

SA TR ISO/IEC 24027:2022
ISO/IEC TR 24027:2021

STANDARDS
Australia



Technical Report

Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making



SA TR ISO/IEC 24027:2022

This Australian Technical Report® was prepared by IT-043, Artificial Intelligence. It was approved on behalf of the Council of Standards Australia on 03 October 2022.

This Technical Report was published on 14 October 2022.

The following are represented on Committee IT-043:

- Australian Chamber of Commerce and Industry
- Australian Computer Society
- Australian Healthcare and Hospitals Association
- Australian Human Rights Commission
- Australian Information Industry Association
- Australian Institute of Company Directors
- Australian Institute of Health & Safety
- Australian Payments Network
- Australian Securities and Investments Commission
- CHOICE
- Consult Australia
- Consumers Federation of Australia
- Data61 (CSIRO)
- Ethics, AI and ADM Professional Group
- NSW Data Analytics Centre
- National Association of Testing Authorities Australia
- Queensland AI Hub
- Royal Australian and New Zealand College of Radiologists
- Therapeutic Goods Administration (TGA)
- University of Melbourne
- University of New South Wales
- University of Technology Sydney
- Western Sydney University

This Technical Report was issued in draft form for comment as DR SA TR ISO/IEC 24027:2022.

Keeping Standards up-to-date

Ensure you have the latest versions of our publications and keep up-to-date about Amendments, Rulings, Withdrawals, and new projects by visiting:

www.standards.org.au

Technical Report

Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making

First published as SA TR ISO/IEC 24027:2022.

COPYRIGHT

© ISO/IEC 2022 — All rights reserved
© Standards Australia Limited 2022

All rights are reserved. No part of this work may be reproduced or copied in any form or by any means, electronic or mechanical, including photocopying, without the written permission of the publisher, unless otherwise permitted under the Copyright Act 1968 (Cth).

Preface

This Technical Report was prepared by the Standards Australia Committee IT-043, Artificial Intelligence.

The objective of this document is to address bias in relation to AI systems, especially with regards to AI-aided decision-making. In this document, the term bias is defined as a systematic difference in the treatment of certain objects, people, or groups in comparison to others, in its generic meaning beyond the context of AI or ML.

Measurement techniques and methods for assessing bias are described, with the aim to address and treat bias-related vulnerabilities. All AI system lifecycle phases are in scope, including data collection, training, continual learning, design, testing, evaluation, and use.

This document is identical with, and has been reproduced from, ISO/IEC TR 24027:2021, *Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making*.

As this document has been reproduced from an international document, a full point substitutes for a comma when referring to a decimal marker.

Australian or Australian/New Zealand Standards that are identical adoptions of international normative references may be used interchangeably. Refer to the online catalogue for information on specific Standards.

The terms “normative” and “informative” are used in Standards to define the application of the appendices or annexes to which they apply. A “normative” appendix or annex is an integral part of a Standard, whereas an “informative” appendix or annex is only for information and guidance.

Contents

Preface	ii
Foreword	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
3.1 Artificial intelligence	1
3.2 Bias	2
4 Abbreviations	3
5 Overview of bias and fairness	3
5.1 General	3
5.2 Overview of bias	3
5.3 Overview of fairness	5
6 Sources of unwanted bias in AI systems	6
6.1 General	6
6.2 Human cognitive biases	7
6.2.1 General	7
6.2.2 Automation bias	7
6.2.3 Group attribution bias	8
6.2.4 Implicit bias	8
6.2.5 Confirmation bias	8
6.2.6 In-group bias	8
6.2.7 Out-group homogeneity bias	8
6.2.8 Societal bias	9
6.2.9 Rule-based system design	9
6.2.10 Requirements bias	10
6.3 Data bias	10
6.3.1 General	10
6.3.2 Statistical bias	10
6.3.3 Data labels and labelling process	11
6.3.4 Non-representative sampling	11
6.3.5 Missing features and labels	11
6.3.6 Data processing	12
6.3.7 Simpson's paradox	12
6.3.8 Data aggregation	12
6.3.9 Distributed training	12
6.3.10 Other sources of data bias	12
6.4 Bias introduced by engineering decisions	12
6.4.1 General	12
6.4.2 Feature engineering	12
6.4.3 Algorithm selection	13
6.4.4 Hyperparameter tuning	13
6.4.5 Informativeness	14
6.4.6 Model bias	14
6.4.7 Model interaction	14
7 Assessment of bias and fairness in AI systems	14
7.1 General	14
7.2 Confusion matrix	15
7.3 Equalized odds	16
7.4 Equality of opportunity	16
7.5 Demographic parity	17

This is a free preview. Purchase the entire publication at the link below:

[Product Page](#)

-
- Looking for additional Standards? Visit Intertek Inform Infostore
 - Learn about LexConnect, All Jurisdictions, Standards referenced in Australian legislation
-