



**SLOVENSKI STANDARD**  
**kSIST-TS FprCEN/CLC ISO/IEC/TS**  
**12791:2024**  
**01-januar-2024**

---

**Informacijska tehnologija - Umetna inteligenca - Obravnava neželene pristranskosti pri nalogah strojnega učenja klasifikacije in regresije (ISO/IEC DTS 12791:2023)**

Information technology - Artificial intelligence - Treatment of unwanted bias in classification and regression machine learning tasks (ISO/IEC DTS 12791:2023)

iTeh Standards  
(<https://standards.iteh.ai>)

Technologies de l'information - Intelligence artificielle - Traitement des biais indésirables dans les tâches d'apprentissage automatique de classification et de régression (ISO/IEC DTS 12791:2023)

[kSIST-TS FprCEN/CLC ISO/IEC/TS 12791:2024](https://standards.iteh.ai/catalog/standards/sist/cd1844f3-0267-4c74-b6-6955122101/sist-ts-fprcen-clc-iso-iec-ts-12791-2024)

**Ta slovenski standard je istoveten z: FprCEN/CLC ISO/IEC/TS 12791**

---

**ICS:**

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

<b>kSIST-TS FprCEN/CLC ISO/IEC/TS 12791:2024</b>	<b>en,fr,de</b>
--	-----------------



FINAL  
DRAFT

TECHNICAL  
SPECIFICATION

ISO/IEC DTS  
12791

ISO/IEC JTC 1/SC 42

Secretariat: ANSI

Voting begins on:  
2023-11-09

Voting terminates on:  
2024-02-01

## Information technology — Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks

*Technologies de l'information — Intelligence artificielle — Traitement des biais indésirables dans les tâches d'apprentissage automatique de classification et de régression*

iTeh Standards

(<https://standards.iteh.ai>)

Document Preview

[ksist-ts FprCEN/CLC ISO/IEC/TS 12791:2024](https://standards.iteh.ai/catalog/standards/sist/cd1844f3-0267-4c74-9c86-6c9f554a321f/ksist-ts-fprcen-clc-iso-iec-ts-12791-2024)

<https://standards.iteh.ai/catalog/standards/sist/cd1844f3-0267-4c74-9c86-6c9f554a321f/ksist-ts-fprcen-clc-iso-iec-ts-12791-2024>

ISO/CEN PARALLEL PROCESSING

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.



Reference number  
ISO/IEC DTS 12791:2023(E)

© ISO/IEC 2023

**iTeh Standards**  
**(<https://standards.iteh.ai>)**  
**Document Preview**

[KSIST-TS FprCEN/CLC ISO/IEC/TS 12791:2024](https://standards.iteh.ai/catalog/standards/sist/cd1844f3-0267-4c74-9c86-6c9f554a321f/ksist-ts-fprcen-clc-iso-iec-ts-12791-2024)

<https://standards.iteh.ai/catalog/standards/sist/cd1844f3-0267-4c74-9c86-6c9f554a321f/ksist-ts-fprcen-clc-iso-iec-ts-12791-2024>



**COPYRIGHT PROTECTED DOCUMENT**

© ISO/IEC 2023

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](mailto:copyright@iso.org)  
Website: [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

