

团 体 标 准

T/ISC XXX—XXXX

金融智能体系统可观测性技术规范

Technical Specification for Observability of Financial AI Agent Systems

在提交反馈意见时，请将您知道的相关专利与支持性文件一并附上。

（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 次

前 言	IV
引 言	V
金融智能体系统可观测性技术规范	1
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
3.1 金融智能体系统 financial AI agent system	1
3.2 可观测性 observability	2
3.3 观测信号 telemetry signal	2
3.4 链路追踪 trace	2
3.5 审计留痕 audit trail	2
3.6 工具调用 tool invocation	2
3.7 人工接管 human takeover	2
4 符号和缩略语	错误！未定义书签。
5 总体要求	2
5.1 基本原则	2
5.2 建设目标	2
5.3 分层要求	3
6 总体要求可观测性对象与观测信号	错误！未定义书签。
6.1 可观测性对象	3
6.2 日志要求	3
6.3 指标要求	3
6.4 链路追踪要求	3
6.5 事件要求	4
7 数据采集与关联要求	4
7.1 标识规则	4
7.2 数据采集要求	4
7.3 上下文关联要求	4
7.4 模型调用观测要求	4
7.5 工具调用观测要求	5
7.6 人工接管观测要求	5
8 运行可视与分析要求	5
8.1 可视化要求	5
8.2 告警要求	5
8.3 问题定位要求	5
8.4 运行分析要求	6
9 审计追溯要求	6

9.1 审计对象	6
9.2 留痕要求	6
9.3 追溯要求	6
10 安全与合规要求	7
10.1 数据安全要求	7
10.2 隐私保护要求	7
10.3 合规管理要求	7
参 考 文 献	8

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：

引 言

随着大语言模型、智能体编排框架、工具调用技术和金融数字化基础设施快速发展，金融智能体系正在客户服务、营销运营、投研辅助、风控合规、运营管理等场景中逐步落地。与传统信息系统相比，金融智能体系具有模型驱动、动态规划、多轮交互、跨系统调用和执行链路长等特征，系统运行状态更复杂，异常成因更隐蔽，责任边界更难界定。

当前，行业在金融智能体系建设过程中，普遍面临观测对象不统一、日志与指标口径不一致、跨组件链路难以贯通、工具调用行为不可还原、模型与业务执行过程缺乏统一审计留痕等问题，难以满足系统稳定运行、风险识别、合规检查和事后追溯要求。

为规范金融智能体系可观测性建设，统一关键观测对象、观测信号、数据采集、关联追踪、分析展示和审计追溯要求，制定本文件。

金融智能体系统可观测性技术规范

1 范围

本文件规定了金融智能体系统可观测性的总体要求、可观测性对象、观测信号、数据采集与关联、运行可视与分析、审计追溯以及安全与合规要求。

本文件适用于银行、证券、基金、保险及其他金融机构金融智能体系统的规划、设计、开发、部署、运行、运维、审计和评估，也适用于为金融机构提供大模型服务、智能体平台、工具系统、运行支撑平台及相关技术服务的单位。

本文件适用于单智能体、多智能体、基于大模型工作流的智能体系统，以及与外部业务系统联动执行的金融智能体系统。

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 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

