Technical Report

ISO/IEC TR 17903

First edition
2024-05

Information technology — Artificial intelligence — Overview of machine learning computing devices

Technologies de l'information — Intelligence artificielle — Aperçu des dispositifs informatiques pour l'apprentissage automatique

Reference number
ISO/IEC TR 17903:2024(en)

© ISO/IEC 2024



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Abbreviated terms ... **5**
5 ML computing device concepts ... **5**
5.1 Processing ... 5
5.2 Computing ... 5
5.3 Device ... 6
5.4 Infrastructure ... 6
5.5 Service ... 6
5.6 Performance ... 7
5.7 Computing device ... 7
6 ML computing device characteristics ... **7**
6.1 Datatypes ... 7
6.1.1 General ... 7
6.1.2 Effectiveness and efficiency ... 8
6.2 ML operators ... 8
6.2.1 General ... 8
6.2.2 Effectiveness and efficiency ... 9
6.3 Memory access and addressing mechanisms ... 10
6.3.1 General ... 10
6.3.2 Effectiveness and efficiency ... 10
6.4 Scheduling ... 10
6.4.1 General ... 10
6.4.2 Effectiveness and efficiency ... 11
6.5 Topologies ... 12
6.5.1 General ... 12
6.5.2 Effectiveness and efficiency ... 12
6.6 Streams ... 13
6.6.1 General ... 13
6.6.2 Effectiveness and efficiency ... 13
6.7 Buffering mechanisms ... 13
6.7.1 General ... 13
6.7.2 Effectiveness and efficiency ... 14
6.8 Cache mechanisms ... 14
6.8.1 General ... 14
6.8.2 Effectiveness and efficiency ... 14
6.9 Data exchange mechanisms ... 15
6.9.1 General ... 15
6.9.2 Effectiveness and efficiency ... 15
7 Approaches and measures for performance optimization ... **16**
7.1 Approaches ... 16
7.1.1 Overview ... 16
7.1.2 Computing resource-level ... 16
7.1.3 Enabling software-level ... 16
7.2 Measures ... 17
Annex A (informative) **Relationships between ML computing device-related definitions** ... **19**
Bibliography ... **21**

© ISO/IEC 2024 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

© ISO/IEC 2024 – All rights reserved

Introduction

As an important approach for realizing artificial intelligence (AI), machine learning (ML) has been applied to and has improved productivity in multiple domains (e.g. education, finance, environment protection). ML computing devices can be essential to the development and deployment of many types of AI systems.

An ML computing device can have a set of characteristics, including supported datatypes, ML operators, buffer settings, access and share mechanisms, memory access and addressing mechanisms, virtualization and sharing mechanisms, job scheduling mechanisms, topologies, data exchange mechanisms and memory interoperability mechanisms. Use and setting of these characteristics can affect the overall performance of an ML computing device. The performance of an ML computing device used to develop and deploy an AI system can be crucial to business effectiveness and efficiency.

This document surveys and provides information to AI stakeholders to assist them in understanding the representative characteristics of ML computing devices. In Clause 5, ML computing device-related concepts are discussed. Clause 6 summarizes ML computing device characteristics. In Clause 7, existing approaches for optimizing ML computing devices' performance are discussed.

© ISO/IEC 2024 – All rights reserved

Information technology — Artificial intelligence — Overview of machine learning computing devices

1 Scope

This document surveys machine learning (ML) computing devices, including the following:

— ML computing device terminology and characteristics;

— existing approaches to the setting and use of characteristics for optimizing ML computing device performance.

The information provided in this document is relevant for organizations of all types and sizes.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 22989, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

ISO/IEC 23053, *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989 and ISO/IEC 23053 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
artificial intelligence computing
AI computing
processing that leverages infrastructures to enable the set of methods or automated entities that together build, optimize and apply an AI model

3.2
machine learning computing
ML computing
processing that leverages infrastructures to train and execute ML models

3.3
infrastructure
hardware and software environment to support computer system and software design, development, and modification

Note 1 to entry: Network facilities can be also infrastructure working with hardware and software.

© ISO/IEC 2024 – All rights reserved