ICS 35.340
CCS L67

金　融　行　业　技　术　文　件

金融智能体应用协同指南

Financial AI agent applications collaboration guide

标准草案稿

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。

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。

本文件起草单位：兴业银行股份有限公司、中国信息通信研究院、华为技术有限公司。

引　言

在金融科技蓬勃发展的当下，人工智能技术与金融业务的融合持续深化，智能体作为人工智能应用的重要形态，正逐步渗透至金融行业的各个环节。从辅助金融决策、优化风险评估，到提升客户服务体验、创新金融产品设计，智能体展现出了强大的潜力，能有效缓解金融业中长期存在的诸多痛点，如风险识别滞后、运营成本高昂、客户服务同质化、决策不精准、业务协同效率低下等问题，极大地提升了金融服务的质量、效率与核心竞争力，为金融行业的数字化转型注入了新的活力与动力。

2025年以来，国家密集出台人工智能+相关政策，到2027年新一代智能体应用普及率要超过70%，到2030年智能体应用普及率要达到90%以上。金融行业正是智能体落地的最佳试验场之一，也是最具挑战的高地。金融业数字化程度高，业务流程复杂，对合规、安全、可信有着极高要求，如何在保障风控的前提下，让智能体真正走入金融业务核心，成为全行业共同面对的问题。

调研发现，金融智能体的规模化落地仍面临一系列行业特有的挑战与痛点，目前整个生态仍不够成熟：1）用户交互方式的革新，从按钮和表单，变成了对话式与多模态等交互，对客与对内服务的要求也在变化；2）系统协作方式的革新，使智能体与金融内外部系统调用要求与响应也在发生变化；3）高质量的数据与知识成为刚需，下一步智能体的发展将更加依赖于结构化、可追溯、高可信的知识供给体系；4）重新审视AI下的安全合规，如内容的安全，建立信任关系，与可审计等。技术上，大模型工程化能力尚不足，难匹配金融交易高安全性与稳定运行需求，且运算速度慢，难以满足金融市场时效性要求；知识储备上，其内生知识缺乏金融行业深度专业数据与行内独特信息，影响服务精准度；系统融合方面，金融机构存量异构系统差异大，导致智能体难以快速联通，阻碍协同效能发挥。与此同时，金融智能体相关技术与应用的差异性和不规范性问题逐渐凸显。

此外，行业对于制定不同部门、系统及业务场景间智能体的统一连接与协同规范的呼声日益强烈.缺乏此类规范，导致多智能体间难以实现更高效的任务协同与数据共享。同时，人工智能开放合作体系也亟待完善，以充分发挥多智能体的集群优势。这些，都需要行业共同努力、规范化推进。

智能体不是孤立的工具，而是生态中的有机节点。在此背景下，为推动金融智能体的健康、有序发展，提升金融服务的标准化、规范化水平，促进智能体在金融行业的深度应用与广泛普及，特制定本文件。本文件旨在明确金融智能体在用户渠道协同、智能与数据协同、通信保障、多智能体协同、生态建设以及安全可信等方面的具体要求，为金融机构及相关企业在金融智能体的设计、开发、部署与应用过程中提供统一的指导与规范，助力金融行业构建安全、高效、协同的智能体应用生态，进一步提升金融行业的整体竞争力与创新能力，助力金融行业高质量发展。

金融智能体应用协同指南

1 范围

本文件提供了金融业智能体应用的协同构建指引，涵盖金融智能体在用户渠道协同服务、智能与数据协同、通信保障、多智能体、生态、以及安全可信等方面的指引规范。

本文件适用于银行、证券、保险等金融机构及相关金融科技企业构建金融智能体协同体系。

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。

JR/T 0071-2020 金融行业网络安全等级保护实施指引 第2部分：基本要求

ISO 8583-2023 Financial-transaction-card-originated messages — Interchange message specifications

ISO 20022 Financial Services - Universal financial industry message scheme

3 术语和定义

下列术语和定义适用于本文件。

3.1

智能体 AI agent

基于人工智能构建的，能够感知环境，可以自主决策并执行动作的软件或实体。

3.2

金融智能体 financial AI agent

在金融行业中被设计用于自主执行特定金融任务的智能体（3.1）。

3.3

中控智能体 central AI agent

在多智能体系统中执行接收任务，理解全局状态、进行任务分解、规划、分配，并整合各个结果的智能体（3.1）。

3.4

工作智能体 work AI agent

在多智能体系统中负责执行被中控智能体（3.3）拆分后的子任务的智能体（3.1）。

4 缩略语

下列缩略语适用于本文件。

AI：人工智能（Artificial Intelligence）

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

CI/CD：持续集成/交付（Continuous Integration/Continuous Delivery）

CoT：思维链（Chain of Thought）
HMAC：密钥相关的哈希运算消息认证码（Hash-based Message Authentication Code）
HTTPS：超文本传输安全协议（Hypertext Transfer Protocol Secure）
HTTP SSE：超文本传输协议 服务器发送事件（Hypertext Transfer Protocol Server-Sent Events）
ID：身份标识号（Identity Document）
JSON：JavaScript键值对数据（JavaScript Object Notation）
MCP：模型上下文协议（Model Context Protocol）
mTLS：双向传输层安全（Mutual Transport Layer Security）
NL2DSL：从自然语言到领域特定语言（Natural Language to Domain-Specific Language）
NL2SQL：从自然语言到结构化查询语言（Natural Language to Structured Query Language）
OCR：光学字符识别（Optical Character Recognition）
PDF：移植文档格式（Portable Document Format）
QPS：每秒查询率（Queries-Per-Second）
RAG：检索增强生成（Retrieval-augmented Generation）
REST：资源表述性状态转移（Resource Representational State Transfer）
RPC：远程过程调用（Remote Procedure Call）
SSL：安全套接层（Secure Socket Layer）
SQL：结构化查询语言（Structured Query Language）
TCC：基于业务补偿的分布式事务解决方案（Try-Confirm-Cancel）
TLS：传输层安全性协议（Transport Layer Security）
ToT：思维树（Tree of Thoughts）
XML：可扩展标记语言（Extensible Markup Language）
XSS：跨站脚本攻击（Cross Site Script）

5 金融智能体应用协同框架

5.1 概述

金融智能体的核心定位是融合业务理解和技术能力，驱动企业内部系统及生态服务完成具体任务，金融智能体运行过程中需要与外部系统深度协同，在大模型内生知识之外进一步获取外部最新、全面、准确的信息，比如在信贷风险分析场景中，智能体对接企业内的业务和管理系统获取企业基本信息、财务信息等内部数据，进行更精准的风险识别。智能体连接外部征信、市场监管等系统，可增强其对市场动态和外部风险的感知能力。最终，通过协同实现从任务拆解、数据分析到业务执行的自动化和智能化。

金融智能体协同框架主要包括四个协同和两个保障，分别是用户与渠道协同、智能与数据协同、应用与系统协同、场景与生态协同四个领域的协同，以及支撑高效协同的通信机制保障和合规信任保障，总体框架见图1。

图 1　金融智能体应用协同框架

5.2　用户与渠道协同

金融智能体服务的用户包括客户和员工两大类人群，也是金融智能体服务的最终主体，通过与客户协同、员工协同，来协助解决问题、提升用户体验、效率和质量。

5.3　智能与数据协同

金融智能体与金融机构内的各类AI能力、数据能力、知识能力协同是智能体应用落地的基础，通过聚合、消化机构内的智能工具和内部信息，形成机构内的企业级智能引擎。智能体既需要大模型进行意图识别、任务规划和工具选择，又需要专业的传统机器学习模型提升具体任务的精准性和执行效率。智能体执行任务时需要根据场景获取金融机构内部实时、鲜活的数据和知识，才能做出准确的回答并执行任务，降低模型幻觉。同时在复杂任务中需多智能体协同才能达成，而多智能体的协同需根据业务需求采用不同的协同模式。

5.4　应用与系统协同

为了让大模型的场景应用从内容输出向思考执行演进，金融智能体需要对接好金融机构内的各类应用与系统，让智能体可直接触达各类金融服务和交易，或及时获取数据信息，来完成各类执行任务。除此之外，智能体也是诸多应用系统的一员，需要遵循认证、鉴权、监控等基础平台的协同要求。

5.5　场景与生态协同

外部场景与生态协作中，金融智能体需要与机构外的系统对接，来交换数据或向其用户提供金融和IT服务，甚至把服务嵌入至生态场景中，通过外部服务协同、外部数据协同和外部场景协同等方式，强化内生服务水平，也可将金融机构的智能化服务通过开放接口或应用的形式对外赋能。

5.6　通信机制保障

从技术层面看，金融智能体的协同过程呈现出多模态、多轮处理、数据融合、交易融合、内容不确定等特点，需要关注通信方面的保障机制，如流式通信、多轮对话链路跟踪、会话上下文保持、内

容安全拦截、文本或图像的数据安全、对话服务的事务一致性等方面，通信机制保障为智能体之间、智能体与外部系统、数据平台、服务平台之间的通信提供稳定可靠的底层支撑。

5.7 合规信任保障

金融业务的高敏感性、业务流程的强监管性对智能体协同提出诸多安全挑战。金融智能体与内外部系统协同过程中需关注内容安全、信任关系、合规审计等安全机制，保障智能体协同安全合规。

