ICS XX.XXX
CCS X XX

团　　体　　标　　准

T/SZAS XXX—XXXX

大模型金融应用技术规范

Technical Specification for Large Model in Financial Application

（征求意见稿）

XXXX-XX-XX 发布　　XXXX-XX-XX 实施

深圳市标准化协会　　发 布

目　　次

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。

本文件由深圳国家金融科技测评中心有限公司提出。

本文件由深圳市标准化协会归口。

本文件起草单位：

本文件主要起草人：

大模型金融应用技术规范

1 范围

本文件规定了大模型金融应用的资源池要求、金融优化训练数据要求、金融大模型要求、系统组件要求、服务平台要求、大模型金融应用要求和大模型金融应用处理性能要求。

本文件适用于金融机构开展大模型金融应用活动，也可为金融机构或第三方机构对大模型金融应用评估提供参考。

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 35273 信息安全技术 个人信息安全规范

GB/T 45225 人工智能 深度学习算法评估

GB/T 45288.1 人工智能 大模型 第1部分：通用要求

GB 45438 网络安全技术 人工智能生成合成内容标识方法

GB/T 45654 网络安全技术 生成式人工智能服务安全基本要求

GB/T 45674 网络安全技术 生成式人工智能数据标注安全规范

JR/T 0171 个人金融信息保护技术规范

JR/T 0221 人工智能算法金融应用评价规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

大模型训推资源 large model training and inference resources

大模型训练推理所需的算力资源、存储资源、网络资源等资源。

3.2

金融优化训练数据 financial fine-tuning data

用于优化训练金融大模型，提升金融大模型特定金融应用处理能力的数据。

3.3

生成合成数据 generate or synthesize data

根据数据内容、格式、内在逻辑等要求，通过一定规则生成或合成的数据，作为金融优化训练数据来源之一，用于大模型训练。

3.4

准入控制 access control

通过审查评估等方式，防止不安全或不满足预期的数据、模型、应用等控制对象的使用及部署运行。

3.5

基础模型 foundation model
通过大规模数据预训练而构建的通用模型。

3.6

人工智能体 ai agent
能够感知环境并采取行动以实现特定目标的实体。

3.7

关键词库 keyword database
通过关键词匹配，用于不良违法信息、注入攻击等检测识别。

3.8

安全风险验证库 safety risk validation database
包含已知的注入攻击等漏洞，用于进行攻击识别或开展金融大模型安全风险评估。

4 缩略语

下列缩略语适用于本文件。
API：应用程序接口（Application Programming Interface）
TTFT：首个令牌时间（Time to First Token）
TPS：每秒事务数（Transactions Per Second）
QPS：每秒查询数（Queries Per Second）

5 资源池要求

5.1 大模型训推资源要求

大模型训推资源要求如下：
a) 应保障生产环境大模型训推资源与非生产环境大模型训推资源的逻辑和物理隔离；
b) 应保障大模型训推资源的基本安全，包含但不限于网络安全、主机安全、容器安全；
c) 大模型训推资源应支持弹性与横向扩展能力；
d) 应确保通信带宽满足训推需求，包括但不限于构建RDMA高速互联网络或节点内高速互联网络；
e) 大模型训推资源应符合GB/T 45288.1中5.1的资源池要求。

5.2 大模型训推资源虚拟化要求

大模型训推资源虚拟化要求如下。
a) 大模型训推资源应通过虚拟化方式实现资源的统一管理，包含但不限于调度、回收、监控、运维；
b) 资源虚拟化不应影响生产环境与非生产环境训推资源的逻辑和物理隔离；
c) 虚拟化应支持按照人工智能加速卡数量进行实例规格划分；

d) 虚拟化应支持人工智能加速卡的显存及算力切分；
e) 应保障实例之间的安全隔离，实例应至少为用户级。

5.3 调度策略要求

5.3.1 任务分级机制

任务分级机制要求如下：
a) 宜建立任务分级机制，以优先保障关键大模型金融应用不受影响；
b) 若生产环境大模型训推资源不足且存在两个及以上的大模型金融应用时，宜建立任务分级机制以优先保障关键核心业务不受影响。

5.3.2 资源调度策略

资源调度策略要求如下：
a) 应实现负载均衡等动态资源调度策略；
b) 若建立任务分级机制，应通过优先级配置、资源预留、资源抢占等方式，优先保障关键核心业务可靠性。

