果取决于数据质量、检索质量、模型配置和验证机制，不能替代人工核验。

6.2.2 法律 AI 工具甄选与评估

评估对象应能够：

a）区分通用型人工智能工具与法律垂直领域人工智能工具的功能特点与适用边界；

b）根据任务类型（检索、生成、审查、分析等）选择相匹配的工具；

c）从输出准确性、法源库覆盖范围、数据合规性、部署方式（公有云/私有化）、结果可追溯性等方面对工具进行评估；

d）识别工具服务协议中与数据收集、使用、留存相关的风险条款。

6.2.3 数据安全与伦理合规判断

评估对象应能够：

a）遵守执业保密义务，准确识别委托关系中敏感信息的范围；

b）能在使用人工智能工具前，根据业务目的、数据敏感程度和工具安全属性，判断数据使用的必要性、授权范围和最小使用量，实施输入前脱敏；能识别并避免将法律业务敏感信息输入未经安全评估、

未经授权或不符合机构数据管理要求的公共人工智能工具；能对人工智能输出进行检查，识别并处置其中可能存在的敏感信息泄露、重新识别或不当扩散风险；

数据脱敏的强度应根据业务目的、数据敏感程度和安全风险确定。脱敏处理不替代合法处理依据、必要授权或保密义务审查；无法将风险控制在允许范围内的，不应输入相关工具。

c）了解人工智能生成内容相关的著作权、人格权风险及内容标识要求；

d）把握人工智能使用中的执业伦理边界，明确人工智能输出不能替代执业人员的专业判断与责任承担；

e）了解生成式人工智能服务相关监管规则的基本要求。

人工智能生成内容仅作为法律业务辅助材料。涉及法律意见、诉讼或仲裁策略、合同最终文本、对外正式文件、客户答复、司法或行政机关提交材料的，应由具备相应专业能力和权限的法律从业人员进行独立审查、确认和负责。

评估结果不替代法律职业资格或业务授权，不免除评估对象依法、依约或依职业规范承担的专业责任。

6.3 维度二：法律业务场景应用能力

6.3.1 法律检索与分析

评估对象应能够：

a）将复杂法律问题拆解为检索要点，设计检索策略与提示词；

b）使用人工智能工具辅助判例检索、法规梳理与类案归纳；

c）对人工智能返回的案例、法条进行溯源核验与交叉验证，包括引注真实性、法规效力状态与时效性；

d）识别并纠正检索结果中的虚构案例、失效法规与错误援引。

6.3.2 法律文书生成与审查

评估对象应能够：

a）运用提示词工程方法（角色设定、任务分解、示例引导、约束条件设定等），指导人工智能生成合同、诉讼文书、法律意见书等文书初稿；

b）对人工智能生成文书进行事实核对、法律适用审查与逻辑校验；

c）对生成文书进行纠错与精修，使其达到可交付标准；

d）利用人工智能辅助开展合同条款比对与风险点筛查。

6.3.3 案情拆解与逻辑推理辅助

评估对象应能够：

a）利用人工智能整理案情事实，制作时间线、法律关系与人物关系梳理；

b）借助人工智能梳理证据链条、识别证据缺口；

c）利用人工智能模拟对方立场开展对抗推演，寻找辩护或诉讼策略切入点；

d）对人工智能推理结论保持独立判断，验证其逻辑链条的完整性与可靠性。

6.3.4 业务流程自动化与效能优化

评估对象应能够：

a）将人工智能应用于邮件撰写、会议纪要、尽调清单整理等重复性工作；

b）建立常用提示词模板与标准化工作流程；

c）评估人工智能应用对工作效率与成果质量的实际影响，合理分配人机任务；

d）对自动生成的邮件、会议纪要、尽调清单、任务分派、风险提示及其他业务成果进行必要的人工复核，不得将未经核验的人工智能输出直接作为对外文件、法律意见、诉讼策略或最终业务结论。