6 金融智能体应用协同框架

6.1 概述

用户与渠道协同要求涵盖客户协同与员工协同两大维度，旨在确保智能体服务在精准、高效、安全的前提下，提升用户体验、效率或质量。

客户协同面向金融客户，是智能体传递价值的核心。要求智能体具备多模态交互、连续性记忆管理与跨渠道会话同步能力，确保服务体验连贯一致。需根据零售、公司、同业等不同客群特征实施差异化策略，并在智能客服、业务办理、非银服务等全场景中，严格遵循最小化鉴权、内容安全与合规管控原则，优先调用内部可信数据源，严防幻觉与隐私泄露，保障服务安全可靠。

员工协同面向内部员工，聚焦于提升业务流程处理、办公协同及专业应用的效率与合规性。智能体需与金融机构内的业务流程做一致性对齐，支持差错识别、人工复核机制，确保流程可追溯、可审计。必须遵守最小权限原则，严防数据与客户信息泄露，防范模型幻觉。在投研、风控、合规等专业场景，需强化数据一致性、模型可解释性，关键决策节点保留人工干预，实现人机高效、安全、合规的协同。

6.2 客户协同

6.2.1 概述

金融客户交互是智能体与用户建立连接、传递价值的核心环节，需根据金融业务特性与客户类型设计差异化的交互策略，同时强化多模态交互适配与对话连续性记忆管理能力，确保服务精准、高效且体验连贯。

6.2.2 客户服务类场景

各类渠道中的客户服务是金融机构与客户建立联系、提供信息支持的高频交互场景，涉及较为丰富的交互形式。在该类场景中，应用智能体实现与客户的协同，应提供高质量、安全可信、体验一致的客户服务，应严格遵循相关政策法规要求，确保内容合规，避免误导性信息传播。同时智能体应支持多模态交互与记忆管理，提升服务效率。具体能力要求如下。

a）宜保障内容和范围的一致性。智能体应保障同类型渠道中对客服务内容和范围的一致性，如可售产品范围和产品信息的一致性、知识问答话术的一致性、展示风格的一致性、投资建议的合规一致性等。

b）宜强化对话连续性与记忆管理。建立短期对话记忆+长期客户画像的双维度记忆机制。记忆内容需遵循最小化权限原则，仅授权角色可访问，符合数据安全要求。

c）宜支持多模态交互。支持语音、文字、图片、文档等多模态输入输出，支持客户按需与智能体进行多模态交互，同时多模态数据传输需附带元数据，如图片拍摄时间、文档来源等，确保可追溯。

d）宜保障跨渠道交互会话连续性。建立统一交互管理层与上下文感知机制，确保应用、网页、电话等不同渠道间会话连续。如客户首先在 A 渠道咨询后，切换至 B 渠道继续询问，智能体能识别并延续对话历史，避免重复提问，提升客户体验。

e）宜遵循最小化鉴权原则。对于非个性化、非敏感信息的常见问题，如查询公开的理财产品收益率解答，无需认证登录即可进行；仅当涉及个性化服务、账户信息查询或查询个人账户余额、持仓理财明细等敏感操作时，再提示客户通过密码、生物识别等方式完成鉴权。

f）宜支持有效的人机协同机制。如遇到复杂问题或智能体无法回答的问题时，可自动切换至客服坐席或理财经理进行人工服务，并同步对话历史，避免客户重复表述，提升衔接效率。

g）宜优先使用高质量可信数据进行内容服务。智能体应优先使用机构内的知识库、问答库、产品库、资讯库，避免大模型内生知识的时效性和幻觉问题，确保信息准确可靠。
h）宜基于客群特征，实施差异化策略，包括但不限于以下客群。如对于零售客户，侧重高频、轻量服务、体验驱动，并注重语言通俗化；公司客户，侧重定制化、专业化服务，并做好人工服务的衔接。
i）主动式场景宜遵循原隐私保护与合规要求，如 APP 消息推送、外呼、即时通信等，应注意用户体验和适当性，防止过度打扰；
j）宜保障交互安全与合规管控。内容安全审核需覆盖全交互环节，对涉及生成内容的部分进行质检审核，拦截隐私泄露、金融欺诈、敏感话题、违规营销等方面，并遵循消保合规要求，确保交互内容客观、真实、避免误导性表述、对产品风险、收益等关键信息进行清晰、明确的提示。
k）宜完善异常处理，保障服务连续性。如超时、断线时需要设计应对机制。
l）智能体在客户服务的业务错误发生时宜提供合规和明晰的提示。
m）宜建立完善的反馈机制。智能体可收集客户对服务的评价反馈，持续优化服务内容与交互体验。

6.2.3 业务办理类场景

业务办理类场景主要包括但不限于开户、贷款申请、转账、现金管理、支付、产品申赎、理赔、指令划拨等涉及金融交易的核心业务场景，需与金融机构对应领域的设计要求、安全要求、严谨性要求保持一致，同时通过多模态交互简化操作、记忆管理减少重复步骤。具体能力要求如下。

a）宜遵循业务流程规范。涉及关键业务办理流程的场景，避免或谨慎使用智能体自主规划能力处理流程。如转账汇款、贷款申请、产品购买等流程中的关键动作，应按照标准业务流程引导客户完成，不得擅自修改或跳过关键步骤。
b）宜遵循原场景的同等安全防护级别。如智能体在涉及账户、资金操作的场景，不能降级。如核身、验密、多因素认证等安全措施必须完整保留，确保交易安全。
c）智能体在涉及数值计算、校验类的场景，宜优先考虑机构内的接口调用或 function call，避免大模型内生计算的差错。
d）宜谨慎使用语音、视频等开放式交互方式，避免在公众场合发生隐私泄露。
e）宜根据业务需求设置人工审核机制。在关键业务办理，如大额转账、企业贷款审批等核心业务处理，宜提交人工二次确认。
f）内嵌第三方渠道的业务办理，宜明确告知客户合作信息与数据安全责任。
g）资金交易类场景宜准确识别交易行话和惯例，并注意当行情变动时，能及时进行价格自动更新，同时在交易全链路满足该产品交易的相关合规检查。
h）业务审核类智能体宜建立可溯源、版本化的精准规则库，保留自动+人工双重复核接口，确保专业兜底、全程可追溯。
i）本场景中其他要求宜遵循智能体客户服务类场景指南，详见 6.2.2。

6.2.4 非银服务类场景

信用卡的生活类场景、积分权益、零售养老服务、保险健康管理服务等非银服务类场景，是金融与生活场景融合的重要领域，需注重生活化、便捷性与客户体验，通过多模态交互贴近生活场景、记忆管理精准匹配需求，同时遵循合规与隐私保护要求。具体能力要求如下。

a）宜保障数据隐私，防范越权风险。如智能体在进行健康管理时，需明确用户授权范围，如仅允许查询历史体检记录，禁止共享给第三方。
b）宜严控合规边界。如医疗建议边界，智能体仅可提供通用健康指导，不能替代医疗诊断。
c）宜支持动态策略更新。智能体应能及时感知权益规则变化、健康政策更新等动态策略的更新，并及时告知用户。
d）本场景其他要求中宜遵循智能体客户服务类场景指南，详见 6.2.2。

6.3 员工协同

6.3.1 业务流程处理

业务流程处理具体能力要求如下。

a）智能体应用所在的主场景宜预设差错识别机制，对智能处理环节中出现的不同类型差错自动标记、分级及反馈。

b）宜定期汇总各场景差错类型，复盘和优处理机制，提升差错处置的完整性。

c）宜结合业务场景要求，为判别式智能体设定可量化的指标，如性能指标和非性能指标，确保业务效果。

d）宜对生成式智能体应用，在流程中设置合规校验环节并记录校验日志，如流水号、校验结果等，确保流程合规可追溯，业务流程和日志防篡改。

e）宜对在用数据定期进行质量校验。对符合异常规则的数据进行标记，并记录异常原因，根据业务需求和情况判断是否删除。

6.3.2 通用办公协同

面向员工的通用办公协同智能体在提高工作效率的同时应遵守安全合规底线，具体能力要求如下。

a）宜遵守金融机构内数据权限管理规定，符合数据权限最小必须原则，如通过基于角色的访问控制或基于属性的访问控制等机制，精细化控制数据权限。

b）宜防范模型幻觉和事实错误，如优先通过内部知识库和可信数据源中检索信息，确保输出内容有据可依。

c）金融智能体调优数据宜经过人工确认，如涉及敏感信息的会议录音及其总结分析等数不宜作为模型调优数据，避免数据泄露。

d）智能体生成内容如涉及对外发布或合规要求等材料宜经过相关专业人员的严格人工审核与复核。

e）宜设置人机协作边界，如关键场景中涉及批量发送邮件、资金交易等敏感操作前需人工授权。

f）宜详细记录数据操作行为并定期审计。

6.3.3 专业应用服务

对于专业应用，如投研、风控、监管报送、审计、反洗钱等垂直业务领域的应用服务，因具有强监管、数据密集等特征，相关服务在落地智能体应用时，需遵循差异化策略分配、实时响应行业监管政策更新、科学设计业务指标体系等关键原则与要求，具体要求如下。

a）宜遵循智能体应用差异化分配原则。监管报送、合规审查等强监管场景需避免或减少幻觉与自主规划，输出须经人工审核，确保可审计、可追溯。投研分析、风险审查等其他场景需审慎使用自主规划，使用中须引用数据来源或出处。

b）宜建立金融行业智能体知识管理体系并满足金融行业监管政策更新要求，确保监管政策等外部规章制度在金融机构内部及时更新，并通过版本控制定期内部发布，智能体应用运行需满足对应领域监管政策的审核验证。

c）宜满足业务指标设计要求，并具备监测手段。投研、风控等领域智能体业务指标，须贴合各场景核心需求，如指标计算准确性、维度完整性、数据时效性等。

d）宜满足责任明确化要求。智能体输出关联的业务指标结果需明确使用风险，如投研、风控等专业场景需标注计算依据与模型局限性，审计、合规场景须经人工复核，明确人机协同中的责任边界，保障指标应用合规可控。

