



中华人民共和国国家标准

GB/T 42018—2022

信息技术 人工智能 平台计算资源规范

Information technology—Artificial intelligence—
Platform computing resource specification

2022-10-12 发布

2023-05-01 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	3
5 概述	4
6 技术要求	5
6.1 物理计算资源	5
6.2 虚拟计算资源	8
7 测试方法	9
7.1 通用特性	9
7.2 物理计算资源	9
7.3 虚拟计算资源	12
附录 A (资料性) 人工智能平台物理计算资源测试方法与技术要求的对应关系	13
附录 B (资料性) 物理计算资源测试系统的配置	17
附录 C (规范性) 训练和推理测试模型	20
参考文献	21

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、上海依图网络科技有限公司、杭州海康威视数字技术股份有限公司、西北工业大学、浪潮软件科技有限公司、国防科技大学、上海商汤阡誓科技有限公司、北京智芯微电子科技有限公司、北京电信规划设计院有限公司、三菱电机(中国)有限公司、青岛海尔洗衣机有限公司、英特尔(中国)有限公司、飞腾信息技术有限公司、中车沈阳机车车辆有限公司、杭州中奥科技有限公司、平安科技(深圳)有限公司、北京清能互联科技有限公司。

本文件主要起草人：董建、张琦、鲍薇、曹晓琦、李斌斌、赵春昊、周智强、符海芳、杨刚、姚远、王功明、徐洋、杨绍武、史殿习、吴庚、刘勇、毕盛楠、范科峰、马珊珊、郝守勤、马骋昊、丁晓鹏、孙宁、王海宁、夏磊、谷潇聪、殷世军、宋文林、陆韵、贾一君、李冰、杨雨泽、赵江、汪洋、江易。

引 言

人工智能平台是为人工智能应用提供各类资源的软硬件系统,是各类人工智能应用(如计算机视觉、自然语言处理、声音处理等)实现的基础。组成人工智能平台时,需使用人工智能物理计算资源和虚拟计算资源。物理计算资源是指计算设备实体,包含人工智能服务器、人工智能加速卡、人工智能加速模组等。虚拟计算资源则是基于实体计算资源,经过抽象并在一定程度上屏蔽异构性后,形成的逻辑计算资源。

本文件旨在为人工智能平台的建设提供标准依据,对可供组成人工智能平台常见的物理计算资源和虚拟计算资源提出相关技术参数的基础共性要求,对物理计算资源提出测试方法。

信息技术 人工智能 平台计算资源规范

1 范围

本文件规定了面向机器学习的人工智能平台物理计算资源(包含人工智能服务器、人工智能加速卡、人工智能加速模组)和虚拟计算资源的技术要求,描述了物理计算资源的测试方法。本文件不对资源管理和调度提出要求。

本文件适用于面向机器学习的人工智能平台的设计和测试。

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 9813.3—2017 计算机通用规范 第3部分:服务器

GB/T 17235.1—1998 信息技术 连续色调静态图像的数字压缩及编码 第1部分:要求和指南

GB/T 20090.2—2013 信息技术 先进音视频编码 第2部分:视频

GB/T 41867—2022 信息技术 人工智能 术语

ISO/IEC 15948:2004 信息技术 计算机图形和图像处理 便携式网络图形:功能规范[Information technology—Computer graphics and image processing—Portable Network Graphics (PNG): Functional specification]

ITU-T H.264—2021 通用视听服务高级视频编码(Advanced video coding for generic audiovisual services)

ITU-T H.265—2021 高效视频编码(High efficiency video coding)

3 术语和定义

GB/T 41867—2022 界定的以及下列术语和定义适用于本文件。

3.1

人工智能平台 artificial intelligence platform

为人工智能应用提供各类资源的软硬件系统。

3.2

人工智能平台计算资源 artificial intelligence platform computing resource

在人工智能平台中,用于处理人工智能计算任务的硬件和软件。

注:人工智能平台计算资源包含物理计算资源和虚拟计算资源。

[来源:ISO/IEC 30145.3—2020,3.1.6,有修改]