



# 中华人民共和国国家标准化指导性技术文件

GB/Z XXXXX—XXXX

## 人工智能操作系统通用技术要求

General technical requirements for artificial intelligence operating system

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

国家市场监督管理总局  
国家标准化管理委员会 发布



目 次

前言 ..... III

引言 ..... V

1 范围 ..... 1

2 规范性引用文件 ..... 1

3 术语和定义 ..... 1

4 缩略语 ..... 2

5 人工智能操作系统框架 ..... 2

6 内核要求 ..... 2

7 AI 运行库要求 ..... 3

8 AI 服务要求 ..... 3

    8.1 概述 ..... 3

    8.2 感知服务 ..... 3

    8.3 交互服务 ..... 4

    8.4 模型服务 ..... 4

    8.5 智能体管理 ..... 4

    8.6 工具管理 ..... 4

    8.7 技能管理 ..... 5

    8.8 上下文管理 ..... 5

    8.9 记忆管理 ..... 6

    8.10 跨设备协同 ..... 6

    8.11 数据平台服务 ..... 6

9 系统智能体要求 ..... 6

    9.1 智能化能力 ..... 6

    9.2 智能体调用 ..... 7

    9.3 系统管理能力 ..... 8

10 智能开发套件要求 ..... 8

11 兼容性要求 ..... 9

12 安全要求 ..... 9

    12.1 身份鉴别 ..... 9

    12.2 智能体运行安全 ..... 9

    12.3 技能安全 ..... 9

    12.4 工具安全 ..... 10

    12.5 安全沙箱 ..... 10

    12.6 模型安全 ..... 10

    12.7 行为审计 ..... 10

参考文献 ..... 11

## 前 言

本文件为规范类指导性技术文件。

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。

本文件由全国信息技术标准化技术委员会（SAC/TC 28）提出并归口。

本文件起草单位：中国电子技术标准化研究院、华为终端有限公司、中兴通讯有限公司、小米通讯技术有限公司、统信软件技术有限公司、天翼云科技有限公司、OPPO广东移动通信有限公司、麒麟软件有限公司、中移（苏州）软件技术有限公司、维沃移动通信有限公司、荣耀终端股份有限公司、南京大学、北京交通大学。

本文件主要起草人：王威伟、焦廉洁、苗宗利、董建、杨磊、张希璐、史浩、孟眉、刘东、胡冲、李建慧、王磊、马承青、唐杨、姜少涛、李腾、郭爽、张胜举、高立发、徐曼、冯洋、汪群博、王键、谢南淦、徐立锋、刘海涛、李鹤、刘铭然、李根、张超、李永强、曹宇琼、李辰淑、安佳勇、高歌、李晨。



## 引 言

操作系统作为调用底层硬件、承载上层应用的关键枢纽，是数字基础设施的“魂”。操作系统作为连接物理世界与数字世界的第一入口，更是智能生态的核心载体。当前，随着大模型技术的快速演进，人工智能操作系统正从AI in OS加快迈向AIOS，全球人工智能操作系统市场迎来前所未有的发展机遇。人工智能正驱动操作系统从底层硬件资源的管理者，转变为算法、模型、数据的智能调度中枢，我国人工智能操作系统角色定位、能力边界与生态逻辑正在被重新定义。然而，人工智能操作系统技术及标准化工作处于起步阶段，人工智能操作系统概念泛化与认知分化并存，业界对人工智能操作系统的内涵外延、功能架构、演进方向尚未统一，不利于凝聚创新合力、降低产业链上下游协同成本。在此背景下，加快制定该标准，对规范产业发展、防控安全风险、抢占行业发展主动权具有重要意义。





# 人工智能操作系统通用技术要求

## 1 范围

本文件确立了人工智能操作系统框架，规定了人工智能操作系统（以下简称“操作系统”）的内核、AI运行库、AI服务、系统智能体、智能开发套件、兼容性和安全等要求。

本文件适用于人工智能操作系统的设计、开发、测试等，也为终端生成商、操作系统开发者、应用开发者等提供参考。

## 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/Z 185.7 人工智能 智能体互联 第7部分：智能体工具调用

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**人工智能终端 artificial intelligence terminal**