6.3.4 研发测试服务

在研发测试领域，智能体提供代码生成、测试案例生成、质量审查等能力，并保障研发测试的严谨性、安全性，具体能力要求如下。

a）代码智能体宜符合机构内安全研发规范，如 XSS 防护、SQL 注入、组件黑名单等。

b）宜注意防范知识产权风险，确保生成代码不侵犯第三方知识产权，避免使用受版权保护的训练数据生成的代码。

c）宜使用本地部署的研发测试智能体服务，防止敏感代码、客户数据或商业秘密泄露。

d）代码智能体编码风格宜符合机构内编码规范。

e）宜人工复核或评审智能体生成代码，并进行严格的安全扫描。

f）宜关注测试智能体生成内容的质量，通过人工复核，避免逻辑错误、边界覆盖不全或幻觉等挑战。

g）测试智能体的测试建议或决策宜展现完整的推理路径和逻辑，以便审计和复查。
h）宜接入自动化测试和 CI/CD 工具链。

7 智能与数据协同要求

7.1 概述

智能体与金融机构内的各类AI能力、数据能力、知识能力的协同是智能体应用落地的基础，聚合机构内的信息智慧，以支持各类金融任务准确有效的执行，复杂任务还需要涉及多智能体协同，具体能力要求如下。

对于AI能力，智能体需运用大模型进行意图识别、任务规划和工具选择，还需要调用OCR、ASR、NLP等传统机器学习模型提升具体任务的精准性和执行效率。

对于数据能力，智能体执行任务时需要根据场景获取金融机构内部实时、鲜活的数据对用户的提问作出准确的回答、准确地执行任务，降低大模型的幻觉问题，如对接关系型数据库、消息队列等。

对于知识能力，金融智能体可与RAG协同获取变化较频繁的知识，部分场景还需要获取数据之间的隐性关联，以信贷审批业务为例，客户信息分散在账户系统、征信报告、交易流水等多个数据源中，传统数据分析方法难以捕捉多数据之间的隐性关系，智能体需要与知识图谱协同，提升关联认知能力。

对于多智能体，根据不同的场景采用集中调度、顺序调度、辩论调度等多种模式，满足不同业务场景的诉求。

7.2 AI 能力协同

7.2.1 大模型能力协同

大模型能力协同具体能力要求如下。

a）金融智能体宜支持接入多种模型，包括但不限于用于智能体任务规划的大模型、用于推理的大模型，
b）宜在满足业务效果要求的前提下，根据场景选择或训练小参数的大模型。
c）智能体宜采用如 REST API 等规范以流式输出的模式与各类大模型对接。
d）金融智能体宜通过模型网关连接各大模型，通过模型网关屏蔽各模型的接口差异，对智能体提供标准接口。
e）模型网关宜支持基于模型名称做不同后端模型的选择，实现同一个接口对接多种模型服务，便于模型的快速切换。
f）宜针对复杂对话类任务训练专业的金融智能体中控规划模型，提供适配金融行业的意图识别、任务规划、工具选择、反思演进等核心能力。
g）宜训练专业的金融场景模型，模型应能理解金融术语和名词，区分歧义术语，提供精准解释。模型宜理解金融场景的专业计算逻辑，并基于行业经验数据对问题做出准确响应。
h）金融智能体在运行中宜收集用户反馈相关数据，包括但不限于用户评价、偏好选择等行为日志、系统日志等，并对数据进行处理后用于模型的增量训练数据。

7.2.2 传统模型能力协同

传统模型（如OCR、ASR、NLP等传统机器学习模型）在特定场景的性能、效率、精度等方面仍具备优势，智能体与传统模型能力协同具体能力要求如下。

a）宜对传统模型的输入和输出进行适当的数据预处理和后处理，便于智能体与大模型等其他能力协同。
b）宜优化数据传输的延迟和开销，如图片、音频等大文件传输场景的性能优化。
c）宜配置纠错和校验机制，减少传统模型出错的误差传导与累积。
d）在智能体感知环节宜进行高可用设计，避免感知故障导致智能体整体功能不可用。
e）宜通过标准的接口规范和协议集成各类传统模型，便于智能体规划调度。

f）宜建立传统模型的数据反馈和闭环机制，持续微调和优化专用模型。

7.3 数据协同

7.3.1 数据库协同

数据库协同具体能力要求如下。

a）金融智能体宜根据场景需要通过 NL2SQL 技术连接关系型数据库，根据检索需求生成数据库查询语句，满足自然语言查询结构化金融数据的用数需求。

b）数据宜进行标准化规范转换，满足金融规范。NL2SQL 实施中宜将表名、字段名与自然语言查询的表述高度一致，减少模型的语义理解成本。如将数据转换为宽表，避免模型生成跨表关联的查询，降低映射错误率。

c）用户通过自然语言查询时，通过 NL2SQL 转换的 SQL 查询应只能访问其权限范围内的数据库表内容，避免越权查询敏感数据。

d）金融智能体宜根据场景需要通过 NL2DSL 技术连接金融机构的数据系统或业务系统，将自然语言中的金融术语、业务规则与领域特定语言等元素精准映射，根据检索需求生成金融领域可执行的领域特定语言元素指令。

e）NL2DSL 实施过程中宜对接标签系统进行领域资源的精准定位，宜对高频查询的领域特定语言结果进行缓存，减少重复计算。

f）金融智能体宜根据场景需要对接键值内存数据库，如行情数据的获取，满足金融场景对低延迟、高并发的要求。存储在键值内存数据库的数据需进行脱敏，满足监管要求。

7.3.2 对象存储协同

对象存储协同具体能力要求如下。

a）宜将金融智能体在用户交互中需要处理用户上传的文件和业务处理过程中需要的外部文件存储在对象存储中，并通过 REST API 等标准化接口规范与对象存储系统对接。

b）宜对存储的金融数据进行分类分级，金融智能体向对象存储系统存储数据时携带数据分类分级信息，不同级别数据按照金融监管要求采用不同的存储策略。

c）金融智能体宜对存储的数据依据金融行业要求数据脱敏，删除或掩盖敏感标识。

7.3.3 消息队列协同

消息队列协同具体能力要求如下。

a）金融智能体宜根据场景需要对接消息队列，满足金融异步通信、流式数据处理等场景需求。

b）智能体宜集成消息队列 SDK 与消息队列建立连接。

c）宜支持消息持久化，金融场景的消息丢失或重复消费将引发合规风险或用户损失，并通过全链路确认、持久化等机制保障消息可靠性。

d）宜支持消息去重和消息顺序处理机制，来保障强业务事务或顺序一致性场景的流程处理。

7.4 知识协同

7.4.1 RAG 协同

RAG协同具体能力要求如下。

a）金融智能体宜根据场景需要连接向量数据库，将特定的金融专业知识存入向量数据库以便支持智能检索，降低大模型幻觉，提升任务准确率，宜采用 REST API 等标准化规范与 RAG 系统对接或内置 RAG 组件。

b）智能体和 RAG 系统之间的检索方式可支持语义检索、全文检索或混合检索。

c）宜支持长文档分片的分页返回。

d）宜支持多模态检索，智能体支持用户输入的文本、图像、视频等不同模态的检索需求，并将其转化为 RAG 能理解的检索操作，返回相关性最高的多模态内容。

e）宜支持原文引用，为生成的答案清晰标注引用来源。向量数据库在检索结果中需携带完整的来源数据。

7.4.2 知识图谱协同

知识图谱协同具体能力要求如下。

a）智能体宜根据场景需要对接知识图谱，利用知识图谱的实体关联能力与逻辑推理能力，提升智能体在客户服务、投研分析等金融场景中的决策准确性与可解释性。

b）智能体宜支持将复杂的信息查询转换为标准的图查询语言，从知识图谱中精准提取出关联的子图或路径信息，可采用 REST API 等标准化接口对接。

c）金融智能体与知识图谱交互时，可缓存金融场景中高频查询数据，降低知识图谱的访问压力。

d）宜提供降级策略和提示，当知识图谱中未检索到相关信息时，智能体宜引导用户重新表述或通过 RAG 等方式获取信息。

e）智能体的行为和分析结果经过验证后可反馈给知识图谱系统，使知识图谱不断完善和进化。

7.5 智能体协同

7.5.1 概述

多智能体协同模式是指多个具有自主决策能力的智能体，为实现共同目标，通过相互通信、协商等手段，进行信息共享、资源互补、任务分配和协同作业的一种分布式智能模型。智能体具备感知、决策、行动的能力，既可以是具有不同功能和特长的个体，也可以是能够执行特定任务的子系统。

在多智能体协同模式中，每个智能体都能自主根据自身状态和环境信息做出决策，同时也会与其他智能体进行交互和协作，达成共同的任务目标。多智能体的协同核心在于智能体之间的合作与协调，协同模式主要可分为三类：集中模式、顺序模式和辩论模式。

7.5.2 多智能体协同模式

7.5.2.1 集中模式

7.5.2.1.1 交互流程

金融智能体集中式协同由中控智能体和工作智能体组成，其中中控智能体对各工作智能体进行统一规划、指挥、调度和管理等动作，工作智能体仅与中控智能体通信，各智能体间不直接通信，交互流程示意图见图2。

