

ICS 35.240  
CCS L 70



# 中华人民共和国国家标准

GB/T 45907—2025

## 人工智能 服务能力成熟度评估

Artificial intelligence—Service capability maturity assessment

2025-06-30 发布

2025-06-30 实施

国家市场监督管理总局  
国家标准管理委员会 发布



## 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 服务能力框架 .....	2
5.1 概述 .....	2
5.2 基础能力 .....	3
5.3 业务能力 .....	4
6 成熟度等级及评估分值 .....	5
6.1 成熟度等级 .....	5
6.2 成熟度评估分值 .....	6
7 成熟度评估方法 .....	13
7.1 能力项权重 .....	13
7.2 计算方法 .....	13
7.3 成熟度等级判定 .....	14
附录 A (资料性) 部分场景能力项权重 .....	15
A.1 家用智能设备场景 .....	15
A.2 智能医疗诊疗场景 .....	17
A.3 工业设备预测性维护场景 .....	19





## 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：中国电子技术标准化研究院、小米通讯技术有限公司、广州赛西标准检测研究院有限公司、浪潮云信息技术股份公司、哈尔滨工业大学、北京中数睿智科技有限公司、上海商汤智能科技有限公司、杭州海康威视数字技术股份有限公司、中国科学院自动化研究所、中国科学院软件研究所、浪潮软件科技有限公司、腾讯科技(深圳)有限公司、中国电子工业标准化技术协会、浪潮电子信息产业股份有限公司、中国医学科学院生物医学工程研究所、玉养信息科技(上海)有限公司、威麟信息技术开发(上海)有限公司、美的集团(上海)有限公司、工业和信息化部电子第五研究所。

本文件主要起草人：徐洋、马珊珊、马骋昊、叶珩、朱亚军、周珏嘉、周钢、郑佳佳、杨沐昀、韩涵、吴庚、彭剑峰、朱贵波、孟令中、杨雨泽、卢顺、聂简荻、沈芷月、孙传兴、黄现翠、李照川、王功明、代威、陈庆帅、赵梦芳、黄胜华、邸贺亮、蒲江波、朱兆颖、翁家良、蔡亚森、黄林轶。





# 人工智能 服务能力成熟度评估

## 1 范围

本文件提出了人工智能服务能力框架,规定了成熟度等级及评估分值,并描述了评估方法。

本文件适用于对服务提供商提供的人工智能服务能力的成熟度评估,以及服务能力框架中能力域、能力项的单项评估。

## 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 41867 信息技术 人工智能 术语

## 3 术语和定义

GB/T 41867 界定的以及下列术语和定义适用于本文件。

### 3.1

#### 能力子项 **capability sub-item**

能力子域中构成成熟度的若干特性指标单项。

### 3.2

#### 能力子项权重 **factor of capability sub-item**

对应能力子项在所述能力子域评估中的权重因子。

### 3.3

#### 服务可用性 **service availability**

服务客户发起服务请求后,服务可访问的时间占总服务时间的比例。

注:可用性的计算是在一系列预定义的时间段中,服务可用时间之和占预定义时间段之和的比例,可排除允许的服务不可用时间。

### 3.4

#### 人工智能服务能力 **artificial intelligence service capability**

人工智能服务所能提供的存储、计算、数据结构化处理与决策、机器学习等基础能力和智能语音、自然语言理解、智能对话、智能图像视频分析等业务能力。

## 4 缩略语

下列缩略语适用于本文件。

API:应用编程接口(Application Programming Interface)

CPU:中央处理器(Central Processing Unit)



GPU:图形处理器(Graphic Process Unit)

SDK:软件开发工具包(Software Development Kit)

## 5 服务能力框架

### 5.1 概述

人工智能服务能力框架由两个能力域组成,每个能力域代表成熟度等级评估中主要的领域划分,由一组相关的能力项组成。每个能力项代表该能力域中构成成熟度水平的若干技术能力领域,每个能力项由一组相关的能力子项组成。人工智能服务能力框架见表1。

表1 人工智能服务能力框架

能力域	能力项	能力子项
基础能力	存储	存储管理
		多样性
		分布式可扩展
		其他
	计算	数据计算
		集群效率
		稳定性
		可扩展性
		灵活性
		统计监控
机器学习	数据结构化处理与决策	稳定性
		可用性
		时效性
		可解释性
		通用性
	全生存周期	全生存周期
		训练性能
		兼容性
		灵活性
		易用性
		算法库

表 1 人工智能服务能力框架 (续)

能力域	能力项	能力子项
业务能力	智能语音	语音识别
		语音合成
		声纹识别
	自然语言理解	词法、句法分析
		信息抽取
		信息检索
		语义推理
		情感分析
		知识图谱
		闲聊
	智能对话	知识型问答
		任务导向对话
		情感对话
		启发式对话
		多模态对话
	智能图像视频分析	识别能力
		理解能力
		业务扩展性
		服务稳定性