6.3.5 法律业务场景中的安全合规与输出校验

评估对象应能够：

a）向人工智能工具输入数据前实施数据脱敏，避免敏感信息输入未经授权的公共AI工具；

b）业务任务执行过程中保持操作留痕（提示词记录、AI输出原文、人工修改记录）；

c）对AI输出进行法源核验、时效性核查和适用范围审查；

d）按任务风险等级采取相应核验深度，重大事项独立人工复核；

e）识别AI输出中的敏感信息泄露风险并处置。

7 能力等级划分

7.1 等级设置

能力等级由低到高划分为三级：初级（应用级）、中级（进阶级）、高级（专家级）。高等级要求涵盖低等级要求。

申请人可直接申请与其能力相适应的等级，不以取得较低等级评估证明为前提。

评估机构可对与申请等级相关的实践成果、工作经历或其他能力证明材料作出具体规定，并应在申请前公开；不得将取得较低等级评估证明作为申请前提。

各等级能力要求应通过可观察行为、任务成果或工作记录判定。评估机构应将7.5的分级要求转化为评分点，明确不符合、基本符合和充分符合要求的表现及其分值区间，并记录实际证据。

7.2 初级（应用级）

评估对象应达到以下要求：

a）能在评估规则允许的提示范围内，使用常见法律人工智能工具完成基础性任务；除统一提供的任务说明外，不依赖考官对具体操作步骤、提示词内容或结论进行实质性指导；

b）掌握基本提示词写法，能对人工智能输出进行基础核对；

c）具备基本的数据安全与保密意识，不发生敏感信息违规输入；

d）了解人工智能的局限性，发现明显错误时能停止使用或寻求协助。

7.3 中级（进阶级）

评估对象应达到以下要求：

a）能独立在法律检索、文书生成、案情分析等多种业务场景中熟练使用人工智能工具；

b）熟练运用提示词工程方法开展任务分解与迭代调优；

c）能对人工智能输出进行系统性校验（法源核验、交叉验证），识别幻觉并纠错；

d）能根据业务需求评估、选择工具，并落实数据脱敏与合规操作；

e）能总结形成个人提示词模板与标准化工作流程。

7.4 高级（专家级）

评估对象应达到以下要求：

a）能设计面向复杂法律业务的人机协同解决方案；

b）能制定所在机构的人工智能使用规范、风险防控制度与输出复核流程；

c）能在具备相应专业能力和业务权限的范围内复核人工智能辅助成果，建立并验证分级复核机制；

d）能培训、指导团队成员，主导工具选型与私有化部署评估；

e）能持续跟踪人工智能技术与监管动态，更新知识体系与机构规范。

7.5 能力要素与等级要求的对应关系

各等级均应覆盖两个维度的八个能力要素，具体要求见表1。高级要求包含中级和初级的基础能力，中级要求包含初级的基础能力。评估任务可按专业领域设计，但不得删去能力要素。等级认定应同时满足本条覆盖要求和8.6的评分、最低分及否决项条件。

表 1　能力要素与等级要求对应表

能力要素	初级（应用级）	中级（进阶级）	高级（专家级）
技术基本原理认知	说明生成结果可能出错及知识时效局限，识别明显幻觉。	解释典型幻觉及检索增强生成的局限，选择相应核验方法。	分析复杂方案中的模型、数据和检索风险，制定并验证质量控制措施。
法律 AI 工具甄选与评估	按任务要求选择允许使用的工具，识别基本功能和输入限制。	比较工具准确性、法源覆盖、数据处理和协议风险，说明选型依据。	组织工具选型及部署方式评估，设计验证任务、准入条件和持续监测要求。
数据安全与伦理合规判断	识别敏感信息、授权及保密要求，说明责任边界。	判断复杂任务的数据必要性、授权范围、协议风险及处理措施。	制定机构使用规则、责任分工及风险处置机制，并说明其适用边界。
法律检索与分析	提取基础检索要点，核对案例、法条来源及效力，发现明显错误。	独立拆解法律问题，开展溯源和交叉验证，纠正错误援引。	设计复杂问题的检索与核验方案，处置法源冲突、不确定性和重大遗漏。
法律文书生成与审查	用基本提示词生成简单文书初稿，核对事实、法源和格式。	独立分解任务、迭代提示词，系统审查文书并提出专业修订。	设计复杂文书生成和分级审查流程，验证质量并明确交付与复核权限。
案情拆解与逻辑推理辅助	整理基本事实、时间线和关系，区分已知事实与推测。	梳理证据链、发现缺口，验证推理并比较不同立场。	组织复杂案情分析及对抗推演，识别关键假设、证据局限和决策风险。
业务流程自动化与效能优化	按给定流程完成重复性任务，保留人工检查环节。	形成提示词模板和工作流程，比较使用前后的质量与效率。	设计团队人机分工及异常处置流程，验证效果并指导他人使用。
法律业务场景中的安全合规与输出校验	按规则脱敏、留痕和核对输出，无法确认时停止采信并提交复核。	独立实施数据处理、法源核验、风险分级和纠错，形成可追溯记录。	建立并验证机构复核机制，处置泄露、错误扩散及复核失效等复杂风险。

