

# ISO/IEC TR 24027:2021-11 (E)

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

---

<b>Contents</b>		<b>Page</b>
	<b>Foreword</b> .....	<b>v</b>
	<b>Introduction</b> .....	<b>vi</b>
<b>1</b>	<b>Scope</b> .....	<b>1</b>
<b>2</b>	<b>Normative references</b> .....	<b>1</b>
<b>3</b>	<b>Terms and definitions</b> .....	<b>1</b>
	3.1 Artificial intelligence.....	1
	3.2 Bias.....	2
<b>4</b>	<b>Abbreviations</b> .....	<b>3</b>
<b>5</b>	<b>Overview of bias and fairness</b> .....	<b>3</b>
	5.1 General.....	3
	5.2 Overview of bias.....	3
	5.3 Overview of fairness.....	5
<b>6</b>	<b>Sources of unwanted bias in AI systems</b> .....	<b>6</b>
	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
<b>7</b>	<b>Assessment of bias and fairness in AI systems</b> .....	<b>14</b>
	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
7.6	Predictive equality .....	17
7.7	Other metrics .....	17
<b>8</b>	<b>Treatment of unwanted bias throughout an AI system life cycle .....</b>	<b>17</b>
8.1	General .....	17
8.2	Inception .....	17
8.2.1	General .....	17
8.2.2	External requirements .....	18
8.2.3	Internal requirements .....	19
8.2.4	Trans-disciplinary experts .....	19
8.2.5	Identification of stakeholders .....	19
8.2.6	Selection and documentation of data sources .....	20
8.2.7	External change .....	20
8.2.8	Acceptance criteria .....	21
8.3	Design and development .....	21
8.3.1	General .....	21
8.3.2	Data representation and labelling .....	21
8.3.3	Training and tuning .....	22
8.3.4	Adversarial methods to mitigate bias .....	23
8.3.5	Unwanted bias in rule-based systems .....	24
8.4	Verification and validation .....	24
8.4.1	General .....	24
8.4.2	Static analysis of training data and data preparation .....	25
8.4.3	Sample checks of labels .....	25
8.4.4	Internal validity testing .....	25
8.4.5	External validity testing .....	25
8.4.6	User testing .....	26
8.4.7	Exploratory testing .....	26
8.5	Deployment .....	26
8.5.1	General .....	26
8.5.2	Continuous monitoring and validation .....	26
8.5.3	Transparency tools .....	27
	<b>Annex A (informative) Examples of bias .....</b>	<b>28</b>
	<b>Annex B (informative) Related open source tools .....</b>	<b>31</b>
	<b>Annex C (informative) ISO 26000 – Mapping example .....</b>	<b>32</b>
	<b>Bibliography .....</b>	<b>36</b>