具备主动感知理解、多模态交互、智能化服务和学习进化等能力，完成特定任务的终端产品。

注1：智能化任务处理流程一般涉及感知、规划、决策、执行、学习等环节。

注2：人工智能终端由软件和硬件组成，软件部分包括人工智能模型、智能化应用、操作系统、用户界面、端云协同接口等；硬件部分包括通信模块、处理器、内部存储、外围输入/输出(IO)器件、显示等。

注3：本文件主要指移动终端、微型计算机等产品，可穿戴、机器人等不同设备可进行相应裁剪。

[来源：GB/Z 177.1—2026：3.1，有修改]

### 3.2

**人工智能操作系统 artificial intelligence operating system**

融合大模型、智能体等人工智能技术，具备感知、规划、执行、记忆、学习等智能化服务能力的系统软件。

注：人工智能操作系统负责对人工智能终端的软硬件资源、异构算力资源、数据资源、任务执行进行管理、调度与安全控制，为应用、智能体等分配资源，并提供资源访问接口与运行环境支撑。

### 3.3

**智能体 agent**

感知环境，并主动采取行动以实现特定目标的实体。

注1：本文件中的智能体特指人工智能智能体，一般为软件系统。

注2：具备感知、交互、执行、决策、记忆能力，能够理解用户意图、持续学习优化自身行为。

[来源：GB/Z 185.1—2026：3.1，有修改]

### 3.4

**系统智能体 system agent**

是人工智能操作系统的重要组成部分，获取操作系统信息，具备任务分发、资源调度、系统管理等能力的智能体。

### 3.5

**词元 token**

词元是人工智能语言模型中，文本经过切分或字节级编码后得到的最小离散符号单元。

3.6

工具 tool

用于执行特定操作或服务、硬件驱动程序或软件模块。

[来源：ISO/IEC 23000-15：2016，3.5.4，有修改]

3.7

标识 Identifier

与数字、非数字或抽象实体（如图书、图像、报告、元数据记录或事件）关联的字符序列。

注：本文特指用于在操作系统环境中区分用户、智能体、应用、技能、工具、设备等且具有唯一性的标识。

[来源：ISO 24619:2011(en), 3.2.1，有修改]

