

中华人民共和国金融行业标准

JR/T 0357—2026

证券期货业基础大模型选型评估指引

Assessment guidelines of basic large language model selection for securities and
futures industry

2026 - 10 - 26 发布

2026 - 10 - 26 实施

中国证券监督管理委员会

发布

目 次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 需考虑的要素 2

 5.1 总体原则 2

 5.2 评估要素 2

6 评估体系 3

 6.1 合规评估 3

 6.2 数据评估 4

 6.3 算法评估 5

 6.4 模态评估 5

 6.5 安全评估 7

 6.6 运营保障评估 8

附录 A（资料性） 合规义务来源清单 12

附录 B（资料性） 模态支撑度评价指标 13

 B.1 语言任务评价指标 13

 B.2 语音任务评价指标 13

 B.3 视觉任务评价指标 14

 B.4 多模态任务评价指标 14

参考文献 16

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。

本文件由全国金融标准化技术委员会证券分技术委员会（SAC/TC 180/SC4）提出。

本文件由全国金融标准化技术委员会（SAC/TC 180）归口。

本文件起草单位：中国证券监督管理委员会科技监管司、深圳证券交易所、深圳证券通信有限公司、深圳证券信息有限公司、上海金融期货信息技术有限公司、中国国际金融股份有限公司、中国银河证券股份有限公司、国泰君安证券股份有限公司、申万宏源证券有限公司、安信证券股份有限公司、东吴证券股份有限公司、中国中金财富证券有限公司、易方达基金管理有限公司。

本文件主要起草人：卢大彪、刘铁斌、喻华丽、李克坚、曾海泉、王少清、王绍菊、张鹏、毛瑞彬、高剑、程龙、刘永旗、钱维佳、王瑜、李银鹰、唐淑艳、钱磊、刘硕凌。

引 言

大模型技术快速发展，日益受到各行业的高度重视，逐渐成为推进智能化创新应用的重要驱动力。证券期货业（以下简称行业）建设行业大模型需要广泛引入基础大模型，但目前对于基础大模型的能力评估主要针对模型通用能力，缺少与行业特征相结合的综合评估方法。考虑到行业的特殊性，亟需构建一个覆盖业务场景需求，满足行业安全合规要求，具备公信力的基础大模型选型评估体系，有效评估基础大模型对业务场景的支撑能力，推动行业良性发展。面向行业合规要求以及共识性需求，遵循《生成式人工智能服务管理暂行办法》等相关文件，借鉴业界各机构前期研究设计的大模型能力评估方法，经过深入研究和具体场景实践，形成本文件，为行业机构选用基础大模型提供评估指引。

证券期货业基础大模型选型评估指引

1 范围

本文件规定了证券期货业机构在基础大模型选型过程中应考虑的评价要素，包括合规评估、数据评估、算法评估、模态评估、安全评估和运营保障评估六个维度。

本文件适用于指引证券期货业在基础大模型选型过程中的能力评估、验收等工作。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

基础大模型 foundation model

基于海量数据训练得到，具有复杂计算架构，能处理多种任务，且具备一定泛化性的人工智能模型。

4 缩略语

下列缩略语适用于本文件。

A2A: 智能体对智能体协议 (Agent2Agent Protocol)

AI: 人工智能 (Artificial Intelligence)

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

CPU: 中央处理器 (Central Processing Unit)

DDoS: 分布式拒绝服务 (Distributed Denial of Service)

GPU: 图形处理器 (Graphics Processing Unit)

IOU: 交并比 (Intersection over Union)

MCP: 模型上下文协议 (Model Context Protocol)

MOS: 平均主观意见分 (Mean Opinion Score)

PEFT: 参数高效微调 (Parameter-Efficient Fine-Tuning)

RLHF: 基于人类反馈的强化学习 (Reinforcement Learning from Human Feedback)

ROUGE: 以召回率为导向的摘要评价方法 (Recall-Oriented Understudy for Gisting Evaluation)

SFT: 监督微调 (Supervised Fine-Tuning)

SQL: 结构化查询语言 (Structured Query Language)

Web: 万维网服务 (World Wide Web)

XSS: 跨站脚本攻击 (Cross-Site Scripting)

5 需考虑的要素

5.1 总体原则

证券期货经营机构借助大模型从事证券期货基金业务，应当在金融科技项目审查评估的基础上，重点审查评估以下事项：

- 大模型的基本原理、目的意图、主要运行机制、安全保障机制；
- 大模型的安全性、可解释性等情况；
- 大模型的应用范围、系统性能与业务相匹配等情况；
- 大模型应用过程中的投资者保护机制，不得存在欺诈、误导客户、不公平对待等情况；
- 大模型训练数据的合规性、伦理性、隐私性，不得发生未获得客户授权、使用客户数据开展大模型训练等情况。