6 金融优化训练数据要求

6.1 金融优化训练数据通用要求

金融优化训练数据通用要求如下：
a) 在收集金融优化训练数据用于金融大模型优化训练前，应明确金融优化训练数据的来源、构建方式、数量级以及整体不良信息率等要求；
b) 金融优化训练数据类型应至少包含质量优化训练数据以及对抗性数据，以提高金融大模型对特定应用处理能力和鲁棒性，适用时还应包含用于理解相关安全合规要求、识别金融不良信息的对齐数据；
c) 应明确质量优化训练数据样本要求，以保障质量优化训练数据在内容、数据结构、文件格式、多样性等方面与大模型金融应用处理场景一致；
d) 宜建立统一的不可信优化训练数据来源库，对于优化数据收集阶段、准入评估阶段、使用阶段等过程中发现的不可信数据源进行记录，避免再次采集及使用；
e) 若大模型金融应用处理场景支持结构化类型输入数据，宜优先选择结构化数据类型作为优化训练数据，以提高金融大模型对特定结构化数据的理解及输出内容准确性。

6.2 金融优化训练数据来源要求

6.2.1 开源数据

开源数据作为金融优化训练数据来源的要求如下：
a) 应遵守开源授权协议或相关授权文件；
b) 应优先选择权威性强、行业认可度高的开源数据集；
c) 选择权威性或行业认可度不高的开源数据集时，应通过统计分析、异常检测、人工查验等方式评估并保障所选金融优化训练数据质量。

6.2.2 自采数据

自采数据作为金融优化训练数据来源的要求如下：

a） 应对采集来源、采集方式、采集用途等信息进行记录；

b） 若采集来源为互联网数据，不应采集他人已明确不可采集的信息；

注1：声明不可采集的方式包括但不限于robots协议等。

c） 若采集来源为互联网数据，应优先选择权威性强、行业认可度高的数据来源；

d） 对于数据可信度不高的数据来源应先进行人工审查，若存在较多可疑或不可信数据不应进行采集，不可信数据来源宜纳入到不可信数据来源库中；

e） 若采集来源为金融机构的业务数据，应确保没有引入核心数据、重要数据、敏感数据或对不可泄露数据进行删除、匿名化等预处理操作；

f） 若需对采集后的数据进行标注处理，应符合GB/T 45674的要求；

g） 宜对标注数据的语料来源、标注方式、使用用途等信息进行记录。

6.2.3 生成合成数据

生成合成数据作为金融优化训练数据来源的要求如下：

a） 若大模型金融应用处理的输入数据能够基于规则进行生成合成的，可基于生成合成方式构建金融优化训练数据；

b） 应明确生成合成数据规则，包括字段类型、字符集、随机占比、处理逻辑、约束条件等规则要求；

c） 不应将核心数据、重要数据、敏感数据引入到生成合成数据中。

6.2.4 商业数据

商业数据作为金融优化训练数据来源的要求如下：

a） 采购或合作形式获取第三方数据前，应先要求数据提供方或数据处理方对金融优化训练数据或产品的来源合法合规、质量、安全等方面作出承诺并提供相关证明材料，并对证明材料进行审核，未通过审核的不应使用；