GB/T 41867—2022 信息技术 人工智能 术语

GB/T 28827.1—2022 信息技术服务 运行维护 第1部分：通用要求

GB/T 43208.1—2023 信息技术服务 智能运维 第1部分：通用要求

GB/T 37736—2019 信息技术 云计算 云资源监控通用要求

GB/T 37938—2019 信息技术 云资源监控指标体系

GB/T 45288.1—2025 人工智能 大模型 第1部分：通用要求

GB/T 45288.2—2025 人工智能 大模型 第2部分：评测指标与方法

GB/T 45654—2025 网络安全技术 生成式人工智能服务安全基本要求

GB/T 45907—2025 人工智能 服务能力成熟度评估

GB/T 46347—2025 人工智能 风险管理能力评估

3 术语和定义

GB/T 41867—2022界定的术语和定义适用于本文件。

3.1

金融智能体系统 financial AI agent system

应用于金融领域，基于人工智能模型进行感知、理解、规划、决策，并能够调用工具、服务或系统资源以执行相关任务的软件系统。

3.2

可观测性 observability

通过系统产生的日志、指标、链路追踪和事件等观测数据，对系统运行状态、行为过程和异常原因进行分析、定位和追溯的能力。

3.3

观测信号 telemetry signal

用于表征系统运行状态、行为过程或异常情况的数据。

注：观测信号包括日志、指标、链路追踪和事件等。

3.4

链路追踪 trace

对请求、任务或业务过程在多个组件、多个服务或多个执行阶段之间传播路径及处理状态的记录。

3.5

审计留痕 audit trail

对系统关键操作、关键决策、关键调用及相关责任主体进行记录并保存审计证据的过程和结果。

3.6

工具调用 tool invocation

金融智能体系统为完成任务而对外部工具、接口、服务、数据库或业务系统发起的操作行为。

3.7

人工接管 human takeover

在金融智能体系统运行过程中，由人工对异常、风险或特定业务环节进行干预、确认、中止或续办的行为。

4 总体要求

4.1 基本原则

金融智能体系统可观测性建设应遵循以下原则：

- a) 全面性。覆盖模型调用、任务编排、工具调用、外部系统交互、人工接管等关键环节；
- b) 一致性。统一观测口径、字段定义、时间基准和标识规则；
- c) 关联性。实现跨组件、跨阶段、跨系统的上下文贯通和链路关联；
- d) 及时性。能够及时发现异常、触发告警并支持快速定位；
- e) 可追溯性。关键行为和关键决策应可回放、可核验、可审计；
- f) 安全性。观测数据采集、存储、传输、访问和使用应满足安全与合规要求。

4.2 建设目标

金融智能体系统可观测性应满足以下目标：

- a) 能够识别系统关键运行对象和关键业务链路；
- b) 能够采集反映系统状态和行为过程的基础观测信号；
- c) 能够支持故障发现、问题定位、影响分析和责任追溯；
- d) 能够满足内部管理、风险控制、合规检查和审计复盘需要。

4.3 分层要求

金融智能体系统可观测性宜覆盖以下层次：

- a) 模型层，包括模型请求、响应、时延、令牌消耗、拒答、异常中断等；
- b) 智能体层，包括任务接收、上下文处理、决策规划、记忆调用、执行状态等；
- c) 工具层，包括工具识别、调用参数、返回结果、失败原因、重试行为等；
- d) 业务层，包括业务节点流转、状态变化、结果确认、人工接管等；
- e) 基础设施层，包括算力、网络、存储、容器、队列、数据库等资源状态。

5 可观测对象及观测信号要求

5.1 可观测性对象

金融智能体系统至少应对以下对象进行观测：

- a) 用户请求；
- b) 会话上下文；
- c) 模型调用过程；
- d) 提示输入与输出结果；
- e) 任务规划与执行过程；
- f) 工具调用过程；
- g) 外部业务系统交互过程；
- h) 人工接管与人工确认过程；
- i) 系统资源与运行环境。

5.2 日志要求

日志应满足以下要求：

- a) 记录关键执行过程、关键状态变化、关键异常信息和关键安全事件；
- b) 采用统一字段结构，至少包括时间、系统标识、组件标识、会话标识、请求标识、事件类型、事件内容、结果状态；
- c) 日志内容应支持检索、过滤、聚合和关联分析；
- d) 日志时间应统一时钟基准；

5.3 指标要求

指标应满足以下要求：

- a) 能够反映系统可用性、性能、稳定性、成功率、异常率和资源使用情况；
- b) 应区分系统级指标、组件级指标和业务级指标；
- c) 应支持按时间窗口聚合统计；
- d) 指标定义应统一名称、口径、单位和采样频率。

5.4 链路追踪要求

链路追踪应满足以下要求：

- a) 对一次完整请求或任务执行过程进行全链路标识；
- b) 能够记录跨服务、跨组件、跨工具和跨系统的调用关系；
- c) 能够记录各阶段开始时间、结束时间、处理结果和异常信息；

d) 在多智能体协同场景下，能够反映主智能体与子智能体之间的调用关系。

5.5 事件要求

事件应满足以下要求：

- a) 对重要状态变化、异常触发、策略拦截、人工接管、权限变更等进行记录；
- b) 事件应具有明确分类、分级、来源和处置状态；
- c) 事件应支持与日志、指标和链路信息关联。

6 数据采集与关联要求

6.1 标识规则

金融智能体系统应建立统一标识规则，至少包括：

- a) 系统标识；
- b) 组件标识；
- c) 会话标识；
- d) 请求标识；
- e) 任务标识；
- f) 链路标识；
- g) 工具调用标识；
- h) 人工接管标识。

6.2 数据采集要求

观测数据采集应满足以下要求：

- a) 采集方式应与系统架构相适配；
- b) 应覆盖关键节点，不应影响核心业务正常运行；
- c) 采集频率应满足问题发现和分析需要；
- d) 采集失败、上报失败和数据缺失情况应可识别。