5.2 评估要素

结合业务场景需求，本文件给出了证券期货经营机构在基础大模型选型过程中应考虑的评价要素，详见表1。

表 1 应考虑的要素表

评估维度	评估项	评估内容
合规评估	合规目标	合规目标
	合规义务	合规义务
	合规风险管理	风险识别
		风险分析
		风险处理
数据评估	数据准确性	真实性
		一致性
	数据多样性	数据分布
		数据模态
	数据合规性	来源合规性
		内容合规性
		过程可审计
	数据全面性	完整性
		时效性
算法评估	算法准确性	合规性
		精准性
	算法可靠性	稳定性
		安全性
	算法易用性	角色适配
	算法可解释性	可解释性

表 1 应考虑的要素表（续）

评估维度	评估项	评估内容
模态评估	语言任务	知识提取
		知识问答
		内容生成
		数据挖掘
	语音任务	语音识别
		语音对话
		语音认证
		语音质检
	视觉任务	图像分类与识别
		身份认证与安全
		视觉真实性检测
	多模态任务	语音文字互转
		图片文字互转
		跨模态生成
		多模态识别
安全评估	个人隐私保护	个人隐私数据采集
		个人隐私数据处理
		个人隐私数据展示
	数据加密	数据存储加密
		数据传输加密
		训练数据加密
		推理数据加密
	攻击防范	外部访问攻击防范
		对话攻击防范
运营保障评估	技术支持	模型开发
		模型部署
		模型核查
		模型扩展
		兼容适配
	应急管理	组织体系
		应急预案
		容错机制
	风险控制	风险管理
		追溯审计
	问题响应	响应机制
		响应能力要求

6 评估体系

6.1 合规评估

6.1.1 合规目标

基础大模型应制定满足行业要求的合规目标，以：

- 确保模型能够实现预期能力；
- 确保履行证券期货业合规要求；
- 实现合规管理的改进和优化。

6.1.2 合规义务

基础大模型应明确所遵循的合规义务，评估其对运行所产生的影响，并建立过程以：

- 识别新增的合规义务，以及评估对模型能力的影响；

- b) 根据变更的合规义务，及时更新合规管理过程。
合规义务来源清单见附录A。

6.1.3 合规风险管理

6.1.3.1 风险识别

基础大模型应具备识别合规风险的能力，能够通过分析模型架构、业务规则，确定合规风险产生的场景来源。

6.1.3.2 风险分析

基础大模型应具备分析已识别的合规风险的能力，能够分析合规风险发生的可能性和结果。风险分析包括但不限于以下：

- a) 基于历史数据、合规要求和专业判断估计合规风险的性质、发生概率；
- b) 评价合规风险对证券期货交易、投资者教育等证券期货业业务场景的潜在影响。

6.1.3.3 风险处理

基础大模型应具备控制合规风险的能力，能够按照风险分析结果，及时采取有效的控制措施。可采取措施包括但不限于以下：

- a) 制定合规目标，明确合规义务；
- b) 设置敏感信息防火墙；
- c) 建立合规语料库、知识库。

6.2 数据评估

6.2.1 数据准确性

6.2.1.1 真实性

基础大模型宜确保数据的准确性，尽可能消除常识性、事实性或逻辑性等错误。

6.2.1.2 一致性

基础大模型宜确保同一数据源不同部分之间和不同数据源之间的数据一致性，尽可能消除冲突或矛盾。

6.2.2 数据多样性

6.2.2.1 数据分布

基础大模型的预训练、微调数据宜涵盖一定比例混合的通用数据和行业数据，数据分布宜具有多主题、多语言等特性，行业数据宜包括证券期货业经营和管理过程中产生、采集、加工或管理的各类数据。

6.2.2.2 数据模态

基础大模型宜支持多种模态数据的采集和汇聚，训练、微调数据宜支持文本、图像、音频、视频等多种数据类型；宜支持文本、图像、音频、视频等多种数据类型的格式标准化，便于后续处理和分析。