## 5.2 基础能力

### 5.2.1 存储

存储能力应包括：

- a) 存储管理：具备数据存储系统的整体管理能力，包括存储容量和数据条目数量等；
- b) 多样性：具备可处理多种类型数据和多种开发语言的能力；
- c) 分布式可扩展：具备分布式架构的集群系统业务负载感知和降低运维复杂度的能力；
- d) 其他：具备存储服务效率、设备故障预警和报警能力、存储数据恢复能力、接口便捷性和服务可用性等方面的能力。

### 5.2.2 计算

计算能力应包括：

- a) 数据计算：具备超大规模分布式计算、批量计算、流式计算、图计算、机器学习计算和边缘计算等复杂计算任务的能力；
- b) 集群效率：具备简单高效地提供大规模集群的部署、调度、扩展和管理服务的能力；
- c) 稳定性：分布式架构部署的稳定性，包括故障自动恢复、快速迁移和无故障运行时长等能力；

- d) 可扩展性:具备计算资源扩容和扩展接口的能力;
- e) 灵活性:具备根据业务变化和负载情况灵活调整的能力,以及能提供便捷易用的自动化工具,具备多种开发框架、控制台、API 和 SDK 的能力;
- f) 统计监控:具备数据流、数据收发状态的统计监控和报警触发的能力。

#### 5.2.3 数据结构化处理与决策

数据结构化处理与决策业务服务能力应包括:

- a) 稳定性:具备数据处理模型对抗噪声和数据攻击的能力;
- b) 可用性:系统与用户的互协作能力;提供数据分析结果,并给出可用决策建议的能力;在维护和调整时的实时服务能力等;
- c) 时效性:具备实时性和变化感知的能力;
- d) 可解释性:具备数据处理和决策过程中模型的可解释性与流程的透明性;
- e) 通用性:具备跨领域、覆盖多个场景的迁移能力和可扩展度。

#### 5.2.4 机器学习

机器学习支撑能力应包括:

- a) 全生存周期:具备包含数据框架(采集、筛选、预标注、标注和脱敏等)、特征探索、模型开发(训练、评估和预测)、模型部署、模型管理和应用发布等全生存周期过程的能力;
- b) 训练性能:具备高性能的运算能力;
- c) 兼容性:具备兼容其他训练模型代码的能力,包括主流的 API 和训练模型代码等;
- d) 灵活性:具备灵活分配计算资源的能力,包括训练时灵活分配计算资源的时间颗粒度和灵活的计量计费机制等;
- e) 易用性:操控人工智能服务所使用工具具备可直观显示的拖拽布局编程服务,组合各种数据源、组件、算法、模型和评估模块的能力;
- f) 算法库:提供插件式算子注入,具备多类算子及算法、时空分析和频繁项挖掘等能力。

### 5.3 业务能力

#### 5.3.1 智能语音

智能语音业务服务能力应包括:

- a) 语音识别:具备噪声场景中唤醒词识别和将语音转换为文本的能力;
- b) 语音合成:具备将文本转换为语音的能力;
- c) 声纹识别:具备通过声音判别说话者身份的能力。

#### 5.3.2 自然语言理解

自然语言理解业务服务能力应包括:

- a) 词法、句法分析:具备词法分析、句法分析和语义分析的能力;
- b) 信息抽取:具备从文本、音频、图像或视频等数据中抽取信息的能力;
- c) 信息检索:根据用户需求,查找出所需信息的能力;
- d) 语义推理:根据已有的知识,用规则或模型,扩展或推理现有知识的能力;
- e) 情感分析:通过对自然语言文本的分析,识别文本作者或文本所描述的个体的情感的能力;

f) 知识图谱:具备构建知识图谱并利用知识图谱进行知识计算和推理等方面的能力。

### 5.3.3 智能对话

智能对话业务服务能力包括:

- a) 闲聊:闲聊对话中具备上下文衔接和提高用户粘性的能力;
- b) 知识型问答:具备为用户提供知识内容、知识搜索等答疑解惑服务的能力;
- c) 任务导向对话:具备响应并引导用户的需求,达成任务的能力;
- d) 情感对话:具备与人对话,结合主题、情景等考虑对人进行情感支持的能力;
- e) 启发式对话:具备在用户不知道如何做或者能做什么时,给用户提示,引导用户继续对话,考虑信息的利用程度、自主性以及学习的能力;
- f) 多模态对话:具备在对话中可同时考虑语音、视频、动作、事件输入,以及表情、动作、语音等多模态输出,考虑模态数量、模态选择策略、学习的能力。

