TECHNICAL REPORT

ISO/IEC TR 24368

First edition
2022-08

Information technology — Artificial intelligence — Overview of ethical and societal concerns

Reference number
ISO/IEC TR 24368:2022(E)

© ISO/IEC 2022



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2022

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

© ISO/IEC 2022 – All rights reserved

Contents

Page

Foreword ..... **v**
Introduction ..... **vi**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Overview **3**
4.1 General 3
4.2 Fundamental sources 4
4.3 Ethical frameworks 6
4.3.1 General 6
4.3.2 Virtue ethics 6
4.3.3 Utilitarianism 6
4.3.4 Deontology 6
5 Human rights practices **7**
5.1 General 7
6 Themes and principles **8**
6.1 General 8
6.2 Description of key themes and associated principles 8
6.2.1 Accountability 8
6.2.2 Fairness and non-discrimination 9
6.2.3 Transparency and explainability 9
6.2.4 Professional responsibility 10
6.2.5 Promotion of human values 10
6.2.6 Privacy 11
6.2.7 Safety and security 11
6.2.8 Human control of technology 12
6.2.9 Community involvement and development 12
6.2.10 Human centred design 13
6.2.11 Respect for the rule of law 13
6.2.12 Respect for international norms of behaviour 13
6.2.13 Environmental sustainability 14
6.2.14 Labour practices 14
7 Examples of practices for building and using ethical and socially acceptable AI **15**
7.1 Aligning internal process to AI principles 15
7.1.1 General 15
7.1.2 Defining ethical AI principles the organization can adopt 15
7.1.3 Defining applications the organization cannot pursue 15
7.1.4 Review process for new projects 15
7.1.5 Training in AI ethics 16
7.2 Considerations for ethical review framework 16
7.2.1 Identify an ethical issue 16
7.2.2 Get the facts 16
7.2.3 List and evaluate alternative actions 17
7.2.4 Make a decision and act on it 17
7.2.5 Act and reflect on the outcome 17
8 Considerations for building and using ethical and socially acceptable AI **17**
8.1 General 17
8.2 Non-exhaustive list of ethical and societal considerations 17
8.2.1 General 17
8.2.2 International human rights 18
8.2.3 Accountability 18

© ISO/IEC 2022 – All rights reserved

8.2.4 Fairness and non-discrimination 18
8.2.5 Transparency and explainability 18
8.2.6 Professional responsibility 19
8.2.7 Promotion of human values 20
8.2.8 Privacy 20
8.2.9 Safety and security 20
8.2.10 Human control of technology 21
8.2.11 Community involvement and development 21
8.2.12 Human centred design 21
8.2.13 Respect for the rule of law 21
8.2.14 Respect for international norms of behaviour 22
8.2.15 Environmental sustainability 22
8.2.16 Labour practices 22
Annex A (informative) **AI principles documents** **23**
Annex B (informative) **Use case studies** **33**
Bibliography **42**

© ISO/IEC 2022 – 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).

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 https://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*.

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 2022 – All rights reserved

Introduction

Artificial intelligence (AI) has the potential to revolutionise the world and carry a plethora of benefits for societies, organizations and individuals. However, AI can introduce substantial risks and uncertainties. Professionals, researchers, regulators and individuals need to be aware of the ethical and societal concerns associated with AI systems and applications.

Potential ethical concerns in AI are wide ranging. Examples of ethical and societal concerns in AI include privacy and security breaches to discriminatory outcomes and impact on human autonomy. Sources of ethical and societal concerns include but are not limited to:

— unauthorized means or measures of collection, processing or disclosing personal data;

— the procurement and use of biased, inaccurate or otherwise non-representative training data;

— opaque machine learning (ML) decision-making or insufficient documentation, commonly referred to as lack of explainability;

— lack of traceability;

— insufficient understanding of the social impacts of technology post-deployment.

AI can operate unfairly particularly when trained on biased or inappropriate data or where the model or algorithm is not fit-for-purpose. The values embedded in algorithms, as well as the choice of problems AI systems and applications are used for to address, can be intentionally or inadvertently shaped by developers' and stakeholders' own worldviews and cognitive bias.

Future development of AI can expand existing systems and applications to grow into new fields and increase the level of automation which these systems have. Addressing ethical and societal concerns has not kept pace with the rapid evolution of AI. Consequently, AI designers, developers, deployers and users can benefit from flexible input on ethical frameworks, AI principles, tools and methods for risk mitigation, evaluation of ethical factors, best practices for testing, impact assessment and ethics reviews. This can be addressed through an inclusive, interdisciplinary, diverse and cross-sectoral approach, including all AI stakeholders, aided by International Standards that address issues arising from AI ethical and societal concerns, including work by Joint Technical Committee ISO/IEC JTC 1, SC 42.

