ICS 35. xxx
CCS Lxx

团　　体　　标　　准

T/ISC XXX—XXXX

智能体交互身份与访问控制安全能力要求

Security Capability Requirements for Identity and Access Control in Agent Interaction

（征求意见稿）

2026-04-17

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

中 国 互 联 网 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。

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提出并归口。

本文件起草单位：中国信息通信研究院、北京邮电大学、北京金山云网络技术有限公司、天翼安全科技有限公司

本文件主要起草人：郭雪、卫斌、马铭洋、李忠权、景韩愈

引　言

在数字经济加速演进的当下，智能体作为人机交互的新型载体，正深度融入金融、医疗、交通等各类场景，成为连接用户需求与被调用方服务的核心枢纽。智能体依托分布式算力支撑与灵活部署优势，推动各类服务向智能化、便捷化升级，其深度被调用方正重塑数字服务生态。

随着智能体交互场景的不断丰富，其跨端跨云的访问模式也带来了新的安全风险。智能体与被调用方、用户间的身份核验漏洞、权限授予不规范、授权流程不透明等问题，可能引发身份冒用、权限滥用、数据泄露等安全事件，既威胁用户隐私与被调用方的安全，也制约着智能体产业的健康发展。身份安全是智能体交互安全体系的基石，而完善的身份与访问控制机制则是保障交互全流程安全的核心支撑，更是防范上述风险的关键抓手。

为规范智能体交互过程中的安全行为，防范各类安全风险，本部分标准（智能体交互身份与访问控制安全能力要求）立足端云协同场景下智能体交互的安全核心需求，构建智能体交互访问控制安全框架，明确身份安全、授权管理、权限控制、身份标识管理、凭证生命周期管控及审计追溯等核心要求。本标准通过界定用户、智能体、被调用方三方的安全责任与操作规范，提供可落地的安全指引与参考流程，旨在筑牢端云协同智能体交互的安全防线，为智能体相关被调用方系统的设计、开发、部署及运行提供清晰的安全基线，从源头防范身份与访问相关安全风险，同时兼顾实用性与实施成本，为规范行业健康发展提供依据。

智能体交互身份与访问控制安全能力要求

1 范围

本标准规定了在智能体（Agent）参与的交互过程中，对智能体、用户和被调用方等身份进行认证和访问控制的安全要求。适用于智能体的设计、开发、部署及运行阶段的身份认证和权限管理。

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 31168-2014 信息安全技术 云计算服务安全能力要求

GB/T 32400-2015 信息技术 云计算 概览与词汇

GB/T 20281-2020 信息安全技术 防火墙安全技术要求和测试评价方法

3 术语和定义

下列术语和定义适用于本文件。

3.1 智能体 agent

具备感知环境、处理信息、执行任务和自主决策能力的软件系统，能够模拟人类智能完成特定或复杂任务。智能体通常包括模型推理、知识管理、多模态交互等模块，支持连续学习与个性化服务。

3.2 智能体交互 agent interaction

智能体与被调用方、用户之间为达到特定目的，协同作业而进行的访问和操作过程。

注1：智能体交互场景包括智能体通过接口、操作系统权限、通信协议等方式代理用户访问被调用方的数据、调用被调用方的功能等场景。

注2:被调用方包括APP、其他智能体等。

4 符号和缩略语

下列符号和缩略语适用于本文件。

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

ID 身份标识号（Identity Identifier）

IP 互联网协议（Internet Protocol）

5 智能体交互身份与访问控制安全要求框架

5.1 框架模型

智能体交互身份与访问控制安全框架由用户、智能体、被调用方三方协同构建，核心围绕身份认证、权限管控、安全审计全流程开展。框架明确被调用方与用户需分别对智能体完成授权准入，在全交互周期中，由被调用方持续对用户、智能体及被调用方的身份合法性、权限适配性进行动态验证与安全管控，同步落实身份凭证生命周期管理与操作审计追溯，全面保障用户信息安全、被调用方数据安全及交互过程合规性。智能体交互身份与访问控制安全要求框架如图1所示。