b） 在使用商业数据前，应先签订具有法律效力的交易合同、合作协议等材料；

c） 商业数据的使用不应超过合同、协议的使用范围。

6.3 金融优化训练数据准入控制要求

金融优化训练数据若用于生产环境或对金融业务处理有影响的金融大模型优化训练时，应对金融优化训练数据实施准入控制，准入控制要求如下：

a） 应建立合理的准入控制机制，对金融优化训练数据实施准入控制；

b） 准入控制机制应至少包含金融优化训练数据入库前的数据质量及安全评估，评估内容及要求如下：

1） 评估质量优化训练数据与大模型金融应用在内容、格式、全面性、正负样本分布等方面的一致性；

2） 通过数据统计分析、异常数据检测、人工审查等方式评估优化训练数据集是否存在重复、矛盾、偏见、歧视性、误导性、不良信息等影响金融大模型优化训练效果的有害数据；

3） 评估质量优化训练数据是否存在核心数据、重要数据、敏感数据以及泄露后会造成业务合规风险的数据；

4） 若应用于智能客服等公共服务类场景，应评估是否满足相关监管机构及行业要求。

6.4 金融优化训练数据安全管理及维护要求

入库后的金融优化训练数据安全管理要求如下：

a) 应建立安全有效的金融优化训练数据访问控制机制并遵循最小权限原则；
b) 应对入库后的金融优化训练数据增删改查等操作进行严格管控，避免金融优化训练数据被恶意篡改、删除或泄露；
c) 应定期对金融优化训练数据集进行审查及维护，对大模型金融应用所需的新增术语、概念、逻辑，以及监管要求等数据进行更新并记录，以保持金融优化训练数据良好的适宜性和时效性。

7 金融大模型要求

7.1 基础模型选择要求

基础模型选择要求如下：

a) 宜优先选择已备案或行业认可度高的大模型作为基础模型；
b) 若直接使用基础模型应用于智能客服等公共服务类应用场景，应选择已备案大模型并满足金融行业相关要求；
c) 所选基础模型应具备良好的输入信息理解及指令遵循能力，具备理解输入信息并根据格式约束、语义约束等约束条件进行信息处理的能力；
d) 所选基础模型通过系统提示词优化、知识库配置等应用部署后，准确度不宜低于50%；
e) 若大模型金融应用需要使用广泛的金融知识、数学知识、深度推理能力、复杂任务拆解能力、基于超长上下文的理解等增强能力时，应对基础模型所需的专项能力进行评估或参考权威的评估结果，以确保所选的基础模型具备良好的金融应用处理能力；
f) 若大模型金融应用输入数据不可控，没有人工录入或审核等环节，所选基础模型要求如下：
   1） 应具备良好的鲁棒性，能够识别和处理用户输入中的错别字、语法错误等异常数据；
   2） 具有抵御常见的提示注入等攻击能力。
g) 在基础模型处理能力满足要求前提下应优先选择参数量小的大模型以降低推理成本、提高响应速度，进一步保障业务连续性需求；
h) 宜持续或定期关注可选基础模型，当可选基础模型处理能力有重大提升时，宜选择更适宜的基础模型进行重新优化、开发、部署及应用；
i) 使用开源模型或商业模型作为基础模型前，应进行合规性审查，防止超出许可约定范围或未经授权的情况下使用。

7.2 金融大模型优化要求

金融大模型优化要求如下：

a) 应使用入库后的金融优化训练数据进行金融大模型优化训练；
b) 应关注金融优化训练数据对金融大模型处理效果的影响；
c) 宜在算力资源空闲的时段进行金融大模型优化训练，以降低对大模型金融应用的影响；
d) 宜对金融大模型所使用的金融优化训练数据版本信息进行记录，便于追溯和审计。

7.3 金融大模型准入控制要求

优化训练后的金融大模型若应用于生产环境或对金融业务处理有影响时，应对金融大模型实施准入控制，金融大模型是否满足金融应用准入要求应根据11.1的大模型金融应用准入控制要求进行评估。

7.4 金融大模型安全管理要求

金融大模型安全管理要求如下：

a) 应对准入后的金融大模型文件进行安全存储并建立安全有效的访问控制机制，避免金融大模型原始文件被恶意篡改或泄露；
b) 应通过完整性校验等机制保障金融大模型在部署运行阶段中的完整性和一致性，防止部署运行阶段被恶意篡改或泄露；
c) 宜建立模型卡，对模型基本信息、评估结果、使用的金融优化训练数据、采用的优化策略、金融应用情况等信息进行详细记录。

7.5 金融大模型升级更新要求

金融大模型升级更新要求如下：

a) 当对已准入的金融大模型进行优化训练或修改配置参数等操作时，应重新进行金融大模型准入控制流程；
b) 金融大模型的升级更新及部署不应对金融业务处理造成中断性等影响，若无法避免影响应采用人工接入等方式降低影响。