4 缩略语

下列缩略语适用于本文件。

AI：人工智能（artificial intelligence）

AIP：智能体互联协议（agent interconnection protocol）

API：应用程序编程接口（application programming interface）

A2A：智能体与智能体（agent2agent）

CLI：命令行界面（command line interface）

CPU：中央处理器（central processing unit）

GPU：图形处理器（graphics processing unit）

GUI：图形用户界面（graphical user interface）

IDE：集成开发环境（integrated development environment）

IO：输入输出（input output）

NPU：神经处理单元（neural processing unit）

ReAct：推理行动（reasoning acting）

WLAN：无线局域网（wireless local area network）

5 人工智能操作系统框架

人工智能操作系统框架见图1。本文件对基础库、基础服务、应用开发基础套件不作规定。

注：人工智能操作系统各项功能的实现以硬件为条件，当硬件具备相应能力时，操作系统可实现对应功能；移动终端、微型计算机、可穿戴等不同设备可进行相应裁剪。

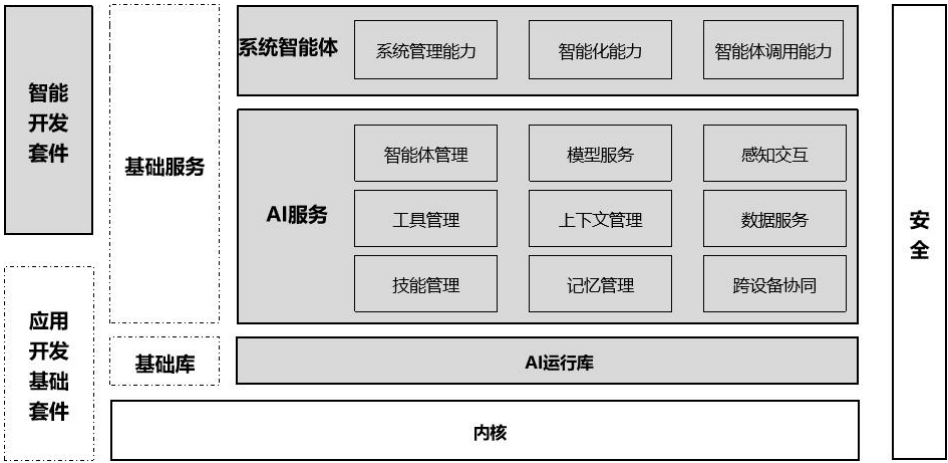


图1 人工智能操作系统框架

6 内核要求

内核应符合下列要求：

- a) 具备异构算力的硬件抽象与驱动框架，通过统一的硬件访问接口，提供硬件状态查询与中断处理能力；
- b) 具备按需裁剪与动态加载能力，支持按场景仅加载内核核心模块，动态加载异构算力驱动框架、AI 运行时所需的支撑组件等；
- c) 具备内存映射、动态分配与回收等能力；
- d) 具备传感器、存储、通信等硬件驱动管理能力；
- e) 提供内核空间隔离、访问控制、数据和代码保护等基础安全机制，包括进程隔离、内存保护、驱动完整性校验、系统资源保护等；
- f) 提供基于硬件信任根的可信启动机制，保护人工智能大模型及关键组件不被篡改；
- g) 提供并发访问与异步 IO 的文件系统及存储驱动，通过数据预取与缓存加速机制，支撑人工智能模型、向量索引、知识库等资源的高效加载、缓存与访问；
- h) 支持跨进程通信与设备互联，提供确定性调度机制，满足从基础响应到意图驱动任务链触发的连续调度需求；
- i) 提供不同优先级任务的抢占式调度机制，根据优先级和运行状态动态分配资源，对时延敏感型 AI 高优先级任务获得毫秒级响应；
- j) 具备多模型并发执行时的算力与内存资源调度能力；
- k) 具备动态电源管理机制，根据任务实时负载、优先级和资源占用情况，动态调节设备工作状态。

## 7 AI 运行库要求

AI 运行库与内核驱动框架之间通过明确定义的边界接口进行通信，应符合下列要求：

- a) 推理运行时不直接访问内核特权资源，模型加载、推理会话与硬件调用遵循操作系统权限管控与审计要求；

注：CPU 指令集中的特权资源访问控制是在处理器上运行的程序对内存区域、IO 端口及特殊指令等资源的访问控制。

- b) 具备主流神经网络交换格式的模型解析与推理执行能力；
- c) 支持面向 CPU、GPU、NPU 等异构处理器的算子库、模型运行时；
- d) 具备面向大语言模型的推理运行时，包括分词器、动态批处理、显存调度及并发请求管理机制等；
- e) 提供向量存储与检索的运行支持，包括向量的高效存储、索引构建（如基于图的索引、基于量化的倒排索引等）、相似度检索能力，为检索增强生成等应用提供基础数据服务；
- f) 支持端侧资源受限场景下的轻量化模型加载、推理、执行能力，在容器等环境的条件下独立运行；
- g) 具备智能体运行时的资源隔离功能；
- h) 实现任务分解、计划编排与执行，支持 ReAct 或工作流等执行模式，并支持智能体间的协同。

## 8 AI 服务要求

### 8.1 概述

操作系统应提供接口，建立智能体与 AI 服务之间的通信机制。

### 8.2 感知服务

感知服务应符合下列要求：

- a) 通过外联设备、内外部传感器、数据采集模块、识别模块或互联网等方式，感知、采集、识别用户生物特征、用户输入内容、用户行为、硬件状态、软件信息、外联设备、环境信息等输入信息；
- b) 对多模态输入信息进行理解与跨模态上下文语义关联，具备复杂意图识别能力。

### 8.3 交互服务

交互服务应符合下列要求：

- a) 具备多模态内容的实时生成、自适应结果输出能力；
- b) 具备交互策略的自主优化与演进能力；
- c) 具备多模态交互过程中的全链路状态管理能力。
- a) 提供多模态交互处理能力的开放接口，支持上层应用、智能体或外部系统调用，向外联设备输出文本、语音、图像、视频、触觉反馈、动作指令、设备控制信号等，并根据设备能力与交互场景自适应调整输出形式。