Page

<b>Foreword</b> .....	<b>iv</b>
<b>Introduction</b> .....	<b>v</b>
<b>1 Scope</b> .....	<b>1</b>
<b>2 Normative references</b> .....	<b>1</b>
<b>3 Terms and definitions</b> .....	<b>1</b>
3.1 General.....	1
3.2 Artificial intelligence.....	3
3.3 Bias.....	4
3.4 Testing.....	5
<b>4 Abbreviated terms</b> .....	<b>7</b>
<b>5 Treating bias in the AI system life cycle</b> .....	<b>7</b>
5.1 Inception.....	7
5.1.1 Stakeholder needs and requirements definition.....	7
5.1.2 Procurement.....	7
5.1.3 Stakeholder identification.....	8
5.1.4 Selection and documentation of data sources.....	9
5.1.5 Integration with risk management.....	9
5.1.6 External change.....	10
5.1.7 Acceptance criteria.....	10
5.2 Design and development.....	11
5.2.1 Data representation and labelling.....	11
5.2.2 Labelling sufficiency.....	11
5.2.3 Labelling quality.....	11
5.2.4 Training data.....	11
5.2.5 Methods for preventing identified risks.....	11
5.3 Verification and validation.....	12
5.3.1 General.....	12
5.3.2 Static testing of training data.....	12
5.3.3 Dynamic testing.....	13
5.4 Re-evaluation, continuous validation, operations and monitoring.....	14
5.5 Disposal.....	14
<b>6 Techniques to address unwanted bias</b> .....	<b>14</b>
6.1 Algorithmic and training techniques.....	15
6.1.1 General.....	15
6.1.2 Pre-trained models.....	15
6.2 Data techniques.....	16
<b>7 Handling bias in a distributed AI system life cycle</b> .....	<b>16</b>
<b>Annex A (informative) Life cycle processes map</b> .....	<b>18</b>
<b>Annex B (informative) Potential impacts of unwanted bias on different types of specific user</b> .....	<b>19</b>
<b>Bibliography</b> .....	<b>21</b>

## ISO/IEC DTS 12791:2023(E)

### 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 documents 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](http://www.iso.org/directives) or [www.iec.ch/members\\_experts/refdocs](http://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](http://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](http://www.iso.org/iso/foreword.html). In the IEC, see [www.iec.ch/understanding-standards](http://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](http://www.iso.org/members.html) and [www.iec.ch/national-committees](http://www.iec.ch/national-committees).

## Introduction

This document describes appropriate steps that can be taken to treat unwanted bias during the development or use of AI systems.

This document is based on ISO/IEC TR 24027<sup>[1]</sup> and provides treatment techniques in accordance with the AI system life cycle as defined in ISO/IEC 22989:2022, Clause 6, and ISO/IEC 5338:—<sup>1)</sup><sup>[2]</sup>. The treatment techniques in this document are agnostic of context. This document is based on the types of bias described in ISO/IEC TR 24027<sup>[1]</sup>.

This document describes good practises for treating unwanted bias and can help an organization with the treatment of unwanted bias in machine learning (ML) systems that conduct classification and regression tasks. The techniques in this document are applicable to classification and regression ML tasks. This document does not address applicability of the described methods outside of the defined ML tasks.

This document does not contain organizational management and enabling processes related to an AI management system, which can be found in ISO/IEC 42001:—<sup>2)</sup><sup>[3]</sup>.

[Annex A](#) provides a cross-reference between the life cycle stages and the clauses of this document.

# iTeh Standards (<https://standards.iteh.ai>) Document Preview

[ksist-ts FprCEN/CLC ISO/IEC/TS 12791:2024](https://standards.iteh.ai/catalog/standards/sist/cd1844f3-0267-4c74-9c86-6c9f554a321f/ksist-ts-fprcen-clc-iso-iec-ts-12791-2024)

<https://standards.iteh.ai/catalog/standards/sist/cd1844f3-0267-4c74-9c86-6c9f554a321f/ksist-ts-fprcen-clc-iso-iec-ts-12791-2024>

- 
- 1) Under preparation. Stage at the time of publication: ISO/IEC FDIS 5338:2023.
  - 2) Under preparation. Stage at the time of publication: ISO/IEC FDIS 42001:2023.





# Information technology — Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks

## 1 Scope

This document describes how to address unwanted bias in AI systems that use machine learning to conduct classification and regression tasks. This document provides mitigation techniques that can be applied throughout the AI system life cycle in order to treat unwanted bias. This document is applicable to all types and sizes of organization.

## 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 5259-2:—<sup>3)</sup>, *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 2: Data quality measures*

ISO/IEC 5259-4:—<sup>4)</sup>, *Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework*

ISO/IEC 22989:2022, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

ISO/IEC/IEEE 29119-3:2021, *Software and systems engineering — Software testing — Part 3: Test documentation*

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989:2022 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 General

#### 3.1.1

##### **consumer vulnerability**

state in which an individual can be placed at risk of harm during their interaction with a service provider due to the presence of personal, situational and market environment factors

[SOURCE: ISO 22458:2022, 3.5]

3) Under preparation. Stage at the time of publication: ISO/IEC DIS 5259-2:2023.

