

ICS 35.240
CCS L 70



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

GB/T 46345-2025

Artificial intelligence - Interface of deep learning compiler

人工智能 深度学习编译器接口

Issued on: October 5, 2025

Implemented on: October 5, 2025

**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.Overview	2
5.1 Technical Architecture	2
5.2 Technical Process	3
5.3 Interface Testing	4
6.Diagram generation module interface	4
6.1 Overview	4
6.2 Loading the computational graph	5
6.3 Calculation Graph Editing	6
7.1 Overview	6
7.2 Graph Optimization	6
7.3 Diagram Splitting	6
7.4 Descending Figure	7
8.Diagram scheduling module interface	8
8.3 Subgraph Scheduling	9
18 References	24

Foreword

This document complies with the provisions of GB/T 1.1-2020 "Standardization Work Guidelines Part

1.Structure and Drafting Rules of Standardization Documents". Drafting. Please note that some content in this document may involve patents. The issuing organization of this document assumes no responsibility for identifying patents. This document was proposed and is under the jurisdiction of the National Information Technology Standardization Technical Committee (SAC/TC28). This document was drafted by: China Electronics Technology Standardization Institute, Shanghai Artificial Intelligence Innovation Center, Huawei Technologies Co., Ltd., and Beijing Baiyunshan Pharmaceutical Holdings Co., Ltd. Baidu SenseTime Technology Co., Ltd., Shanghai SenseTime Technology Co., Ltd., Peking University, Beijing Academy of Artificial Intelligence, Shanghai Artificial Intelligence Industry

Association Industry Association, Shanghai Suiyuan Technology Co., Ltd., Shanghai Biren Technology Co., Ltd., Qingdao Port International Co., Ltd., Shenzhen CESI Information Technology Co., Ltd. Technology Co., Ltd., Inspur Electronic Information Industry Co., Ltd., China Mobile Communications Group Co., Ltd., Zhejiang Dahua Technology Co., Ltd. The company, Shanghai Tianshu Zhixin Semiconductor Co., Ltd., Hangzhou Hikvision Digital Technology Co., Ltd., Zhejiang Lab, and Pingtoug (Shanghai) Semiconductor Technology Co., Ltd., Kunlun Core (Beijing) Technology Co., Ltd., Cambricon Technologies Corporation Limited, Shenzhen Cloudwalk Technology Co., Ltd. Joint-stock limited companies, ZTE Corporation, Shanghai Instrument & Electronics (Group) Co., Ltd., Shanghai Intelligent Computing Technology Co., Ltd., Beijing Daxing International Technology Co., Ltd. Learn from Changsha Institute of Computing and Digital Economy, Shanghai Wenyao Information Technology Co., Ltd., and Information System Integration of NARI Technology Co., Ltd. Branch offices, Shanghai Artificial Intelligence Research Institute Co., Ltd., iFlytek Co., Ltd., China Electric Power Research Institute Co., Ltd., China Mobile Group Integrated Circuit Co., Ltd., China Southern Power Grid Co., Ltd. Ultra-High Voltage Transmission Company, Shanghai Mobile Communication Technology Co., Ltd., China Mobile Xiong'an Information & Communication Technology Co., Ltd., Nanjing Nanrui Ruiteng Technology Co., Ltd., China Coal Information Technology (Beijing) Co., Ltd., and Yixingzhi Energy Technology (Guangzhou) Co., Ltd., China Telecom Corporation Limited Chongqing Branch, and Beijing Haohan Deep Information Technology Co., Ltd. The main drafters of this document are. Dong Jian, Zhang Chengxing, Yang Yuze, Xu Yang, Pei Zhilin, Yang Heng, Hu Xiaoguang, Yang Chao, Zhong Pu, Li Xiaoru, and Ma Shanshan. Lu Shun, Men Chunlei, Ding Ruiquan, Wu Geng, Zhao Chunhao, Wang Sishan, Wu Yuzhen, Guo Yiyun, Yu Chao, Zhang Yunfei, Guo Zhenhua, Wang Bin, Mei Jingqing Gao Wanqi, Nie Jiandi, Shen Zhiyue, Xing Feng, Tang Yinan, Sun Yue, Niu Hongxing, Bai Tongxin, Ouyang Jian, Gao Hui, Kong Weisheng, Hu Mingshan, Wang Peng Huang Dandan, Dong Shouyang, Li Aijun, Huang Cheng, Zhang Yibo, Sun Xiaosi, Gou Haipeng, Zhong Kaitao, Wang Zhao, Chen Xi, Qiao Yuping, Yan Minhui, Zhang Tian, Wen Xing Liang Hengkang, Yang Jiali, Wang Lichen, Yang Yongyong, Zhang Bo, Ma Jianhua, Pan Wu, Han Fujun, Zhang Fang, Meng Guiyun, Huang Yanzhe, Yuan Jie, Zhu Jing, Chen Jun Rui Ziwen, Yang Yunfei, Yang Tonghui, Zhang Xiaojuan, Wang Jing, Wang Ning, Tian Kang, Shi Chao, Liu Yun, Wang Chenzi, Lai Su, Zhao Jiajie. Artificial Intelligence Deep Learning Compiler Interface

1 Scope

GB/T 46345-2025 is the Chinese national standard covering the interface between a model and the compiler that turns it into code for a particular accelerator - the intermediate representation, the operator set and its semantics, the optimisation passes exposed, and the target description, so that a model is not tied to one vendor's toolchain. First edition, already in force. Issued on 5 October 2025, it has been in force since 5 October 2025.

This document specifies the graph generation module interface, graph transformation

module interface, graph scheduling module interface, and domain-specific syntax of the deep learning compiler. The requirements for the functions and input/output parameters of the language interface and operator generator interface are described, along with the corresponding testing methods. This document applies to the design and implementation of deep learning compilers, and also provides a reference for the integration and application of deep learning compilers.

2 Normative references

The contents of the following documents, through normative references within the text, constitute essential provisions of this document. Dated citations are not included. For references to documents, only the version corresponding to that date applies to this document; for undated references, the latest version (including all amendments) applies. This document.

GB/T 41867-2022 Terminology for Information Technology and Artificial Intelligence

3 Terms and Definitions

The terms and definitions defined in GB/T 41867-2022, as well as the following terms and definitions, apply to this document.

3.1 Deep learning A method for creating rich hierarchical representations by training neural networks with many hidden layers.

Note. Deep learning is a subset of machine learning. [Source: GB/T 41867-2022, 3.2.27]

3.2 Deep learning compiler Tools for optimizing and compiling source code for deep learning models and domain-specific languages.

Note. Deep learning compilers convert deep learning models or domain-specific language source code into executable code for AI acceleration chips, enabling efficient... Training and reasoning.

3.3 Integrated circuit components with a computing microarchitecture adapted to artificial intelligence algorithms, capable of performing computational processing for artificial intelligence applications. [Source: GB/T 41867-2022, 3.1.5]

3.4 A form of encoding used by deep learning compilers in the process of converting source code into executable code.

Note. Intermediate representations are used during the conversion process to show the structure and semantic information of the program, so as to facilitate further optimization and code generation.