### 8.4 模型服务

模型服务符合下列要求：

- a) 应提供统一的模型服务接口，支持智能体调用模型能力；
- b) 应具备模型信息的注册查询功能；
- c) 应具备模型推理指标分析功能，包括推理时延、输出准确率等；
- d) 应具备端侧模型的便捷部署、升级管理功能，支持文本生成、语义转化、重排序及多模态处理等模型类型；
- e) 应具备云侧模型服务凭证的安全存储能力，在请求云侧模型服务时，按需将凭证注入请求中；
- f) 应具备云侧模型使用量与配额管理功能，宜支持按调用方、模型实例等维度，统计词元消耗量等指标；
- g) 应具备模型路由能力，根据任务需求、模型能力、实例负载及可用性等，自动将推理请求路由至最合适的模型实例。

### 8.5 智能体管理

智能体管理符合下列要求。

- a) 应具备智能体安装与卸载功能，符合下列要求：
  - 1) 应具备智能体描述信息的声明、查询功能，智能体描述信息包括标识、名称、版本、描述、提供方、默认输入类型、默认输出类型、技能等；
  - 2) 应具备智能体安装、卸载等功能，卸载时终止该智能体的运行实例并清除服务注册信息，经用户确认后清除本地存储数据；
  - 3) 宜与智能体服务公共平台连接，并进行信息查询、核验等。
- b) 应具备智能体运行功能，并符合下列要求：
  - 1) 提供智能体间通信机制，如 A2A、AIP 等；
  - 2) 提供智能体发现机制，支持智能体按能力描述进行查找与匹配；
  - 3) 具备智能体运行时的资源隔离功能；
  - 4) 依据任务类型、前后台状态等维度建立智能体任务调度机制，保障紧急通信等高优先级任务；
  - 5) 具备智能体在后台持续运行的功能；
  - 6) 具备智能体等运行实例的状态监测、异常告警、恢复等功能，采集内存等资源占用指标；
  - 7) 具备暂停或结束智能体运行的功能；
  - 8) 提供事件总线，供智能体、工具、技能、会话管理等通过发布、订阅方式，交换相关事件信息。

### 8.6 工具管理

工具管理符合下列要求。

- a) 工具清单维护：
  - 1) 应具备工具声明等功能；
  - 2) 应具备工具下载等功能；

- 3) 工具描述应包括名称、版本、功能说明、输入参数模式、输出格式、调用方式、权限声明等信息；
  - 4) 应具备工具清单及描述信息的查询、增加、更新、注销等功能；
  - 5) 宜将常用能力进行工具化封装，供其他智能体或其他应用调用，如媒体摄录、通话、短信、日历、时钟、备忘录、相册、浏览器、应用商店等；
  - 6) 宜将外联设备的操作能力进行工具化封装，供其他智能体或其他应用调用，如智能家居控制、短距通信设备、打印扫描设备、显示投屏设备、穿戴设备等。
- b) 工具调用管理：
- 1) 应提供统一的访问接口进行工具查询与调用；
  - 2) 应具备工具调用的权限管理功能；
  - 3) 应提供工具调用链路记录，包括调用时间、请求参数、返回结果、执行耗时及调用方身份等；
  - 4) 应对工具运行环境进行沙箱隔离，单个工具故障不影响其他工具的正常运行；
  - 5) 宜限制资源使用上限与执行超时阈值。

## 8.7 技能管理

技能管理符合下列要求。

- a) 技能维护：
- 1) 应具备技能的下载、安装、注册、查询、更新、删除等功能；
  - 2) 应具备技能的版本升级和管理功能，记录技能的版本号和更新记录；
  - 3) 应对技能进行安全性扫描与校验；
  - 4) 宜对技能依赖关系进行声明与校验，包括与工具、其他技能、运行环境及模型的依赖关系。
- b) 技能调用管理：
- 1) 应具备技能的发现与按需激活功能，允许智能体根据用户任务意图自动匹配并加载相关技能；
  - 2) 应具备技能调用记录功能；
  - 3) 应对技能质量进行评估与反馈。