8 系统组件要求

8.1 开源软件要求

开源数据、开源大模型、开源服务平台、开源知识库、开源标识工具等大模型开发、部署、应用所涉及的开源软件，应满足行业关于开源软件引入、使用、管理等相关要求。

8.2 协议接口要求

协议接口要求如下：

a) 应建立统一的大模型金融应用接口平台，用于进行接口的新增、审核、访问控制及删除，以实现接口的统一管理；
b) 应对接口的使用情况进行监控，若发现异常调用情况应采取安全措施；
c) 若应用于生产环境或对金融业务处理有影响时，不应通过API等方式直接调用外部服务，也不应向外部机构直接提供服务接口；
d) 若采用人工智能体等由模型自主选择并调用接口服务的方式，宜使用MCP等统一接口协议以加强各类大模型对接口的理解和调用能力。

8.3 知识库要求

8.3.1 知识库内容要求

知识库内容要求如下：

a) 应支持数据库作为知识数据源；
b) 应支持结构化文档作为数据源，包含但不限于xls、csv文件；
c) 知识库增添数据应满足最小必要原则，只上传大模型金融应用所需的最少数据集，为保障检索效率及准确性，每个知识库的token数量不宜超过10万；
d) 应定期对知识库内容进行审查及维护，以保持知识库内容的良好适宜性和时效性；
e) 若知识库对应的大模型金融应用不再使用，应及时删除相关知识库；
f) 宜对添加到知识库的数据进行预处理、删除无意义、非必要冗余信息，确保知识内容的精炼，可行时宜通过结构化转换，以提高知识库检索准确率及效率。

8.3.2 知识库构建要求

知识库构建要求如下：

a) 知识库应支持常见文档格式解析，如pdf、txt、doc、docx、xlsx、ppt、pptx、eml、jpg、png、csv、md等；
b) 嵌入模型的选型应考虑其在金融领域知识表征的准确性，所选的嵌入模型应具备良好的语义理解能力，能够有效捕捉金融术语、语义关系及差异性；
c) 应采用语义相关的文本分块策略，并合理设置分块大小与块间重叠比例，以保证信息单元的完整性，提升检索效果和模型理解能力；
d) 对于实体和关系复杂的知识，宜通过知识图谱等技术进行构建，以保障深层次语义关联检索的准确性，提升复杂金融场景下的问答和推理能力。

8.3.3 知识库安全要求

知识库安全要求如下：

a) 应在知识库内容添加前进行内容审核，确保新增内容不包含不良及虚假信息；
b) 知识库应实现知识的精细化隔离及访问控制，确保机构内外部使用者无法通过知识库直接访问或金融大模型间接检索等方式，获得原有渠道无法获得的信息；
c) 应建立知识内容溯源机制，记录每一条知识的来源、创建者、修改历史、审核记录和发布时间，以便于追溯和审计；
d) 在建立知识库的过程中，若涉及个人信息的收集、使用、存储和共享，应符合GB/T 35273、JR/T 0171的相关要求；
e) 若用于智能客服等金融公共服务类场景，知识库中的数据不应包含核心数据、重要数据以及敏感数据。

8.4 标识工具要求

标识工具要求如下：

a) 若用于智能客服等金融公共服务类场景，标识工具应用效果应符合GB/T 45438的要求；
b) 若大模型金融应用的输出内容在后续流程中需进行显示提醒，以便于进行重点审核及验证，应对模型的输出内容进行显示标识，显示标识工具应用效果应符合GB/T 45438中的显示标识要求；
c) 显式标识及隐式标识工具的可靠性、性能应经过验证并满足业务处理场景需求。

8.5 关键词库及安全风险验证库要求

若用于智能客服等金融大模型输入数据不可控的场景，应通过关键词库或安全风险验证库，加强模型输入侧的攻击抵御能力，关键词库及安全风险验证库要求如下：

a) 关键词库应至少涵盖GB/T 45654的附录A内容；
b) 应建立违法不良关键词库及安全风险验证库安全机制，避免相关数据泄露后造成的特定攻击绕过风险；
c) 应对关键词库进行分类管理，类型可分为敏感词、风险提示词、业务引导词、禁用词等；
d) 应对关键词触发后的响应机制进行分级，可根据其风险程度设定不同响应级别，如“拒绝回答”“警示提醒”“转人工服务”等；
e) 应建立关键词库及安全风险验证库更新机制，更新周期不宜超过1个月，在金融政策重大调整等情况下，应在7天内完成更新，更新方式应包括内部来源更新及外部来源更新：

