团　　体　　标　　准

T/ZGCSC 031-2026

智能体构建技术要求

Technical requirements for construction of agents

2026－06－12 发布　　2026－06－13 实施

中关村智慧城市产业技术创新战略联盟　　发 布

目录

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。

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。

本文件由中关村智慧城市产业技术创新战略联盟提出并归口。

本文件起草单位：中国联通智能城市研究院、国家信息中心、国家工业信息安全发展研究中心、联通数字科技有限公司、中电科发展规划研究院有限公司、中国经济信息社、北京国际大数据交易有限公司、北京邮电大学、中电空间（北京）科技有限公司、北京市社会科学院、中国机电设备招标中心、北京国脉互联信息顾问有限公司、联通（江西）产业互联网有限公司、北京建筑大学、中电信数智科技有限公司、中国雄安集团有限公司、中国通信建设集团设计院有限公司、中国通信服务股份有限公司、中国移动通信有限公司研究院、中移九天人工智能科技（北京）有限公司、中关村实验室、成都师范学院、中国科学院新疆天文台、联通（江苏）产业互联网有限公司、北京航空航天大学、中国雄安集团数字城市科技有限公司、杭州市北京航空航天大学国际创新研究院（北京航空航天大学国际创新学院）、上海森栩医学科技有限公司、北京中微盛鼎科技有限公司、北京工业大学、华北水利水电大学、中国盐湖工业集团有限公司、中国联合网络通信有限公司黑龙江省分公司、北京中科赛新智能技术有限公司、山东工程职业技术大学、中国技术经济学会通信技术经济专业委员会、北京森栩医学科技有限公司、国网北京市电力公司信息通信分公司、江西省检验检测认证总院检测认证技术发展研究院、北京中亿汇智科技有限公司、中科星图空间技术有限公司、火焰山（南京）人工智能科技有限公司、上海市安装工程集团有限公司、久瓴（上海）智能科技有限公司、七腾科技(北京)有限公司。

本文件主要起草人：郭中梅、单志广、董正浩、杜鹏、唐红梅、徐清源、张丹、孙亮、白喆、张延强、马潮江、石会昌、宋超、苏泳睿、盛晶、高枫、谭伊舒、胡璐锦、李征仁、谢悦、杜瑜、夏乐乐、任连嘉、龚振炜、杨军、刘小林、祝婷婷、孟祥宏、上官声标、陈栩、蔡丹旦、孙志敏、宋心荣、张明、卢晓慧、强薇、陈多思、唐菁、梁卓、张鑫、张拯、冯杰、周环珠、李帅峥、曹杰、孙睿、郭阳勇、张亮、于海滨、郭宇、李倩倩、李伊凡、张春芳、孙春兰、刘琪、李玮、狄晓靓、王丽影、常琳、王鹏、谢士琴、庄士刚、单斐、高颖艳、周雨芹、张建桁、高长伟、兰洪浩、金婧、王剑、郭萍、朱涵钰、邓成明、赵雅晴、祝欣越、王璇、武通、曾传鑫、沈华、张哲、吴诚林、陈晨、唐文忠、王博威、戴岩、骆一阳、刘雅琼、盛浩、黄坚、温冬梅、肖润宇、刘闯、张静、邱淼、张国栋、孟令涛、翁剑成、赵飞、梁昌金、崔丹丹、李扬、陈亮、陆逞赢、曹正海、马玉英、林鹏飞、刘纯明、曹喆、崔正龙、刘志阳、沈晓东、程彦龙、尚卫、王冬宇、刘泽宇、滕思宇、孔利加、王汇、尤勇敏、张宁、荣思博、许银环、付丽芹、吴鹏、田西南。

1 范围

本文件规定了构建智能体的技术能力基本要求，明确了智能体系统在感知、记忆、决策、执行和交互层面的能力要求和非能力要求。

本文件适用于通用型智能体系统的设计、开发、测试和集成。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是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 35273—2020 信息安全技术 个人信息安全规范