8　评估方法

8.1　通则

评估应采用表2规定的相应等级评估方式组合和权重，不得任意省略其中的评估方式。

高级评估不单独设置理论知识测试，但应在案例分析、成果评审和答辩中考察人工智能基本原理、工具风险判断、数据安全、执业伦理及监管要求等内容。

表 2　各等级评估方式组合及占比

评估方式	初级	中级	高级
理论知识测试	30%	20%	—
场景实操演练	70%	60%	40%
案例分析	—	20%	20%
成果评审与答辩	—	—	40%
合计	100%	100%	100%

8.1.1 评估方式与能力维度的对应关系

各评估方式应按照表3覆盖相应的能力维度和能力要素。评估机构可根据申请等级、法律业务领域和评估任务特点设置具体评分项目，但应满足以下要求：

a）每个等级的八个能力要素均应至少设置一个具有对应等级要求的评分项目；

b）维度一和维度二均应有可观测的评分证据；

c）数据安全、保密义务、执业伦理、输出校验和人工复核等要求，应在相关评分项目中设置明确的评分点；

d）每一评分项目只能计入一种评估方式和一个能力要素；

e）同一评分表现不得通过不同评分项目重复计分。

表 3　评估方式与能力要素覆盖关系

评估方式	维度一覆盖的能力要素	维度二覆盖的能力要素
理论知识测试	技术基本原理认知、法律AI工具甄选与评估、数据安全与伦理合规判断	—
场景实操演练	数据安全与伦理合规判断	法律检索与分析、法律文书生成与审查、案情拆解与逻辑推理辅助、业务流程自动化与效能优化、法律业务场景中的安全合规与输出校验
案例分析	技术基本原理认知、数据安全与伦理合规判断	法律检索与分析、法律文书生成与审查、法律业务场景中的安全合规与输出校验
成果评审与答辩	技术基本原理认知、法律AI工具甄选与评估、数据安全与伦理合规判断	法律检索与分析、法律文书生成与审查、案情拆解与逻辑推理辅助、业务流程自动化与效能优化、法律业务场景中的安全合规与输出校验

表3规定的是评估方式对能力维度和能力要素的覆盖关系，不改变表2规定的评估方式权重和附录A规定的能力要素权重。

8.2　理论知识测试

理论知识测试适用于初级和中级评估，采用单选题、多选题、判断题等客观题形式，主要考察人工智能基本原理、工具选择、数据安全、保密义务、执业伦理、输出校验及相关监管要求。

初级理论知识测试重点考察基本概念、基本风险和基本操作要求；中级理论知识测试重点考察复杂场景下的工具判断、风险识别和核验方法。题库应定期复核和更新，并采用随机抽题组卷方式。

8.3　场景实操演练

场景实操演练应在评估机构提供的受控环境中进行，并使用经评估机构审核的人工智能工具和经脱敏处理的案例数据。评估机构应在评估开始前明确以下事项：

a）允许使用的人工智能工具、账号和功能；

b）是否允许访问互联网、法律数据库或其他外部知识库；

c）是否允许使用个人设备、个人账号或外部插件；

d）数据输入、输出保存、操作记录和屏幕录制要求；

e）禁止使用的工具、资料和辅助方式。

评估对象应在规定时间内完成任务。评分应包括任务完成质量、提示词设计、工具选择、输出校验、法律专业判断、操作合规性和成果表达等内容。

8.4 案例分析

评估机构应提供与案例分析相关的事实材料、任务说明、适用法域、法律法规版本或检索范围，并可提供含有预设问题的人工智能输出材料。预设问题可包括虚构案例、错误法条、失效法规、法域适用错误、事实遗漏、逻辑跳跃和敏感信息泄露风险等。