图 1 智能体交互身份与访问控制安全要求框架

5.2 基本要求

a）智能体交互过程中，应保障身份安全和权限安全，未经身份认证和权限授予的智能体，不得代理用户访问被调用方；

b）智能体在代理用户访问被调用方前，应获得用户的授权，并保障用户身份和权限的安全，防止用户和权限身份被冒用；

c）智能体在访问被调用方前，应通过被调用方的身份认证和授权；

d）智能体在访问被调用方前，应当满足透明化要求，并以可识别、可验证的方式主动声明其作为智能体的身份。声明方式应符合被调用方公开的接口规范或接入管理要求，不得以伪装用户主体的方式实施访问；

e）智能体在通过标准化接口调用的方式与被调用方协作完成任务时，应确保接口调用安全，防止接口被未经授权访问。

6 智能体交互身份安全要求

包括对智能体身份的认证、用户身份的认证以及被调用方身份的认证。

6.1 智能体身份认证

a）智能体应和被调用方协商身份认证方式，如采用基于数字证书的身份凭证、API 密钥等认证方式；

b）身份通过被调用方认证的智能体，才允许与被调用方进行交互；身份未通过被调用方认证的智能体，无法访问被调用方，被调用方可直接拒绝访问；

c）智能体应对自身的数字证书等身份凭证信息、凭证有效期等进行安全保护，防范身份凭证的泄露风险；

d）应支持与被调用方交互前，通过预设协议协商认证方式，选用数字证书、API 密钥、令牌（Token）等主流方式，适配双方需求；

e）应支持对身份凭证及有效期加密存储、权限管控，定期核查有效性，防范窃取、篡改与滥用；

f）应满足优先选择安全性适配场景的认证方式，高交互风险场景禁用弱认证方式（如简单密钥明文传输）；

g）应支持身份凭证自动更新机制，临近有效期前主动发起更新，避免因凭证失效中断合法交互；

h）应支持认证失败后给出模糊提示，仅告知 “认证失败”，不泄露具体失败原因（如凭证错误、有效期过期）；

i）应满足跨域交互场景下，采用统一认证网关校验身份，避免智能体重复认证。

6.2 用户身份认证

a）被调用方应支持对用户的多种认证机制，如账户密码、短信验证码、生物识别等；

b）对于敏感权限操作（如金融支付），被调用方可考虑对用户进行二次身份认证。如用户通过智能体发起对被调用方的某些敏感权限操作时，被调用方可先验证用户的指纹（生物识别），再验证用户手机的动态验证码，二次验证通过后才允许智能体获取该操作权限；