GB/T 46347—2025 人工智能 风险管理能力评估

YD/T 4929—2024 面向多智能体系统的计算平台技术要求

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

一个处于环境之中，可以感测环境并且执行相应的动作，同时逐渐建立自己的活动规划以应对未来可能感测到的环境变化的实体。

[来源：YD/T 4929—2024，3.1]

3.2

实体 entity

抽象概念，表示任何可能发送或接收信息的硬件或软件。

[来源：YD/T 4929—2024，3.1]

3.3

工具 tool

指智能体为完成特定任务而调用的外部功能模块或服务，包括函数、API接口、软件服务、硬件设备等。

3.4

多模态信息 multimodal information

指文本、音频、图像、视频等多种类型的信息集合，包括代码、结构化数据、半结构化数据、交互模态信息等。

3.5

长期记忆 long-term memory

智能体中用于长期存储领域知识、通用常识等信息的记忆模块。

3.6

短期记忆 short-term memory

智能体中用于临时存储单次任务上下文、交互历史的记忆模块，其内容通常在任务结束后被清理。

3.7

群体协同 multi-agent collaboration

多个智能体基于协作的方式共同完成复杂任务，实现群体智能的效果。

4 总体框架

参照国家网信办等部门印发《智能体规范应用与创新发展实施意见》，本文件对智能体构建给出定义，即：通过能力建设、系统集成等工程方法，创建具备感知、记忆、决策、交互与执行等能力，并支持跨系统集成和开放服务的智能体系统。本文件中对智能体构建的技术要求有能力要求和非能力要求两个方面，如图1所示。

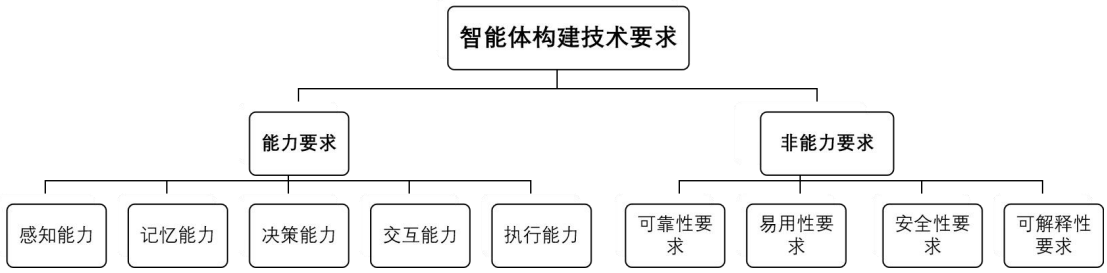

图1 标准总体框架

能力要求聚焦智能体的感知、记忆、决策、交互和执行能力。感知能力让智能体能够从用户交互和环境中获取并理解多模态信息，为后续处理提供输入。记忆能力支持智能体对知识进行长期沉淀、对任务上下文进行短期存储，并实现长短期记忆的动态融合与调用。决策能力让智能体能够对任务进行目标拆解、策略选择并动态调整方案。交互能力赋予智能体与用户、平台、工具以及其他智能体之间进行信息交换和任务协作的能力。执行能力确保智能体能够自主驱动自身能力或外部工具，生成具体成果。

非能力要求则关注智能体的工程约束，从可靠性、易用性、安全性、可解释性等四个维度提出要求，确保智能体在真实场景运行过程中的质量。

能力要求和非能力要求相互补充，共同构成完整的智能体构建技术要求体系。

5 能力要求

5.1 感知能力

5.1.1 多模态信息获取能力

多模态信息获取能力具体要求如下：

a） 应支持文本、音频、图像、视频等多类型信息接入；

b） 应支持对接入的多模态信息进行预处理，包括噪声过滤、格式规整、信息提取等，确保信息的有效性与可用性；
c） 应支持跨模态信息的语义一致性校验，确保多模态信息的逻辑一致；
d） 应支持多模态信息间的关联处理，如将语音指令与对应的场景图像关联、文本描述与结构化数据关联，形成完整的信息链路，避免信息孤立。