### 5.3.4 智能图像视频分析

智能图像视频分析业务服务能力应包括:

- a) 识别能力:具备至少一项识别服务能力,如人脸识别、物体识别和文字识别等;识别过程中防攻击的能力,如防呈现攻击;
- b) 理解能力:具备准确和高效理解图像内容信息的能力;
- c) 业务扩展性:具备不同设备能力情况下的多种不同要求的服务业务的能力;
- d) 服务稳定性:具备图像视频稳定性的能力,包括容错和负载均衡等。

## 6 成熟度等级及评估分值

### 6.1 成熟度等级

人工智能服务能力成熟度等级分为 5 级:初始级、发展级、稳健级、优秀级和卓越级,见表 2。

每个成熟度等级表明人工智能服务能力所达到的成熟度水平。人工智能服务能力的改进和成熟度提升是通过渐进的方式来实现的,较高的成熟度等级涵盖了低于其成熟度等级的全部要求。

表 2 人工智能服务能力成熟度分级

等级	特征
初始级	人工智能服务提供商能提供基本的人工智能服务,基础能力域的评分满足最低的要求,业务能力域的评分符合最低的要求
发展级	人工智能服务提供商能提供较多的人工智能服务,基础能力域的评分满足中等的要求,业务能力域的评分符合最低的要求
稳健级	人工智能服务提供商能提供稳定的人工智能服务,基础能力域的评分满足中等的要求,业务能力域的评分满足中等的要求
优秀级	人工智能服务提供商能提供优秀的人工智能服务,基础能力域的评分满足最高的要求,业务能力域的评分满足中等的要求
卓越级	人工智能服务提供商能提供卓越的人工智能服务,基础能力域的评分满足最高的要求,业务能力域的评分满足最高的要求

## 6.2 成熟度评估分值

### 6.2.1 基础能力

#### 6.2.1.1 存储

存储能力项的能力子项成熟度评估分值见表 3,如未达到表 3 中最低评分要求或不具备该能力项功能的应评为 0 分。

表 3 存储能力项成熟度评估分值

能力子项	能力要求	分值
存储管理	高:具备 5 种或以上异构数据的采集能力、EB 级分布式存储平台;具备千亿以上数据条目的表格存储系统	5
	中:具备 2 种或以上异构数据的数据采集技术、较强能力的存储平台	3
	低:具备 1 种结构数据的采集技术,受限能力的存储平台	1
多样性	高:具备 5 种或以上的自由存储系统,包括关系数据库、键值存储服务、文件存储服务、结构化存储服务、弹性消息队列等;具备 5 种或以上语言 SDK 包括 Java、Python、Go、JavaScript、Node.js 等语言	5
	中:具备 2 种或以上的存储系统以及 2 种或以上语言 SDK	3
	低:具备 1 种存储系统及 1 种特定语言 SDK	1
分布式可扩展	高:具备分布式架构的集群系统,系统能够很方便增加或者减少服务器节点个数,以应对业务负载的变化,并且这样的操作是自动化的,减少运维负担	5
	中:具备分布式架构的集群系统,但无法自动化地根据业务负载调整服务器个数,运维成本适中	3
	低:不具备分布式架构的集群系统,服务器节点变化对于用户影响较大,运维成本较高	1
其他	高:提供安全高效的存储服务,自动故障迁移与负载均衡,具备底层资源双可用区支撑,数据一致;系统给用户提供简单易懂的库和接口,方便用户使用	5
	中:具备安全的存储服务,可快速处理服务器故障	3
	低:可感知服务器故障,无法提供安全高效的存储服务	1