1） 内部来源更新主要是对于未能通过原有关键词库及安全风险验证库等技术检测识别到的非法输入及输出内容，应经过个人信息脱敏处理后纳入安全风险验证库中，若涉及关键词触发的安全风险，还应纳入关键词库中；
2） 外部来源更新主要是根据金融市场、监管要求、可信公开渠道等信息，进行必要的关键词库及安全风险验证库更新。

f） 关键词的定义与匹配规则应明确、精准，避免因泛化或歧义导致的高误报率，影响正常用户交互，宜建立关键词规则的测试与验证流程；
g） 应保障关键词库及安全风险验证库的全面多样性，适用时不同环境、不同大模型金融应用场景可共用并维护一套关键词库及安全风险验证库；
h） 对关键词库的所有增、删、改操作均应记录日志，日志内容应包括操作人、操作时间、变更内容及审批记录，确保所有变更可审计、可追溯。

9 服务平台要求

服务平台要求如下：

a） 大模型服务平台应支持大模型金融应用所需的大模型管理、知识库管理等基本功能、宜支持大模型金融应用管理、API接口管理等辅助功能；
b） 大模型服务平台的使用及管理均应经过身份认证，禁止使用默认或弱密码；
c） 大模型服务平台的关键参数配置需经审核后，由管理人员进行操作；
d） 不应使用默认或非必要端口，防止外部攻击风险；
e） 若基于大模型服务平台进行金融大模型升级更新应符合8.5的金融大模型升级更新要求。

10 大模型金融应用要求

10.1 大模型金融应用准入控制要求

在开发环境中开发的金融大模型应用若部署运行在生产环境或对金融业务处理有影响时，应建立合理的金融大模型应用准入控制机制，准入前应经过金融机构内部或第三方机构评估，评估内容在适用时应至少包括输出内容准确性、输出内容可解释性、金融应用安全性以及金融应用处理性能，评估环境所涉及的金融大模型、知识库、系统提示词、服务平台、验证数据类型等在适用时应与实际金融大模型应用保持一致。

10.1.1 输出内容准确性评估

输出内容准确性评估要求如下：

a） 应构建与金融应用场景数据处理类型及分布一致的验证数据集，以更合理验证大模型金融应用在实际业务场景中的表现；
b） 验证数据集不应从金融优化训练数据集中直接抽取，适用时可在构建金融优化训练数据集时同步构建验证数据集；
c） 所构建的验证数据集应具备完整性、正确性、多样性，应通过人工审查等方式保障验证数据集质量；
d） 应根据大模型金融应用类型使用合理的评价指标，宜通过多个指标进行综合分析以发现金融应用处理缺陷，例如分类任务指标可包括准确率、召回率、精确率、F1分值等；

e） 应根据金融大模型应用类型建立合理的准入基准，例如分类任务中非资金类大模型金融应用输出内容准确率应大于90%，资金类大模型金融应用输出准确率应大于95%；
f） 对于无法通过评价指标进行评价的大模型金融应用，应通过裁判员大模型、人工评价、专家评审等方式确定大模型金融应用是否符合准入要求。

10.1.2 输出内容可解释性评估

当大模型金融应用的输入数据及输出结果之间不存在显然易见的因果关系或难以通过人工审核直接判断时，应建立合理的可解释性准入基准，对金融大模型给出的必要详细解释进行评估，评估维度至少包括解释内容的完整及正确性，金融大模型给出的解释内容适用时应包括：

a） 推理过程描述；
b） 推理及决策依据文件描述；
c） 决策合理性解释；
d） 其他决策的不合理性解释；
e） 决策公平性解释。

10.1.3 金融应用安全性评估

若应用于智能客服等公共服务类场景，应建立合理的金融应用安全准入基准，对大模型金融应用安全进行评估，评估内容应包括：

a） 输出内容不包含核心数据、重要数据以及敏感数据；
b） 适用时输出内容应有显式和隐式标识，标识效果应符合GB/T 45438的要求；
c） 完成生成式人工智能应用或功能登记；
d） 满足其他监管机构及行业要求。