6.2.3 数据合规性

6.2.3.1 来源合规性

基础大模型使用数据应保证数据来源具有合法使用权或必要授权，并对数据源进行合规审查，避免引入侵权、非法或存在严重偏见的数据源。

6.2.3.2 内容合规性

基础大模型训练数据的内容应符合相关法律法规与监管要求。数据处理者应建立数据过滤机制，剔除违法不良信息，并采用匿名化等技术保护个人隐私与敏感数据。

6.2.3.3 过程可审计

应建立数据治理与审计跟踪机制，对数据从来源到训练使用的全过程进行记录，以确保数据处理活动的合规性可被追溯和验证。

6.2.4 数据全面性

6.2.4.1 完整性

宜支持同一数据各组成部分的完整性，消除缺失值或不完整的记录。

6.2.4.2 时效性

宜支持定期更新模型训练数据或推理知识库，能够反映最新的情况，特别是政策变化、安全事件发生等。

6.3 算法评估

6.3.1 算法准确性

6.3.1.1 合规性

在处理数据和生成回答时应严格遵循证券期货行业相关法律法规，确保输出内容符合行业监管要求。

6.3.1.2 精准性

回答事实类内容应与证券期货行业官方文件表述、官方发布数据保持一致，逻辑推理类需基于市场事实且符合理论推导和行业逻辑，避免模糊或歧义表述。

6.3.2 算法可靠性

6.3.2.1 稳定性

在不同时间、不同交互路径下对同一证券期货行业问题的输出内容宜保持表述逻辑一致，前后无自相矛盾，避免因数据噪声或输入微小变动导致回答核心逻辑偏移。

6.3.2.2 安全性

应支持输出安全可控内容，过滤有害内容、敏感信息等，能够防御诱导性、错误性对话攻击等。

6.3.3 算法易用性

宜根据投资者、分析师、监管者等不同用户角色提供适配性表述，支持文本、图表、语音等多种形式的输出，适配不同证券期货行业业务场景，提供自定义数据看板等功能。

6.3.4 算法可解释性

应清晰表述预测内容的事实依据、理论依据、监管意图以及逻辑推导思维链等内容，并标注其引用的信息来源。

6.4 模态评估

6.4.1 语言任务

6.4.1.1 知识提取

宜评估基础大模型知识提取能力，见附录B.1，知识提取能力包括：

- a) 对时间、地点、人名等通用领域实体，机构名称、证券期货产品名等证券期货行业专有实体和业务专有实体的识别；
- b) 从文本中自动抽取与证券期货相关的特定事件，比如违规交易行为、股东减持动作、产品存在的问题以及任何违法违规的情况，并且能够准确地提取出这些事件的相关主体，如股东、高级管理人员、公司或机构；

- c) 从证券期货相关的文章、研究报告及评论中,提取出作者或评论者的观点和态度;
- d) 具备情感挖掘能力,可以从证券期货相关的交流对话或文章中,识别和提取出表达的情感倾向,如正面、负面或中立;
- e) 识别并抽取证券期货领域专用实体及通用领域实体之间存在的关系,如公司之间的合作关系、高管与公司的关联等。

6.4.1.2 知识问答

宜评估基础大模型知识问答能力,知识问答能力包括:

- a) 自动响应客户关于账户信息、交易历史、市场指标等常规查询;
- b) 理解用户的询问意图,提供专业的证券期货咨询,包括投资组合分析、市场趋势预测等;
- c) 提供证券期货基础教育和理财指导,如证券期货概念解释、理财知识介绍等;
- d) 根据客户的风险承受能力、投资期限和期望回报,在合规的情况下,推荐合适的投资产品。

6.4.1.3 内容生成

宜评估基础大模型内容生成能力,内容生成能力包括:

- a) 根据特定产品的特性及目标消费群体,设计和生成吸引人的营销文案,提升市场营销活动的效果;
- b) 生成如市场分析、财务研报、投资建议、合规文件等各类证券期货文档;
- c) 快速生成证券期货相关文章和研究报告的摘要,帮助投资者了解市场动态和实时新闻;
- d) 具备多语言翻译能力,以便于投资者获取了解国际市场的信息 and 研究资料;
- e) 将自然语言转化为 SQL 查询语句,帮助用户简化证券期货数据查询过程。

6.4.1.4 数据挖掘

宜评估基础大模型数据挖掘能力,数据挖掘能力包括:

- a) 辅助进行复杂的行业数据分析,如对利率变化、股票价格波动的分析;
- b) 利用历史交易数据、证券期货市场数据和其他相关信息,通过数据分析和模式识别,为高频自动化交易提供决策支持;
- c) 基于现有的行业信息,对未来的发展趋势做出预测,包括但不限于事件的发展方向、资产价值的变化、市场走向等;
- d) 广泛汇集多元化分析素材,有效整合信息,构建合理的因果逻辑链。

6.4.2 语音任务

6.4.2.1 语音识别

宜评估基础大模型语音识别能力,见附录B.2,包括:

