



中华人民共和国国家标准

GB/T 45288.3—2025

人工智能 大模型 第3部分：服务能力成熟度评估

Artificial intelligence—Large-scale model—
Part 3:Service capability maturity assessment

2025-01-24 发布

2025-01-24 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	IV
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 概述	2
5.1 大模型服务类型	2
5.2 服务能力框架	2
6 评估指标	3
6.1 大模型平台	3
6.2 大模型开发定制	7
6.3 大模型推理及运营	9
7 成熟度分级规则	11
7.1 成熟度等级	11
7.2 能力要求	12
8 成熟度评估方法	13
8.1 评分方法	13
8.2 评估域权重	13
8.3 计算方法	13
8.4 成熟度等级判定	14

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。

本文件是 GB/T 45288《人工智能 大模型》的第 3 部分，GB/T 45288 已经发布了以下部分：

- 第 1 部分：通用要求；
- 第 2 部分：评测指标与方法；
- 第 3 部分：服务能力成熟度评估。

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、华为技术有限公司、浪潮云信息技术股份公司、清华大学、华为云计算技术有限公司、中国科学院自动化研究所、北京百度网讯科技有限公司、深圳市腾讯计算机系统有限公司、菲特(天津)检测技术有限公司、北京奇虎科技有限公司、北京航空航天大学、国能信息技术有限公司、麒麟合盛网络技术股份有限公司、上海市人工智能行业协会、上海燧原科技股份有限公司、阿里云计算有限公司、平头哥(上海)半导体技术有限公司、上海计算机软件技术开发中心、浙江大华技术股份有限公司、青岛海信电子技术服务有限公司、上海人工智能研究院有限公司、南方电网人工智能科技有限公司、航天信息股份有限公司、广东电网有限责任公司、北京大学长沙计算与数字经济研究院、北京大学、中国科学院软件研究所、蚂蚁科技集团股份有限公司、中国移动通信集团有限公司、马上消费金融股份有限公司、深圳云天励飞技术股份有限公司、深圳思谋信息科技有限公司、北京格灵深瞳信息技术股份有限公司、中国南方电网有限责任公司超高压输电公司、北京软件产品质量检测检验中心有限公司、中国电力科学研究院有限公司、上海文镭信息科技有限公司、浪潮软件科技有限公司、浪潮电子信息产业股份有限公司、浪潮软件集团有限公司、中电科大数据研究院有限公司、上海商汤智能科技有限公司、中国电信集团有限公司、科大讯飞股份有限公司、中国电信股份有限公司北京研究院、中移(苏州)软件技术有限公司、中国科学院新疆理化技术研究所、杭州海康威视数字技术股份有限公司、上海文镭信息科技有限公司、西北工业大学、云知声智能科技股份有限公司、北京工业大学、北京智芯电子科技有限公司。

本文件主要起草人：徐洋、马珊珊、于超、王莞尔、董建、陶建华、曹晓琦、鲍薇、黄现翠、马骋昊、郑佳佳、郑子木、朱贵波、王金桥、刘静、汪群博、杨旭、马同森、靳伟、刘海涛、曹彬、张向征、任海峰、刘祥龙、刘艾杉、张旭、陈曦、赵春昊、蒋燕、梅敬青、彭骏涛、张艺伯、陈敏刚、孔维生、刘微、刘常昱、宋海涛、任正国、邵彦宁、刘佳宁、周昊、杨超、孟令中、孙曦、金镛、李宽、王志芳、吕江波、胡全一、王宁、王志刚、孔昊、莫文昊、仲凯韬、王珂琛、刘璐、张天霖、蒋慧、刘敬谦、刘威辰、高建清、孟建、舒珏淋、商兴宇、李旭东、杨雅婷、钟凯伦、仲凯韬、张涛、梁家恩、刘峥、郑哲、武姗姗。

引 言

大模型已成为人工智能发展的重要技术手段,在引领产业变革中发挥重要作用,国内外人工智能相关机构相继研究开发百余种大模型产品和评测榜单,导致用户难以有效评价人工智能产品的技术水平和服务能力。GB/T 45288 旨在规定通用大模型的技术要求、评测指标和服务能力,拟由五个部分构成。

- 第 1 部分:通用要求。目的在于确立大模型的参考架构,规定通用技术要求。
- 第 2 部分:评测指标与方法。目的在于确立大模型的评测指标,描述评测方法。
- 第 3 部分:服务能力成熟度评估。目的在于给出大模型服务能力成熟度等级及评估方法。
- 第 4 部分:计算机视觉大模型。目的在于定义计算机视觉大模型的概念和功能,规定技术要求和测试方法。
- 第 5 部分:多模态大模型。目的在于定义多模态大模型的概念和功能,规定技术要求和测试方法。

人工智能 大模型

第3部分：服务能力成熟度评估

1 范围

本文件给出了大模型服务能力框架和评估指标,描述了大模型服务能力成熟度等级划分及评估方法。

本文件适用于服务提供方和需求方对大模型平台、模型定制及推理运营服务的能力进行全面评估,也适用于指导大模型服务能力的规划、设计和实现。

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 42018—2022 信息技术 人工智能 平台计算资源规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

大模型平台 large-scale model platform

为开发或使用大模型提供各类资源的软硬件平台。

注：大模型平台不包含大模型。

3.2

大模型服务 large-scale model service

开发、应用大模型及大模型系统的服务,以及以此为手段提供支持需方业务活动的服务。

注：大模型系统是大模型与大模型平台的集成,是与大模型服务相关的活动、过程等的集合。

3.3

工具链 toolchain

用于支撑大模型开发、定制及应用的软件集合。

3.4

指令 instruct

由大模型输入和输出信号构成的二元组。

注：如自然语言处理的大模型,指令以提问文本和回答文本成对呈现。

4 缩略语

下列缩略语适用于本文件。

AI:人工智能(Artificial Intelligence)