



中华人民共和国国家标准

GB/T 45079—2024

人工智能 深度学习框架多硬件平台 适配技术规范

Artificial intelligence—Technical specification for deep learning framework
adaption to multi-hardware platform

2024-11-28 发布

2024-11-28 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
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 环境要求	2
5.1 概述	2
5.2 训练框架与硬件平台适配环境要求	2
5.3 推理框架与硬件平台适配环境要求	3
6 适配接口要求	3
6.1 概述	3
6.2 训练场景适配接口要求	4
6.3 推理场景适配接口要求	8
7 功能要求	10
7.1 训练场景适配功能要求	10
7.2 推理场景适配功能要求	10
8 测试方法	11
8.1 环境测试方法	11
8.2 接口测试方法	11
8.3 功能测试方法	12
附录 A (资料性) 训练基础模型及评价指标	13
附录 B (资料性) 推理模型及评价指标	14
参考文献	15

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、北京百度网讯科技有限公司、浪潮电子信息产业股份有限公司、深圳云天励飞技术股份有限公司、上海壁仞科技股份有限公司、中国科学院软件研究所、上海燧原科技股份有限公司、北京智芯微电子科技有限公司、浙江大华技术股份有限公司、上海商汤智能科技有限公司、南京南瑞瑞腾科技有限责任公司、平头哥(上海)半导体技术有限公司、上海天数智芯半导体有限公司、上海市人工智能行业协会、龙芯中科(合肥)技术有限公司、上海计算机软件技术开发中心、青岛海信电子技术服务有限公司、杭州海康威视数字技术股份有限公司、中国铁建股份有限公司、中铁第五勘察设计院集团有限公司、广电运通集团股份有限公司、北京航天自动控制研究所、中国移动通信集团有限公司、南方电网人工智能科技有限公司、西南科技大学、美的集团(上海)有限公司、罗克佳华科技集团股份有限公司、北京大学、天津(滨海)人工智能创新中心、中国南方电网有限责任公司、上海文镭信息科技有限公司、北京声智科技有限公司、北京大学长沙计算与数字经济研究院、北京电子数智科技有限责任公司。

本文件主要起草人：徐洋、马艳军、马骋昊、吴韶华、董建、高铁柱、王志芳、丁瑞全、胡晓光、杨雨泽、董乾、王思善、刘勇、孔维生、张行程、石超、高慧、余雪松、赵春昊、鲍薇、马珊珊、李斌斌、张强、陈文捷、刘微、彭剑峰、李栋、郑中、郭振华、黄宇恒、王丽娜、秦日臻、梁寿愚、孟令中、俞文心、方贵明、蔡亚森、李玮、何源宏、杨超、田涛、林志达、林克全、芮子文、陈孝良、吴岳。

人工智能 深度学习框架多硬件平台 适配技术规范

1 范围

本文件规定了在训练和推理场景下,深度学习框架适配多硬件平台的技术要求,描述了相应的测试方法。

本文件适用于支持训练和推理功能的深度学习框架与多硬件平台完成适配,以及深度学习框架与硬件的适配效果评价,也适用于指导人工智能软硬件适配过程。

注:本文件不涉及硬件平台的技术要求。

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

深度学习框架 deep learning framework

实现对人工智能算法开发、封装、数据调用以及计算资源使用的软件库。

3.2

多硬件平台 multi-hardware platform

包含多种人工智能加速处理器的可提供人工智能计算能力的硬件系统。

3.3

适配多硬件平台 multi-hardware platform adaption

深度学习框架可将多硬件平台作为计算资源完成深度学习模型训练与推理任务的活动。

3.4

计算图 computational graph

用来表示数学函数,由节点和连接构成的有向图。

注1:节点表示数学运算,即算子。

注2:连接表示数学运算之间的依赖关系。

注3:一个连接联通起始节点和终止节点。

[来源:ISO/IEC/IEEE 24765:2017,3.1762.1,有修改]

3.5

整图 graph

用于描述某个深度学习特定任务的计算过程,由一系列算子和张量组成的完整的计算图。