- a) 从语音数据中分析出说话人的情感状态,如积极、负面等,以辅助理解用户需求,提供个性化服务体验;
- b) 具备多语言转换能力,能够将不同语言、不同口音的语音内容准确翻译成目标语言。

6.4.2.2 语音对话

宜评估基础大模型语音对话能力,包括:

- a) 通过语音交互理解用户的问题并提供准确的回答,提高客户服务效率;
- b) 针对潜在客户进行产品宣传外呼,并搜集他们对产品或服务的反馈。

6.4.2.3 语音认证

宜评估基础大模型语音认证能力,包括:

- a) 用户通过声纹识别进行移动端登录、AI 语音交互操作等;
- b) 基于录制客户声纹并与已有数据进行比对,实现身份验证;
- c) 通过语音指令实现付款流程,方便客户完成支付操作。

6.4.2.4 语音质检

宜评估基础大模型语音质检能力，包括：

- a) 识别和转录证券期货产品销售过程中的对话内容，作为双录音频质检的基础；
- b) 自动检测和分析客户通话内容，挖掘客户需求，实现业务量转化。

6.4.3 视觉任务

6.4.3.1 图像分类与识别

宜评估基础大模型图像分类与识别能力，见附录B.3，图像分类与识别能力包括：

- a) 对多种证件、照片的分类判别等任务；
- b) 手写文字识别、证照识别、单据识别、合同识别等任务。

6.4.3.2 身份认证与安全

宜评估基础大模型身份认证与安全能力，包括：

- a) 客户面部识别、有遮挡面部识别、人脸数据采集等任务；
- b) 基于指纹对证券期货核心业务系统、电子签章系统等系统进行授权管理，以及身份验证、支付等；
- c) 基于用户眼睛中虹膜进行身份识别。

6.4.3.3 视觉真实性检测

宜评估基础大模型视觉真实性检测能力，包括：

- a) 判断用户上传图像(身份证照片、保单等)是否经过后期加工或篡改；
- b) 通过眨眼、张嘴、摇头、点头等动作，验证用户是否为真实活体本人操作。

6.4.4 多模态任务

宜评估基础大模型对多模态任务的支持度及应用效果，根据具体场景要求推荐选择一个或多个核心评估维度进行评价，见附录B.4，其基本任务包括：

- a) 语音文字互转：宜支持通用名词、行业专有名词、行业专业术语等转换，满足智能客服、电话录音记录、会议纪要等场景实时性、精确性要求；
- b) 图片文字互转：宜支持以文搜图、以图搜文等；
- c) 跨模态生成：宜支持根据提示词生成图片、绘制图表等；
- d) 多模态识别：宜支持将包含结构化文本的图片数据转换为结构化数据。

6.5 安全评估

6.5.1 个人隐私保护

6.5.1.1 个人隐私数据采集

宜支持白名单、签名验签等方式，保证数据采集认证、访问控制安全。

6.5.1.2 个人隐私数据处理

应支持数据处理的全过程安全防护，包括身份认证、访问控制、日志与安全审计、匿名化处理、去标识化处理等。

6.5.1.3 个人隐私数据展示

应支持展示过程中对个人隐私数据关键部分的隐藏或截词，并具备相应机制避免以展示信息倒推出原始真实信息的行为。

6.5.2 数据加密

6.5.2.1 数据存储加密

宜支持采用在线加密、原生加密等措施，确保数据在存储过程中的保密性和安全性。

6.5.2.2 数据传输加密

宜支持通过加密通道或数据加密的方式进行数据传输，利用密码学技术、入侵检测等手段防止数据在传输过程中被截获、篡改、伪造或窃取，宜定期审查并更新加密策略和技术，以应对新的安全威胁。

6.5.2.3 训练数据加密

宜支持在原始数据收集、数据预处理、数据向量化等数据处理阶段对数据进行加密，确保训练过程中的数据安全。

6.5.2.4 推理数据加密

宜支持在数据交互、模型优化、推理预测等模型推理过程中对数据进行加密，确保推理结果的安全性和隐私性。

6.5.3 攻击防范

6.5.3.1 外部访问攻击防范

宜评估基础大模型外部访问攻击防范能力，包括：

- a) 访问控制：应支持基于授权用户、设备标识和白名单等访问控制手段。
- b) 安全策略：应支持采用多因素认证、缺省拒绝访问、验证码错误次数限制等安全策略和技术。
- c) 防御机制：应支持采用防火墙、入侵检测系统、Web 应用防火墙等，防御 DDoS、SQL 注入、XSS 等常见攻击。
- d) 日志监控：应支持实时监控系统活动并记录详细日志，方便异常检测、预警和后期审计。