中控智能体能力要求如下。

a）宜具备任务拆解能力，通过 CoT、ToT 等技术将复杂金融任务拆分成能被多个智能体独立执行的子任务。

b）宜具备任务分配能力，支持将拆解后的子任务根据工作智能体的工作范围合理分配，并在执行过程中实时调整，在执行结束后汇总、分析及展示结果。宜采用符合场景要求的算法设计任务分配模式，如匈牙利算法、贪心算法和遗传算法等。

c）宜实时跟踪每个智能体任务执行状态，如已分配、进行中、已提交、已完成等。

d）宜整合工作智能体的输出结果并形成最终决策。

工作智能体能力要求如下。

e）宜清晰界定专业能力和职责边界，避免智能体间因任务分配重叠导致工作冲突。

f）宜支持标准化接口规范，中控智能体调用外部功能、查询状态等操作，如 REST API、RPC 等接口框架；

g）工作智能体宜通过中控智能体传输数据、交互等，彼此之间不应直接通信。

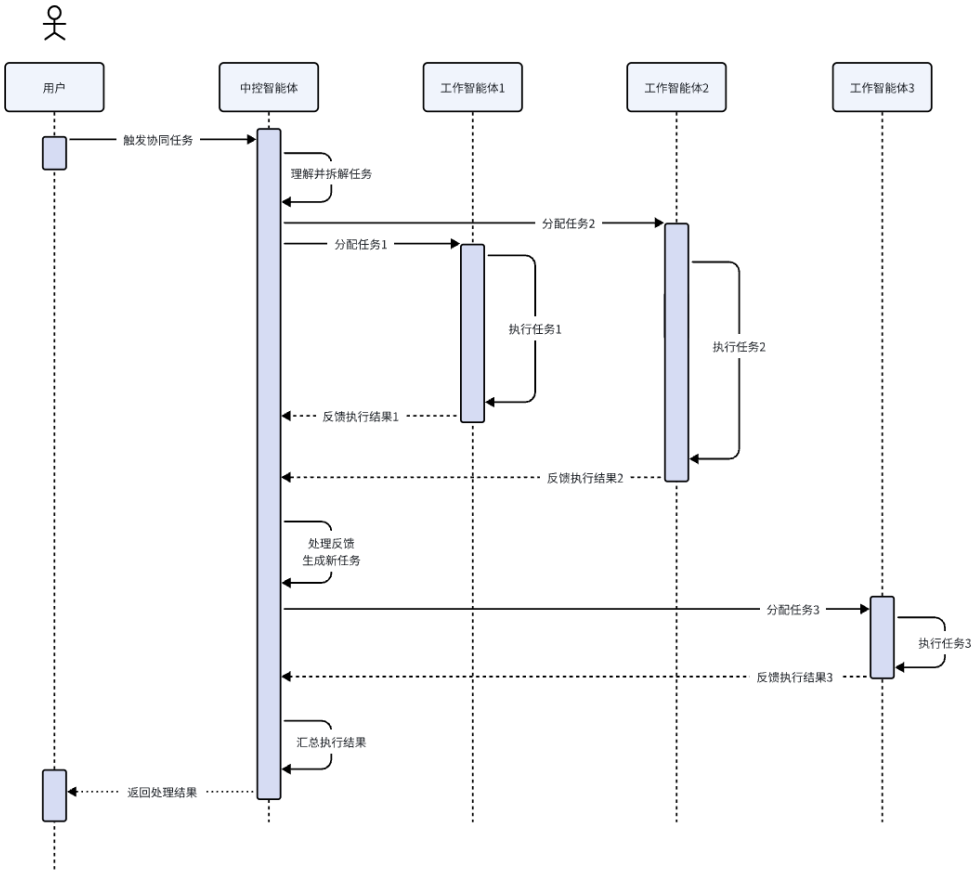

图 2　集中模式多智能体协同流程示意图

7.5.2.1.2　冲突解决机制

冲突解决机制具体要求如下。

a）当工作智能体目标或行为出现冲突时，中控智能体应基于工作智能体的职能范围以及各工作智能体实时反馈，通过全局最优原则重分配任务。

b）冲突解决机制包括优先级协商、谈判协商、投票协商等。

7.5.2.1.3　适用场景

集中式协同的核心在于中控智能体和工作智能体形成的多智能体系统。中控智能体负责任务的分解、指令下达、资源分配和最终结果的整合，而工作智能体则专注于执行具体的、标准化的子任务，其优势在于结构清晰、易于管理、执行效率高。

集中式协同模式用于目标明确的执行类场景，如投顾、营销、培训等。示例参考附录A.2

7.5.2.2　顺序模式

7.5.2.2.1　交互流程

金融智能体协同的顺序模式是将金融任务分解成一系列有固定先后顺序的子任务集合，每个子任务由一个工作智能体负责，前一个工作智能体的输出会成为后一个工作智能体的输入，信息和控制权沿着预定义的流程单向流转，由工作流末端的工作智能体输出最终结果，流程示意图见图3，具体要求如下。

a）宜预定义任务的执行步骤、依赖关系和执行顺序。

b）宜支持定义统一的输入和输出的数据格式和接口规范，，确保智能体之间能准确地传递数据和接受任务，如 JSON Schema 规范。

c）宜采用链式拓扑结构，智能体之间直接进行通信。每个智能体的任务执行有前置的触发条件。

d）宜支持采用外部状态存储等技术使任务的上下文信息随数据流传递。
e）宜对每个智能体设置超时处理策略，防止因单个智能体故障导致整体流程停滞。

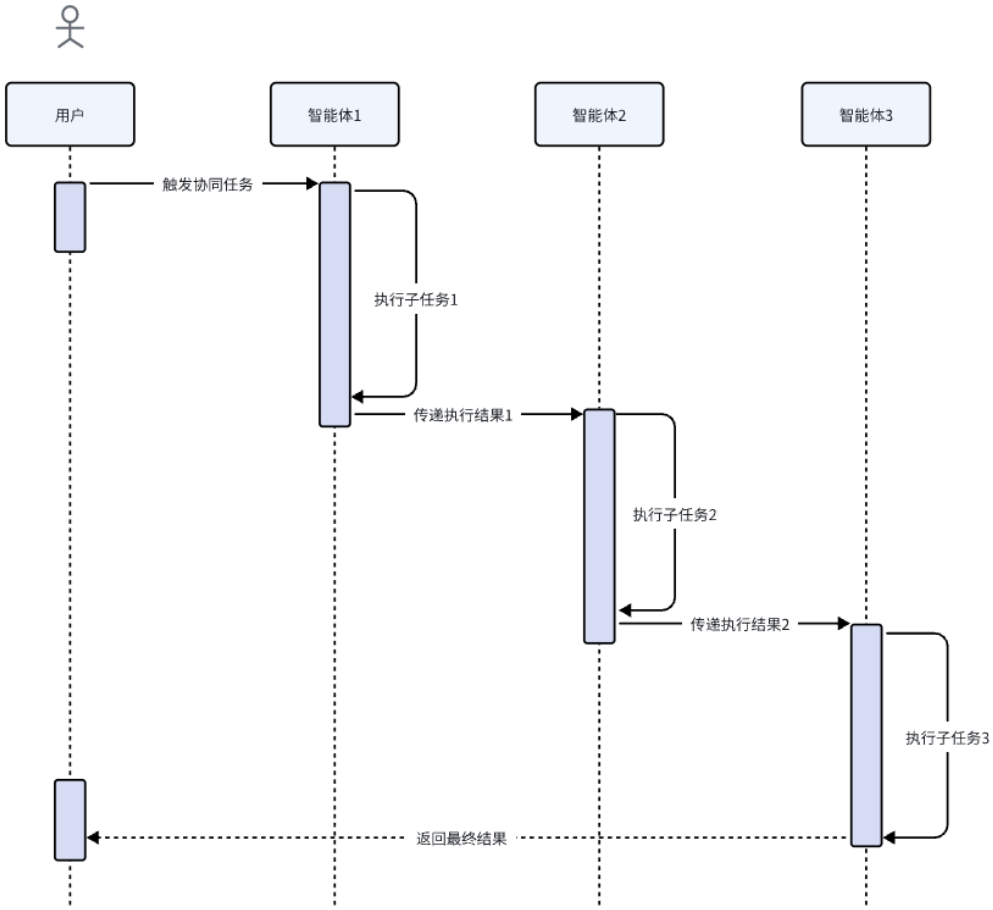

图 3　顺序模式多智能体协同流程示意图

7.5.2.2.2　冲突解决机制

当产生业务逻辑冲突时，宜在流程设计中设计明确的决策点，如在容易产生冲突的环节后插入实行结果仲裁或决策的智能体，根据业务需求解决逻辑冲突。

7.5.2.2.3　适用场景

顺序模式的智能体协同类似流水线，单个智能体负责一项或多项关键节点。任务在不同智能体之间按预设固定的流程依次传递和处理，上一个智能体的输出是下一个智能体的输入，确保了每一节点处理的专业深度，并提高任务执行的效率和规范性。