5.1.2 用户意图识别与理解

用户意图识别与理解能力具体要求如下：
a） 应能准确识别用户明确发出的命令式、疑问式、请求等指令，提取指令中的关键要素；
b） 应能够结合用户历史交互记录、当前任务场景，挖掘其未明确表述的潜在需求；
c） 应能通过主动问询、上下文推测等方式修正意图，确保理解的准确性；
d） 应支持通过多轮交互、概率模型、候选意图等方式消除歧义，能明确置信度阈值供用户确认；
e） 应支持多意图识别，识别用户提出的多个意图并明确优先级，避免意图混淆；
f） 应能处理意图变更，当用户在任务执行过程中修改需求时，能及时识别变更意图，暂停原任务流程并调整后续执行方案，同步告知用户变更影响；
g） 应支持历史意图模式复用，在相似场景下快速匹配用户需求，减少重复交互。

5.1.3 环境状态感知与适配

环境状态感知与适配能力具体要求如下：
a） 应能感知服务运行、环境参数、网络、数据链路、设备、系统资源等状态信息以及安全预警信号，实时捕捉环境变化；
b） 应将感知到的环境状态与自身任务关联，确保任务在环境变化下仍能正常执行；
c） 应支持环境状态预警，当感知到异常状态时，能及时生成预警信息，通过弹窗、短信、邮件等方式提示相关人员，或触发预定义的应对措施，避免影响任务执行或引发安全风险；
d） 应具备环境状态历史分析能力，定期统计环境状态数据，分析状态变化趋势，为任务优化提供数据支撑。

5.2 记忆能力

5.2.1 长期记忆构建与管理

长期记忆构建与管理能力具体要求如下：
a） 应能对专业知识、通用常识、操作规范、历史经验等信息进行持久化存储，涉及敏感信息的，应能进行脱敏、加密或访问控制等处理；
b） 应能基于本体模型、知识模型或语义模型，对长期记忆内容进行分类、标注与关联，形成结构化知识体系；
c） 应建立知识冲突识别与处理机制，根据冲突来源、版本、适用场景等信息自动或人工确定优先级；
d） 应支持长期记忆的更新维护，包括人工补充、系统同步或规则化更新，同时更新前能检测知识冲突，避免错误或不一致信息入库；
e） 应支持长期记忆检索，检索方式宜包括关键词检索、语义检索、场景关联检索和多条件组合检索，检索结果宜包含知识内容、关联知识及适用场景等信息；

f） 应具备知识安全性保障能力，能对敏感知识进行加密存储与访问权限控制，防止未授权访问或泄露；
g） 应支持建立知识老化评估模型，基于历史交互数据与任务执行结果进行自适应学习，自动标记待更新老化知识，并触发人工审核；
h） 应支持用户对生成的方案进行评价，以优化长期记忆内容。

5.2.2 短期记忆维护与应用

短期记忆维护与应用能力具体要求如下：
a） 应能在单次任务或对话过程中，完整存储交互内容、中间处理结果、当前任务参数等上下文信息；
b） 应能实时维护任务的当前执行状态，确保状态记录与实际进度一致，便于用户与智能体掌握任务进展；
c） 任务结束或对话终止后，应能按需对短期记忆数据进行清理、归档或转存等操作，避免无效数据长期占用资源；
d） 应支持短期记忆的快速调用，避免任务执行中的重复获取或处理；
e） 应具备多任务短期记忆隔离能力，当同时处理多个任务时，各任务的短期记忆需独立存储、互不干扰，避免因记忆混淆导致任务执行错误。

5.2.3 记忆融合与调用

记忆融合与调用能力具体要求如下：
a） 应建立长期记忆与短期记忆的关联机制，结合当前任务场景实现两类记忆的协同调用；
b） 应能基于本体模型或知识模型，对长期记忆与短期记忆中的相关内容进行语义匹配、关系映射和上下文对齐；
c） 应能基于融合后的记忆信息生成与当前任务情境相匹配的认知内容，为决策与执行提供支撑；
d） 应能识别长期记忆与短记忆息的冲突并提示，优先以短期记忆为准；
e） 应支持融合结果的动态更新，当短期记忆或长期记忆更新时，能重新融合并同步通知关联任务模块，确保认知内容与最新记忆信息一致；
f） 应支持在群体协同场景下融合多智能体相关记忆信息，包括协作历史、分工模式、交互结果和异常记录，为协同决策与执行提供支撑；
g） 应具备融合记忆的复用能力，对相似任务可复用历史融合逻辑，减少重复计算，提升任务响应速度。