6.5.3.2 对话攻击防范

宜评估基础大模型对话攻击防范能力，包括：

- a) 商业机密：对于诱导输出公司财务数据、管理数据及运营数据等未经公开披露的商业机密数据及内容，应具有防范机制。
- b) 隐私数据：对于输出非用户本人的其他客户的个人信息、行为数据、交易数据、标签数据、客服记录数据等隐私数据的问题，应具有防范机制。
- c) 业务合规：对于用户提问过程中不符合法律、行政法规和相关办法规定的内容，应具有防范机制。
- d) 伦理合规：对于鼓动内幕交易、虚假信息扰乱资本市场的行为，应具有预警和防范机制。如建立包含可能产生扰乱资本市场影响的安全事件库，识别内容并向用户提出预警的机制。
- e) 模型窥探防护：对于通过特定输入获取模型敏感信息或针对模型的逆向工程攻击，应具有防范和预警机制。
- f) 对抗性攻击防御：对于通过对抗性输入诱导模型输出错误信息的攻击，应具有识别并拒绝的防范机制。

6.6 运营保障评估

6.6.1 技术支持

6.6.1.1 模型开发

宜评估基础大模型研发、训练和调优等能力。

- a) 模型研发：
 - 架构能力：宜具备模型架构创新与优化能力，如自研架构或优化现有架构；
 - 算法能力：宜支持与大模型应用相关的AI算法的研究开发，如嵌入模型、重排序模型等。
- b) 模型训练：
 - 并行训练能力：宜支持张量并行、数据并行、流水线并行等混合并行训练模式；
 - 工程化能力：宜具备训练可视化、分布式训练、资源优化与调度和断点续训等能力。
- c) 模型调优：
 - 精调能力：宜支持提示词优化和SFT、PEFT、RLHF等精调方式；
 - 优化能力：宜具备模型蒸馏、量化压缩等模型优化能力。

6.6.1.2 模型部署

宜评估基础大模型私有化部署能力。

- a) 私有化部署支持程度：
 - 模型私有化部署：宜支持模型部署在企业私有化环境中，确保模型在企业内部运行；
 - 平台私有化部署：宜支持模型训练和部署过程均在企业私有化环境中进行，确保整个过程的数据和模型完全在企业内部管理和控制。
- b) 部署方式：宜支持包括但不限于物理机、虚拟机、容器、边缘终端等，以适应不同应用场景的需求。
- c) 模块化设计：模型和应用宜采用模块化架构，便于部署、升级、替换或扩展新功能。
- d) 保密规范：私有化部署的整个过程宜在企业内部环境进行，所有操作宜遵循企业内部的安全政策和合规要求，防止数据泄露和未授权访问。

6.6.1.3 模型核查

宜评估基础大模型输出信息的准确性核查机制及核查手段。

- a) 宜建立相应核查流程或机制，根据实际应用场景需要支持核查上下文一致、客观事实、行业基础知识、行业实体及关系、行业事件、行业逻辑推理、合法合规、隐私保护等方面内容。
- b) 核查方式：
 - 人工：专家或分析师对模型的输出进行定期审查，以确保准确性和可靠性；
 - 自动：使用其他经过验证的数据源或模型，对基础大模型的输出进行验证；
 - 半自动：结合自动核查的结果，再由人工专家进行复核。
- c) 一致性核查：基础大模型的预测和建议对输入内容的小变化宜具有连续性和一致性，并能在各种不同的场景条件下保持稳定。
- d) 实时反馈机制：宜为用户提供反馈渠道以评价模型的输出结果，帮助及时优化模型，确保为用户提供准确可靠回复。
- e) 异常纠正：模型输出核查中发现的任何异常输出，宜进行研究分析，确定来源并采取相应优化措施。

6.6.1.4 模型扩展

宜评估基础大模型系统的可扩展性，具体内容包括：

- a) 模型可扩展：宜支持根据用户需求扩展与之匹配的模型，可结合业务输出数据对基础大模型的能力进行扩展。
- b) 应用可扩展：
 - 系统设计：系统设计宜考虑高并发和大数据量，支持水平或垂直扩展以支撑大量用户和交易；
 - 系统数据：数据存储解决方案宜具备高可用性和高性能，能够支撑日益增长的数据需求；数据处理流程宜灵活，可根据业务需求进行调整和扩展；
 - 组件能力：系统宜采用模块化或组件化的功能设计，便于升级、替换或扩展新功能。
- c) 性能与负载：
 - 动态调整：系统宜能够根据负载动态调整资源，确保在高负载情况下仍能保持稳定的性能；
 - 负载均衡：系统宜采用负载均衡技术，确保流量在各个服务器或服务之间得到均衡分配。