## 8.8 上下文管理

上下文管理符合下列要求：

- a) 应具备上下文构建能力，支持将会话历史、记忆库召回、环境感知数据、提示词及模板等，按需组织为结构化消息序列；
- b) 应实现上下文窗口的动态容量管理，依据模型能力、任务需求等，对上下文总量进行词元粒度的计量与约束，在窗口受限时实现内容重要性排序与自动裁剪，并记录裁剪日志；
- c) 应具备上下文语义摘要、压缩、自动优化等能力，包括会话历史关键信息提取、冗余信息检测与剔除等；
- d) 应提供子智能体派生场景的上下文继承与隔离机制，子智能体继承父智能体的任务目标与上下文，同时建立独立的上下文空间；
- e) 应提供上下文内容的数据生命周期管理，支持上下文信息的存储、归档与安全清除；
- f) 应具备上下文访问控制与安全隔离能力，防止跨用户、跨智能体、跨会话的越权访问；
- g) 应具备上下文状态的保存与恢复能力，支持跨设备的会话同步及任务迁移时的上下文还原；
- h) 宜具备长周期上下文的保持能力，使智能体在跨会话、跨任务周期中能够继承并复用关键上下文信息；
- i) 宜提供文本、音频、图像、传感器数据等多模态上下文的语义关联与协同；
- j) 宜提供与检索增强生成系统的接口，支持将外部知识库、向量数据库中的相关文档片段动态检索并注入上下文窗口，并实现检索结果与上下文的融合去重与优先级排序。

## 8.9 记忆管理

记忆管理符合下列要求：

- a) 应具备多模态数据在端侧进行记忆融合处理的能力；
- b) 应具备记忆分析能力，内置轻量化模型对融合后的记忆数据进行预处理与分析，对分析结果进行总结，形成记忆标签库并分类管理；
- c) 应提供记忆安全沙箱隔离，为不同应用及智能体分配独立记忆空间，进行隔离执行，防止跨应用数据泄露；
- d) 应提供记忆数据生命周期管理，自定义数据失效规则，过期数据自动擦除，避免冗余积累；
- e) 应对共享的记忆数据实施脱敏处理，防止逆向推断用户身份；
- f) 宜提供记忆管理界面，展示记忆用途、记忆标签、记忆状态等，支持用户删除或导出记忆数据。

## 8.10 跨设备协同

跨设备协同符合下列要求：

- a) 应具备多链路聚合与智能路由能力，支持异构网络自适应，支持如 GB/T 46803(所有部分)、蓝牙、Wi-Fi Direct 或 NFC 等规定的通信协议，确保在复杂环境下的多模态设备连接和跨网络切换；
- b) 应根据跨设备的原子化能力描述，构建设备能力模型；提供轻量级分布式注册机制，保障动态拓扑下服务的实时同步与状态一致性；
- c) 应具备任务的细粒度拆分与动态调度，根据负载、能耗及隐私要求，自动将任务调度至端、边、云执行，实现多设备的协同执行；
- d) 应支持不同设备的智能体进行协商，包括任务协同、状态互通、统一决策等；
- e) 应通过端到端加密方法，实施设备双向认证与用户意图确认；
- f) 跨设备传输应仅共享必要数据并脱敏，协同组件在独立安全沙箱中运行，限制底层资源访问，防止恶意代码横向扩散；
- g) 跨端执行任务时，应在任务发起前完成目标端的执行环境校验，确保安全策略随任务下发后再执行；
- h) 跨端执行任务宜通过系统智能体来调度决策，支持跨设备协同的任务迁移、上下文一致、状态同步。

## 8.11 数据平台服务

数据平台服务应支持不同应用、智能体之间交换数据和任务上下文，并符合下列要求：

- a) 提供接口，支持应用、智能体、用户自主决定和配置数据接入、开放等策略，包括开放范围、开放频次、有效时间、访问主体等；
- b) 具备跨应用沙箱的数据安全汇集、数据存储、语义加工与受控流转等能力，未被授权的应用或智能体无权访问其他应用数据。