评估对象应识别相关问题，说明问题性质及可能造成的法律、业务或合规后果，提出核验方法、纠正措施和必要的人工复核方案。评分不应仅依据结论是否与参考答案一致，还应评价问题识别的完整性、核验方法的合理性和论证过程的可复核性。

8.5 成果评审与答辩

高级申请人应提交本人参与完成的人工智能应用实践成果或解决方案，并接受专家组评审和答辩。成果应能够支持表1高级要求的评价；覆盖不足的能力要素应由其他评估任务补足。

8.5.1 提交材料

申请人应提交任务背景与目标、适用法域及资料版本、工具选择依据、人机分工、关键提示词与操作记录、人工智能原始输出、人工修订与核验记录、最终成果及效果说明。团队成果应说明本人职责、贡献和可核验的佐证。材料应符合9.3的数据要求；不能安全提供的原始材料应以经审核的模拟材料或其他可核验证据替代。

8.5.2 评价内容

成果评审应至少评价：任务目标与方案适配程度；事实、法源及法律分析的准确性；工具选型与人机分工的合理性；数据处理、保密和权限控制；输出核验、人工纠错及复核机制；效果验证和成果可复核性。未实施的解决方案应提交模拟验证或测试结果，不得将预期效益作为已实现效果计分。

8.5.3 答辩要求

答辩应核验申请人的实际贡献，要求其解释关键选择、核验路径和专业判断，并回应事实变化、法源失效、输出错误或信息泄露等情境问题。评估机构应统一同等级答辩的基本流程、时间范围和追问规则，保留问题、回答要点及评分依据。

8.5.4 评分与复核

评估机构应事先制定成果评审与答辩评分表，分别设置成果评价和答辩评分点，明确其分值、能力要素归属、等级要求及证据。两部分均应计分，合并为表2中的一种评估方式；同一表现不得重复计分。评分不应以材料包装、机构资源或工具性能替代对个人能力的评价。

专家组应覆盖法律实务、人工智能应用及安全合规方面的专业能力，按9.2独立评分并复核。缺乏证据支持的能力表现不得获得相应分值；涉及伪造、抄袭等情形的，按8.6.3处理。

8.6 评分与合格判定

评分采用"两条线、一票否决"的体系。每一评分项目均须明确满分、实际得分、所属评估方式和所属能力要素，同一评分项目从两个角度分别汇总，独立使用，不叠加计算：

——综合得分线：按评估方式归集评分项目，结合表2权重计算综合得分，用于判定总体是否合格；

——维度得分线：按能力要素归集评分项目，结合附录A权重计算维度得分，用于判定是否存在严重偏科。

此外，评估过程中触发否决项的，直接判定不合格。

8.6.1 综合得分

第一步，计算各评估方式得分：将同一评估方式下全部评分项目的实际得分求和，除以满分求和，换算为百分制：

评估方式得分＝该评估方式下实际得分之和÷该评估方式下满分之和×100

第二步，按表2规定的权重加权求和：

综合得分＝Σ（各评估方式得分×表2对应权重）

综合得分满分为100分。

8.6.2 维度得分

第一步，计算各能力要素得分：将同一能力要素下全部评分项目的实际得分求和，除以满分求和，换算为百分制：

能力要素得分＝该能力要素下实际得分之和÷该能力要素下满分之和×100

第二步，按附录A规定的权重计算维度内加权平均：

维度得分＝Σ（维度内各能力要素得分×附录A对应权重）÷维度总权重

其中，维度一总权重为40%，维度二总权重为60%。能力要素得分已经换算为百分制，维度加权平均不再乘以100。维度得分满分为100分，用于各等级的维度最低分判定，不计入综合得分。

8.6.3 合格判定

各等级评估综合得分达到70分及以上，且同时满足表4规定的条件，方可判定为合格：

表 4　各等级合格判定条件

判定项	初级	中级	高级
综合得分	≥70	≥70	≥70
评估方式单项最低分	各评估方式得分均≥50	各评估方式得分均≥50	场景实操演练≥60、案例分析≥60、成果评审与答辩≥60
能力维度最低分	维度一≥60且维度二≥60	维度一≥60且维度二≥60	维度一≥60 且维度二≥60
否决项	不得触发	不得触发	不得触发