© ISO/IEC 2022 – All rights reserved

Information technology — Artificial intelligence — Overview of ethical and societal concerns

1 Scope

This document provides a high-level overview of AI ethical and societal concerns.

In addition, this document:

— provides information in relation to principles, processes and methods in this area;

— is intended for technologists, regulators, interest groups, and society at large;

— is not intended to advocate for any specific set of values (value systems).

This document includes an overview of International Standards that address issues arising from AI ethical and societal concerns.

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*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989 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
agency
ability to define one's goals and act upon them

[SOURCE: ISO/TR 21276:2018, 3.6.2]

3.2
bias
systematic difference in *treatment* (3.13) of certain objects, people, or groups in comparison to others

[SOURCE: ISO/IEC TR 24027:2021, 3.2.2, modified — Removed Note to entry.]

3.3
data management
process of keeping track of all data and/or information related to the creation, production, distribution, storage, disposal and use of e-media, and associated processes

[SOURCE: ISO 20294:2018, 3.5.4, modified — Added "disposal" to definition.]

© ISO/IEC 2022 – All rights reserved

3.4
data protection
legal, administrative, technical or physical measures taken to avoid unauthorized access to and use of data

[SOURCE: ISO 5127:2017, 3.13.5.01, modified — Removed Note to entry.]

3.5
equality
state of being equal, especially in status, rights or opportunities

[SOURCE: ISO 30415:2021, 3.9, modified — Removed "outcome" from definition.]

3.6
equity
practice of eliminating avoidable or remediable differences among groups of people, whether those groups are defined socially, economically, demographically or geographically

3.7
fairness
treatment (3.13), behaviour or outcomes that respect established facts, societal norms and beliefs and are not determined or affected by favouritism or unjust discrimination

Note 1 to entry: Considerations of fairness are highly contextual and vary across cultures, generations, geographies and political opinions.

Note 2 to entry: Fairness is not the same as the lack of *bias* (3.2). Bias does not always result in unfairness and unfairness can be caused by factors other than bias.

3.8
cognitive bias
human cognitive bias
bias (3.2) that occurs when humans are processing and interpreting information

Note 1 to entry: Human cognitive bias influences judgement and decision-making.

[SOURCE: ISO/IEC TR 24027:2021, 3.2.4, modified — Added "cognitive bias" as preferred term.]

3.9
life cycle
evolution of a system, product, service, project or other human-made entity from conception through retirement

[SOURCE: ISO/IEC/IEEE 12207:2017, 3.1.26]

3.10
organization
company, corporation, firm, enterprise, authority or institution, person or persons or part or combination thereof, whether incorporated or not, public or private, that has its own functions and administration

[SOURCE: ISO 30000:2009, 3:10]

3.11
privacy
rights of an entity (normally an individual or an organization), acting on its own behalf, to determine the degree to which the confidentiality of their information is maintained

[SOURCE: ISO/IEC 24775-2:2021, 3.1.46]

© ISO/IEC 2022 – All rights reserved

3.12
responsibility
obligation to act or take decisions to achieve required outcomes

Note 1 to entry: A decision can be taken not to act.

[SOURCE: ISO/IEC 38500:2015, 2.22, modified — Changed “and” to “or” and added Note to entry.]

3.13
treatment
kind of action, such as perception, observation, representation, prediction or decision

[SOURCE: ISO/IEC TR 24027:2021, 3.2.2, modified — Changed Note to entry to term and definition.]

3.14
safety
expectation that a system does not, under defined conditions, lead to a state in which human life, health, property, or the environment is endangered

[SOURCE: ISO/IEC/IEEE 12207:2017, 3.1.48]

3.15
security
aspects related to defining, achieving, and maintaining confidentiality, integrity, availability, accountability, authenticity, and reliability

Note 1 to entry: A product, system, or service is considered to be secure to the extent that its users can rely that it functions (or will function) in the intended way. This is usually considered in the context of an assessment of actual or perceived threats.

[SOURCE: ISO/IEC 15444-8:2007, 3.25]

3.16
sustainability
state of the global system, including environmental, social and economic aspects, in which the needs of the present are met without compromising the ability of future generations to meet their own needs

[SOURCE: ISO/Guide 82:2019, 3.1, modified — Removed Notes to entry.]

3.17
traceability
ability to identify or recover the history, provenance, application, use and location of an item or its characteristics

3.18
value chain
range of activities or parties that create or receive value in the form of products or services

[SOURCE: ISO 22948:2020, 3.2.11]

4 Overview

4.1 General

Ethical and societal concerns are a factor when developing and using AI systems and applications. Taking context, scope and risks into consideration can mitigate undesirable ethical and societal outcomes and harms. Examples of areas where there is an increasing risk for undesirable ethical and societal outcomes and harms include the following[24]:

— financial harm;

© ISO/IEC 2022 – All rights reserved