ICS 35.020
CCS L 70

NATIONAL STANDARD OF THE PEOPLE'S REPUBLIC OF CHINA

GB/T 45401.2-2025

Artificial intelligence - Scheduling and cooperation for computing devices - Part 2: Framework for distributed computing

人工智能 计算设备调度与协同 第2部分：分布式计算框架

Issued on: March 28, 2025 **Implemented on: March 28, 2025**

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

Table of Contents

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.

This document is part 2 of GB/T 45401 "Scheduling and coordination of artificial

intelligence computing equipment". GB/T 45401 has been published The following parts.

- Part 1: Virtualization and scheduling;

- Part 2: Distributed computing framework.

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 article was proposed and coordinated by the National Information Technology Standardization Technical Committee (SAC/TC28).

This document was drafted by: China Electronics Standardization Institute, Huawei Technologies Co., Ltd., Beijing University of Aeronautics and Astronautics, Chinese Academy of Sciences Software Research Institute, Huawei Cloud Computing Technology Co., Ltd., Alibaba Cloud Computing Co., Ltd., Beijing Baidu Netcom Technology Co., Ltd., Inspur Electronics Information Industry Co., Ltd., Shanghai SenseTime Intelligent Technology Co., Ltd., Peking University Wuhan Artificial Intelligence Research Institute, Shanghai Artificial Intelligence Industry Co., Ltd.

Industry Association, China Mobile Communications Group Co., Ltd., Institute of Computing Technology, Chinese Academy of Sciences, iFLYTEK Co., Ltd., Peking University, Shenzhen Yuntian Lifei Technology Co., Ltd., Shanghai Tianshu Zhixin Semiconductor Co., Ltd., Beijing Biren Technology Development Co., Ltd., Hangzhou Hai Kangwei Digital Technology Co., Ltd., China Southern Power Grid Artificial Intelligence Technology Co., Ltd., Loongson Technology Co., Ltd., Suzhou Deng Lin Technology Co., Ltd., Zhejiang Dahua Technology Co., Ltd., Ant Technology Group Co., Ltd., Guoke Chushi (Chongqing) Software Co., Ltd.

Company, Guangdong Power Grid Co., Ltd., Guangdong Broadcasting Corporation Limited, Shanghai Computer Software Technology Development Center, Shanghai Wenyuxin Information Technology Co., Ltd., BOE Technology Group Co., Ltd., and Tianjin (Binhai) Artificial Intelligence Innovation Center.

The main drafters of this document are: Cao Xiaoqi, Dong Jian, Yang Yuze, Bao Wei, Xu Yang, Yu Chao, Li Binbin, Wang Waner, Luan Zhongzhi, Zhu Yixin, Dong Qian, Meng Lingzhong, Zheng Zimu, Wu Tao, Tian Xiaoli, Zhang Yaqiang, Ma Shanshan, Ma Chenghao, Zhao Chunhao, Wu Geng, Cao Xi, Wang Yuwei, Wu Ting, Yang Chao, Wang Zhifang, Yu Xuesong, Ding Ruiquan, Ye Tingqun, Dong Zhaojie, Ma Wanyue, Dai Jun, Kong Weisheng, Guo Zhihui, Luo Yongjun, Yan Yuping, Chen Haomin, Yang Bo, Chen Minggang, Niu Keke, Zhong Kaitao, Jiang Xingqun, and Shi Dianxi.

Introduction

With the continuous development of AI computing, the deployment and use of computing devices that carry AI applications have become distributed and full-scale.

The same AI computing task often requires the collaboration of multiple computing devices to provide services for users in different regions and types.

It is necessary to rationally utilize and allocate computing equipment resources of different forms, and clarify the necessary technical architecture, capability requirements and interfaces.

Etc., provide a reference framework and evaluation system for products, and alleviate the current situation of horizontal collaboration and fragmentation of different forms of artificial intelligence computing devices.

GB/T 45401 "Scheduling and coordination of artificial intelligence computing devices" is planned to consist of two parts.

- Part 1: Virtualization and Scheduling, aims to establish the architecture of the virtualization and scheduling system for artificial intelligence computing devices and stipulate the technical requirements Request and corresponding test methods.

- Part 2: Distributed computing framework, which aims to establish the architecture of distributed computing for artificial intelligence computing devices and specify functions and performance Technical requirements,define the distributed computing collaboration interface.

AI computing equipment scheduling and coordination Part 2: Distributed Computing Framework

1 Scope

This document establishes the architecture of distributed computing for artificial intelligence computing devices, specifies functional and performance technical requirements, and defines distributed computing It is a collaborative interface.

This document is applicable to the design, development, and testing of distributed artificial intelligence computing systems.

2 Normative references

GB/T 41867

GB/T 42018-2022

GB/T 45401.1-2025

GB/T 45280-2025

GB/T 45087-2024

3 Terms and definitions

GB/T 41867, GB/T 42018-2022, GB/T 45401.1-2025 and GB/T 45280-2025 and the following terms The following terms and definitions apply to this document.

3.1 [Source. ISO /IEC 2382 2015, 2178059, modified]

A machine learning method or process that uses the Internet as the preferred communication carrier to complete the same machine learning task on different subsystems. Learning tasks.

Note 1.The Internet includes the local Internet and the wide area Internet.

Note 2.Distributed machine learning is divided into distributed training and distributed reasoning according to the different types of machine learning tasks.

3.2 federated [machine] learning

A machine learning method or process that enables multiple participants to collaboratively build and use machine learning models without exposing the participants' private data.

3.3 Incremental learning

A multi-stage adaptive learning approach in which knowledge acquired in a precursor phase is transformed into a suitable form for subsequent