具体条件如下：

a）初级和中级评估中，各评估方式得分均不低于50分，且维度一和维度二得分均不低于60分；

b）高级评估中，各评估方式得分均不低于60分，且维度一和维度二得分均不低于60分；

c）未触发本文件规定的否决项。

评估中出现下列情形之一的，直接判定为不合格：

a）将未脱敏的真实客户涉密信息、未公开案件材料或其他依法、依约应当保密的信息输入评估环境之外的公共人工智能工具，或者将法律业务敏感信息输入未经安全评估、未经授权或不符合机构数据管理要求的公共人工智能工具；

b）对评估任务中明确设置的关键虚构案例、错误法源、失效法规或其他重大事实错误未予识别，直接将其作为法律依据或专业结论使用，并导致任务结论发生实质性错误；

c）违反评估环境管理要求，擅自接入未经许可的工具、账号、数据源或外部人员；

d）存在抄袭、替考、伪造操作记录、篡改评估成果或其他违反评估纪律的行为。

否决项应由至少两名评估人员共同确认，并形成书面记录。评估对象对否决项认定有异议的，可依照9.6提出申诉。

评估结果不合格的申请人可按评估机构事先公开的规则申请补考或重新评估。规则应明确申请间隔、次数、评估范围及成绩保留条件，并对同类申请人一致适用；补考不得降低本文件规定的能力要求和合格条件。

9 评估实施要求

9.1 评估机构

评估机构应具备独立公正性，具备题库与场景库的开发维护能力及安全的评估实施环境，宜配备法律与人工智能复合背景的专业团队。

评估机构应在首次实施或评分规则发生实质变化时，组织专家论证和试评，检查分级任务难度、评分证据、评分一致性及判定结果的合理性，并保留记录。试评发现本文件权重或分数线需要调整的，应形成标准修订建议，不得在正式评估中自行变更表2、附录A及8.6规定的参数。

9.2 评估人员

评估人员应具备法律实务、人工智能应用或相关安全合规专业能力，并经评分培训。主观评分应由至少两名人员独立完成。评估机构应在实施前明确分差复核阈值、汇总方法及复核程序；超过阈值或对合格结论、否决项认定存在分歧的，应组织复核并记录理由。与申请人或成果存在利害关系的人员应回避。

9.3 评估环境

a）场景实操演练应使用经脱敏处理的案例数据、经审核的人工智能工具和受控的评估环境；

b）不应将真实未公开案件材料、未经授权的客户资料、商业秘密或其他法律业务敏感信息用于评估；

c）评估机构应在评估开始前明确允许使用的工具、账号、数据源、网络环境、外部资料和辅助设备；

d）实操过程应留存必要的操作记录、提示词记录、人工智能输出、人工修改内容、最终成果、评分记录和复核记录。涉及屏幕录制的，应事先告知评估对象并采取相应的信息保护措施；

e）相关评估资料保存期限自评估结果形成之日起不少于3年。法律法规、监管要求或评估机构规则规定更长期限的，从其规定；

f）评估机构应对评估资料采取访问控制、加密、脱敏、备份和权限审计等安全措施，评估结束后按照规定期限删除、归档或安全销毁相关数据。

9.4 评估流程

评估流程包括：申请与资格审核、评估实施、评分与复核、结果告知、证书或评估报告出具、档案管理。

申请与资格审核应至少包括以下内容：

a）申请人提交评估申请及评估机构评估规则要求的相关材料，并明确申请等级；

b）评估机构对申请人的身份信息、申请材料、申请等级及其他规定条件进行审核；

c）评估机构向申请人告知评估方式、评估内容、评估时间、评估环境、评分规则、合格判定要求及申诉渠道；

d）对不符合申请条件或申请材料不完整的，评估机构应一次性告知补正要求或不予受理的理由。

各等级申请条件和材料要求应符合7.1，并在申请前公开。评估机构不得临时增加未公开条件，不得要求申请人先取得较低等级证明。

9.5 结果管理

评估机构应根据评估结果出具评估报告。评估报告应包括评估对象基本信息、评估等级、各评估方式得分、综合得分、维度一和维度二得分、各能力要素得分、否决项触发情况、评估日期、评估机构及评估人员信息和能力改进建议。