#### 6.2.1.2 计算

计算能力项的能力子项成熟度评估分值见表 4,如未达到表 4 中最低评分要求或不具备该能力项功能的应评为 0 分。

表 4 计算能力项成熟度评估分值

能力子项	能力要求	分值
数据计算	高:具备基于云计算虚拟化技术的超大规模分布式计算的通用系统;具备批处理计算、流式计算、图计算和机器学习计算等复杂任务;具备大规模机器学习系统和异构深度学习系统	5
	中:具备大规模分布式计算的通用系统;具备批量计算、流式计算、图计算和机器学习计算等复杂任务	3
	低:不具备大规模的数据计算能力,无法支持复杂度较高的数据计算任务	1
集群效率	高:能简单高效地提供大规模集群的部署、调度、扩展和管理服务	5
	中:能较高效地提供较大规模集群的部署、调度、扩展和管理服务,但便捷性有限	3
	低:能提供初步规模集群的部署、调度、扩展和管理服务,便捷性较低	1
稳定性	高:具备底层资源双区域可用性部署支撑,使故障自动恢复、快速迁移,可提供连续稳定高可用服务	5
	中:部分故障能自动恢复和快速迁移,可提供连续稳定基础服务	3
	低:故障不能自动恢复和迁移,不能提供连续稳定基础服务	1
可扩展性	高:具备低时延(如秒级水平)扩容,提供简洁的预定义扩展接口,用户可从较低计算规模扩展到较大规模,而运行应用程序的方式并不会因此变得复杂	5
	中:具备较低时延(多于秒级水平)扩容,提供扩展接口	3
	低:扩容时延较长,不提供扩展接口	1
灵活性	高:能根据业务变化和负载情况,对计算资源进行弹性伸缩;提供便捷易用的自动化工具,具备批处理计算、流式计算等多种常用框架和 Web 控制台、API 和 SDK 等	5
	中:能根据业务变化和负载情况之一对计算资源进行调整;提供自动化工具,具备部分常用框架和 Web 控制台、API 和 SDK 等	3
	低:不能完全根据业务变化和负载情况对计算资源进行调整;没有自动化工具,只具备个别常用框架、Web 控制台、API 或 SDK	1
统计监控	高:能量化整体数据流的数据质量情况,实时统计各模块的数据延迟和收发状态,提供数据流全链路监控,在数据异常和服务异常时能根据规则触发报警	5
	中:能量化部分数据流的数据质量情况,能够统计各模块的数据延迟和收发状态,提供数据流部分链路监控,在数据异常或服务异常出现时可触发报警	3
	低:无法完善量化数据流的数据质量情况,没有完善统计各模块的数据延迟和收发状态,提供数据流少量链路的监控,当个别数据异常和服务异常时有报警机制	1

#### 6.2.1.3 数据结构化处理与决策

数据结构化处理与决策能力项的能力子项成熟度评估分值见表 5,如未达到表 5 中最低评分要求或不具备该能力项功能的应评为 0 分。

表 5 数据结构化处理与决策能力项成熟度评估分值

能力子项	能力要求	分值
稳定性	高:系统能完全管理且容易配置;针对噪声污染、对抗数据攻击等情形,所提供的数据智能分析业务稳定,输出结果一致	5
	中:基本可管理,配置有一定难度,仅能针对噪声污染提供有限的业务稳定性	3
	低:完全不可管理,配置困难,所提供的数据智能分析业务仅能在较为理想的条件下提供业务,存在鲁棒性问题	1
可用性	高:能提供高准确率(大于或等于 95%)、低错误率(小于或等于 5%)、高灵敏度(大于或等于 95%)、高特异性(大于或等于 95%)的分类精度;决策科学合理,可行性很高,易于接受;具备用户在数据处理与决策的全过程中实时交互,同时系统可为每位用户建立自适应的协同模型;服务可在不暂停服务的同时开展完全的系统维护和算法、模型等模块更新	5
	中:能提供较高准确率(大于或等于 80%, 小于 95%)、较低错误率(大于 5%, 小于或等于 20%)、较高灵敏度(大于或等于 80%, 小于 95%)、较高特异性(大于或等于 80%, 小于 95%)的分类精度;决策基本合理、可行,可以接受;在大部分环节中(大于 60%, 小于 100%)具备用户在数据处理与决策中与系统交互,系统可为部分用户建立自适应的协同模型;大于 50%但小于 100%的系统维护和算法、模型等模块更新操作时,无需暂停服务	3
	低:仅能提供较高准确率(大于或等于 90%, 小于 95%),但可能出现灵敏度(小于 80%)或特异性较低情形(小于 80%),影响用户对分类结果的使用;决策不合理,不可行,无效,无法接受;仅在较少环节(小于 20%)具备用户在数据处理与决策中与系统交互,系统不具备为用户建立自适应协同模型;系统维护和算法、模型等模块更新时需暂停服务	1
时效性	高:快速反应,在一定的时间限制和要求内能够快速处理并给出结果,并能感知变化,如方案已过时,可根据新的情况及时调整	5
	中:及时反应,在一定的时间限制和要求内能够及时处理并给出结果,保持实施期间的相对稳定性	3
	低:延迟反应,在处理和决策时拖延时间,决策的效益低,甚至会因问题的性质发生根本变化,而使决策无效	1
可解释性	高:数据处理流程和分析结果具有完全的可解释性和可预测性,处理和决策过程可理解,业务可解释性高,可呈现为供有资质第三方审查的文本(超过 90%),不存在黑盒	5
	中:数据处理流程存在无法预判的随机分支,处理和决策过程可基本理解,业务可基本解释,决策模型方法基本可解释说明(超过 50%),但更换同领域其他数据集时,90%以上分析结果可预测	3
	低:数据处理流程可解释性差,处理和决策过程难理解,业务无法解释,决策行为很难甚至无法解释说明(低于 50%)	1
通用性	高:面向一个领域的服务易于迁移至另一业务领域,同时能够综合多种智能处理和决策方法,可处理多个场景的知识,并做出有效决策	5
	中:仅能针对某些共性领域开展稳定可靠的服务,适用某些场景的知识处理,针对某些特定的问题可做出有效决策	3
	低:仅针对某一特定领域提供服务或只适用于单个特定场景	1