10.1.4 金融应用处理性能评估

应对大模型金融应用处理性能进行评估，大模型金融应用处理性能准入基准应满足 11 的大模型金融应用处理性能要求。

10.2 大模型金融应用安全管理要求

10.2.1 大模型金融应用通用安全管理要求

大模型金融应用通用安全管理要求如下：

a） 应建立有效的大模型金融应用安全管理机制，未经准入控制不应上线对金融业务处理有影响的大模型金融应用或将大模型金融应用部署运行在生产环境；
b） 应对准入后的大模型金融应用系统组件、金融大模型、系统提示词、配置信息等影响大模型金融应用效果的软件、模型、配置信息进行记录、安全存储并建立安全有效的访问控制机制，避免被恶意篡改或泄露；
c） 应通过完整性校验等机制保障大模型金融应用在部署运行阶段中的完整性和一致性，防止部署运行阶段被恶意篡改或泄露；
d） 不应使用默认或非必要端口。

10.2.2 人工智能体金融应用安全管理要求

人工智能体金融应用安全管理要求如下：

a） 不应授权人工智能体访问与应用无关的知识库或业务数据；

b） 不应授权人工智能体调用与应用无关的API或服务；
c） 不应授权人工智能体通过工具调用、接口协议等任何方式进行敏感操作，敏感操作如下：
1） 对金融业务等原始数据或文件进行修改、删除等操作；
2） 对系统、平台、组件、大模型等配置信息进行添加、修改、删除等操作；
3） 执行脚本、软件等可运行程序；
4） 影响金融业务系统或金融业务处理的其他操作。
d） 当人工智能体需要进行敏感操作时，应经过人工审查，确认无误后通过人工或其他可控方式进行操作；
e） 应保障人工智能体能够调用的工具及服务不存在恶意代码或行为。

10.3 大模型金融应用运行要求

10.3.1 输入内容审查

若应用于资金类或高风险大模型金融应用，应对大模型金融应用输入内容质量及安全进行审查，以降低内外部攻击风险，提高输出内容可信度，输入内容审查要求如下：

a） 输入内容质量审查应包括输入内容的完整性及真实性、数据结构的正确性、文件格式的符合性等质量审查点；
b） 输入内容安全审查应包括是否含有注入攻击特征、不良违法信息、篡改痕迹、恶意意图等安全审查点；
c） 为提高审核效率，适用时宜通过注入攻击识别、不良违法信息检测、文件格式校验等技术检测方式辅助判定，但最终的审核结果应交由人工审核及判断并保留审查人员、审查结果等信息；
d） 若输入内容来源自机构内部人员录入等已经过审查的可信渠道，质量审查及安全审查点应在人工录入等上一环节中进行；
e） 人工审查发现质量或安全问题的，应采取人工修改、中断等处理方式并保留相关处理日志，防止恶意攻击或不符合要求的输入内容传输至金融大模型；
f） 在人工审查过程中若发现关键词库或安全风险验证库等技术检测未能识别到的注入攻击、恶意意图等非法输入内容时，应进行记录并定期更新至关键词库或安全风险验证库。

10.3.2 输出内容使用

若大模型金融应用对金融业务处理有影响时，金融大模型输出内容仅能作为输入数据处理的中间参考结果用于提高人工处理效率，最终输出内容或决策应由人工决定，输出内容使用要求如下：

a） 大模型金融应用输出内容作为中间参考结果应重点关注内容质量及安全；
b） 输出内容质量应关注输出内容完整性、数据结构正确性、文件格式符合性等检查点；
c） 输出内容安全应关注是否含有不良违法信息、虚假信息、偏见及歧视性、逻辑错误等检查点；
d） 为提高内容检查效率，适用时宜通过不良违法信息检测、输出内容完整性校验、文件格式校验等技术检测方式辅助检查；
e） 若检查无误后可直接采纳输出内容；
f） 若发现输出内容存在质量或安全问题，应采取人工修改、中断等处理方式，防止输出内容对后续业务流程造成影响；
g） 宜对检查过程中发现的大模型金融应用输出质量及安全问题进行记录，为大模型金融应用优化分析及升级更新提供数据支撑。

10.3.3 输出内容监控

若大模型金融应用对金融业务处理有影响时，应对金融大模型输出内容整体表现进行监控，输出内容监控要求如下：

a) 应对大模型金融应用处理准确率、安全风险等指标进行监控；
b) 应对大模型金融应用处理准确率、安全风险等监控指标设置预警线，若触发预警线，应及时采取相应措施、如对大模型金融应用进行下架、分析、优化处理，并根据11.4的大模型金融应用升级更新要求进行升级更新。