评估机构可向通过评估的人员出具相应等级的评估证明或证书。有效期、继续教育及复评条件应根据技术变化、岗位风险和评估用途确定并事先公开。发生重大法律法规变化、岗位职责变化或使用风险显著变化时，可按公开规则组织复评。

评估机构应对评估申请材料、过程记录、评分记录、复核记录、申诉处理记录和评估结果进行归档管理。

9.6 监督与申诉

评估机构应建立评估监督、投诉和申诉制度，并公开申诉受理方式、受理范围、申请期限、所需材料、办理流程和处理时限。

评估对象对评估过程、评分结果或否决项认定有异议的，可在结果告知之日起10个工作日内提出申诉，并提交相关事实、理由和证明材料。参与原评估或与申诉事项存在利害关系的人员应当回避。

评估机构应在收到申诉材料后5个工作日内决定是否受理；对予以受理的申诉，应在受理之日起15个工作日内作出处理决定。情况复杂需要延长处理期限的，应告知申诉人，延长期限原则上不超过15个工作日。

申诉处理结果应形成书面记录，并向申诉人反馈。申诉期间，原评估结果原则上继续有效；经复核确认评估结果存在错误的，评估机构应及时更正评估报告或评估证明。

附录 A

（规范性）

能力评估指标体系与权重

表A.1所列能力要素权重适用于三个等级，分级要求见7.5；权重仅用于8.6.2的维度得分计算，各评估方式的综合得分权重见表2。

表 A.1　能力评估指标体系与权重

一级指标（维度）	二级指标（能力要素）	主要评估要点	权重
维度一：AI 基础认知与工具判断能力（40%）	技术基本原理认知	生成机制、模型局限性、幻觉成因与危害、知识时效性、RAG 等概念	12%
	法律 AI 工具甄选与评估	工具分类与适用边界、任务匹配、数据合规与部署方式评估、协议风险识别	10%
	数据安全与伦理合规判断	保密义务、输入前脱敏、AIGC 知识产权风险、执业伦理边界、监管规则	18%
维度二：法律业务场景应用能力（60%）	法律检索与分析	检索策略设计、结果溯源核验、交叉验证、错误识别与纠正	13%
	法律文书生成与审查	提示词工程应用、文书起草质量、纠错精修、合同风险筛查	18%
	案情拆解与逻辑推理辅助	事实梳理、证据链分析、对抗推演、独立判断	12%
	业务流程自动化与效能优化	重复性工作提效、模板与工作流建设、人机任务分配	7%
	法律业务场景中的安全合规与输出校验	脱敏处理、工具使用边界、操作留痕、法源核验、人工复核、结果使用限制	10%

本附录权重用于维度内加权平均，不再对已换算为百分制的能力要素得分重复进行百分制换算。

表2的评估方式权重与本附录的能力要素权重用于不同的汇总计算，不应相加或重复加权。

评分项目的能力要素归属、拆分和证据使用应符合6.1.1。

安全合规知识与工具风险判断计入维度一；业务操作中的数据处理、核验留痕和复核行为计入维度二的安全合规与输出校验要素。其他业务要素评价相应业务成果和分析质量，同一合规表现不得再次计分。

能力维度得分用于各等级的维度最低分判定及结果分析，不计入综合得分。

附录 B

（资料性）

典型场景评估任务示例

本附录为任务和计算示例，其中任务时长、交互轮次和错误数量不构成统一要求。评估机构应依据7.5设置同等级难度相当的任务，并事先明确任务条件。

B.1　合同审查与纠错（中级场景实操示例）

任务描述：提供一份经脱敏处理的采购合同及人工智能初审意见，其中预设3处错误的风险判断和1处虚构或错误援引的法条。要求评估对象在60分钟内完成复核，并出具书面修订意见。

评分要点：对预设错误的识别完整率；事实、法源及合同条款的核验情况；修订意见的专业性、准确性和可操作性；操作过程的合规性。

B.2　类案检索与幻觉识别（案例分析示例）

任务描述：提供某一明确法域、明确法律关系和明确检索截止日期下的类案检索报告，以及与该报告对应的部分法律法规和裁判文书来源。其中人工智能输出含2个虚构案例、1个已失效法规引用及1处裁判观点概括错误。要求评估对象指出问题，说明核验路径、资料来源和判断依据，并给出修正后的分析结论。

