

JR

中华人民共和国金融行业标准

JR/T XXXXX—XXXX

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

（征求意见稿）

XXXX—XX—XX 发布

XXXX—XX—XX 实施

中国证券监督管理委员会

发布

目 次

前言 III

引言 IV

1 范围 0

2 规范性引用文件 0

3 术语和定义 0

4 缩略语 0

5 需考虑的要素 0

6 评估体系 1

6.1 合规评估 1

6.1.1 合规目标 1

6.1.2 合规义务 2

6.1.3 合规风险管理 2

6.2 数据评估 2

6.2.1 数据准确性 2

6.2.2 数据多样性 2

6.2.3 数据全面性 3

6.3 算法评估 3

6.3.1 算法准确性 3

6.3.2 算法可靠性 3

6.3.3 算法易用性 3

6.3.4 算法可解释性 3

6.4 模态评估 3

6.4.1 语言任务 3

6.4.2 语音任务 4

6.4.3 视觉任务 5

6.4.4 多模态任务 5

6.5 安全评估 5

6.5.1 个人隐私保护 5

6.5.2 数据安全 6

6.5.3 风险防范 6

6.6 运营保障评估 6

6.6.1 技术支持 7

6.6.2 应急管理 8

6.6.3 风险控制 9

6.6.4 问题响应 9

附录 A（资料性） 合规义务来源清单 11

附录 B（资料性） 模态支撑度评价指标 12

参考文献.....	15
-----------	----

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范》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：

本文件主要起草人：

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的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，形成本标准，为行业单位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

结合海量通用数据大规模预训练得到的能够处理多种领域任务，具备强大的通用表示能力和迁移学习能力的人工智能模型。

4 缩略语

- API：应用程序编程接口（Application Program Interface）
MOS：平均主观意见分（Mean Opinion Score）
Rouge：以召回率为导向的摘要评价方法（Recall-Oriented Understudy for Gisting Evaluation）
SQL：结构化查询语言（Structured Query Language）
DDoS：分布式拒绝服务（Distributed Denial of Service）
XSS：跨站脚本攻击（Cross-Site Scripting）

5 需考虑的要素

本文件给出了证券期货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	技术指标
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
数据评估	数据准确性	风险处理
		真实性
	数据多样性	一致性
		数据分布
		数据模态

评估维度	评估项	技术指标
	数据全面性	数据来源审查
		完整性
		时效性
算法评估	算法准确性	合规性
	算法可靠性	精准性
		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
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	个人隐私数据管理
		个人隐私数据处理
		个人隐私数据展示
	数据安全	数据存储加密
		数据传输加密
		训练数据加密
		推理数据加密
	风险防范	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
		模型安全
		模型扩展
	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国家、行业在人工智能领域的合规目标，以：

- a) 确保模型能够实现预期能力；
- b) 确保履行证券期货业合规要求；
- c) 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，并评估其对运行所产生的影响，并建立过程以：

- a) 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

基础大模型宜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

基础大模型宜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) 评价合规风险对资本市场、证券交易、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

基础大模型宜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，尽可能的消除冲突或矛盾。

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。

6.2.2.3 数据来源审查

在基础大模型训练、微调等开发过程中使用的数据来源，能够追溯数据授权与许可。

注：涉及个人隐私的数据，数据采集需要经过用户授权。

6.2.3 数据全面性

6.2.3.1 完整性

宜确保同一数据的不同部分完整性，消除缺失值或不完整的记录。

6.2.3.2 时效性

宜确保能够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宜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

回答事实类内容宜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

宜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

基础大模型宜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

宜评估基础大模型对语言类任务的支持度及应用效果，涵盖行业知识提取、行业知识问答、行业内容生产、行业交易辅助等任务，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B.1。

6.4.1.1 知识提取

宜评估基础大模型知识提取能力，知识提取能力包括：

- a) 对时间、地点、人名、机构名等通用领域实体，证券公司名、基金公司名、证券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，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和研究资料；
- 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。