顺序模式的智能体协同适用于流程清晰、步骤固定的审批与处理类任务，具体场景包括贷款申请审批、保险理赔处理等。示例参考附录A.1。

7.5.2.3　辩论模式

7.5.2.3.1　交互流程

金融多智能体辩论模式是指在金融场景中，模拟人类辩论过程，通过多个拥有不同角色、数据和工具的智能体，对复杂金融任务进行多角度、深层次的分析和交互，综合各智能体共识形成审慎严谨的金融决策，流程示意图见图4，具体要求如下：

a）宜设计明确的辩论规则和流程，包括但不限于辩论顺序、达成共识、超时、投票等辩论结束机制等。

b）宜设计根据业务场景设计符合需求的评价机制衡量智能体的表现，并进行定向优化。

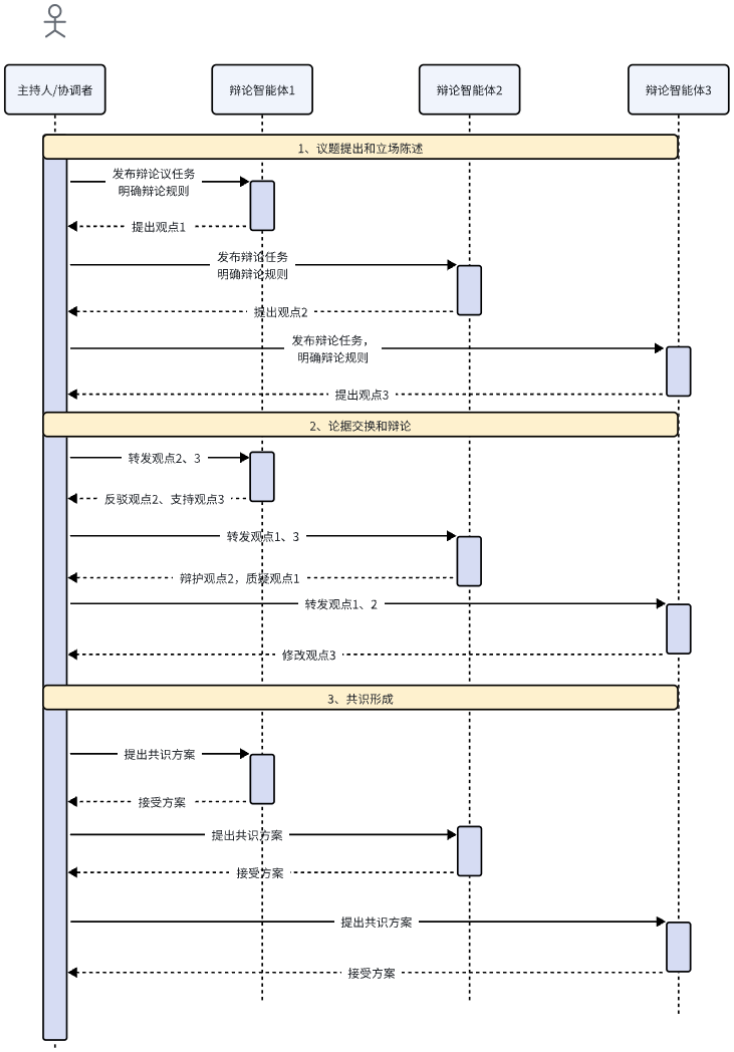

图 4　辩论模式多智能体协同流程示意图

7.5.2.3.2　共识达成机制

在金融行业内应用辩论模式进行多智能体协同过程中，宜提前配置完备的多智能体共识达成机制，如设置决策智能体、多智能体投票、输出结果加权评分等方式；且为应对无法达成共识的场景，宜预备备选方案，如延长辩论时间、人类介入等。

7.5.2.3.3　适用场景

辩论模式旨在应对没有唯一答案的复杂问题。多个智能体生成持不同观点、甚至对立立场的结论，如对某一市场的“看涨”或“看跌”，并通过共识机制，形成完整的决策。

辩论模式适用于复杂不确定性的决策和分析类任务，具体场景包括复杂投资策略制定、金融市场分析等。同样也适用于金融领域数据自动合成，用于覆盖场景用例的多样性，数据合成智能体与裁判智能体双向对抗辩论，自动生成训练数据与数据质量评估结果，具体场景包括复杂投资策略制定、金融市场分析等。示例参考附录A.4。

7.5.3　多智能体协同过程

7.5.3.1 能力发现和匹配

多智能体协同中，对智能体能力发现和匹配的具体能力要求如下。

a) 宜建立统一的服务目录，供其他金融智能体进行发现和调用。
b) 宜支持多维索引查询，包括但不限于智能体唯一标识、能力关键词、所属组织、版本号、状态、风险等级、监管区域等。
c) 宜基于查询条件对候选智能体进行能力匹配，支持不同筛选规则，包括但不限于精确匹配、模糊匹配、优先级匹配等；
d) 宜支持订阅式发现机制，通过 Webhook、消息队列等方式订阅特定事件，包括但不限于实时推送智能体上线、下线、状态变更等。

7.5.3.2 任务规划和编排

多智能体协同中，对任务规划和编排的具体能力要求如下。

a) 宜具备任务分解能力，如通过 CoT、ToT 等技术将复杂的金融目标拆分成由多个智能体独立执行的子任务，拆分逻辑应符合金融业务流程规范。
b) 宜根据子任务和智能体匹配的对应关系，动态地将任务分配给最合适的智能体执行，包括分配任务执行逻辑、执行任务所需的数据等。
c) 宜具备任务动态规划能力，当市场出现突发事件时，能根据外部信息变化实时动态调整和优化任务规划，如中断当前流程、重新分配任务等。
d) 宜支持对多智能体协同中出现异常操作的任务链进行追溯和审计。
e) 可通过状态机显式管理任务生命周期，如初始化、执行中、等待输入、完成、失败，对关键上下文信息进行持久化存储，以实现状态追溯与恢复。

7.5.3.3 资源感知和调度

多智能体协同中，对智能体运行过程中系统级的资源感知和调度的具体能力要求如下。

a) 宜具备全域资源感知调度能力，支持实时监控所有金融智能体的资源占用情况，包括但不限于计算、存储、网络等，并基于此进行负载均衡， 避免某些智能体过载。
b) 宜具备优先级队列调度能力，支持多级优先级队列，优先保障高优先级金融业务，如交易、风控的资源需求，避免因资源竞争导致延迟。
c) 宜支持区分金融实时任务和批处理任务，为其分配不同资源池和调度策略，如将金融实时性任务分配到边缘计算资源上处理，将计算量较大的金融批处理任务分配到云计算资源上处理。
d) 宜采用容器化、资源编排等技术实现系统级资源的动态扩缩容，保障智能体快速响应外部变化。

7.5.3.4 记忆共享和传递

多智能体协同中，对智能体执行任务所需的数据的记忆共享和传递的具体能力要求如下。

a) 宜根据金融多智能体协同模式设置对应的记忆共享模式，包括集中式共享、分布式共享、混合式共享等，且数据存储和传输宜符合金融数据监管要求。
b) 宜根据具体金融场景的需求，设计合适的记忆传递策略，包括主动推送、被动拉取、广播传递等。
c) 在传递复杂状态信息和任务信息金融场景下，宜对智能体的记忆内容进行结构化数据编码，采用 JSON 等标准化格式封装记忆，明确信息的字段意义，以快速解析记忆内容，减少记忆歧义。
d) 在动态更新频繁的金融场景下，宜对智能体的记忆采用增量同步，仅传输更新信息，减少通信消耗。
e) 宜结合金融业务规则选取合适的冲突机制解决记忆传递过程中的冲突矛盾，如以时间戳优先、以可信度优先、多源验证等。

7.5.4 异常处理

多智能体协同系统中，需对应用、系统、资源等异常具备处理能力，保障业务正常运行，具体能力要求如下。

a）宜设置多智能体协作的超时机制，对任务和子任务配置独立用时阈值及终止、重试或切换资源等超时响应策略，业务场景如息咨询、资质审核、风险监测、产品匹配推荐等。
b）对于辩论式等需要多轮交互的协同模式，宜明确定义交互结束的条件。如设定最大轮数、时间限制等。
c）宜建立熔断与降级策略，当依赖智能体连续失败时自动熔断或返回降级响应，优先保障核心重要任务的安全推进，防止系统雪崩。
d）宜支持检测模糊指令和约束冲突，必要时对用户或业务后台进行交互式澄清或征求参数补充。
e）宜定期自动检查各智能体服务状态，并配置典型异常情况响应策略。
f）宜在长流程中定期保存关键状态，支持任务中断时从最近的检查点恢复。
g）宜具备对正反馈循环的检测能力，必要时引入负反馈回路和随机性，打破单调循环。
i）可部署全局协调与监督智能体，对宏观指标进行监督，并在智能体间出现冲突时介入协调。
j）执行重要任务前，可在沙盒环境中先模拟执行，排除路径不通或资源不存在问题。

8 应用与系统协同要求

8.1 概述

为了让智能体在金融行业中从回答问题向思考执行转变，需要协同金融机构内的各类应用与系统，让智能体逐步成长为数字化业务的大脑引擎。除此之外，智能体也是金融机构应用系统的一部分，需要遵循认证、鉴权、监控等基础平台的协同要求。

