ISO/IEC ~~DIS~~FDIS 12792~~:2024(en)~~

ISO/IEC JTC 1/SC 42

Secretariat: ANSI

Date: ~~2024-07-02~~2025-05-19

Information technology — Artificial intelligence — Transparency taxonomy of AI systems

Technologies de l'information — Intelligence artificielle — Taxonomie pour la transparence des systèmes d'IA

FDIS stage

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

Published in Switzerland

Contents

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*~~.~~, in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/CLC/JTC 21, *Artificial Intelligence*, in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

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

The objectives of this document include:

— ~~—~~improving trustworthiness, accountability and communication among different AI stakeholders, including partners in a supply chain, customers, users, society and regulators, by establishing a consistent terminology around transparency of AI systems;

— ~~—~~providing AI stakeholders with information about different elements of transparency with their relevance and possible limitations to different use cases and target audience;

— ~~—~~serving as a basis for developing technology-specific, industry-specific or region-specific standards for transparency of AI systems.

Transparency for AI systems is the property of a system that stakeholders receive relevant information about the system. This can include information on items such as system features, limitations, data, system design and design choices (see ISO/IEC 22989:2022, 5.15.8).

Transparency enables relevant stakeholders to have access to information, so they can better understand how an AI system is developed, deployed and can be used. For example, this allows an AI customer (such as an AI user) of an AI system to determine if it is appropriate for their situation and supports an AI auditor in assessing if the system complies with conformity requirements.

A standardized transparency taxonomy of AI systems helps people with different backgrounds to better understand each other by using the same terminology. This in turn supports an improved understanding of AI systems, and provides a foundation for developing interoperable and coherent transparency related standards.

This document is structured as follows:

— 5 ~~Clause 5~~ provides an overview of this document and describes the concept of transparency of AI systems;

— 6 ~~Clause 6~~ discusses how transparency needs can vary depending on the AI system context and on the stakeholders involved;

— 7 ~~Clause 7~~ discusses transparency items that describe the context of the AI system;

— 8 ~~Clause 8~~ pertains to describing transparency disclosures at the AI system level;

— 9 ~~Clause 9~~ focuses on documenting the internal functioning of the system;

— 10 ~~Clause 10~~ offers guidance on the documentation of datasets as stand-alone items.

AI systems often affect society and the environment (see 7 ~~Clause 7).~~). However, it is also true that society and the environment can affect the performance of an AI system. This can include various aspects, such as:

— ~~—~~environmental conditions that affect the functioning and longevity of an AI system;

— ~~—~~introduction or reinforcement of unwanted bias;

— ~~—~~organizational practices that can cause poor outcomes;

— ——effects such as concept drift (concept drift occurs when there is a change in the functional relationship between the model's input and output) and data drift (data drift is change in the statistical properties and characteristics of the input data);

— ——formation of unwanted feedback loops (e.g. negative reinforcement of discriminatory patterns), which can be especially problematic in the case of continuous learning.

While these are important items for consideration, a detailed analysis is not provided in this document. Some of these aspects are further detailed in ISO/IEC 23894, ISO/IEC 27701, and ISO/IEC 29134.