SLOVENSKI STANDARD
SIST-TP EN ISO/IEC/TR 24029-1:2024

01-maj-2024

Umetna inteligenca (UI) - Ocenjevanje robustnosti nevronskih omrežij - 1. del: Pregled (ISO/IEC TR 24029-1:2021)

Artificial Intelligence (AI) - Assessment of the robustness of neural networks - Part 1: Overview (ISO/IEC TR 24029-1:2021)

Informationstechnik - Künstliche Intelligenz - Bewertung der Robustheit neuronaler Netze (ISO/IEC TR 24029-1:2021)

Intelligence artificielle (IA) - Évaluation de la robustesse des réseaux de neurones - Partie 1: Vue d'ensemble (ISO/IEC TR 24029-1:2021)

Ta slovenski standard je istoveten z: **CEN/CLC ISO/IEC/TR 24029-1:2023**

ICS:

35.020	Informacijska tehnika in tehnologija na splošno	Information technology (IT) in general

SIST-TP EN ISO/IEC/TR 24029-1:2024 **en,fr,de**

TECHNICAL REPORT

RAPPORT TECHNIQUE

TECHNISCHER REPORT

CEN/CLC ISO/IEC/TR 24029-1

December 2023

ICS 35.020

English version

Artificial Intelligence (AI) - Assessment of the robustness of neural networks - Part 1: Overview (ISO/IEC TR 24029-1:2021)

Intelligence artificielle (IA) - Évaluation de la robustesse des réseaux de neurones - Partie 1: Vue d'ensemble (ISO/IEC TR 24029-1:2021)

Informationstechnik - Künstliche Intelligenz - Bewertung der Robustheit neuronaler Netze (ISO/IEC TR 24029-1:2021)

This Technical Report was approved by CEN on 17 December 2023. It has been drawn up by the Technical Committee CEN/CLC/JTC 21.

CEN and CENELEC members are the national standards bodies and national electrotechnical committees of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Türkiye and United Kingdom.

CEN-CENELEC Management Centre:
Rue de la Science 23, B-1040 Brussels

Contents

Page

European foreword

The text of ISO/IEC TR 24029-1:2021 has been prepared by Technical Committee ISO/IEC JTC 1 "Information technology" of the International Organization for Standardization (ISO) and has been taken over as CEN/CLC ISO/IEC/TR 24029-1:2023 by Technical Committee CEN-CENELEC/ JTC 21 "Artificial Intelligence" the secretariat of which is held by DS.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN-CENELEC shall not be held responsible for identifying any or all such patent rights.

Any feedback and questions on this document should be directed to the users' national standards body. A complete listing of these bodies can be found on the CEN and CENELEC websites.

Endorsement notice

The text of ISO/IEC TR 24029-1:2021 has been approved by CEN-CENELEC as CEN/CLC ISO/IEC/TR 24029-1:2023 without any modification.

TECHNICAL REPORT

ISO/IEC TR 24029-1

First edition
2021-03

Artificial Intelligence (AI) — Assessment of the robustness of neural networks —

Part 1: Overview

Reference number
ISO/IEC TR 24029-1:2021(E)

COPYRIGHT PROTECTED DOCUMENT

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

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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see patents.iec.ch).

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*.

A list of all parts in the ISO/IEC 24029 series can be found on the ISO and IEC websites.

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.

Introduction

When designing an AI system, several properties are often considered desirable, such as robustness, resiliency, reliability, accuracy, safety, security, privacy. A definition of robustness is provided in 3.6. Robustness is a crucial property that poses new challenges in the context of AI systems. For example, in AI systems there are some risks specifically tied to the robustness of AI systems. Understanding these risks is essential for the adoption of AI in many contexts. This document aims at providing an overview of the approaches available to assess these risks, with a particular focus on neural networks, which are heavily used in industry, government and academia.

In many organizations, software validation is an essential part of putting software into production. The objective is to ensure various properties including safety and performance of the software used in all parts of the system. In some domains, the software validation and verification process is also an important part of system certification. For example, in the automotive or aeronautic fields, existing standards, such as ISO 26262 or Reference [2], require some specific actions to justify the design, the implementation and the testing of any piece of embedded software.

The techniques used in AI systems are also subject to validation. However, common techniques used in AI systems pose new challenges that require specific approaches in order to ensure adequate testing and validation.

AI technologies are designed to fulfil various tasks, including interpolation/regression, classification and other tasks.

While many methods exist for validating non-AI systems, they are not always directly applicable to AI systems, and neural networks in particular. Neural network systems represent a specific challenge as they are both hard to explain and sometimes have unexpected behaviour due to their non-linear nature. As a result, alternative approaches are needed.

Methods are categorized into three groups: statistical methods, formal methods and empirical methods. This document provides background on these methods to assess the robustness of neural networks.

It is noted that characterizing the robustness of neural networks is an open area of research, and there are limitations to both testing and validation approaches.