c）智能体应将被调用方对用户进行身份认证的方式提供给用户进行操作和授权（如拉起被调用方的用户身份认证界面等）；

d）被调用方应支持提供账户密码、短信验证码、生物识别、硬件令牌等多种认证机制，兼顾安全与易用性；

e）被调用方应满足金融支付、账户修改等高敏感操作落实二次认证，可采用 “生物识别 + 动态验证码” 组合方式；

f）智能体应支持拉起被调用方原生认证界面，清晰提示操作步骤，引导用户完成认证授权，不存储、代输用户信息；

g）智能体与被调用方应支持对用户认证信息加密存储，严禁明文留存，定期清理临时认证缓存数据；

h）被调用方应满足对弱认证方式（如纯密码）设置复杂度要求，同步提醒用户强化认证等级；

i）被调用方应支持认证会话超时控制，超时后自动失效，需用户重新完成认证方可继续操作；

j）被调用方应满足识别异地、异设备等异常认证行为，触发额外校验机制，降低被盗用风险。

6.3 智能体身份标识安全

a）智能体应具备身份标识，身份标识可由开发者 ID、智能体产品 ID、版本号、实例 ID 等组成，确保智能体身份在多开发者、多版本、多实例场景下的唯一性；

b）智能体的身份凭证应和身份标识一一映射，并对身份标识进行加密存储，避免明文泄露。

6.4 被调用方身份认证

a）应满足为每个被调用方分配全局唯一标识；

b）应支持为被调用方配置独立身份凭证（如 API Key），并存储于安全介质（如密钥管理服务或加密环境变量）；

c）应满足智能体调用被调用方时，验证被调用方凭证的有效性；

d）应支持限制被调用方认证方式（如仅允许使用带有效期的 JWT）；

e）应支持被调用方停用或升级时，旧版本凭证在 ≤7 天内逐步失效。

7 智能体交互访问控制安全要求

7.1 被调用方对智能体的授权要求

智能体在访问被调用方前，应向被调用方申请权限并获得被调用方的授权。只有通过授权的智能体，才允许访问被调用方。

7.1.1 权限申请流程

a）智能体需通过被调用方提供的权限申请渠道提交申请，申请材料应明确权限用途、使用场景、预期调用频率、开发者资质证明（如营业执照、ICP 备案）、安全能力说明（如智能体的权限管控机制、数据加密措施）等内容；

b）被调用方在审核通过后，应向智能体颁发相关的权限授权凭证（如包括权限列表、有效期、调用限制）；

c）被调用方审核不通过的，智能体可向被调用方咨询原因（如权限与智能体功能无关、安全能力不达标）；

d）智能体的权限变更应重新获得被调用方的审核。智能体的权限被撤销后，被调用方应保障智能体无法使用已撤销的权限再次访问被调用方。

7.1.2 权限安全管控和风险监测

a）被调用方可根据自身安全需要，将权限进行分级划分，明确不同类型的智能体可申请的权限范围。例如：可按敏感权限（如转账、支付、修改密码）、一般权限（如查询公开信息、调用基础功能）等进行划分；

b）被调用方可对智能体的访问频率进行限制；

c）被调用方可根据自身安全需要，对智能体授予相应的权限级别；对智能体的权限授予应坚持最小必要原则，防范权限的滥用风险；

d）被调用方为区分自然人用户访问或智能体访问，可按照自身安全管理要求设置合理、必要的核验措施。智能体在访问被调用方时，应当遵循被调用方的核验规则，不得通过模拟用户行为、伪造交互事件、自动生成用户响应或其他方式绕过核验措施；

e）被调用方可建立智能体权限滥用等风险监测机制，对智能体的权限调用行为进行安全分析，如发现“未授权访问”、“超范围调用”等异常行为，可停止智能体权限的使用；

f）被调用方可对智能体权限调用情况进行日志记录，日志内容可包括调用时间、调用者、权限类型、操作内容等。

7.2 用户对智能体的授权要求

智能体在代理用户访问被调用方前，应向用户申请权限并获得用户的授权。同时用户应向智能体授予满足实现用户代理任务（如读取被调用方数据，操作被调用方功能）的最小权限范围。

7.2.1 用户授权交互要求

a）智能体向用户申请权限时，可通过可视化等方式展示所需的权限详情，包括“权限名称、权限用途、风险等级、授权有效期选项（如单次有效、7 天有效、长期有效）”；

b）用户授权应通过显式操作完成（如点击“确认授权”按钮、录入指纹），禁止智能体实施默认勾选或隐性授权；

c）智能体需支持细粒度权限控制（如“仅允许智能体查看近 3 天的日程，不允许修改”），允许用户授予智能体的权限需与操作意图严格匹配；

d）智能体应在获取用户授权后，记录用户对权限授予的情况。

7.2.2 授权撤回机制要求

a）智能体应提供用户授权撤回机制，以方便用户撤回已授予的权限。撤回后智能体应立即失去对应的权限，智能体不应利用用户已撤回的权限访问被调用方。

7.2.3 权限匹配推荐要求

a）智能体可具备权限智能推荐功能，根据用户的操作意图自动推荐最小必要权限。例如，用户要求“帮我预订明天的酒店”，智能体仅推荐“酒店查询、房型选择、订单创建”权限，不推荐“酒店会员信息修改、历史订单删除”权限。