#### 6.2.1.4 机器学习

机器学习能力项的能力子项成熟度评估分值见表 6,如未达到表 6 中最低评分要求或不具备该能力项功能的应评为 0 分。

表 6 机器学习能力项成熟度评估分值

能力子项	能力要求	分值
全生存周期	高:具备包含数据框架(采集、筛选、预标注、标注、脱敏等)、特征探索、模型开发(训练、评估和预测)、模型部署、模型管理、应用发布等全生存周期过程	5
	中:具备以上所述 80% 或以上的生存周期过程	3
	低:只具备以上所述 50% 或以上的生存周期过程	1
训练性能	高:具备超高性能 GPU 运算,支持大容量内存以上的单机运算能力;具备数据并行和模型并行、单机多卡和多机多卡的分布式训练;同时具备 CPU 服务器集群、虚拟机集群计算能力	5
	中:具备较高性能 GPU 运算,支持中等容量内存以上的单机运算能力;部分具备数据并行和模型并行、单机多卡和多机多卡的分布式训练;部分具备 CPU 服务器集群、虚拟机集群计算能力	3
	低:不具备高性能 GPU 运算能力,只有简单的并行计算能力;不能完善的具备 CPU 服务器集群、虚拟机集群计算能力	1
兼容性	高:支持不少于五种主流深度学习框架的标准 API 接口,确保与广泛使用的样本代码高度兼容。同一套模型代码能够在不同的云平台环境下无缝迁移与执行	5
	中:能够良好地集成至少两种主流深度学习框架的标准 API 接口,确保与多数常用样本代码的兼容性	3
	低:仅支持一种深度学习框架的标准 API 接口,不兼容其他样本代码	1
灵活性	高:具备按需申请和分配 CPU、内存和 GPU 资源	5
	中:在部分维度上具备按需申请和分配 CPU、内存和 GPU 资源	3
	低:不具备按需申请和分配 CPU、内存和 GPU 资源	1
易用性	高:具备 5 种或以上简单易用的命令行工具,可通过 API、SDK 或者 Web 控制台使用云深度学习服务;具备易用的方式提交云端训练,具备 5 种或以上主流深度学习框架和超参数自动调优等功能;具备可直观显示的拖拽布局编程服务,组合各种数据源、组件、算法、模型和评估模块	5
	中:具备 2 种或以上易用的命令行工具,可通过 API、SDK 或者 Web 控制台中的某一种使用云深度学习服务;具备云端训练;具备部分功能的可直观显示	3
	低:所使用的工具易用性不高,无法通过 API、SDK 或者 Web 控制台使用云深度学习服务;不具备云端训练;不具备可直观显示	1
算法库	高:具备提供插件式算子注入能力;具备 70 种以上类型的算子及算法,具备时空分析和频繁项挖掘的功能	5
	中:具备 30 种以上类型的算子及算法	3
	低:只具备 5 种以下类型的算子及算法	1

## 6.2.2 业务能力

### 6.2.2.1 智能语音

智能语音能力项的能力子项成熟度评估分值见表 7,如未达到表 7 中最低评分要求或不具备该能力项功能的应评为 0 分。

表 7 智能语音能力项成熟度评估分值

能力子项	能力要求	分值
语音识别	高:在不同信噪比状态下,识别率高(大于 90%)	5
	中:在不同信噪比状态下,识别率较高(小于或等于 90%,大于或等于 80%)	3
	低:在不同信噪比状态下,识别率较低(小于 80%)	1
语音合成	高:分词准确,句读流畅,用户可准确理解,与自然人发声没有显著区别	5
	中:分词较准确,句读较为流畅,用户基本可理解,与自然人发声有一定区别	3
	低:分词不准确,句读不流畅,影响用户理解,与自然人发声有较大区别	1
声纹识别	高:错误接受率小于或等于 0.25%,错误拒绝率小于或等于 1.5%	5
	中:错误接受率小于或等于 0.5%,错误拒绝率小于或等于 3%	3
	低:错误接受率小于或等于 0.75%,错误拒绝率小于或等于 5%	1