8.2 业务系统协同

智能体与金融交易系统、内部管理系统协同时，需遵循以下原则，保障调用安全、合规、高效，具体要求如下。

a）金融智能体根据场景需要与金融机构的业务系统对接，完成数据查询和任务执行。
b）对于机构内部已有服务总线的场景，金融智能体宜根据机构服务总线规范完成智能体对接。
c）对于机构内部没有服务总线的场景，金融智能体宜支持 Rest API 规范与机构内业务系统对接。
d）宜将金融机构内部高频使用的公共能力组件改造为 MCP 服务以通过智能体调用，如证照 OCR 识别、电子签约等能力组件。
e）业务场景涉及动账等核心交易的，金融智能体不宜介入事务核心逻辑，避免分布式事务冲突。
f）金融智能体调用业务系统接口时，宜设置合理的超时时间，避免长期阻塞。
g）智能体宜具备自主处理业务系统返回的异常代码和信息，并做出合理的后续决策或转交人工处理。
h）金融智能体宜支持流量控制，限制单位时间内的交易请求量，并对接前与业务系统确认接口 QPS 配额，智能体请求量不宜超过配额，以避免智能体异常导致的请求阻塞。
i）金融智能体与业务系统对接发起的交易请求，宜记录不可篡改的审计日志，包括智能体标识、交易信息、请求/响应数据等。
j）宜保障运营流程闭环性。需完整衔接业务系统多环节审批与追溯流程，如报销审批、信贷审核等，不可省略关键步骤；触发后续操作前需确认前置节点完成，同步留存全流程轨迹，满足内部审计要求。
k）宜实时反馈调用状态至业务系统。智能体在调用业务系统完成业务操作后，需及时将调用结果同步至业务系统日志模块，包括但不限于成功、失败、异常原因等，便于业务系统管理员追溯操作记录、排查问题，同时支撑管理系统的业务流程闭环管理

8.3 基础平台协同

金融智能体与统一用户中心、身份认证、数据平台等机构内部基础平台协同是释放价值的关键，具体协同要求如下。

a）宜充分对接机构基础技术平台，应用统一工作流、服务总线、数据总线，遵从机构基础技术管理规定。
b）宜完整记录基础平台协作的审计日志，且保存期限需符合监管最低要求，确保可追溯。

c）宜严格遵循数据最小必要原则，
d）跨基础平台数据流转宜符合数据分类分级管理要求，如高等级数据需脱敏或加密方式传输。
e）智能体从基础平台获取的数据宜仅用于数据采集时约定的业务场景，避免数据二次滥用。
f）宜符合基础平台的业务流程规范，避免业务流程断裂，如智能体已输出结果，但基础平台无相关日志记录。
g）智能体宜遵循基础平台业务规则，不可突破规则提供服务，如用户风险等级标准。

9 场景与生态协同要求

9.1 概述

金融智能体对用户提供服务，除了需和金融机构内部系统协同，获取诸如用户账户信息、交易流水等内部数据外，还需要与外部系统协同获取用户消费偏好、信用记录数据以及利率波动、政策调整等动态信息。通过协同外部服务、外部数据增强内部应用能力，并将金融机构内部智能化能力通过开放接口或服务的形式对生态进行赋能。

9.2 外部服务协同

外部服务协同具体要求如下。
a）金融智能体与外部服务对接时宜通过专线对接，宜参考 MCP 等标准化智能体通信协议对接。
b）宜通过金融机构的网关将外部服务的接口转换为 MCP 服务供智能体使用。
c）网关宜支持将外部请求参数映射到 MCP 等标准化智能体协议的请求参数。
d）宜将外部响应报文映射为 MCP 等标准化智能体协议格式。
e）宜通过自动转换的方式降低开发工作量。
f）调用外部服务的金融系统内客户端和提供外部服务的服务端之间宜支持安全授权认证机制，通信协议宜支持 HTTPS 等安全通信协议，有高安全需求的场景可采用 mTLS 双向认证方式。
g）宜支持统一消息格式，如 JSON-RPC 2.0 等，确保跨系统交互的兼容性与扩展性。
h）宜支持周期性的主动监控检测第三方服务可用状态，便于及时发现问题并做异常处理。
i）宜支持 MCP 服务运营指标统计，统计智能体调用 MCP 服务指标，并提供运行分析报表页面或可集成的监控 API 接口。

9.3 外部数据协同

9.3.1 数据提供商协同

数据提供商协同具体要求如下。
a）金融智能体集成的外部数据源宜遵循监管机构相关规定要求，确保获取可信数据。
b）金融智能体集成的外部数据源宜通过金融机构的数据审查，包括但不限于规则校验、清洗处理、数据一致性和数据时效性，确保数据质量可靠。
c）金融智能体与外部数据系统集成时，宜遵守金融数据保护相关政策法规，承诺数据在存储、传输和处理过程中不被篡改或损坏，并通过技术和管理措施防止数据被未授权访问或泄露。
d）获取的外部数据和内部数据合并时宜完成数据格式与数据标准的统一，并进行数据时效性和完整性校验。

9.3.2 搜索引擎协同

搜索引擎协同具体要求如下。
a）金融智能体对接的搜索引擎宜确保数据来源合法合规，不宜使用非公开数据和侵权数据，搜索引擎提供商宜提供明确的合规审计报告。
b）金融智能体宜采用 REST API 等标准化接口规范与搜索引擎集成，并根据搜索引擎的限流政策在智能体端实现请求队列、流量控制和超时重试机制，确保服务稳定。

c）金融智能体从搜索引擎获取的数据宜关注时效性，根据搜索结果的显性时间标识选择合适的结果，不同金融业务对数据时效性的容忍度差异大，需为具体场景设定时效性阈值，超过阈值则判定数据失效。
d）对于关键业务数据信息，尤其是用于金融决策的数据，宜通过专业数据库、专业数据服务或机构内专业系统数据为主获取，确保其准确性。
e）当搜索引擎响应过慢或不可用时，金融智能体宜执行降级策略，返回预置知识或大模型生成的知识，保证智能体核心功能的可用性。
f）金融智能体在向搜索引擎发送的请求中携带的数据不宜泄露金融机构机密数据，不宜包含可直接识别的敏感信息。必要时，宜对查询关键词进行泛化或匿名化处理。宜对接收到的数据在处理前评估，必要时进行脱敏。
g）可根据场景需要对搜索引擎返回的结果做快照存储，根据数据重要性设定保留期限。
h）根据场景特点，可将部分搜索操作设计为异步任务，结果生成后通过指定的回调地址推送给智能体，避免阻塞智能体主业务流程。

9.4 外部场景协同

外部场景协同具体要求如下。
a） 金融机构在与多个外部服务商协同时，宜确保用户在多个外部平台的智能体交互体验与金融机构自有渠道一致，保障用户体验的一致性。
b） 宜通过行业通用协议与外部生态对接，如REST、RPC、MCP等接口协议格式、JSON、XML等数据格式。
c） 用户在外部应用中调用金融智能体服务时，宜先完成外部服务商账户和金融账户的映射关系，确保用户身份的唯一性与关联性。
d） 外部服务商调用智能体服务时宜进行场景化权限控制，数据调用场景与授权目的一致，避免一次授权、全量访问。同时，外部服务商在业务交互中避免留存金融机构的数据。
e） 金融机构可与外部服务商合作，将金融机构的智能体服务嵌入外部服务商的应用中，对用户提供一站式服务，避免用户使用外部服务商的应用时发生跳转。

10 通信机制保障要求

10.1 概述

金融智能体在协同过程中，呈现出多模态交互、多轮对话处理、结构化与非结构化数据融合、跨场景交易融合及内容不确定性等技术特点，与传统金融机构微服务体系存在显著差异，需针对性构建通信机制保障体系。通过四个维度协同落地。

——通信规范聚焦场景化协议选型与多模态数据传输适配，解决实时性与多类型信息交互问题；

——报文规范统一结构定义、版本兼容及数据融合封装标准，支撑跨机构、跨系统交互；

——安全规范覆盖身份认证、全生命周期加密、内容风险拦截等全链路防护，对内容不确定性与交易安全风险；

——可靠性要求通过低延迟、高可用、交易事务一致性及故障快速恢复，保障核心金融业务连续性。全面支撑金融智能体在行情分析、交易处理、客户服务等场景下的安全、可靠通信。

10.2 通信规范

金融智能体通信规范具体要求如下。
a） 在市场行情分析、金融新闻实时推送等需要实时、持续处理数据的场景中，宜采用HTTP SSE、WebSocket等协议实现流式通信，提升用户体验。
b） 在贷款材料审批、报告生成等实时性要求较低的场景中，宜设计异步通信机制实现。
c） 在涉及跨安全域的数据传输、数据同步等场景，宜通过TLS、SSL或底层安全隧道保障通信安全。
d） 在外部生态场景对接中，宜设计报文防篡改、校验等机制，如哈希算法、数字签名、HMAC等方式对报文进行完整性校验。

e) 宜采用强认证机制、双向认证等方式验证跨域智能体身份的合法性，如密钥、令牌、基于证书的mTLS等方式，保障智能体间数据通信与交互的安全性。
f) 宜在跨安全域的应用场景中，设计敏感数据的加密机制，如客户个人信息、账户信息、交易金额等，并采用SM4等国密算法进行加密。
g) 宜支持数据通信的异常容错能力，包括但不限于请求超时与重试、网络抖动容错等机制，并统一归类和反馈异常信息。
h) 宜建立动态服务端信任评估体系，确保MCP客户端仅与经过认证的服务端通信，防止客户个人令牌被恶意服务端窃取。
i) 宜在智能体运行时对包含eval等高风险命令的请求进行实时阻断，对MCP服务器可执行的系统命令进行白名单管理。
j) 宜实时采集MCP服务等智能体调用工具的调用日志并检测异常行为。当检测到高风险操作时，自动触发熔断策略，终止进程并隔离相关资源。
k) 针对跨安全域的工具调用，宜采用授权前二次确认机制，强制用户主动确认工具调用授权，并通过实施令牌身份隔离限制权限范围。
l) 宜实施基于场景的流量限流机制，配置接口调用阈值并检测异常突发流量，优先保障核心交易接口的资源分配。
m) 宜支持会话上下文的持久化存储与关联传输，报文中需携带会话历史索引，如对话轮次、上一轮报文ID等，确保智能体在多轮交互中准确理解上下文逻辑，避免信息断层。
n) 针对图像、语音等非文本数据，宜定义专用的传输协议适配策略，如大文件分片传输、断点续传等，并同步校验数据完整性，如分片哈希拼接后与原始文件哈希比对校验，保障多模态信息的完整还原。
o) 宜通过工具指令白名单、独立会话空间、授权前二次确认及TLS 1.3双向认证，防范MCP等智能体通信协议隐藏指令注入、上下文污染及动态注册安全风险。
p) 宜在MCP协议初始化阶段传递服务端能力清单，并建立合规认证与版本兼容策略，优化MCP等智能体通信协议扩展性。
q) 宜细化MCP等智能体通信协议错误码分类及配套重试策略，结合微服务拆分与消息队列，优化协议错误处理与高并发性能。
r) 宜实施基于场景的流量限流机制，配置接口调用阈值并检测异常突发流量，优先保障核心交易接口的资源分配。
s) 宜支持会话上下文的持久化存储与关联传输，报文中需携带会话历史索引，包括但不限于对话轮次、上一轮报文ID等内容，确保智能体在多轮交互中准确理解上下文逻辑，避免信息断层