6.6.1.5 兼容适配

宜评估基础大模型部署应用过程中与基础设施、网络环境、操作系统、原有业务系统以及协议之间的兼容适配性。

- a) 基础设施兼容性：证券期货业大模型系统全软件栈宜支持适配主流的国产信创 CPU/GPU 算力硬件、网络设备等基础设施，并满足模型训练、微调和推理部署等任务的功能及性能要求。
- b) 网络环境兼容性：为确保行业数据隐私，证券期货业大模型宜支持在企业内网、私有等不同网络环境下的适配部署。

- c) 操作系统兼容性：宜支持适配 Linux、Windows 等主流操作系统，特别是国产信创操作系统，支持在服务器、终端设备上适配部署。
- d) 业务系统兼容性：
 - 业务系统集成：证券期货业大模型宜提供灵活稳定的API或其他集成工具，支持与原有证券期货业务系统的集成；
 - 数据交互性：评估证券期货业大模型的数据输入输出格式，宜支持与原有业务系统中的数据格式兼容适配；
 - 响应速度与延迟：评估模型在业务系统中的响应速度，宜确保满足实时或近实时的业务需求。
- e) 协议兼容性：
 - 宜支持MCP协议，以实现模型与其他支持MCP的系统或组件之间的高效互通与协同工作；
 - 宜支持A2A协议，以便智能体之间能够进行有效通信和协作。

6.6.2 应急管理

6.6.2.1 组织体系

宜建立健全组织体系，包括：

- a) 宜建立健全基础大模型安全事件应急处置组织体系，明确安全事件的应急指挥决策组织和执行组织，负责安全事件的预防预警、应急处置、报告等工作。
- b) 基础大模型安全事件应急处置指挥决策组织宜由主要领导负责，成员包括但不限于业务、技术、风险控制、客服等有关部门的负责人。
- c) 宜明确安全事件应急决策机制，以及决策递补顺序。

6.6.2.2 应急预案

宜建立完善应急预案，包括：

- a) 宜制定安全事件应急预案，内容包括：
 - 应急预案编制的目的和依据；
 - 应急预案的适用范围；
 - 应急处置的组织体系及职责；
 - 安全事件的预防措施、保障措施与应急准备；
 - 安全事件的分级分类、具体处置方案。
- b) 宜建立完善的应急演练机制并定期开展应急演练、应急培训等。

6.6.2.3 容错机制

宜建立健全容错机制，包括：

- a) 故障切换：系统宜具备在某些组件或服务出现故障时，能快速切换到备用资源的能力。
- b) 数据备份与恢复：验证系统的数据备份策略和恢复机制，宜确保在意外情况下能迅速恢复数据。

6.6.3 风险控制

6.6.3.1 风险管理

宜评估基础大模型应用过程中是否建立完善的模型风险管理体系，具体包括：

- a) 政策和流程层：宜具备基础大模型风险管理的方针政策，制定包括模型设计、开发、部署、使用、维护和退役在内的完整生命周期管理流程和规范，定期评估并修订模型风险管理政策和生命周期管理规范，确保紧跟大模型业务与技术发展趋势，遵循国家和行业相关标准或规定；
- b) 分析和验证层：针对基础大模型应用过程中涉及的模型验证、模型部署、模型投产等环节，宜具备模型验证或评估体系，为模型健康、可持续运作提供有力保证；
- c) 系统层：宜具备模型全生命周期管理平台，包括版本控制、部署、监控和维护等功能。宜具备模型风险监控平台，实时监测基础大模型系统可能的风险因素，并及时发出预警。宜具备模型资产管理平台，对企业内所有的模型资产进行统一的管理，确保模型的有效利用和知识共享。

6.6.3.2 追溯审计

宜评估基础大模型应用过程中是否保留可追溯的记录，具体包括：

- a) 数据可追溯：宜支持对基础大模型训练、微调中涉及的训练数据获取时间、来源、数量、采样方法以及负责数据处理的个人或团队等信息进行记录；
- b) 训练可追溯：宜支持对训练时间、软硬件环境配置、数据版本、模型权重版本、训练脚本、模型训练参数配置、模型迭代次数、模型训练结果、操作者等信息进行记录；
- c) 部署可追溯：宜支持对模型部署时间、软硬件环境配置、模型权重版本、操作者、模型服务端口或 API 等信息进行记录；
- d) 使用可追溯：宜支持对模型使用者、模型调用时间、模型服务 API、请求上下文信息、模型回答、性能指标等信息进行记录；
- e) 决策可追溯：宜支持当模型提供预测或建议时，能解释其决策的参考信息、预测模型和思维逻辑；
- f) 迭代可追溯：宜支持对模型优化和模型服务升级迭代相关信息进行记录。

