



中华人民共和国国家标准

GB/T 46350—2025

信息技术 云计算 智能云服务通用要求

Information technology—Cloud computing—General requirements of
AI cloud service

2025-10-05 发布

2026-02-01 实施

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

前言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 能力框架 2

6 基础设施要求 3

 6.1 资源 3

 6.2 资源管理 4

 6.3 加速 5

 6.4 任务管理和调度 6

7 模型开发要求 6

 7.1 数据处理 6

 7.2 模型构建 8

 7.3 模型管理 9

 7.4 模型部署 11

 7.5 模型推理 11

 7.6 数据闭环 12

 7.7 大模型提示工程 12

 7.8 大模型指令调优 12

 7.9 MLOps 工作流 12

 7.10 LMOps 工作流 13

8 模型服务要求 13

 8.1 计算机视觉类 13

 8.2 语音类 13

 8.3 自然语言处理类 14

 8.4 多模态类 14

9 人工智能应用开发要求 14

 9.1 开发环境 14

 9.2 应用开发框架 14

 9.3 应用组件 14

 9.4 应用模板 14

 9.5 知识管理 15

 9.6 应用管理 15

参考文献 16

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：北京百度网讯科技有限公司、中国电子技术标准化研究院、华为云计算技术有限公司、浪潮云信息技术股份公司、天翼云科技有限公司、中移(苏州)软件技术有限公司、广州市品高软件股份有限公司、中国人民解放军国防科技大学、阿里云计算有限公司、深信服科技股份有限公司、北京华胜天成科技股份有限公司、中兴通讯股份有限公司、济南浪潮数据技术有限公司、中移雄安信息通信科技有限公司、中移(杭州)信息技术有限公司、北京天融信网络安全技术有限公司、湖南麒麟信安科技股份有限公司、京东科技信息技术有限公司、上海商汤智能科技有限公司、浪潮电子信息产业股份有限公司、中国移动通信有限公司研究院、北京银信长远科技股份有限公司、安超云软件有限公司、中国烟草总公司湖南省公司、电子科技大学、中国科学院信息工程研究所、杭州谐云科技有限公司、荣联科技集团股份有限公司、哈尔滨工业大学、清华大学、北京国信网联科技有限公司、甘肃同兴智能科技发展有限责任公司、中航材导航技术(北京)有限公司、广东粤电信息科技有限公司、深圳聚瑞云控科技有限公司、广东车卫士信息科技有限公司、深圳市易连汇通科技有限公司、宝德计算机系统股份有限公司、北京大道云行科技有限公司、北京厚普医药科技有限公司、广州赋安数字科技有限公司、深圳市五兴科技有限公司、海天地数码科技(北京)有限公司、桂林理工大学、江西云眼视界科技股份有限公司、北京金智达管理顾问有限公司、澳克多普有限公司、广脉科技股份有限公司、北京智慧易科技有限公司、四川汉唐云分布式存储技术有限公司、苏州明逸智库信息科技有限公司、广东博今信息技术集团有限公司。

本文件主要起草人：宋飞、陈行、贾超、王志鹏、杨丽蕴、玄凌博、谢永康、刘增志、查丽、赵华、张敏、郑佳佳、杨明、苗子聪、郭旻、徐军、林冬艺、黄继承、包涵、王意洁、吴涛、朱松、刘晨、魏胜平、梁钢、李响、亓开元、张坤、姚飞、宁溯、张风雷、喻涵、张朝璐、雷晓锋、杨鹏举、邓旺波、刘辰、韩冬、吴庚、代文斌、鲁璐、叶丰华、范亚梅、毛浣渠、季寒德、鞠万里、郭晓、许刚、邹瞰、李益文、罗光春、陈爱国、边鹏旭、王翔宇、王向东、王静、刘奇旭、吴晓清、谭儒、李伟泽、徐冠群、乔霖、陈志峰、王程安、王瑞敏、李幸、闫纪红、黄双喜、杨思维、高宇、曹建农、詹越、孙沛、王林军、胡周达、林禧桐、郭文艺、曾俭、张云霞、刘爱贵、张芳、梁华、赵亚云、董广明、刘刚、罗军、王文迪、谢晓兰、唐涛、赵本金、王巍、赵淑飞、关涛、聂琨林、王亮、彭黎。

信息技术 云计算 智能云服务通用要求

1 范围

本文件确立了智能云服务能力框架,规定了智能云服务基础设施、模型开发、模型服务、人工智能应用开发等方面的要求。

本文件适用于指导智能云服务的设计、开发、部署、使用和测评。

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 32400—2015 信息技术 云计算 概览与词汇

3 术语和定义

GB/T 32400—2015 界定的以及下列术语和定义适用于本文件。

3.1

云计算 cloud computing

通过网络将可伸缩、弹性的共享物理和虚拟资源池以按需自服务的方式供应和管理的模式。

注:资源包括服务器、操作系统、网络、软件、应用和存储设备等。

[来源:GB/T 32400—2015,3.2.5,有修改]

3.2

云服务 cloud service

通过云计算(3.1)已定义的接口提供的一种或多种能力。

[来源:GB/T 32400—2015,3.2.8]

3.3

智能云服务 AI cloud service

支撑或提供人工智能能力的云服务。

注:人工智能能力指模型开发、模型服务、人工智能应用开发等。

4 缩略语

下列缩略语适用于本文件。

AI:人工智能(Artificial Intelligence)

API:应用程序接口(Application Programming Interface)

ASIC:专用集成电路(Application Specific Integrated Circuit)

CCIX:缓存一致性的加速器互联(Cache Coherent Interconnect for Accelerators)

CPU:中央处理器(Central Processing Unit)