# 9 系统智能体要求

## 9.1 智能化能力

### 9.1.1 感知

应具备多模态感知能力，感知并获取用户信息、终端信息、环境信息等多维度数据，以支撑任务响应和执行。

### 9.1.2 认知

认知应符合下列要求：

- a) 具备意图识别能力，从用户的自然表达或行为模式，结合上下文、设备状态及环境等，推断其潜在需求；

- b) 具备多步骤任务自主规划能力，将用户的复合意图需求自动拆解为可行的操作序列。

### 9.1.3 执行

执行应符合下列要求：

- a) 对系统基本功能软件的操作，包括拨打电话、记事、发送短信、设定闹钟、设置系统日历、设置通讯录等；
- b) 支持用户通过自定义自然语言指令整合多步系统设置操作，如“早安”指令一键开启 WLAN、关闭静音等多步操作；
- c) 提供定时任务与任务触发机制，允许用户预设特定时间或特定条件满足时，自动执行预设任务；
- d) 具备多任务并行执行能力，在不干扰用户当前操作的前提下独立运行，各任务完成或需要用户介入时主动提示用户；
- e) 具备执行过程的可追溯性，记录每次任务的决策依据、操作步骤与执行结果，支持用户事后查看与审计，并提供任务回滚或纠错能力。
- f) 支持用户实时查看任务当前所处的状态，包括待执行、执行中、暂停、已完成、异常中断、超时等，允许用户执行暂停、继续、终止或重启等操作；
- g) 具备任务执行状态判断能力，若任务执行异常，应提示用户介入。

### 9.1.4 记忆

记忆应符合下列要求：

- a) 具备短期记忆能力，将用户单个会话上下文内容进行临时存储，并具备短期记忆固化能力，对短期记忆进行语义理解、分类或分层；
- a) 具备长周期的持久化存储的长期记忆能力，对会话历史记录、身份信息、偏好习惯、历史交互上下文及常用指令等长期记忆数据，进行归纳、总结与推理，形成稳定的个人化用户偏好与规律的衍生记忆。

### 9.1.5 学习

应通过用户反馈、外部知识更新、自我优化等方式，对模型和策略进行调整，优化任务执行能力。

## 9.2 智能体调用

### 9.2.1 智能体组件

应提供统一的系统智能体组件注册与调用机制，允许工具通过系统智能体提供的组件接口进行适配，开放功能供系统智能体调用。

### 9.2.2 调用模型

应支持AI模型配置，宜提供端侧本地处理模式，用户隐私数据优先在本地推理，仅在必要时经用户授权后上传云端。

### 9.2.3 调用工具

调用工具应符合下列要求：

- a) 应理解用户意图，生成工具调用需求并调用一个或多个工具；
- b) 应具备直接调用单工具、多工具链式编排、无依赖工具的并行调用能力，根据任务自动规划工具调用序列，调用失败时自动重试与备选工具降级；
- c) 应支持外部工具调用等协议，并符合 GB/Z 185.7 的规定；
- d) 显式获取用户许可后，宜提供 GUI 导航等调用方式；
- e) 显式获取用户许可后，宜支持浏览器调用，具备页面解析、理解、交互、执行等能力。

注：建立明确的授权范围，涉及敏感、高危或资产类操作时经用户确认。

#### 9.2.4 技能接入

技能接入应符合下列要求：

- a) 提供技能接入框架，支持用户或应用按需接入技能；
- b) 对技能质量进行评估与反馈，并对技能进行迭代优化；
- c) 支持技能的渐进式加载，启动时仅加载技能的名称与描述等元信息用于路由匹配，触发时加载技能正文指令，执行时按需读取附属脚本与资源，以最大限度节省上下文窗口。

#### 9.2.5 智能体间协作

应提供与其他智能体的协作机制，并符合下列要求：

- a) 支持智能体协作协议，如 AIP、A2A 等；
- b) 将任务分解并分配给不同的智能体并行或串行执行；
- c) 接收其他智能体任务协作；
- d) 根据协作需求选择点对点、群组或混合等方式。

#### 9.2.6 跨设备调用能力

宜具备与其他设备（如手机、平板、穿戴、车机、智慧屏等）的任务协同能力。

### 9.3 系统管理能力

#### 9.3.1 系统设置

应提供系统设置功能，包括开关网络与蓝牙、调节音量、显示亮度、文件管理等。

#### 9.3.2 系统运维

系统运维符合下列要求：

- a) 应具备系统运维能力，包括执行垃圾清理、应用管理、省电模式/性能模式切换、后台进程管控等资源调度操作，优化系统运行效率；
- b) 应具备系统运行状态监测能力，包括系统资源消耗、应用运行状态、智能体运行状态等；
- c) 宜记录系统智能体执行的系统操作，包括操作时间、操作对象与执行结果等；
- d) 当系统资源不足时，可通过省电模式等方式保障正常运行需求。