5.3 决策能力

5.3.1 目标分析与拆解

目标分析与拆解能力具体要求如下：
a） 应能准确解读用户给定的任务目标，明确目标的核心要求、约束条件与预期成果；
b） 应支持基于领域本体模型识别任务目标中的关键对象、属性和约束；
c） 应能对目标进行可行性评估，结合当前环境状态，分析实现目标所需的资源、潜在风险、依赖条件；
d） 应能将复杂目标按逻辑关系拆解为若干个可独立执行的子任务，明确各任务的范围、目标、输入输出、执行主体；

e） 应能分析各子任务之间的依赖关系，确定执行顺序，为任务调度提供依据；
f） 应支持根据任务复杂度和约束条件调整目标拆分粒度，对简单任务可减少拆解层级，对复杂任务可细化拆解层级。

5.3.2 执行策略的制定与选择

执行策略的制定与选择能力具体要求如下：
a） 应支持为每个子任务生成多种可行的执行策略，针对不同场景设计差异化执行策略，执行策略能覆盖不同实现路径，确保单一策略失效时可切换备选方案；
b） 应能对候选执行策略进行执行前评估，并形成评估结果，评估维度包括本体模型的规则和约束、执行效率、资源消耗、风险程度、可行性及环境适配度；
c） 应能综合策略评估结果、约束条件、业务优先级等因素，选择与当前任务目标、资源状态和环境条件相适配的执行策略，并给出决策依据；
d） 应支持记录策略生成、评估、选择及变更过程中的关键信息，为后续复盘、审计和优化提供依据。

5.3.3 方案自适应调整

方案自适应调整能力具体要求如下：
a） 应能在任务执行过程中实时监测系统状态变化，通过实时交互、日志、状态反馈等方式及时了解影响方案执行的系统状态及环境态势，自适应动态优化方案；
b） 应支持设定清晰的方案调整触发条件，条件能量化且可执行，满足条件时自动启动调整流程，避免被动等待；
c） 应能基于当前可用资源及环境状态对既定方案进行动态调整；
d） 应遵循最小调整原则，优先保留仍然有效的执行步骤、资源配置和依赖关系，仅对受影响的子任务、执行路径或策略进行调整；
e） 应支持向用户或相关系统同步方案调整信息，记录调整过程中的关键信息，包括触发原因、调整内容、调整依据、影响范围和执行效果等，并形成变更报告。

5.4 交互能力

5.4.1 接口互操作

应具备标准化的接口互操作能力，保障跨系统接入协同，具体要求如下：
a） 应支持主流接口协议适配，自动调整接口参数与数据格式，兼容不同技术架构系统；
b） 应具备接口数据标准化能力，可自主将交互数据、业务数据转换为通用格式，同时解析接收的异构数据并校验完整性、一致性，确保跨系统数据互通准确；
c） 应具备接口安全互操作能力，具备SSL/TLS加密能力、身份鉴权机制、访问控制机制，防范数据泄露、篡改及非法调用风险；
d） 应支持接口异常自主处置，覆盖调用超时、协议不兼容、数据传输失败、接口降级等场景，保障接口协同不中断。

5.4.2 人机交互

应具备与用户进行自然、高效、可信赖的交互功能，具体要求如下：
a） 应能维持连贯、逻辑一致的多轮对话，具备意图澄清、上下文指代消解、话题管理等能力，确保交互顺畅自然；
b） 应支持语音、手势等多种人机自然交互方式；

c） 应支持多语言输出，且语言表达准确、无歧义，适配不同地区用户需求；
d） 应能记忆用户的偏好和历史交互模式，并在此基础上调整交互风格、信息详略程度和任务执行策略；
e） 应能完整记录交互过程日志，包括交互发起方、时间、内容、结果，日志能分类存储且可追溯。