7.2.4 用户授权教育

a）智能体需在用户首次授权时，通过图文或视频教程等方式向用户说明 “权限最小化原则”、“授权风险点”等授权相关安全提示。

7.3 访问权限控制

7.3.1 智能体访问权限控制

a）应满足每个智能体仅被授予其任务所需的最小操作权限集，不得拥有超出业务范围的访问能力；

b）应支持智能体权限与其身份标识强绑定，确保其只能调用被明确授权的外部服务；

c）应满足在金融、医疗等高风险场景中，智能体对敏感数据的操作必须经过显式授权审批；

d）应支持多实例部署下的权限隔离，防止同一产品不同实例间相互越权访问；

e）应满足智能体权限变更后实时生效，禁止依赖缓存导致权限延迟更新。

7.3.2 用户访问权限控制

a）应满足在智能体代理用户发起请求时，必须基于原始用户身份进行权限判断；

b）应支持用户仅能访问其本人或被明确授权的数据，禁止跨用户越权操作；

c）应满足在金融支付、医疗健康等场景中，强制校验操作账户与当前用户的一致性；

d）应满足用户权限变更（如角色调整）后立即生效，即使通过智能体发起请求也受新策略约束。

7.3.3 被调用方访问权限控制

a）应满足每个被调用方仅暴露其声明的功能接口，禁止提供未授权的操作入口；

b）应支持被调用方被调用的权限与调用方身份强绑定，仅允许被授权的智能体或用户调用；

c）应满足在敏感操作中，被调用方对接口参数实施合法性与范围校验（如金额、数据类型）；

d）应支持被调用方版本升级时权限收敛，旧版本高危权限不得自动继承至新版本；

e）应满足被调用方停止被调用后，所有调用方立即失去访问能力，返回明确拒绝响应。

8 身份标识安全

8.1 智能体身份标识

a）应支持具备唯一身份标识，确保多场景下可区分；

b）应支持身份标识加密存储，严禁明文留存；

c）应满足身份标识具备不可篡改特性，生成后如需变更，经被调用方校验并同步更新绑定凭证；

d）应支持多实例部署场景下动态分配实例 ID，避免与其他实例标识冲突；

e）应支持身份标识纳入智能体签名信息，供被调用方校验身份合法性与完整性；

f）应满足身份标识变更后，同步更新全链路关联数据，确保标识一致性。

8.2 用户身份标识

a）应满足用户身份标识采用不可预测格式（如 UUID vd）；

b）应支持在日志/监控中脱敏显示，避免暴露手机号等敏感信息；

c）应支持跨系统交互通过标准化协议传递标识；

d）应满足身份标识不得与原始个人敏感信息混存于同一字段。

8.3 被调用方身份标识

a）应满足被调用方身份标识采用固定结构；

b）应支持被调用方标识在注册时声明其认证方式要求；

c）应满足被调用方版本迭代时生成新标识，旧标识 7 天内失效；

d）应支持标识变更后同步更新全链路关联数据。

9 身份凭证生命周期管理

a）应满足按生成、分发、使用、更新、吊销、注销全流程管理身份凭证，每环节留存操作记录；

b）应支持凭证分发采用加密通道传输，避免中途泄露，接收后立即校验完整性与合法性；

c）应满足智能体被注销或失权时，及时吊销其身份凭证，同步删除对应绑定关系；

d）应满足合理设置凭证有效期，超期自动失效，且不支持回溯使用，降低泄露后风险；

e）应支持智能体与被调用方的身份凭证进行加密备份，应急恢复需经双重审批；用户身份凭证不得以任何形式备份或还原；

f）应满足定期清理废弃凭证，建立废弃清单并留存销毁记录，做到全程可追溯。

10 身份认证审计与追溯

a）应支持记录身份认证日志，包含认证时间、主体、方式、结果、IP 地址等信息；

b）应满足审计日志留存至少 90 天，支持按主体、时间范围、操作类型进行查询；

c）应支持多次认证失败时触发临时锁定机制，同步向用户推送异常提醒；

d）应满足审计日志禁止擅自篡改、删除，重要操作附加操作人员身份标识；

e）应支持审计日志格式化输出，适配常见安全分析工具，便于批量排查问题；

f）应满足对日志访问权限严格管控，仅授权人员可查询、导出，防止日志信息泄露。