ICS 35.020
CCS L77

T/ SIA

中 国 软 件 行 业 协 会 团 体 标 准

T/SIA065-2025

智能体行为安全要求

Security requirements for action of artificial intelligence agent

2025-10-10发布　　2025-10-10实施

中 国 软 件 行 业 协 会 发 布

目 录

前 言

智能体是指以智能终端为载体，以大模型为决策核心，为用户提供智能服务的智能应用。它能够感知用户、环境、系统信息，理解用户意图和需求，决策如何为用户选择和提供优质服务，并基于智能体能力执行任务。智能体在提供高效便捷服务的同时，也带来了诸多与用户权益密切相关的风险与挑战。为了贯彻落实《数据安全法》《个人信息保护法》《民法典》《生成式人工智能服务管理暂行办法》等法律法规与政策文件要求，建立起以用户为中心的智能体行为治理体系，中国软件行业协会特起草《智能体行为安全要求》，明确智能体行为安全的基本原则和智能体告知同意、权限申请、行为协作、行为运行和用户权益保障要求等方面的标准要求，规范智能体安全健康发展的同时平衡全产业利益和保障用户合法权益。

本标准按照GB / T 1.1-2020 给出的规则起草，由中国软件行业协会提出并归口。

起草单位：北京航空航天大学、北京交通大学、北京工业大学、北京信息科技大学、北京物资学院、中国科学院自动化研究所、北京中软国际教育科技股份有限公司、软通动力信息技术（集团）股份有限公司、南京云信达科技有限公司、宁波优策信息技术有限公司、北京乐鑫科技有限公司、中油油气勘探软件国家工程研究中心有限公司、上海幕库科技发展有限公司、北京软件和信息服务业协会、上海市软件行业协会、天津市软件行业协会、河北省软件集成电路与人工智能协会、山西软件行业协会、内蒙古软件行业协会、辽宁省软件行业协会、大连软件行业协会、吉林省软件行业协会、黑龙江省软件与信息服务业协会、江苏省软件行业协会、浙江省软件行业协会、安徽省软件行业协会、福建省软件行业协会、山东省软件行业协会、河南省软件行业协会、湖北省软件行业协会、湖南省软件行业协会、广东软件行业协会、广西软件行业协会、海南省软件行业协会、四川省软件行业协会、云南省软件行业协会、陕西省软件行业协会、宁夏信息产业协会、新疆维吾尔自治区软件行业协会、贵州省信息技术服务业协会、新疆生产建设兵团软件行业协会、重庆市软件行业协会、宁波市软件行业协会、厦门市软件行业协会、青岛市软件行业协会、深圳市软件行业协会。

主要起草人：陈宝国、高祥、陈乃月、方娟、陈雯柏、宋燕星、陈波、王晓华、赵文静、黄志荣、王建平、黄景宋、孙鹏远、徐少波、孙斌、许珂、蔡伟、李云芝、姚顺义、赵原、刘显富、王迪、秦健、吕彦伟、徐维科、夏冰莹、杨岚、董先权、施政、陈菲菲、李书利、李智勇、喻晖、吕晖、郑明德、张苗苗、邓小华、李巡生、罗惠芳、白丽梅、刘靓、刘杰、柴冬海、刘谦、金励君、周晓瑜、韩鑫峰、郑飞、毛伟、张然、曾雪征。

本标准的某些内容可能涉及专利，本标准的发布机构不承担识别这些专利的责任。

本标准为首次制定。

智能体行为安全要求

1 范围

本文件确立了智能体行为安全的基本原则，提出了智能体告知同意、权限申请、行为协作、行为运行和用户权益保障要求等内容。

本文件适用于提供智能体设计、开发和运营的各类主体。

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 25069 信息安全技术 术语

GB/T 42884 信息安全技术 移动互联网应用程序（App）生命周期安全管理指南

ISO/IEC 22989 信息技术 人工智能 人工智能概念和术语（Information technology - Artificial intelligence - Artificial intelligence concepts and terminology）

3 术语和定义

GB/T 25069、GB/T 42884界定的以及下列术语和定义适用于本文件。

3.1

智能体 artificial intelligence agent

能够感知和响应环境并能执行操作以完成其目标的自动化实体。

注：本文件仅指运行在移动智能终端、PC终端、智能可穿戴设备上由终端厂商或应用厂商提供的、涉及与第三方APP 协作完成任务的智能体。

[来源： ISO/IEC 22989:2022，3.1.1，有修改]

3.2

应用软件 application

运行在智能终端上向用户提供信息服务的应用软件。

注1：智能终端包括移动智能终端、PC终端以及智能耳机、智能手表等可穿戴设备。

注2：包含智能终端预置应用、小程序、快应用以及互联网信息服务提供者提供的可以通过网站、应用商店等应用分发平台下载、安装、升级的应用程序，简称App。

[来源：GB/T 42884—2023, 3.2，有修改]