10.3.4 运行稳定性保障

若应用于生产环境或对金融业务处理有影响时，应保障大模型金融应用运行的稳定性，要求如下：

a) 资源池资源应能够保障关键核心金融大模型应用运行稳定性；
b) 当出现大模型金融应用中断时应及时通过拉取等方式进行恢复；
c) 应建立大模型金融应用备选方案，在应用处理无法及时恢复时采取应急措施。

10.4 大模型金融应用升级更新要求

大模型金融应用升级更新要求如下：

a) 应建立大模型金融应用升级更新管控机制，保障在受控情况下进行升级更新；
b) 当大模型金融应用需要进行系统组件升级更新、金融大模型升级更新、系统提示词优化、配置信息修改等影响大模型金融应用效果的软件、模型、配置信息更新时，应重新进行大模型金融应用准入控制；
c) 大模型金融应用升级更新前应做好相关备份，以便在发生异常时能够恢复或回退；
d) 大模型金融应用的升级更新及部署不应对金融业务处理造成中断性等影响，若无法避免影响应采用人工接入等方式降低影响。

11 大模型金融应用处理性能要求

大模型金融应用处理性能要求如下：

a) 应依据 JRT/T 0221和 GB/T 45225中关于性能评价的要求，评测大模型金融应用处理性能，评价指标应至少包括TTFT、吞吐量（QPS/TPS）等；
b) 大模型金融应用处理性能指标应满足金融业务处理需求。

附 录 A
大模型金融应用场景示例

大模型金融应用场景示例见表A.1。

表 A.1 大模型金融应用场景示例

一级维度	二级维度	典型任务	大模型金融应用场景
单模态	文本	文本分类	客户情绪识别、主体识别、交易分类、数据分类分级、规范性文件分类、投资产品分类。
		信息抽取	企业财报信息提取、银行流水信息提取、合同条款提取、合规文件条款提取。
		数学推理	流水分析、基金净值计算、风险价值计算、保费精算定价。
		因果推理	授信审查、反电诈核实、违约预测、异常行为检测、市场波动分析、理赔合理性分析、投资决策因果分析、经营状况分析、合规风险分析。
		任务分解	舆情分析任务拆解、尽调报告撰写任务拆解、理赔调查任务拆解、项目开发任务拆解。
		文本问答	对公问答、零售问答、合规问答、授信问答、理赔进度查询、保单条款解释。
		多轮对话	智能客服、在线助手、投资顾问。
		代码理解	代码辅助开发、逻辑漏洞分析、代码审查、交易系统代码解释、API功能说明、脚本分析。
		长文本理解	合同内容理解、财务报表理解、风险报告解读、监管政策解读、风险评估报告总结。
多模态	图像	静态图像分类	进件材料分类、金融票据分类、营业执照主体分类、抵押品类型分类、事故类型分类。
		静态图像关键信息提取	金融票据信息提取、征信报告信息提取、营业执照信息提取、房产证产权信息提取、医疗证明诊断信息提取、图表数据提取。
		图像推理	财务报表图表推理、投资组合绩效图表分析、市场数据图表分析、风险模型图表分析。
		图文匹配分析	合同图像与录入文本匹配分析、身份证图像与申请表单文本匹配分析、抵押品损坏程度与描述匹配分析。
	音频	声纹识别	身份验证、反洗钱检测。
		音频问答	交互式语音应答（IVR）、客户外呼。
		文音检索	客服通话录音检索、理赔语音检索、投资咨询音频检索。
		音频内容摘要	客服通话内容摘要、理赔通话内容摘要、证券交易通话摘要。
	视频	视频分类	事故视频分类。
		视频异常检测	交易场景异常检测、监控异常检测(如画面静止、遮挡)、网点客户异常行为识别。

参 考 文 献

[1] GB/T 9813.3-2017 计算机通用规范 第3部分：服务器
[2] GB/T 45288.2-2025 人工智能 大模型 第2部分：评测指标与方法
[3] GB/T 45288.3-2025 人工智能 大模型 第3部分：服务能力成熟度评估
[4] JR/T 0197-2020 金融数据安全 数据安全分级指南