评分要点：幻觉及其他事实错误的识别完整性；案例、法条和裁判观点的核验方法；对法律法规效力状态和适用范围的判断；结论的准确性、论证完整性和可复核性。

B.3　起诉状起草与提示词调优（初级场景实操示例）

任务描述：给定经脱敏处理的案情材料、当事人基本信息和任务要求，评估对象应在不超过5轮提示词交互的条件下，指导人工智能生成起诉状初稿，并对事实陈述、诉讼请求、法律依据、当事人信息和文书格式进行人工核对和修正。

评分要点：提示词的清晰性、完整性和结构性；任务要求和事实材料的准确传递；对生成内容的事实核对、法源核验和格式修正质量；是否存在将未经核验内容直接采信的情形；是否存在敏感信息不当输入、超范围使用或其他不合规操作。

B.4　中级评估评分计算示例

以下示例展示一名中级考生在各项评估中的得分情况，说明评分项目得分如何汇总为综合得分和能力维度得分，并完成合格判定。

B.4.1　评分项目及原始得分

表 B.1　评分项目明细

评分项目	评估方式	能力要素	满分	实际得分
1	理论知识测试	技术基本原理认知	10	8
2	理论知识测试	法律AI工具甄选与评估	10	7
3	理论知识测试	数据安全与伦理合规判断	10	9

评分项目	评估方式	能力要素	满分	实际得分
4	场景实操演练	法律检索与分析	20	16
5	场景实操演练	法律文书生成与审查	20	15
6	场景实操演练	案情拆解与逻辑推理辅助	20	14
7	场景实操演练	业务流程自动化与效能优化	10	8
8	场景实操演练	法律业务场景中的安全合规与输出校验	30	24
9	案例分析	法律检索与分析	15	12
10	案例分析	法律业务场景中的安全合规与输出校验	15	12

B.4.2 综合得分

第一步，按评估方式归集，计算各评估方式得分：

理论知识测试得分＝(8+7+9)÷(10+10+10)×100＝80

场景实操演练得分＝(16+15+14+8+24)÷(20+20+20+10+30)×100＝77

案例分析得分＝(12+12)÷(15+15)×100＝80

第二步，按表2中级权重加权求和：

综合得分＝80×20%＋77×60%＋80×20%＝78.2

B.4.3 维度得分

第一步，按能力要素归集，计算各能力要素得分：

表 B.2 维度一能力要素得分

能力要素	评分项目	要素得分
技术基本原理认知	项目1	8÷10×100＝80
法律AI工具甄选与评估	项目2	7÷10×100＝70
数据安全与伦理合规判断	项目3	9÷10×100＝90

第二步，按附录A权重在维度内加权平均：

维度一得分＝(80×12%＋70×10%＋90×18%)÷40%＝82

表 B.3 维度二能力要素得分

能力要素	评分项目	要素得分
法律检索与分析	项目4＋9	(16+12)÷(20+15)×100＝80
法律文书生成与审查	项目5	15÷20×100＝75

能力要素	评分项目	要素得分
案情拆解与逻辑推理辅助	项目6	14÷20×100＝70
业务流程自动化与效能优化	项目7	8÷10×100＝80
法律业务场景中的安全合规与输出校验	项目8＋10	(24+12)÷(30+15)×100＝80

维度二得分＝(80×13%＋75×18%＋70×12%＋80×7%＋80×10%)÷60%＝76.5

B.4.4 合格判定

表 B.4 合格判定汇总

判定项	结果	合格标准	是否满足
综合得分	78.2	≥70	✓
理论知识测试得分	80	≥50	✓
场景实操演练得分	77	≥50	✓
案例分析得分	80	≥50	✓
维度一得分	82	≥60	✓
维度二得分	76.5	≥60	✓
否决项	未触发	不得触发	✓

判定结果：合格。

参考文献

[1] GB/T 41867—2022《信息技术 人工智能 术语》

[2] GB/T 46800—2025《生成式人工智能技术应用社会影响评估指南》

[3] 《中华人民共和国律师法》

[4] 《中华人民共和国个人信息保护法》

[5] 《中华人民共和国数据安全法》

[6] 《律师执业管理办法（司法部令第 134 号）》

[7] 《生成式人工智能服务管理暂行办法》

[8] 《互联网信息服务深度合成管理规定》