6.4.2 语音任务

宜评估基础大模型对语音类任务的支持度及应用效果，涵盖语音识别、语音对话、语音认证、语音质检等任务，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B.2。

6.4.2.1 语音识别

宜评估基础大模型语音识别能力，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手机银行登陆、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

宜评估基础大模型对视觉类任务的支持度及应用效果，涵盖行业图像分类与识别、身份认证与安全、视觉真实性检测等任务，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B.3。

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，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B.4，其基本任务包括：

- 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；
- 图片文字互转：宜支持以文搜图、以图搜文等；
- 跨模态生成：宜支持根据提示词生成图片、绘制图表等；
- 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 个人隐私数据管理

- a) 保护周期管理：宜覆盖数据采集、数据传输、数据存储、数据处理、数据交换、数据删除、数据销毁的全生命周期，确保每个环节的数据安全；
- b) 数据采集安全管理：数据采集宜支持采用白名单、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

- a) 身份认证：宜支持对进行数据处理的人员进行身份鉴别，并设置认证失败处理机制，防止暴力破解等攻击行为；
- b) 访问控制：宜支持对数据访问权限进行严格授权和管控，确保只有被授权的人员可以访问和处理数据，遵循最小化权限原则。结合数据分类分级和实际情况，对访问控制粒度进行设定；
- c) 日志与安全审计：宜支持对数据处理行为进行留痕审计，内容包括但不限于操作者、调用接口、数据版本；

- d) 匿名化：数据宜经过彻底的匿名化处理，确保完全无法识别个人身份，经过匿名化处理的数据不再属于个人信息；
- e) 去标识化：数据宜保留个体颗粒度，如采用假名、加密、哈希函数等方式处理数据，但需确保在没有额外信息的情况下无法复原个人信息。

6.5.1.3 个人隐私数据展示

- a) 模糊化：宜通过隐藏或截词部分信息，使个人证券期货信息无法完整显示，从而保护个人信息的安全；
- b) 不可逆：宜确保无法通过展示的样本信息倒推出原始的真实信息，增强数据展示的安全性。

6.5.2 数据安全

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中断、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

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 风险防范

6.5.3.1 外部访问攻击防范

- a) 访问控制：宜支持授权用户、设备和白名单等访问控制手段；
- b) 安全策略：宜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：宜支持采用防火墙、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：宜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

- 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，宜具有防范机制；
- b) 隐私数据：对于诱导输出客户的个人信息、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，宜具有防范机制；
- c) 业务合规：对于用户提出的市场走势预测、市场强观点分析、个股推荐等不合规请求，宜具有防范机制；
- d) 伦理合规：对于鼓动资产转移、电信诈骗等扰乱经济秩序的行为，以及由于财务危机造成的危险行为，宜具有预警和防范机制；
- e) 模型窥探防护：对于通过特定输入获取模型敏感信息或逆向工程模型的攻击，宜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，宜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：

- 1) 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；
- 2) 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 embedding 模型、rerank 模型等。

b) 模型训练：

- 1) 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；
- 2) 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：

- 1) 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：

- 1) 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；
- 2) 平台私有化部署：宜支持模型训练和部署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：

- 1) 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；
- 2) 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；
- 3) 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 模型安全

宜评估基础大模型应用过程中防范攻击的保护措施，具体措施包括：

a) 基础大模型宜支持的外部访问攻击防范措施。

- 1) 访问控制：宜支持授权用户、设备和白名单等访问控制手段；
- 2) 安全策略：宜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；
- 3) 防御机制：宜支持采用防火墙、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；
- 4) 日志监控：宜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。

- b) 基础大模型宜支持防范如下类型的对话攻击。
 - 1) 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，宜具有防范机制；
 - 2) 隐私数据：对于诱导输出客户的个人信息、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，宜具有防范机制；
 - 3) 业务合规：对于用户提出的市场走势预测、市场强观点分析、个股推荐等不合规请求，宜具有防范机制；
 - 4) 伦理合规：对于鼓动资产转移、电信诈骗等扰乱经济秩序的行为，以及由于财务危机造成的危险行为，宜具有预警和防范机制；
 - 5) 模型窥探防护：对于通过特定输入获取模型敏感信息或逆向工程模型的攻击，宜具有防范和预警机制；
 - 6) 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，宜具有识别并拒绝的防范机制。