4) Under preparation. Stage at the time of publication: ISO/IEC DIS 5259-4:2023.

## ISO/IEC DTS 12791:2023(E)

### 3.1.2

#### **disposition**

range of records processes associated with implementing records retention, destruction or transfer decisions which are documented in *disposition authorities* (3.1.3) or other instruments

[SOURCE: ISO 30300:2020, 3.4.8]

### 3.1.3

#### **disposition authority**

instrument that defines the *disposition* (3.1.2) actions that are authorized or required for specified records

[SOURCE: ISO 30300:2020, 3.5.4]

### 3.1.4

#### **management system**

set of interrelated or interacting elements of an *organization* (3.1.5) to establish policies and objectives, as well as processes to achieve those objectives

Note 1 to entry: A management system can address a single discipline or several disciplines.

Note 2 to entry: The management system elements include the organization's structure, roles and responsibilities, planning and operation.

[SOURCE: ISO/IEC 42001:—, 3.4]

### 3.1.5

#### **organization**

person or group of people that has its own functions with responsibilities, authorities and relationships to achieve its objectives

Note 1 to entry: The concept of organization includes, but is not limited to, sole-trader, company, corporation, firm, enterprise, authority, partnership, charity or institution or part or combination thereof, whether incorporated or not, public or private.

Note 2 to entry: If the organization is part of a larger entity, the term "organization" refers only to the part of the larger entity that is within the scope of the AI *management system* (3.1.4).

[SOURCE: ISO/IEC 42001:—, 3.1]

### 3.1.6

#### **sample**

<statistics> subset of a population made up of one or more sampling units

Note 1 to entry: The sampling units can be items, numerical values or even abstract entities depending on the population of interest.

Note 2 to entry: A sample from a normal, a gamma, an exponential, a Weibull, a lognormal or a type I extreme value population is often referred to as a normal, a gamma, an exponential, a Weibull, a lognormal or a type I extreme value sample, respectively.

Note 3 to entry: A different definition of sample is used in ISO/IEC 22989:2022.

[SOURCE: ISO 16269-4:2010, 2.1, modified — <statistics> domain and Note 3 to entry have been added.]

### 3.1.7

#### **operational conditions**

conditions under which an AI system is intended to function

**3.1.8****user**

individual or group that interacts with a system or benefits from a system during its utilization

[SOURCE: ISO/IEC/IEEE 15288:2023, 3.53, modified — Note 1 to entry has been removed.]

**3.1.9****at-risk group**

subset of stakeholders that can be adversely affected by unwanted bias

**3.2 Artificial intelligence****3.2.1****machine learning model****ML model****classification model**

<machine learning> model whose expected output for a given input is one or more classes

[SOURCE: ISO/IEC 23053:2022, 3.1.1, modified — preferred terms have been added, and “machine learning” has been removed from the definition.]

**3.2.2****data quality**

characteristic of data that the data meet the *organization's* (3.1.5) data requirements for a specified context

[SOURCE: ISO/IEC 5259-1:—<sup>5</sup>), 3.4]

**3.2.3****data quality characteristic**

category of data quality attributes that bears on *data quality* (3.2.2)

[SOURCE: ISO/IEC 5259-1:—, 3.5]

**3.2.4****data quality measure**

variable to which a value is assigned as the result of measurement of a *data quality characteristic* (3.2.3)

[SOURCE: ISO/IEC 5259-1:—, 3.7]

**3.2.5****data provenance**

provenance

information on the place and time of origin, derivation or generation of a data set, proof of authenticity of the data set, or a record of past and present ownership of the data set

[SOURCE: ISO/IEC 5259-1:—, 3.16]

**3.2.6****feature**

<machine learning> measurable property of an object or event with respect to a set of characteristics

Note 1 to entry: Features play a role in training and prediction.

Note 2 to entry: Features provide a machine-readable way to describe the relevant objects. As the algorithm will not go back to the objects or events themselves, feature representations are designed to contain all useful information.

[SOURCE: ISO/IEC 23053:2022, 3.3.3]

5) Under preparation. Stage at the time of publication ISO/IEC DIS 5259-1:2023.