4 安全原则

4.1 告知同意

智能体在启用前应明确告知用户其能力边界、操作范围、所申请的操作系统权限、访问用户数据的目的、处理方式和潜在影响等内容，并获得用户同意。

4.2 公开透明

智能体应对用户公开透明其行为执行过程并提供行为执行记录。

4.3 用户可控

智能体应确保相关行为能力支持用户自主开启、关闭或随时终止，第三方APP不应要求额外授权，妨碍用户自主可控。

4.4 最小必要

智能体应确保对用户数据的收集和处理做到最小必要，遵循数据处理端侧优先原则。

4.5 安全保障

智能体应设置行为安全底限，不应代理用户执行涉及用户财产（含虚拟）和人身安全的操作（相关操作举例参见附录A）。

5 安全要求

5.1 告知同意要求

智能体行为应遵循对用户的告知同意要求，包括：

a)智能体在启用前应明确告知用户其能力边界、操作范围、所申请的操作系统权限、访问用户数据的目的、处理方式和潜在影响等内容，并获得用户同意；

b)智能体应控制告知的频率，避免告知频率过高，导致对用户造成困扰；

c)智能体在任务执行过程中需获取新的用户个人信息，扩展操作范围或获得新的系统权限，应再次告知用户并获得用户同意；

d)告知方式宜满足多样性原则，即告知方式宜采取多种形式，更易于用户感知与理解。此外，应考虑到用户的特定类型（例如面向残障人士）和特殊使用场景（例如开车过程中），考虑通过目标个人信息主体可接受的方式进行告知和获得同意。

5.2 权限申请要求

智能体应遵循权限申请要求，包括：

a)智能体应遵循系统现有的应用权限管理体系，在现有应用权限管理体系的管控下，调用相关的系统的权限；

b)智能体应仅申请任务执行行为所必须的最小权限，权限开通应经过用户明示同意；

c)智能体应将用户申请的权限与设备和账号双重绑定，不应在同一终端默认用户跨设备或跨账号权限继承；

d)智能体已申请的权限，可通过终端权限配置管理页面进行查看、变更或撤销。当用户撤销权限后，不应影响其它业务。

5.3 行为协作要求

智能体行为与第三方APP协作完成任务要求，包括：

a)智能体通过端侧或云侧的标准化接口调用的方式与第三方 App 或第三方云服务协作完成任务时，应确保接口调用安全，防止接口被未经授权访问；

b)智能体在进行用户意图识别、通过第三方 App 执行任务时，应严格遵循“用户可控”原则，即智能体应支持用户自主开启、关闭或随时终止通过第三方 App 执行任务，第三方APP不应要求额外授权，妨碍用户自主可控；

c)智能体通过系统能力操作第三方 App 完成任务时，应严格遵循“告知同意”和“安全保障”原则，对于涉及用户敏感操作的行为（参见附录B），还应提供提醒界面并支持用户随时停止智能体行为。

5.4 行为运行要求

智能体行为运行应满足以下要求，包括：

a)智能体应确保仅操作授权范围内的资源，仅收集和处理经过用户同意的用户数据，仅进行用户同意的行为；

b)智能体应采取安全措施防止行为运行能力被恶意网络攻击工具利用；

c)智能体应对用户指令接收和运行过程进行记录，并支持用户回溯查看；

d)智能体行为运行应支持用户可控， 允许用户修改、终止智能体行为和进行人工接管。

5.5 用户权益保障要求

智能体行为应遵循以下用户权益保障要求， 包括：

a)智能体应确保用户数据安全，仅收集和处理完成任务所必需的数据，向用户明确告知收集数据的目的及处理方式并获得用户同意，涉及向第三方共享数据的需额外说明；

b)智能体应支持用户自主查看、删除用户个人信息，不应设置障碍阻碍用户行使权益；

c)智能体开发运营者应建立安全管理机制，确保智能体在开发、部署、升级迭代等全生命周期安全；

d)智能体在处理用户敏感数据时应优先在端侧完成，降低用户敏感数据泄露风险；

e)应建立有效的用户投诉和反馈机制，及时解决智能体行为安全问题。

附 录 A
（资料性）
涉及用户财产（含虚拟）和人身安全的操作

涉及用户财产（含虚拟）和人身安全的操作包括：
账号操作：含账号注销、账号改密码、三方授权操作等；
金融交易：含股票、基金、期货、贷款交易等；
支付操作：含立即支付、免密支付、转账、发红包等；
协议类操作：含用户授权、用户同意、电子签章等；
不可逆的数据删除：恢复出厂设置、清空回收站、彻底删除等；
安防控制类：智能门/门锁/保险箱/车辆解锁。

附 录 B
（资料性）
涉及用户敏感操作的行为

涉及用户敏感操作的行为包括：

a）信息获取和挖掘：包括终端传感器（摄像头、麦克风、定位、光线、陀螺仪）数据、用户数据（相册、联系人、短信、通话记录、闹钟、日程、文件、录音、应用数据）、屏幕信息、运动健康数据、智能家居数据、车辆信息等；

b）应用学习：操作路径抽象建模，记录并学习用户使用应用的方式及其过程中产生的信息：酒店偏好、航司偏好、饮食习惯、偏好店铺等；

c）下单（不含支付）：购物、打车、航旅预定等；

d）社交类：发送即时通信、发送邮件、社区留言、发表评论等；

e）设备控制：控制加湿器、净化、空调、照明、厨房、卫浴、影音娱乐、插座开关、车辆座椅等；

f）健康监测：体重监测、血压监测、睡眠监测等。

参考文献

[1] GB/T 41391—2022 信息安全技术 移动互联网应用程序（App）收集个人信息基本要求
[2] YD/T 6221—2024 移动应用软件个人信息保护要求和评估方法
[3] 中华人民共和国个人信息保护法
[4] 中华人民共和国数据安全法