### 6.2.2.2 自然语言理解

自然语言理解能力项的能力子项成熟度评估分值见表 8,如未达到表 8 中最低评分要求或不具备该能力项功能的应评为 0 分。



表 8 自然语言理解能力项成熟度评估分值

能力子项	能力要求	分值
词法、句法分析	高:词、句识别和语义理解错误率很低(小于 5%)、召回率高(或称命中率不小于 90%),可根据上下文理解复杂句子,响应时间符合自然人类问答习惯,完全不影响语音服务功能和语音交互用户体验	5
	中:词、句识别和语义理解错误率较低(小于 10%)、召回率较高(或称命中率不小于 75%),可理解复杂句子,对语音服务功能和语音交互用户体验有少量的影响	3
	低:词、句识别和语义理解错误率较高(大于或等于 20%),对语音服务功能和语音交互用户体验有明显影响	1
信息抽取	高:能提取关键词,确定实体及实体间的显示语义关系,结合上下文提取实体间的隐含语义关系、背景知识、常识知识	5
	中:能提取关键词,确定实体及实体间的显示语义关系,结合上下文提取实体间的隐含语义关系	3
	低:能提取关键词,确定实体及实体间的显示语义关系	1

表 8 自然语言理解能力项成熟度评估分值(续)

能力子项	能力要求	分值
信息检索	高:符合所有信息需求,根据用户意图对结果做准确排序	5
	中:符合基本和重要信息需求,能够对结果做简单排序	3
	低:符合基本信息需求	1
语义推理	高:可实现通用的逻辑推理,准确率大于或等于 90%,召回率不低于 80%	5
	中:可实现特定领域的逻辑推理,准确率大于或等于 90%,召回率不低于 50%	3
	低:不能进行推理	1
情感分析	高:可准确区分 10 种以上(包括但不限于快乐、着急、愤怒、热情、鄙视、哀伤等)不同的情绪,每一种情绪区分至少 3 个级别(准确率大于 90%)	5
	中:可准确区分(准确率大于 90%)正面、负面或中性的情绪	3
	低:无法相对准确地区分正面、负面或中性的情绪	1
知识图谱	高:具备通用领域和专用领域知识图谱的构建,具备利用知识图谱进行跨领域的知识计算和推理,支持多个语种类别	5
	中:具备专用领域知识图谱的构建,具备利用知识图谱进行知识计算和推理,支持多个语种类别	3
	低:具备专用领域知识图谱的构建,具备利用知识图谱进行知识计算和推理,支持单个语种类别	1

#### 6.2.2.3 智能对话

智能对话能力项的能力子项成熟度评估分值见表 9,如未达到表 9 中最低评分要求或不具备该能力项功能的应评为 0 分。

表 9 智能对话能力项成熟度评估分值

能力子项	评估要求	分值
闲聊	高:平均对话单轮次数高,上下文衔接很流畅	5
	中:平均对话单轮次数中,上下文衔接较流畅	3
	低:平均对话单轮次数低,上下文衔接不流畅	1
知识型问答	高:能在限定和公开的知识库中根据关键词查询答案;能正确表示用户的问题,并自动转换为不同知识库支持的查询语句;根据已有会话,自动识别其中的上下文,从而进行相应的应答;依靠少量标注数据和用户反馈,能逐步学会回答新领域里的问题;能够根据用户问题进行知识图谱扩充和完善	5
	中:能够在限定和公开的知识库中根据关键词查询答案;根据已有的知识图谱识别用户意图,给出相应回复;有利用上下文信息的能力,依靠少量标注数据和用户反馈,能逐步学会回答新领域里的问题	3
	低:只能在限定领域的知识库中根据关键词查找答案	1

表 9 智能对话能力项成熟度评估分值(续)

能力子项	评估要求	分值
任务导向对话	高:能够完成单句指令任务和多轮交互型任务。在多轮交互型任务上,支持带有固定服务目标的对话引导;可在用户目标及其目标间寻找最优目标组合,并以此组合目标为基础,支持基于用户个性化数据的带目标的任务型对话,并能自主选择最佳的交互模态	5
	中:能够完成单句指令任务和多轮交互型任务。在多轮交互型任务上,具有带有固定服务目标的对话引导	2~3
	低:能够完成单句指令任务和多轮交互型任务	1
情感对话	高:在多轮对话中,理解并引导用户情绪,支持视频等多模态交互;在与用户会话过程中,仅根据用户的反馈,自动学会应对不同的用户情绪、主题和场景;可发掘用户面临的问题或挑战,并提出新的观点,并引导用户接受新观点	5
	中:在多轮对话中,能够识别主人情感,能结合当前主题和情景,给予合适应答以达成情感支持目的;可根据标注数据,学会应对不同的用户情绪、主题和场景	3
	低:单轮对话中,能够识别主人的正向、负向、中性的情感,并采用相应情感应答	1
启发式对话	高:根据已标注的数据,产生新的触发条件和引导方式;根据用户画像、具有相近用户画像的用户交互历史,预测用户最可能感兴趣的对话内容;可从已有的交互记录以及用户对历史引导的反馈数据中,自动学会何时以及怎样引导用户最好;可根据用户对引导的反馈,自主调整引导策略	5
	中:根据标注好的数据,学会新的触发条件和引导方式;根据用户画像、具有相近用户画像的用户交互历史,预测用户最可能感兴趣的对话内容	3
	低:根据一定的策略和推荐规则给出问题,能自然过渡到预先设定的目标话题	1
多模态对话	高:符合所有信息需求,能根据用户意图对结果进行准确排序,重要结果置顶	5
	中:符合所有信息需求,能对结果进行简单排序	3
	低:符合基本和重要信息需求	1

