



中华人民共和国国家标准

GB/T 45288.1—2025

人工智能 大模型 第 1 部分：通用要求

Artificial intelligence—Large-scale model—Part 1: General requirements

2025-02-28 发布

2025-02-28 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 参考架构 2

5 通用要求 3

 5.1 资源池 3

 5.2 工具 4

 5.3 数据资源 6

 5.4 模型 6

 5.5 行业应用 7

 5.6 服务平台/组件 7

参考文献..... 8

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1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。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、上海人工智能创新中心、华为云计算技术有限公司、蚂蚁科技集团股份有限公司、清华大学、中国科学院自动化研究所、北京中关村实验室、北京百度网讯科技有限公司、中国铁建股份有限公司、北京奇虎科技有限公司、中国南方电网有限责任公司、中国移动通信有限公司研究院、国家能源投资集团有限责任公司信息技术分公司、杭州联汇科技股份有限公司、北京智源人工智能研究院、阿里云计算有限公司、深圳市腾讯计算机系统有限公司、科大讯飞股份有限公司、国网湖北省电力有限公司、华为技术有限公司、天津大学、中铁第五勘察设计院集团有限公司、上海商汤智能科技有限公司、北京航空航天大学、浪潮云信息技术股份公司、上海市人工智能行业协会、哈尔滨工业大学、西南科技大学、北京大学、西安电子科技大学、北京赛西科技发展有限责任公司、中国科学院软件研究所、北京大学武汉人工智能研究院、北京大学长沙计算与数字经济研究院、青岛海尔科技有限公司、北京格灵深瞳信息技术股份有限公司、北京工业大学、中山大学、中国电信集团有限公司、北京软件产品质量检测检验中心有限公司、北京小米移动软件有限公司、北京智芯微电子科技有限公司、北京世纪好未来教育科技有限公司、杭州海康威视数字技术股份有限公司、昆仑数智科技有限责任公司、浪潮电子信息产业股份有限公司、青岛海信电子技术服务股份有限公司、北京中关村科金技术有限公司、天翼云科技有限公司、浪潮软件科技有限公司、上海燧原科技股份有限公司、马上消费金融股份有限公司、上海天数智芯半导体有限公司、咪咕文化科技有限公司、平头哥(上海)半导体技术有限公司、麒麟合盛网络技术股份有限公司、上海文骐信息科技有限公司、深圳前海微众银行股份有限公司、深圳思谋信息科技有限公司、云知声智能科技股份有限公司、山东浪潮科学研究院有限公司、山东省人工智能研究院、上海计算机软件技术开发中心、上海人工智能研究院有限公司、同方知网数字出版技术股份有限公司、安徽大学、西门子(中国)有限公司、云从科技集团股份有限公司、浙江大华技术股份有限公司、中电信数智科技有限公司、南方电网人工智能科技有限公司、中国移动通信集团有限公司、中移互联网有限公司、湖南科创纺织股份有限公司、中移(苏州)软件技术有限公司、西北工业大学。

本文件主要起草人：董建、徐洋、叶珩、乔宇、鲍薇、曹晓琦、孙曦、陶建华、刘静、王嘉凯、张军、李栋、张向征、余芸、刘伟东、经迪春、赵天成、林咏华、马珊珊、马骋昊、王莞尔、李建欣、熊德意、吴涛、黄超、王士进、彭祥礼、郑中、郑子木、蒋慧、刘祥龙、汪群博、郑佳佳、高东辉、马同森、张天霖、黄现翠、孙传兴、何逸楠、赵春昊、杨沐昀、俞文心、杨超、何刚、郝文建、薛云志、刘艾杉、吴玺宏、刘尚、余甜、刘颖、陈曦、郑若琳、沈芷月、聂简荻、王先庆、王金桥、胡全一、朱贵波、韩红桂、潘恩荣、武姗姗、孔昊、于磊、郑哲、刘子韬、朱江、陈宏志、范宝余、刘微、崔明飞、高鹏军、张峰、梅敬青、曾定衡、宋煜、赵磊、高慧、张旭、仲凯韬、李斌、刘枢、梁家恩、魏子重、舒明雷、陈敏刚、孟令中、王资凯、刘长欣、范存航、生若谷、孙进、孔维生、陈利明、郑桦、赵晓玮、冯俊兰、杨玉宽、孙文庆、朱林、曾杰、钱岭、张涛。

引 言

大模型已成为人工智能发展的重要手段,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《人工智能 大模型》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 1 部分：通用要求

1 范围

本文件确立了大模型的参考架构,规定了大模型的通用要求。

本文件适用于大模型开发、制备、部署和应用。

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

GB/T 42755—2023 人工智能 面向机器学习的数据标注规程

GB/T 45401.1—2025 人工智能 计算设备调度与协同 第 1 部分：虚拟化与调度

3 术语和定义

下列术语和定义适用于本文件。

3.1

大模型 **large-scale model**

大规模深度学习模型 **large-scale deep learning model**

基于大量数据训练得到,具有复杂计算架构,能处理复杂任务,且具备一定泛化性的深度学习模型。

注:大模型的参数量由其功能和模态决定,一般不低于 1 亿。大模型训练使用的数据总量受参数数量的影响,达到收敛的大模型的参数数量的对数与其训练数据总量的对数成正比。

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求方业务活动的服务。

注:常见大模型服务内容包括大模型平台服务、大模型开发定制服务、大模型推理及运营服务。

3.3

任务 **task**

被调度的训练或推理对象。

注:任务用于完成一个相对独立的业务功能。一个任务属于且仅属于一个作业。

[来源:GB/T 25000.23—2019,4.12,有修改]

3.4

微调 **fine-tuning**

为提升机器学习模型预测准确性,使用专门领域数据在大模型上继续训练的过程。

注 1:专门领域数据一般是特定场景的生产数据或合成数据。

注 2:常用的微调方法包括提示词微调、全参微调、参数高效微调等。