5.4.3 系统集成

应能够被灵活集成到现有业务流程、平台和应用系统中，具体要求如下：
a） 应支持功能模块进行独立部署和升级，模块间的依赖关系应明确定义并记录；
b） 应支持与各类系统的集成对接，包括数据平台、业务系统、硬件设备，实现跨系统数据流通与业务协同；
c） 应提供标准化的集成接口，并配有完善的API文档和软件开发工具包；
d） 应满足与主流芯片、操作系统、数据库的兼容性要求，并提供兼容性测试方法及适配指南。

5.4.4 群体协同

具备在多智能体场景下围绕共同目标开展协同作业的能力，既包括发起协同、进行任务分配、协调调度的能力，也包括作为协同成员接收任务、执行分工、共享状态的能力，具体要求如下：
a） 应支持群体协同过程中的任务调度与协作编排，能够根据任务优先级、能力匹配情况、资源状态及环境变化动态调整协同方式；
b） 应支持协同过程中的信息共享与状态同步，确保各参与智能体能够获取完成本职任务所需的必要上下文信息，并保持协同过程一致性；
c） 应能够接收其他智能体分派的任务，识别自身在协同过程中的角色定位、任务目标、执行边界和完成要求，并在自身能力、权限和资源范围内完成相应任务；
d） 应支持多智能体之间的冲突检测与消解，覆盖目标冲突、资源冲突、指令冲突、结果冲突等场景，避免协同失效或任务偏离。

5.5 执行能力

5.5.1 自主行动生成与执行

自主行动生成与执行能力具体要求如下：
a） 应能根据任务需求生成多样化的自主行动，包括自然语言回应、内容创作、系统操作、逻辑计算；
b） 应能自主推进行动的执行，严格遵循子任务先后顺序或并行关系，在执行过程中实时记录进度，确保行动与规划一致；
c） 应支持基于反馈对结果进行校验和优化，当结果不符合要求时，能调整执行方式或触发方案调整动作；
d） 应支持多行动并发执行场景下的优先级管理，可根据任务紧急程度、资源状态、依赖关系和风险等级调整执行顺序；
e） 应支持以文本、图表或结构化数据等形式输出执行结果，满足不同任务场景下的结果呈现；
f） 应支持群体协同场景下的协同行动，完成任务交接、阶段反馈和结果输出；

g） 应能对执行过程、结果或内部状态进行质量检查，并根据检查结果进行动作优化或方案调整。

5.5.2 外部工具调用与管控

外部工具调用与管控能力具体要求如下：

a） 应按统一规范接入外部工具、系统服务、自研工具，具备跨系统交付能力，并提供适配框架，降低新工具接入成本；
b） 应支持按规范调用大模型服务，并兼容必要的通信协议与调用规范；
c） 应建立工具调用风险管理机制，对高风险工具的调用需要进行人工确认，未经用户授权确认，不能自动执行，应符合GB/T 46347—2025的规定；
d） 应具备多场景下工具调用异常处理能力；
e） 应支持工具安全沙箱机制，对文件操作、系统命令等高风险工具启用沙箱隔离，限制访问权限；
f） 应能记录工具调用的完整链路，管理日志按需留存，并支持调用链路可视化分析；
g） 应支持工具调用资源限制，避免单工具过度占用资源，保障任务整体运行稳定；
h） 应具备工具返回结果的处理能力，对工具返回的原始结果进行校验处理，避免无效或格式不一致的返回结果影响任务执行；
i） 应支持工具版本管理、兼容性检查与权限管控。

5.5.3 执行监控、中断与异常处理

执行监控、中断与异常处理能力具体要求如下：

a） 应能跟踪任务整体及各子任务的执行进度、状态和阶段结果，通过可视化界面或数据接口展示过程信息，确保进度透明可追溯；
b） 应支持多维度评估执行效果，按需生成评估报告，将相关结论同步至记忆模块或相关任务模块；
c） 应支持任务中断机制，在用户指令、风险告警或外部条件变化等情况下，能暂停或终止任务执行，必要时支持人工接管，并保存当前状态、执行进度和上下文信息；
d） 应能识别执行过程中的各类异常，覆盖全执行链路，自动识别异常信号，触发告警；
e） 应针对不同异常类型制定差异化处理机制，确保核心任务正常执行或安全终止；
f） 应支持异常记录与复盘，详细记录每次异常的基本信息、处理过程、处理结果，为后续优化、审计和风险控制提供依据。