#### 6.2.2.4 智能图像视频分析

智能图像视频分析能力项的能力子项成熟度评估分值见表 10,如未达到表 10 中最低评分要求或不具备该能力项功能的应评为 0 分。

表 10 智能图像视频分析能力项成熟度评估分值

能力子项	能力要求	分值
识别能力	高:至少具备一项识别服务如人脸识别、物体识别、文字识别等,识别准确率较高	5
	中:至少具备一项识别服务如人脸识别、物体识别、文字识别等,识别准确率一般	3
	低:只具备特定的识别服务,识别准确率较低	1
理解能力	高:能够良好解读图像的内容信息,识别正确率大于或等于 90%,解读效率高、用时短,完全不影响用户体验	5
	中:能够完成解读图像内容信息的功能,识别正确率大于或等于 80%,低于 90%,解读效率和用时适中,只少量影响用户体验	3
	低:无法完成解读图像内容信息的功能,识别正确率明显影响用户体验(低于 80%),解读效率低,用时较长	1

表 10 智能图像视频分析能力项成熟度评估分值(续)

能力子项	能力要求	分值
业务扩展性	高:具备不同设备能力情况下的 5 种或以上不同要求的服务业务,具备基于云平台和近端的视觉服务	5
	中:具备 2 种或以上不同要求的服务业务,仅具备基于近端的视觉服务	3
	低:仅具备对于时延要求不敏感的业务,仅具备基于近端的视觉服务	1
服务稳定性	高:提供由多台高性能服务器组成的集群的服务后台,拥有较高的容错和负载均衡能力,为客户提供稳定高效的图像视频服务	5
	中:提供由数台性能稳定的服务器组成的集群的服务后台,具备基本的容错能力和初步的负载均衡机制,能够为客户提供较为稳定和有效的图像视频处理服务	3
	低:基于单台服务器的服务后台,缺乏专门的容错设计和负载均衡支持,适合对稳定性要求不高、流量较小的场景,为客户提供基础的图像视频服务	1

7 成熟度评估方法

## 7.1 能力项权重

对人工智能服务能力进行成熟度评估时,应根据不同应用场景和不同人工智能服务类型,选择所需的能力项,并调整相适应的评估要素构建不同的能力子项评估要素。部分场景(家用智能设备场景、智能医疗诊疗场景、工业设备预测性维护场景)的能力项见附录A。

## 7.2 计算方法

服务能力成熟度分级可采用评分量化的方式进行。

服务能力成熟度等级得分  $A$  按公式(1)计算。

式中：

A——成熟度等级得分；

*k* ——能力域成熟度得分；

$j$  ——能力域权重。

能力域成熟度得分  $k$  按公式(2)计算。

式中：

*k* ——能力域成熟度得分；

*m*——能力项成熟度得分：

$n$  ——能力项权重。

能力项成熟度得分  $m$  按公式(3)计算。

式中：

$m$  ——能力子项成熟度得分；

$s$  ——能力子项成熟度得分；

$t$  ——能力子项权重。

### 7.3 成熟度等级判定

当评估对象在某一成熟度等级下的成熟度得分超过评分区间的最低分视为符合该成熟度等级要求，反之，则视为不符合。

成熟度实际得分  $A$  与成熟度等级的对应关系见表 11，根据得到的相应的成熟度得分，对照相应服务能力成熟度等级。

表 11 分数与成熟度等级对应关系

等级	成熟度等级得分 $A$
初始级	$1 \leqslant A < 1.5$
发展级	$1.5 \leqslant A < 2.5$
稳健级	$2.5 \leqslant A < 3.5$
优秀级	$3.5 \leqslant A < 4.5$
卓越级	$4.5 \leqslant A \leqslant 5$