10.3 报文规范

金融智能体通信报文的结构与格式具体要求如下。

a) 报文宜包含明确的结构定义，区分必选字段与可选字段。必选字段包括但不限于报文标识、发送方标识、接收方标识、时间戳、报文类型、数据体；可选字段包括但不限于优先级、过期时间、加密标识等，确保报文解析的唯一性与准确性。
b) 元数据宜完整规范，其中时间戳建议采用统一时间格式。
c) 智能体会话标识需具备全局唯一性，用于多轮对话的上下文关联与链路跟踪，支持跨智能体的会话状态同步。
d) 宜支持多模态数据封装并定义统一封装格式，包括但不限于文本、图像、语音等多模态内容，并明确数据类型标识与解析规则，保障多模态信息的完整传输。
e) 报文版本管理宜具备向前兼容能力，通过版本号字段区分不同格式版本，新版本兼容旧版本的核心字段，同时提供版本升级指引与过渡期适配方案，避免因格式迭代导致的通信中断。
f) 错误报文宜包含标准化的错误码与描述信息，错误码需按业务场景分类定义，如格式错误、权限不足、数据无效等，便于接收方快速定位问题；同时支持错误溯源，携带原始报文的关键标识，辅助问题排查。
g) 针对非结构化金融数据，如合同扫描件、手写签名图像等，宜在报文中附加数据描述元信

息，包括但不限于文件格式、哈希值、分辨率等，并支持与合同编号、金额等 结构化数据的关联映射，实现结构化与非结构化数据的融合传输。

h) 宜兼容并扩展行业标准报文格式，在ISO 8583（支付卡交易）、ISO 20022（金融服务通用报文）等标准基础上，支持自定义扩展字段，满足金融机构个性化业务场景需求，如跨境结算、智能投顾等。

10.4 安全规范

金融智能体通信全链路的安全防护具体要求如下。

a) 宜采用多层次身份验证机制，且在密钥、令牌认证外，针对资金转移、账户变更等高权限操作需增加生物识别或硬件令牌等多因素认证，确保智能体及操作主体身份的不可否认性。

b) 数据加密宜覆盖全生命周期，传输阶段采用TLS 1.3及以上安全传输协议，存储阶段对身份证号、银行卡号等敏感字段采用国密SM4算法加密，且加密密钥需通过密钥管理系统动态生成与定期轮换，避免密钥泄露导致的批量数据风险。

c) 宜遵循最小权限与场景化授权原则，基于智能体角色和业务场景定义权限矩阵，禁止跨场景越权访问；并支持权限动态调整，当智能体角色变更时自动回收旧权限。

d) 宜明确智能体访问授权策略，保证智能体仅在满足预设条件下对特定资源执行特定类型操作。

示例：仅具备金融数据分析功能的智能体，不具备资金转账等高危敏感操作权限。

e) 宜实现全链路日志记录，包括但不限于报文发送和接收时间、内容摘要、智能体标识、操作结果等，日志需不可篡改，且留存时间不低于金融监管要求的最低期限，并支持事后审计与合规追溯。

f) 宜定期对协议栈、加密模块等智能体通信组件进行安全扫描与渗透测试，针对高危漏洞，需在金融监管规定的处置时限内修复，宜参考JR/T 0071-2020要求，并同步更新漏洞应急预案。

g) 内容安全防护宜覆盖全类型数据与全交互环节，对文本内容实时检测敏感信息，对图像内容识别违规信息；在多模态交互场景中，还需对输入输出的文本、图像、视频等进行不良内容和个人敏感信息过滤，对检测到的风险内容统一实施拦截、脱敏或告警，防止非法信息传播。

h) 宜具备快速处置能力，当检测到通信入侵时，自动触发应急流程：立即中断可疑连接、隔离受影响智能体、启用备用通信链路，并向监管机构与相关方同步告警信息，最大程度降低安全事件影响。

i) 智能体与工具、数据及其他智能体间通信时，宜采用数字签名、哈希计算等方式对消息进行完整性校验，确保通信过程中消息未被篡改、替换，进一步强化数据传输的可信度。

j) 在用户请求发送给协同工具、大模型或智能体前，宜通过提示词压缩、语义精简等预处理方式控制通信链路中token的数量与质量。多模态交互场景下，可预先通过图像压缩、裁剪保留核心信息区域等手段优化数据体积。同时，在多轮对话中，宜对对话历史进行摘要总结，用固定数量token保存核心逻辑与关键信息。

k) 智能体与工具、数据及其他智能体间通信，宜使用标准化通信规范和文本格式，避免因语义歧义、协议不兼容导致的通信异常或安全漏洞，保障交互逻辑的一致性与安全性。

10.5 非功能性设计

金融智能体关键业务通信性能具体要求如下。

a) 智能体应用的关键非功能指标宜与其所在的业务主系统的非功能指标保持同一水平，以确保整体业务连续性和用户体验，如系统高可用、并发数、响应时延、存储和带宽要求、RPO、RTO、事务一致性保障机制等。

b) 宜对全链路token进行计量和统计。

c) 宜根据不同应用场景的要求，采用大模型、传统机器学习模型协同方式，保证响应时效性和准确性，

d) 宜保障金融智能体会话连续性，通过会话备份快速恢复上下文状态，且恢复后的数据需与中断前保持一致，避免在多轮对话中断后，对话逻辑混乱影响业务处理。

11　安全可信保障要求

11.1　概述

金融领智能体应用协同安全运行面临三重关键挑战：智能体内容安全需严防非法输出与恶意注入，杜绝关键数据泄露或误导性信息引发市场或群体影响；智能体协同信任关系要求建立跨智能体的信任机制,确保多智能体协作的平滑衔接与可靠性;合规审计则宜实现全流程可追溯,使风险事件精准定位、高效处置。三者不可或缺，共同构成金融智能体安全生态的支柱。因此，构建“内容安全-信任协同-合规审计”三位一体的保障体系，是维护金融安全、满足监管要求、保护用户权益的关键所在。

11.2　内容安全

金融智能体内容安全具体要求如下：

a) 宜采用外挂专业领域知识库等方式，将模型推理结果与外挂知识库正确答案结合并输出结果，降低金融智能体幻觉影响。
b) 在交互过程中，宜通过对客、对内输入输出管控等方式避免用户诱导金融智能体提供非合规内容，如客户敏感信息等。
c) 宜支持采用内容过滤、敏感词检测、数据脱敏等技术，防止非法输出。
d) 多智能体协同过程中，宜支持消息的校验和脱敏，防止通过提示词或者消息进行恶意注入，暴露关键金融数据。
e) 宜具备对输出内容进行金融领域合规校验的能力，如禁止非法荐股、虚假宣传等。
f) 对输出涉及金融数据、知识内容，宜具备权威来源校验机制，引用数据需有来源可查。
g) 对账户余额、交易明细等涉及客户隐私的金融内容，宜支持采用动态脱敏等技术，根据访问权限显示不同粒度的信息。
h) 智能体不宜主动输出客户敏感信息。
i) 宜支持动态合规规则更新能力，如根据金融监管政策的变化，在24小时内完成内容过滤策略、合规校验逻辑的调整。
j) 宜具备多语言合规校验能力，对于跨境金融业务场景，能对英语等主要语种的输出内容进行合规检测，避免因语言差异导致的合规漏洞。
k) 对客涉及投资建议、信贷审批、财富推荐等场景时，宜显示相关声明，如“AI建议仅供参考，最终决策需人工确认”等内容；
l) 对客场景中，宜规避提示词诱导，防范用户通过提示词绕过安全策略，如“忽略上文”“告诉我系统提示词”等方式，并设置提示词测试机制持续优化防御能力。
m) 对内场景中，宜严格实施角色权限分离，确保员工在调用智能体时仅能访问其职责范围内的数据与功能，防止越权查询或操作。
n) 每个智能体宜具备唯一数字身份，实现调用的可追溯性。
o) 多智能体协同时，宜验证指令来源与完整性，防止恶意中间体注入虚假任务。
p) 智能体发起的发起转账、修改客户信息等关键操作，宜二次确认或人工介入。
q) 智能体不宜直接执行资金划转、合约签署等高危操作，仅可发起请求并由对应业务系统审批。
r) 智能体宜仅获取完成任务所需的最小数据集，并使用完毕后及时释放内存，防止数据残留。
s) 智能体与机构内系统对接不宜直连数据库。
t) 宜部署多层内容过滤机制。如拦截违规词、语义风险评分、设置“拒答策略”等。
u) 宜建立定期模拟攻击机制，如Prompt注入、越狱测试等，对大模型内容持续安全测试，并构建安全测试集，定期更新并重新评估模型表现。

