

ICS 35.020
CCS L60



**NATIONAL STANDARD OF THE
PEOPLE'S REPUBLIC OF CHINA**

GB/T 45079-2024

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

人工智能 深度学习框架多硬件平台适配技术规范

Issued on: November 28, 2024

Implemented on: November 28, 2024

**Issued by: State Administration for Market Regulation;
Standardization Administration of the PRC**

Table of Contents

1 Scope	1
2 Normative references	1
3 Terms and Definitions	1
4 Abbreviations	2
5 Environmental Requirements	2
5.1 Overview	2
5.2 Training framework and hardware platform adaptation environment requirements.....	2
5.3 Reasoning framework and hardware platform adaptation environment requirements ...	3
6 Adapter interface requirements	3
6.1 Overview	3
6.2 Training scenario adaptation interface requirements	4
6.3 Reasoning scenario adaptation interface requirements	8
7 Functional Requirements	10
7.1 Training scenario adaptation function requirements	10
7.2 Reasoning scenario adaptation function requirements	10
8 Test Methods	11
8.1 Environmental test methods	11
8.2 Interface testing method	11
14 Reference	15

Foreword

This document is in accordance with the provisions of GB/T 1.1-2020 "Guidelines for standardization work Part

1. Structure and drafting rules for standardization documents" Drafting. Please note that some of the contents of this document may involve patents. The issuing organization of this document does not assume the responsibility for identifying patents. This document was proposed and coordinated by the National Information Technology Standardization Technical Committee (SAC/TC28). This document was drafted by: China Electronics Technology Standardization Institute, Beijing Baidu Netcom Technology Co., Ltd., Inspur Electronic Information Industry Co., Ltd. Co., Ltd., Shenzhen Yuntian Lifei Technology Co., Ltd., Shanghai Biren Technology Co., Ltd., Institute of Software, Chinese Academy of Sciences, Shanghai Enflame Technology Co., Ltd., Beijing Zhixin Microelectronics

Technology Co., Ltd., Zhejiang Dahua Technology Co., Ltd., Shanghai SenseTime Energy Technology Co., Ltd., Nanjing NARI Ruiteng Technology Co., Ltd., Pingtoug (Shanghai) Semiconductor Technology Co., Ltd., Shanghai Tianshu Zhixin Semiconductor Co., Ltd., Shanghai Artificial Intelligence Industry Association, Loongson Technology (Hefei) Co., Ltd., Shanghai Computer Software Technology Development Center, Qingdao Hisense Electronic Technology Service Co., Ltd., Hangzhou Hikvision Digital Technology Co., Ltd., China Railway Construction Corporation Limited, China Railway Fifth Survey and Design Institute Group Co., Ltd., China Broadcasting and Television Network Group Co., Ltd., Beijing Aerospace Automatic Control Research Institute, China Mobile Communications Group Co., Ltd., China Southern Power Grid Artificial Intelligence Technology Co., Ltd., Southwest University of Science and Technology, Midea Group (Shanghai) Co., Ltd., Luo Kejia Hua Technology Group Co., Ltd., Peking University, Tianjin (Binhai) Artificial Intelligence Innovation Center, China Southern Power Grid Co., Ltd., Shanghai Wenyu Information Technology Co., Ltd., Beijing Shengzhi Technology Co., Ltd., Peking University Changsha Institute of Computing and Digital Economy, Beijing Electronic Digital Intelligence TECHNOLOGY LIMITED. The main drafters of this document are. Xu Yang, Ma Yanjun, Ma Chenghao, Wu Shaohua, Dong Jian, Gao Tiezhu, Wang Zhifang, Ding Ruiquan, Hu Xiaoguang, Yang Yuze, Dong Qian, Wang Sishan, Liu Yong, Kong Weisheng, Zhang Chengxing, Shi Chao, Gao Hui, Yu Xuesong, Zhao Chunhao, Bao Wei, Ma Shanshan, Li Binbin, Zhang Qiang, Chen Wenjie, Liu Wei, Peng Jianfeng, Li Dong, Zheng Zhong, Guo Zhenhua, Huang Yuheng, Wang Lina, Qin Rizhen, Liang Shouyu, Meng Lingzhong, Yu Wenxin, Fang Guiming, Cai Yasen, Li Wei, He Yuanhong, Yang Chao, Tian Tao, Lin Zhida, Lin Kequan, Rui Ziwen, Chen Xiaoliang, Wu Yue. Artificial intelligence deep learning framework multiple hardware platforms Adaptation technical specifications

1 Scope

GB/T 45079-2024 specifies how a deep learning framework is adapted to run on more than one kind of hardware. The question is not academic in China: export controls have pushed training and inference onto domestic accelerators alongside the incumbent GPUs, and every framework that has to reach them needs a defined boundary between the framework and the vendor's backend, or each new chip means another fork. The standard sets the environment requirements for adapting a training framework and an inference framework to a hardware platform, the adapter interface requirements for the training and inference scenarios, the functional requirements the adaptation must satisfy in each scenario, and the test methods by which conformance and equivalence of results are verified. It took effect on 28 November 2024.

This document specifies the technical requirements for deep learning frameworks to adapt to multiple hardware platforms in training and inference scenarios, and describes the corresponding tests. method. This document is applicable to deep learning frameworks that support training and reasoning functions and multiple hardware platforms to complete

adaptation, as well as deep learning frameworks and The evaluation of hardware adaptation effects is also applicable to guiding the artificial intelligence software and hardware adaptation process.

Note. This document does not cover technical requirements for hardware platforms.

2 Normative references

The contents of the following documents constitute essential clauses of this document through normative references in this document. For referenced documents without a date, only the version corresponding to that date applies to this document; for referenced documents without a date, the latest version (including all amendments) applies to This document.

GB/T 41867 Information Technology Artificial Intelligence Terminology

3 Terms and definitions

The terms and definitions defined in GB/T 41867 and the following apply to this document.

3.1 A software library that enables artificial intelligence algorithm development, packaging, data calls, and computing resource usage.

3.2 multi-hardwareplatform A hardware system that includes a variety of AI acceleration processors that can provide AI computing capabilities.

3.3 The deep learning framework can use multiple hardware platforms as computing resources to complete deep learning model training and reasoning tasks.

3.4 computationalgraph A directed graph consisting of nodes and links used to represent mathematical functions.

Note 1: A node represents a mathematical operation, i.e. an operator. Note

2.Connections represent dependencies between mathematical operations.

Note 3: A connection connects the start node and the end node. [Source: ISO /IEC /IEEE24765.2017, 3.1762.1, modified]

3.5 Graph It is used to describe the computational process of a specific deep learning task, and is a complete computational graph consisting of a series of operators and tensors.