#### 9.3.3 安全防护

应提供安全防护能力，并符合下列要求：

- a) 具备系统配置和用户数据保护能力，如隐私替身、私密保险箱、系统分身等；
- b) 具备异常行为监测能力。

### 10 智能开发套件要求

智能开发套件符合下列要求。

- a) 应提供智能体打包、签名与发布格式，包括元信息、依赖声明、入口代码、资源配置等。
- b) 应具备智能体的版本管理与环境依赖能力。
- c) 应提供面向智能体的开发、测试、调试、仿真、部署、监控等软件开发工具包，具备功能与行为定义、会话与状态管理，版本管理等能力：
  - 1) 应提供技能定义模板、工程构建与发布机制，支持技能能力声明、行为规则配置、业务逻辑编排及数字签名等；
  - 2) 应提供工具定义模板、打包构建及数字签名校验机制，支持工具能力描述、调用逻辑配置、最小权限声明等；
  - 3) 应提供测试工具，覆盖单元测试、集成测试、仿真测试、性能测试、安全测试等。



- d) 应支持智能体和前端组件交互协议的开发，实现和网页端、移动端、IDE 等多种客户端/前端组件的交互通信。
- e) 应提供模型适配器，支持接入不同场景的模型。
- f) 应提供自然语言编程工具，支持对话历史追踪、上下文管理、跨会话长期偏好记忆。
- g) 宜使用智能体编排引擎创建工作流进行多智能体编排，智能体编排引擎支持条件分支、循环、并行与串行执行等多种编排模式，并支持编排模板的复用与分享。

## 11 兼容性要求

兼容性应符合下列要求：

- a) 提供大模型和智能体的移植开发指南；
- b) 支持主流 AI 开发工具链与调试工具，保障开发者生态的一致性；
- c) 操作系统升级保证开发运行环境与服务接口的兼容性，不影响已安装的软硬件、智能体、模型及 AI 应用的正常运行，API 保持向后兼容；
- d) 支持不同厂商的 CPU、GPU、NPU 等异构计算单元；
- e) 支持主流模型格式的导入与运行，提供统一的模型转换工具。

## 12 安全要求

### 12.1 身份鉴别

身份鉴别应符合下列要求：

- a) 操作系统为每个账号分配唯一的账号标识，系统智能体与操作系统用户身份绑定；操作系统用户切换时，系统智能体身份自动切换，不同用户之间的数据进行隔离；
- b) 操作系统识别部署在本系统上的智能体、技能、工具等并为其分配标识，标识与包名、命名空间、部署路径等建立映射关系，并在生存周期内唯一且不变；
- c) 身份注销后，操作系统立即回收所有相关权限，删除相关身份凭证，确保注销的身份不继续执行操作。

### 12.2 智能体运行安全

智能体运行安全应符合下列要求：

- a) 提供智能体访问行为的管理机制，如基于访问数据敏感度、智能体行为风险度等进行分级分类，并实施相适应的访问控制策略；
- b) 按场景向智能体提供受控的传感器数据、位置数据、通讯录、日历等敏感信息，不默认全量接入；
- c) 具备提示词注入、对抗样本、恶意指令、诱导泄露、权限绕过等攻击行为的识别、防护和处置能力，对异常输入、异常上下文和异常指令进行检测、拦截、降级或提示用户确认；
- d) 对智能体的外部工具和技能调用请求进行安全策略校验，防止智能体执行未经授权、违反系统安全策略的行为；
- e) 智能体之间的通信采用加密传输协议，支持端到端加密，防止中间人攻击和数据窃听。

### 12.3 技能安全

技能安全应符合下列要求：

- a) 校验技能完整性、来源可信性、依赖安全性，校验通过后技能进入本地安装目录或运行目录，校验失败时拒绝安装并提示用户确认；
- b) 提供技能行为的持续监测机制，发现异常行为时提示用户处置；
- c) 支持技能在安全沙箱或受控执行环境中运行，限制技能脚本的资源访问范围与执行权限，如代码生成与执行、系统 CLI 调用。