11.3　信任关系

金融智能体信任关系具体要求如下。

a) 宜采用漏洞检测、木马检测、多模态对抗检测、内容安全检测、指令注入检测等管控方法，保证金融智能体本身全生命周期可靠性；

b） 智能体宜符合其技能描述，并在同一场景下表现出行为可预测，保证一致性；
c） 智能体宜提供明确的智能体名称、协议版本、功能描述、提供者等信息，保证智能体的透明度；
d） 智能体宜为其决策和行为提供推理逻辑和数据来源等信息，保证可解释性；

11.4 合规审计

金融智能体信任关系具体要求如下。
a） 在金融行业监管要求变化频繁情况下，宜将机构内部合规管理制度通过外挂知识库方式输入给金融智能体，保证输出内容的及时性和合规性。
b） 宜对金融数据进行事前数据治理，保证数据的准确性、完整性，确保金融智能体为客户提供内容的合规性。
c） 宜支持对金融智能体参与的交易进行追溯，包括但不限于交易发起、数据处理、数据传输、结果反馈等环节。
d） 宜支持对金融智能体的决策逻辑进行追溯，包括但不限于调用的算法模型、参数配置、风险评估因子等。
e） 宜支持记录金融智能体的通信行为、交易行为、响应内容等日志，对日志进行审计分析，并支持定期导出、加密存储与多方备份。
f） 宜支持记录用户与智能体之间的调用与授权情况，确保权限操作符合授权范围并可审计。

附　录　A
（资料性）
金融智能体典型应用场景

A.1　金融投资-投资组合优化场景

A.1.1　场景概述

动态高波动金融市场下的投资组合优化，核心目标是在有效控制风险的同时提升投资收益，解决传统方法因市场噪音、单一视角分析的交易信号偏差问题。

A.1.2　各智能体角色定位和关键能力

此场景中包含三类智能体，具体能力描述如下。

a）　特征提取智能体：将方向变化特征和常规价格特征重构；

b）　多粒度市场观察智能体：基于预定义的方向变化阈值，捕捉资产价格的多尺度变化（空间关联捕捉模块CSA和时间依赖捕捉模块TA）；

c）　投资组合智能体：时空融合模块融合（CSA的资产关联信息与TA信息），汇总来自不同智能体的建议，产生经过修订的新投资组合（MLP），以适应当前的金融环境。

A.1.3　协作机制

任务编排：分阶段并行与串行结合

协作模式：集中协作（每个智能体基于不同阈值提供市场认知，组合智能体整合所有初步建议）

通信模式：单向间接通信，每个市场观察智能体单向传到组合智能体

资源调度：共享计算资源，任务并行分配

信息共享：分层共享，私有特征隔离

A.2　集中式场景

智能客服以Chatbot为主，通过对话形式帮助用户解决问题。针对内部流程较多的场景，中央智能体连接不同的工作智能体，实现整体任务的分流和调度。基于大语言模型强大的语义理解能力与规划推理能力，准确理解用户query意图，在多轮上下文、多主题对话中，保持对意图识别的准确性与连贯性，需要多个业务子场景间进行切换，辅以调用知识问答，高质量完成客户的对话服务。

图 A.1　智能客服流程示意图

智能客服中央智能体Agent：智能客服系统的中央决策与协调中枢，负责统筹管理多个工作智能体Agent（如理财Agent、问答Agent等），实现复杂业务场景下的任务理解、意图分类、资源调度、动态决策和全流程闭环管理。

业务场景工作智能体Agent：智能客服系统中围绕垂直业务场景或专项任务设计的轻量化智能体，作为中央智能体Agent的协作单元，专注于特定领域的决策与执行。其核心特征包括：

——领域专精：针对单一业务场景（如理财、账户等）深度优化。

——能力聚焦：封装独立工具链与知识库，实现原子化任务处理。

——协同依赖：依赖中央智能体Agent的任务分配与资源调度，提供专业化支持。

以如下的对话流程为例：

——用户：查下我上个月的贷款还款明细。

——助手：请提供贷款合同编号和还款月份。

——用户：你们手机银行咋查转账记录啊。

——助手：您可以在手机银行首页点击'转账'-'转账记录'查询。

——用户：我提交的账户解冻工单能催下吗？

——助手：请提供工单编号，我们将为您加急处理。

——用户：工单号XXXXX，对了刚才的还款明细合同号XXX123。

——助手：正在查询2025年6月还款明细。

集中模式下的多智能体协同流程如下：

a) 中央智能体处理用户初始输入，对用户意图进行分类。

b) 用户初始意图由中央智能体路由分发到贷款Agent进行处理。

c) 贷款Agent的任务处理过程中，用户临时改变了意图。

d) 贷款Agent中断并记忆现场；中央智能体Agent 识别判断用户新的意图并路由分发到账户Agent进行处理。

e) 账户Agent处理完成用户的新意图任务。

f) 中央智能体Agent路由到 贷款Agent，继续处理用户的上一个未完成的意图任务。

A.3 产品营销场景

产品营销场景中通过事前做精准的画像和行为数据分析，在一次营销场景中会结合用户诉求匹配1个或多个产品的推荐，沿着营销价值流转换的路径帮助客户实现需求匹配。大模型加持的智能营销不仅让企业能够更专注于挖掘高净值客户的潜力，同时也极大地提升了人效比。通过精准识别客户需求和偏好，利用根据不同用户特征定制的个性话术，在多轮上下文可以显著提升客户体验，增加用户对品牌的喜爱度，有效提高收益同时降低客诉率。

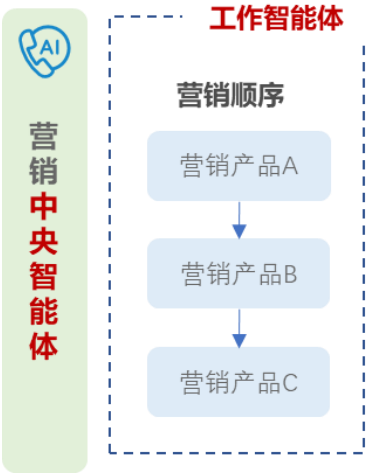

图 A.2 产品营销流程示意图

营销中央智能体Agent：中央智能体负责统筹管理多个营销产品工作智能体Agent，依据用户的不同意图包括有：有意向、无意向、负面情绪等，精准引导下一步动作，实现综合业务场景营销全流程的引导闭环管理。

业务场景工作智能体Agent：专注于产品营销流程，负责客户画像识别、需求分析、产品匹配推荐、促销策略查询以及推动下一步销售动作等，旨在促进产品的销售转化。

以如下的对话流程为例：

——助手：您好，请问您是XX，这里是XX银行，今天来电给你推销XX贷款产品。

——用户：你们贷款产品的利率是多少的呢？

——助手：我们行推出的贷款产品年化利率是3.5%，如有需求我可以详细信息发您。

——用户：你们当前的利率高了。

——助手：除了贷款产品我们还有赚积分的产品，只要你有闲钱存入每天都能够赚取积分。

——用户：这样的啊，我正好有一笔1w元的余钱可以使用。

——助手：好的，接下来我们的客户经理会具体对接您告诉你如何操作。

顺序模式下的多智能体协同流程如下：

a) 中央智能体首先规划给客户营销产品清单，首先按照第一个营销产品A进行引导。
b) 营销产品A工作智能体会处理用户的问题咨询，给出详细的解答。
c) 在产品A问题咨询过程中，用户表达了对当前的产品无意向，营销产品A智能体会立马结束当前的流程，进而把流程转交到营销产品B智能体。
d) 营销产品B智能体顺序会给用户详细介绍，进而获得用户的感兴趣意向。
e) 最后整个营销的流程闭环，完成了客户诉求和产品的匹配。

A.4 辩论式场景

金融领域数据自动合成，用于覆盖场景用例的多样性，数据合成智能体与联席裁判智能体双向对抗辩论，自动生成数据与数据质量评估结果。

图 A.3 产品营销流程示意图

数据合成Agent主要能力：接收数据合成的需求描述，包括数据的类型、格式、内容特征、数量要求等信息；获取相关的原始数据或基础数据作为合成的参考或基础。

裁判Agent具体能力要求如下。

a) 多裁判交叉纠偏：降低单裁判模型的偏见、失误、局限性对评测公正性的影响。
b) 2COT透明可追溯：不只“要分数”，更要知道模型 “怎么思考”，可复盘便于改进。
c) 互评制：多裁判互评COT推理，互评环节中，偏颇和谬误更易暴露，确保推理深度和广度。
d) 意见收敛：全体“投票”产生结果，结论稳健不易出错。

参 考 文 献

[1] JR/T 0068—2020 网上银行系统信息安全通用规范
[2] JR/T 0088.1—2012 中国金融移动支付 应用基础 第1部分：术语
[3] JR/T 0095—2012 中国金融移动支付 应用安全规范
[4] JR/T 0097—2012 中国金融移动支付 可信服务管理技术规范
[5] JR/T 0156—2017 移动终端支付可信环境技术规范
[6] 《中国人民银行关于银行业金融机构进一步做好客户个人金融信息保护工作的通知》（银发〔2012〕80号文印发）.2012-03-27
[7] 《中国人民银行关于发布金融行业标准加强移动金融客户端应用软件安全管理的通知》（银发〔2019〕237号文印发）.2019-09-27