附录 A  
(资料性)  
部分场景能力项权重

#### A.1 家用智能设备场景

家用智能设备场景的能力项权重见表 A.1。

表 A.1 家用智能设备场景能力项权重

能力域	能力项	能力子项	能力子项权重
基础能力 (权重 50%)	存储 (权重 20%)	存储管理	25%
		多样性	25%
		分布式可扩展	20%
		其他	30%
	计算 (权重 30%)	数据计算	20%
		集群效率	20%
		稳定性	20%
		可扩展性	15%
		灵活性	15%
		统计监控	10%
	数据结构化处理与决策 (权重 20%)	稳定性	20%
		可用性	20%
		时效性	20%
		可解释性	20%
		通用性	20%
	机器学习 (权重 30%)	全生存周期	20%
		训练性能	15%
		兼容性	15%
		灵活性	15%
		易用性	15%
		算法库	20%

表 A.1 家用智能设备场景能力项权重(续)

能力域	能力项	能力子项	能力子项权重
业务能力 (权重 50%)	智能语音 (权重 20%)	语音识别	50%
		语音合成	40%
		声纹识别	10%
	自然语言理解 (权重 20%)	词法、句法分析	30%
		信息抽取	15%
		信息检索	15%
		语义推理	15%
		情感分析	15%
		知识图谱	10%
	智能对话 (权重 20%)	闲聊	20%
		知识型问答	20%
		任务导向对话	20%
		情感对话	15%
		启发式对话	15%
		多模态对话	10%
	智能图像视频分析 (权重 20%)	识别能力	50%
		理解能力	20%
		业务扩展性	15%
		服务稳定性	15%
	数据结构化处理与决策 (权重 20%)	稳定性	30%
		可用性	30%
		时效性	15%
		可解释性	10%
		通用性	15%

## A.2 智能医疗诊疗场景

智能医疗诊疗场景的能力项权重见表 A.2。

表 A.2 智能医疗诊疗场景能力项权重

能力域	能力项	能力子项	能力子项权重
基础能力 (权重 50%)	存储 (权重 20%)	存储管理	30%
		多样性	10%
		分布式可扩展	30%
		其他	30%
	计算 (权重 30%)	数据计算	20%
		集群效率	20%
		稳定性	30%
		可扩展性	5%
		灵活性	5%
		统计监控	20%
	数据结构化处理与决策 (权重 20%)	稳定性	20%
		可用性	20%
		时效性	20%
		可解释性	20%
		通用性	20%
	机器学习 (权重 30%)	全生存周期	20%
		训练性能	20%
		兼容性	10%
		灵活性	15%
		易用性	25%
		算法库	10%

表 A.2 智能医疗诊疗场景能力项权重（续）

能力域	能力项	能力子项	能力子项权重
业务能力 (权重 50%)	智能语音 (权重 20%)	语音识别	60%
		语音合成	40%
		声纹识别	0%
	自然语言理解 (权重 20%)	词法、句法分析	30%
		信息抽取	15%
		信息检索	15%
		语义推理	15%
		情感分析	15%
		知识图谱	10%
	智能对话 (权重 15%)	闲聊	5%
		知识型问答	20%
		任务导向对话	35%
		情感对话	15%
		启发式对话	10%
		多模态对话	15%
	智能图像视频分析 (权重 25%)	识别能力	60%
		理解能力	10%
		业务扩展性	5%
		服务稳定性	25%
	数据结构化处理与决策 (权重 20%)	稳定性	50%
		可用性	15%
		时效性	15%
		可解释性	10%
		通用性	10%

### A.3 工业设备预测性维护场景

工业设备预测性维护场景的能力项权重见表 A.3。

表 A.3 工业设备预测性维护场景能力项权重

能力域	能力项	能力子项	能力子项权重
基础能力 (权重 50%)	存储 (权重 20%)	存储管理	25%
		多样性	25%
		分布式可扩展	20%
		其他	30%
	计算 (权重 30%)	数据计算	20%
		集群效率	20%
		稳定性	20%
		可扩展性	15%
		灵活性	15%
		统计监控	10%
机器学习 (权重 30%)	数据结构化处理与决策 (权重 20%)	稳定性	20%
		可用性	20%
		时效性	20%
		可解释性	20%
		通用性	20%
	机器学习 (权重 30%)	全生存周期	15%
		训练性能	25%
		兼容性	15%
		灵活性	10%
		易用性	15%
		算法库	20%



表 A.3 工业设备预测性维护场景能力项权重 (续)

能力域	能力项	能力子项	能力子项权重
业务能力 (权重 50%)	自然语言理解 (权重 15%)	词法、句法分析	30%
		信息抽取	10%
		信息检索	10%
		语义推理	10%
		情感分析	10%
		知识图谱	30%
	智能图像视频分析 (权重 15%)	识别能力	50%
		理解能力	20%
		业务扩展性	15%
		服务稳定性	15%
	数据结构化处理与决策 (权重 70%)	稳定性	50%
		可用性	15% 
		时效性	15%
		可解释性	10%
		通用性	10%