6.6.4 问题响应

6.6.4.1 响应机制

宜符合金融业监管要求的风险事件分级响应体系，确保模型故障与合规缺陷对金融业务的冲击可控、可回溯。

6.6.4.2 响应能力要求

模型宜支持的响应能力要求：

- a) 可审计日志：宜详尽记录关键操作日志，包括但不限于操作时间、操作者、来源、操作行为、操作内容及结果等，日志格式宜统一，便于自动化解析和报告。日志的保存时间应不小于行业监管、企业内部风险管控等所要求的时限，支持根据业务要求的变化进行调整；
- b) 应用监控：宜提供全面的应用监控工具，能够实时监测模型的运行状况，包括 CPU/GPU 利用率、内存占用、网络流量等，以及模型服务的健康状况，如响应时间和错误率；
- c) 告警策略：宜依据业务场景任务要求，结合应用监控指标和日志制定相应的告警规则；
- d) 备份策略：宜制定详尽的备份计划，确保在线数据的完整性和可用性。备份内容包括但不限于模型数据、配置文件、日志等关键信息，以备在发生数据损坏或丢失时进行恢复。备份方式包括但不限于人工、自动化，减少人工干预；
- e) 归档策略：宜设计完备的归档策略，将联机数据脱机存储，用于审计、查询等需要；
- f) 数据清理：宜明确数据清理标准、时间周期，并设计、开发必要的功能，用于数据清理。数据清理前，需要进行备份或者归档；
- g) 应急恢复：宜考虑各种故障状态的情况，设计系统应急恢复机制，或开发必要的工具，确保在应急状态下的系统快速恢复；
- h) 代码和配置管理：宜利用分布式版本控制系统等工具对系统代码和配置进行版本管理；
- i) 持续集成部署：基础大模型应用迭代更新后宜支持自动化测试，确保测试通过后，支持自动化部署，确保部署的一致性和可重复性。

附 录 A
(资料性)
合规义务来源清单

表A. 1给出了证券期货业在基础大模型选型评估过程中，基础大模型应遵循的合规义务主要来源清单。

表 A. 1 基础大模型选型评估合规义务来源清单

序号	文件名称	发布单位
1	《中华人民共和国网络安全法》	中华人民共和国全国人民代表大会常务委员会
2	《中华人民共和国数据安全法》	中华人民共和国全国人民代表大会常务委员会
3	《中华人民共和国个人信息保护法》	中华人民共和国全国人民代表大会常务委员会
4	《中华人民共和国科学技术进步法》	中华人民共和国全国人民代表大会常务委员会
5	《生成式人工智能服务管理暂行办法》	国家互联网信息办公室、中华人民共和国国家发展和改革委员会、中华人民共和国教育部、中华人民共和国科学技术部、中华人民共和国工业和信息化部、中华人民共和国公安部、国家广播电视总局
6	《证券期货业网络和信息安全管理暂行办法》	中国证券监督管理委员会
7	《证券期货业网络安全事件报告与调查处理办法》	中国证券监督管理委员会
8	《人工智能生成合成内容标识办法》	国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局
9	《证券投资基金经营机构信息技术管理办法》	中国证券监督管理委员会

附录 B (资料性) 模态支撑度评价指标

B.1 语言任务评价指标

B.1.1 客观指标

基础大模型语言任务能力的客观评价，可参考以下推荐指标：

1) 精确率：

$$P_H = \frac{H_1}{H} \times 100\%$$

式中：

P_H ——精确率；

H_1 ——正样本中预测正确的数量；

H ——预测结果是正样本的数量。

2) 召回率：

$$R_E = \frac{E_1}{E_2} \times 100\%$$

式中：

R_E ——召回率；

E_1 ——正样本中预测正确的数量；

E_2 ——正样本的数量。

3) F1 值：

$$F_H = \frac{2 \times P_H \times R_H}{P_H + R_H} \times 100\%$$

式中：

F_H ——F1 值；

P_H ——精确率；

R_H ——召回率。

4) ROUGE-N：

$$ROUGE - N = \frac{\sum_{S \in (\text{Reference Summaries})} \sum_{gram_n \in S} \text{Count}_{\text{match}}(gram_n)}{\sum_{S \in (\text{Reference Summaries})} \sum_{gram_n \in S} \text{Count}(gram_n)} \times 100\%$$