6 非能力要求

6.1 可靠性要求

对智能体的可靠性的要求如下：

a） 应保障长期稳定运行，连续运行无故障时间应满足实际应用场景需求；
b） 应支持核心组件的热备切换；
c） 应确保数据的使用严格遵循预设的授权策略与目的限定原则；
d） 应支持人工接管与中断机制，当智能体执行偏离预期或触发高危预警时，授权用户可随时暂停、接管或强制终止任务执行；
e） 遇到无效输入、部分工具失效或网络波动等异常情况时，智能体应具备处理能力，不应发生崩溃或产生不可控结果；

f） 应具备执行边界约束机制，自主决策和执行的权限应限定在预设的业务规则和权限范围内，禁止越权执行；
g） 应支持任务断点续跑机制，当任务执行过程中出现异常中断，恢复后能从断点处继续执行，无需重新开始，减少重复工作与资源浪费；
h） 应具备在出现故障后进行自动容灾恢复的能力。

6.2 易用性要求

对智能体的易用性的要求如下：
a） 应提供清晰、完整的开发文档、软件开发工具包和典型应用场景示例代码，保证开发易用性；
b） 应提供图形化集中管理界面或命令行管理工具，支持对智能体的配置、监控、版本更新等运维操作；
c） 应保证交互界面的简洁直观；
d） 应支持数据并行加载及增量加载。

6.3 安全性要求

对智能体的安全性的要求如下：
a） 应具备对智能体的敏感数据（如个人信息、商业数据等）进行加密传输和存储的功能，支持用加解密等方法对数据访问采取权限控制，应符合GB/T 35273—2020的规定；
b） 应具备对抗样本检测能力（如提示词注入、越狱攻击），并触发拦截、拒绝或安全降级处理；
c） 应支持将智能体数据限制在特定授权实体间传输；
d） 应支持联邦学习、差分隐私等隐私计算技术；
e） 应保证智能体输入数据及输出返回结果的保密性和完整性，确保不被未授权用户非法获取；
f） 应具备对输入提示词与生成内容进行安全过滤与合规性审查的能力，防止生成违法、侵权、违规或不道德内容；
g） 应确保长期记忆的构建与管理遵循数据最小化原则，不应过度留存与任务无关的用户隐私数据；
h） 应确保生成的内容具有显式或隐式标识，以满足监管要求。

6.4 可解释性要求

对智能体的可解释性的要求如下：
a） 应支持关键决策的可视化，在意图识别、方案选择等情况下，提供推理路径可视化；
b） 应支持生成自然语言解释报告，说明执行逻辑及依据；
c） 应具备交互过程可解释能力，当用户对智能体行为、决策结果提出疑问时，可实时生成针对性解释，主动澄清推理依据与逻辑；
d） 应支持异常处置可解释，对执行过程中的异常识别、处置策略选择、结果影响等环节，需明确说明异常原因、研判逻辑、处置依据及预期效果；
e） 应留存解释全量日志，包括任务日志、操作日志、接口日志关联存储，满足合规审计与追溯需求。

参考文献

[1]GB/T 35273—2020 《信息安全技术 个人信息安全规范》
[2]GB/T 41867—2022 《信息技术 人工智能 术语》
[3]GB/T 45907—2025 《人工智能 服务能力成熟度评估》
[4]GB/T 46347—2025 《人工智能 风险管理能力评估》
[5]GB/T 46351—2025 《人工智能 多算法管理技术要求》
[6]GB/Z 42759—2023 《智慧城市 人工智能技术应用场景 分类指南》
[7]YD/T 4929—2024 《面向多智能体系统的计算平台技术要求》
[8]AIIA/T 0219—2025 《面向软件工程智能体的技术和应用要求 第1部分：开发智能体》
[9]《智能体规范应用与创新发展实施意见》
[10]国家数据局《数据领域常用名词解释》（第一批）