## 12.4 工具安全

对来自网页、文档、邮件、插件、知识库或第三方接口的工具调用指令进行识别和隔离，防止其绕过用户意图或系统策略影响智能体行为。

## 12.5 安全沙箱

安全沙箱符合下列要求：

- a) 应支持对智能体调用或生成的相关脚本、代码、CLI 等工具的隔离运行，沙箱对工具的文件读写、网络访问、系统调用、系统服务调用等实施权限约束和隔离管控；
- a) 安全沙箱应具备沙箱逃逸监测和防护能力，防止代码突破隔离边界；
- b) 安全沙箱会话结束时，应清除沙箱内的临时数据和运行状态；
- c) 智能体任务执行时，宜支持任务级授权模式，按任务需要、用户确认结果动态赋予沙箱权限，执行完成后权限即时回收，避免静态永久授权后提权风险。

## 12.6 模型安全

模型安全符合下列要求。

- a) 应对模型输入、模型输出、上下文记忆、工具调用参数和审计日志中的敏感信息进行识别与脱敏处理，防止个人信息、身份凭证、密钥、商业秘密等敏感数据泄露。
- b) 应支持推理模型文件、模型参数、权重及相关配置信息的加密存储，密码算法宜支持 GM/T 0002、GM/T 0003（所有部分）和 GM/T 0004 规定的密码算法。
- c) 应支持终端模型密钥存储在可信环境中，加密模型文件的密钥应存储于安全执行环境或安全存储区域，不与加密模型同域存放。
- d) 当使用云端计算环境推理时，应满足下列要求：
  - 1) 对云端计算节点进行身份和环境认证，包含对令牌发行者证书链的校验、对可信计算服务器凭据证书的验证、周期性会话公钥来源证明等；
  - 2) 传输的会话密钥支持周期性更新，对每个通信会话或特定操作仅使用一次密钥，确保长期密钥仅用于密钥协商，而非直接参与数据加密。
- e) 应建立模型生成内容审计机制，记录合规情况、风险类别、处置方式、输出摘要和相关上下文信息，支持事后追溯、统计分析和策略优化；
- f) 当模型输出触发安全风险时，可根据风险等级采取拒绝生成、降级回答、内容改写、用户确认、人工审核、阻断输出或安全告警等处置措施。

## 12.7 行为审计

行为审计符合下列要求：

- a) 应具备覆盖用户请求、系统调度、智能体生成、工具调用、技能接入、资源访问、应用输出的审计能力，对关键操作过程输出审计日志；
- b) 审计日志应记录用户身份、智能体身份、应用身份、工具标识、技能标识、请求时间、授权状态、调用对象、资源访问范围、指令内容摘要、执行结果、异常信息或时间戳等信息；涉及用户敏感信息、个人信息、商业秘密、提示词内容、模型输入输出等内容的，应采取脱敏、最小化记录、访问控制或留存期限管理等措施；
- c) 应记录用户授权确认信息、风险提示内容、执行前后状态、调用参数摘要、执行结果等，支持事后责任定位和安全事件追溯；
- d) 应提供审计数据的检索、关联分析、告警和合规导出能力，支持按照用户、智能体、应用、资源、时间范围、风险等级和事件类型进行查询分析；
- e) 审计日志应具备防篡改、防删除和完整性校验能力，可采用可信时间戳、数字签名、哈希链、只追加存储或安全日志服务等机制进行保护。

## 参 考 文 献

- [1] GB/Z 177.1—2026 人工智能终端智能化分级 第1部分：参考框架
- [2] GB/Z 185.1—2026 人工智能 智能体互联 第1部分：总体架构
- [3] ISO/IEC 23000-15: 2016 Information technology—Multimedia application format (MPEG -A)  
—Part 15: Multimedia preservation application format
- [4] ISO 24619: 2011 Language resource management—Persistent identification and sustainable access (PISA)
- [5] GB/T 46803（所有部分） 信息技术 基于极化码的低功耗无线通信网络
- [6] GM/T 0002 SM4分组密码算法
- [7] GM/T 0003（所有部分） SM2椭圆曲线公钥密码算法
- [8] GM/T 0004 SM3密码杂凑算法