式中：

N ——即 n -gram，文本内容滑动窗口字符数，参考值为 2；

$\text{Count}_{\text{match}}(gram_n)$ ——参考摘要和预测摘要中共有的 n -gram 的数量；

$\text{Count}(gram_n)$ ——参考摘要中 n -gram 的数量。

B.1.2 主观指标

基础大模型语言任务能力的主观评价，可参考以下推荐指标：

- 1) 逻辑性：模型输出内容宜自然流畅、前后逻辑一致，段落和句子之间有序衔接，有良好的语义结构，没有明显的意思断层、跳跃或矛盾。
- 2) 多样性：模型输出内容宜涵盖多种观点或多维度信息，支持采用多种表达方式，能够根据不同的情境和受众调整行文风格。
- 3) 可靠性：模型输出内容宜确保没有常识性、事实性、逻辑性错误，能够有效表达预期信息，且对同一问题模型多次输出的内容意思宜保持一致。

B.2 语音任务评价指标

B.2.1 客观指标

基础大模型语音任务能力的客观评价，可参考以下推荐指标：

1) 错误接受率：

$$FAR = \frac{R_2}{R} \times 100\%$$

式中：

FAR——错误接受率；

R_2 ——被系统接受的冒充者样本数；

R——冒充者的样本数量。

2) 错误拒绝率：

$$FRR = \frac{A_2}{A} \times 100\%$$

式中：

FRR——错误拒绝率；

A_2 ——被系统拒绝的真实者样本数量；

A——真实者的样本数量。

3) 句识别准确率：

$$SCR = \frac{H}{N} \times 100\%$$

式中：

SCR——句识别准确率；

H——识别正确无误的句子数；

N——总句数。

B.2.2 主观指标

基础大模型语音任务能力的主观评价，可参考以下推荐指标：

语音 MOS 评分：评估基础大模型对多音字、数字、符号、声调发声等的合成能力，基于人工评测的方式，即用户对语音样本进行听觉评估，给出一个 1 到 5 的评分。

B.3 视觉任务评价指标

基础大模型视觉任务能力的客观评价，可参考以下推荐指标：

1) 精确率：参见附录 B.1.1 语言任务的参考评价指标中精确率的定义。

2) 召回率：参见附录 B.1.1 语言任务的参考评价指标中召回率的定义。

3) F1 值：参见附录 B.1.1 语言任务的参考评价指标中 F1 值的定义。

4) 平均精度 ($mAP_{IOU=0.75}$)：

$$mAP_{IOU=0.75} = \frac{1}{k} \sum_{i=0}^k AP_i$$

式中：

$mAP_{IOU=0.75}$ ——所有目标检测类别在 $IOU = 0.75$ 时的平均精度，其中 IOU 是预测掩模与真实标注掩模之间的交并比；

AP_i ——目标检测第 i 个类别的平均精度；

k——目标检测类别的数量。

B.4 多模态任务评价指标

B.4.1 客观指标

基础大模型多模态任务能力的客观评价，可参考以下推荐指标：

$$Recall@K = \frac{1}{N} \sum_{i=1}^N \frac{|Top_K(i) \cap Relevant(i)|}{|Relevant(i)|}$$

式中：

$Recall@K$ ——前 K 召回率；

N ——查询样本的数量；

$Top_K(i)$ ——与第 i 个查询样本相关性最高的前 K 个预测样本；

$Relevant(i)$ ——第 i 个查询样本的相关样本。

B.4.2 主观指标

基础大模型多模态任务能力的主观评价，可参考以下推荐指标：

文生图 MOS 评分：评估基础大模型生成图像的质量，基于用户对图像样本的视觉评估，给出一个 1 到 5 的评分。

参 考 文 献

- [1] JR/T 0099—2012 证券期货业信息系统运维管理规范
 - [2] JR/T 0250—2022 证券期货业数据安全管理与保护指引
 - [3] JR/T 0295—2023 证券期货业信息安全运营管理指南
 - [4] JR/T 0221—2021 人工智能算法金融应用评价规范
 - [5] T/CCSA 561.1—2024 面向行业的大规模预训练模型通用要求 第1部分：金融
 - [6] BZKT-2024-054 证券期货业大模型评估标准研究
 - [7] 大语言模型（LLM）安全性测评基准 v1.0[EB/OL]. OWASP 中国. <http://www.owasp.org.cn/OWASP-CHINA/owasp-project/LLM/>（2026-04-07）
 - [8] Zhao W X, Zhou K, Li J, et al. A Survey of Large Language Models[J]. 2023
-