



中华人民共和国国家标准化指导性技术文件

GB/Z 251—2026

人工智能 智算集群异构人工智能加速器 混合训练技术要求

Artificial intelligence—Technical requirements of heterogeneous AI accelerating
processor collaborative training in intelligent computing cluster

2026-08-27 发布

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 总体架构 2

6 统一异构算网协同管理调度平台功能要求 3

7 统一异构人工智能加速器通信库功能与性能要求 4

8 统一异构人工智能加速器混合训练框架和拆分策略功能要求 4

9 异构人工智能加速器端到端混合训练大模型功能与性能要求 6

附录 A（资料性） 模型类型和参数配置 8

附录 B（规范性） 混合训练效率计算公式 9

前 言

本文件为规范类指导性技术文件。

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：上海壁仞科技股份有限公司、中国电子技术标准化研究院、上海智能算力科技有限公司、中国电信股份有限公司研究院、中兴通讯股份有限公司、北京电子数智科技有限责任公司、中国电力科学研究院有限公司、科华数据股份有限公司、浦江国家实验室、北京智源人工智能研究院、中国移动通信有限公司研究院、中移九天人工智能科技(北京)有限公司、北京大学、北京大学长沙计算与数字经济研究院、中国科学院计算技术研究所、中科寒武纪科技股份有限公司、海光信息技术股份有限公司、上海天数智芯半导体股份有限公司、上海燧原科技股份有限公司、平头哥(上海)半导体技术有限公司、河南昆仑技术有限公司、浪潮电子信息产业股份有限公司、京东科技信息技术有限公司、蚂蚁科技集团股份有限公司、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、南京南瑞瑞腾科技有限责任公司、中国电子技术标准化研究院华东分院、北京智芯微电子科技有限公司、中移物联网有限公司、浙江大华技术股份有限公司、天翼云科技有限公司、浪潮云信息技术股份公司、山东省计算中心(国家超级计算济南中心)。

本文件主要起草人：徐洋、丁瑞全、张琦、李志、邢冯、杨雨泽、孙兆群、秦甘尧、黄琨翠、孙梦宇、黄志兰、黄程、王昱苏、郭文、高昆仑、王岳、全康、王辉、钟普、敖玉龙、刘颖、黄蕾、秦日臻、杨超、李喆涵、乔欢、王煜炜、梁帆、黄岩哲、段爱民、杨朋霖、瞿昊旻、姚涛、王思善、张艺伯、李卓群、刘劲楠、邸贺亮、韩冬、郭智慧、彭剑峰、仲凯韬、石超、王绍飞、严小格、董月文、刘勇、郑哲、孔栋、丁源、孔维生、丁立戈、郑佳佳、逢锦山。

人工智能 智算集群异构人工智能加速器 混合训练技术要求

1 范围

本文件确立了智算集群异构人工智能加速器混合训练的总体架构,规定了混合训练的功能和性能要求。

本文件适用于由不同厂商人工智能加速器组建的异构智算集群。同一厂商不同型号的人工智能加速器组建的异构智算集群参考使用。

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 41867—2022 信息技术 人工智能 术语

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

人工智能加速器 artificial intelligence accelerating processor

人工智能加速芯片 artificial intelligence accelerating chip

具备适配人工智能算法的运算微架构,能够完成人工智能应用运算处理的集成电路。

注:在本文件中,在不引起误解的语境下,将人工智能加速器简称为加速器。

3.2

异构人工智能加速器 heterogeneous artificial intelligence accelerating processor

异构 AI 加速器 heterogeneous AI accelerating processor

不同厂商的人工智能加速器。

3.3

异构人工智能加速器混合训练 heterogeneous AI accelerating processor collaborative training

在异构智能计算集群中,使用异构人工智能加速器协同完成同一个模型的训练任务。

4 缩略语

下列缩略语适用于本文件。

AI:人工智能(Artificial Intelligence)

DP:数据并行(Data Parallel)

EP:专家并行(Expert Parallel)

GDR:GPU 直接访问远程内存(GPU Direct RDMA)