6.6.1.5 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：
 - 1) 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；
 - 2) 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；
 - 3) 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：
 - 1) 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；
 - 2) 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。
- d) 容错与冗余：
 - 1) 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；
 - 2) 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。

6.6.2 应急管理

6.6.2.1 组织体系

- 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

- a) 宜制定安全事件应急预案，内容包括：
 - 1) 应急预案编制的目的和依据；
 - 2) 应急预案的适用范围；
 - 3) 应急处置的组织体系及职责；
 - 4) 安全事件的预防措施、保障措施与应急准备；
 - 5) 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

- 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到的模型验证、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、模型服务端点或 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、可补偿。

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；
- 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：宜利用如 Git 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

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

语言任务推荐多个客观指标和主观指标，推荐指标的计算方法和释义如下，可根据任务实际情况，选择合适的评价指标。

B.1.1 客观指标

1) 准确率：

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

式中：

P_H ——准确率；

H_1 ——正样本预测正确的结果；

H ——正样本预测的结果和预测错误的结果的和。

2) 召回率：

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

式中：

R_E ——召回率；

E_1 ——正样本预测正确的结果；

E_2 ——正样本预测正确的结果和正样本预测错误的和。

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 ——准确率；

R_H ——召回率。

4) ROUGE-N：

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

式中：

N ——即 n-gram，文本内容滑动窗口字节数，参考值为 2；

$Count_{match}(gram_n)$ ——参考摘要和机器生产摘要中共有的 n-gram 的数量；

$Count(gram_n)$ ——参考摘要中 n-gram 的数量。

B.1.2 主观指标

- 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

语音任务推荐多个客观指标和主观指标，推荐指标的计算方法和释义如下，可根据任务实际情况，选择合适的评价指标。

B.2.1 客观指标

- 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测试样本数；

R ——总的冒充者测试样本数。

- 2) 错误拒绝率：

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

式中：

FRR ——错误拒绝率；

A_2 ——被系统拒绝的真实人测试样本数；

A ——总的真实者测试样本数。

- 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

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

视觉任务推荐多个客观指标，推荐指标的计算方法和释义如下，可根据任务实际情况，选择合适的评价指标。

B.3.1 客观指标

- 1) 准确率：参见 B.1.1 语言任务的参考评价指标中准确率的定义；
- 2) 召回率：参见 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 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：

$mAP_{IOU=0.75}$ ——所有实例分割类别在 $IOU = 0.75$ 下的平均精度，其中 IOU 计算时预测掩模与真实标注掩模之间的交并比；

AP_i ——某个目标类别 i 实例分割的平均精度；

k ——目标类别数量。

B.4 多模态任务评价指标

多模态任务评估宜从客观指标和主观指标开展，可根据任务实际情况设置，具体包括：

B.4.1 客观指标

1) 前 10 命中率:

计算每张图像和查询词之间的相似度，按相似度从高到低对所有图像进行排序，评估该排序列表的前10命中率。

$$Rank@k(Q, G) = \frac{1}{N} \sum_{i=0}^{N-1} I\left(\left(\sum_{j=1}^k r(l_{i,j}, q_i)\right) > 0\right)$$

式中:

$Rank@k(Q, G)$ ——查询集合为 Q ，底库集合为 G 时的前 k 位平均命中率;

q_i ——第 i 个查询样本;

$l_{i,j}$ ——查询样本 q_i 的排序列表中的第 j 个样本;

$r(q, g)$ ——样本是否相关，相关返回 1，不相关返回 0;

N ——测试集中的样本总数;

I ——示性函数;

k ——取 10。

B.4.2 主观指标

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] 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
-