6.3 上下文关联要求

上下文关联应满足以下要求：

- a) 同一请求在不同组件间传播时应保持关联标识一致；
- b) 同一任务涉及多个模型调用和工具调用时应支持聚合展示；
- c) 人工接管前后的系统状态应能够关联；
- d) 业务结果应能够回溯至触发该结果的关键执行链路。

6.4 模型调用观测要求

模型调用观测至少应包括：

- a) 模型名称或模型版本；
- b) 调用时间；
- c) 响应时间；
- d) 输入长度或令牌数量；
- e) 输出长度或令牌数量；
- f) 调用结果；

g) 异常类型。

6.5 工具调用观测要求

工具调用观测至少应包括：

- a) 工具名称；
- b) 调用发起方；
- c) 调用时间；
- d) 输入参数摘要；
- e) 返回结果摘要；
- f) 调用状态；
- g) 失败原因；
- h) 重试次数。

6.6 人工接管观测要求

人工接管观测至少应包括：

- a) 接管触发原因；
- b) 接管时间；
- c) 接管人或接管角色；
- d) 接管前任务状态；
- e) 接管处置动作；
- f) 接管结果。

7 运行可视与分析要求

7.1 可视化要求

金融智能体系统应具备可视化展示能力，至少包括：

- a) 系统运行总览；
- b) 关键指标趋势；
- c) 关键事件列表；
- d) 链路追踪视图；
- e) 异常分布视图；
- f) 资源状态视图。

7.2 告警要求

告警机制应满足以下要求：

- a) 基于指标阈值、异常事件、规则命中或链路失败触发告警；
- b) 支持按严重程度分级；
- c) 支持告警抑制、聚合和升级；
- d) 支持与处置流程联动。

7.3 问题定位要求

问题定位应满足以下要求：

- a) 能够从告警快速定位到相关链路、日志和事件;
- b) 能够识别异常发生节点、传播范围和影响对象;
- c) 能够区分模型异常、工具异常、系统异常和业务异常;
- d) 能够支撑事后复盘分析。

7.4 运行分析要求

运行分析至少应支持以下内容:

- a) 成功率分析;
- b) 时延分析;
- c) 失败原因分析;
- d) 模型调用质量分析;
- e) 工具调用稳定性分析;
- f) 人工接管原因分析;
- g) 高风险事件分析。

8 审计追溯要求

8.1 审计对象

下列行为应纳入审计追溯范围:

- a) 模型调用;
- b) 关键任务规划与执行;
- c) 外部工具调用;
- d) 关键业务结果输出;
- e) 权限申请与权限变更;
- f) 人工接管和人工确认;
- g) 高风险策略命中与拦截。

8.2 留痕要求

审计留痕应满足以下要求:

- a) 内容完整;
- b) 时间准确;
- c) 责任可界定;
- d) 记录不可随意篡改;
- e) 保存期限满足监管、内控和审计要求。

8.3 追溯要求

系统应支持以下追溯能力:

- a) 从业务结果追溯到相关请求、链路和调用行为;
- b) 从异常事件追溯到触发原因、影响范围和处置过程;
- c) 从人工接管结果追溯到接管前后系统状态;
- d) 从审计记录追溯到责任主体和关键操作节点。

9 安全与合规要求

9.1 数据安全要求

观测数据安全应满足以下要求：

- a) 按照数据敏感程度实施分类分级管理；
- b) 采取访问控制、传输保护、完整性校验和备份恢复措施；
- c) 观测数据调用和导出应受权限控制并留痕；
- d) 外部共享应经过审批并符合相关要求。

9.2 隐私保护要求

观测数据处理应满足以下要求：

- a) 遵循最小必要原则；
- b) 涉及个人信息时应进行脱敏、匿名化或其他保护处理；
- c) 不应将敏感原文作为默认展示内容；
- d) 调试场景和生产场景应区分管理。
- e) 观测数据不应明文记录口令、完整银行卡号、完整身份证号等敏感信息。

9.3 合规管理要求

金融智能体系统观测性建设应符合金融监管、网络安全、数据安全和个人信息保护等相关要求，并与机构内部风险管理、审计管理和科技管理制度衔接。

参 考 文 献

- [1] NIST AI Risk Management Framework 1.0
 - [2] NIST AI RMF: Generative Artificial Intelligence Profile
 - [3] ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system
 - [4] OpenTelemetry Documentation
-