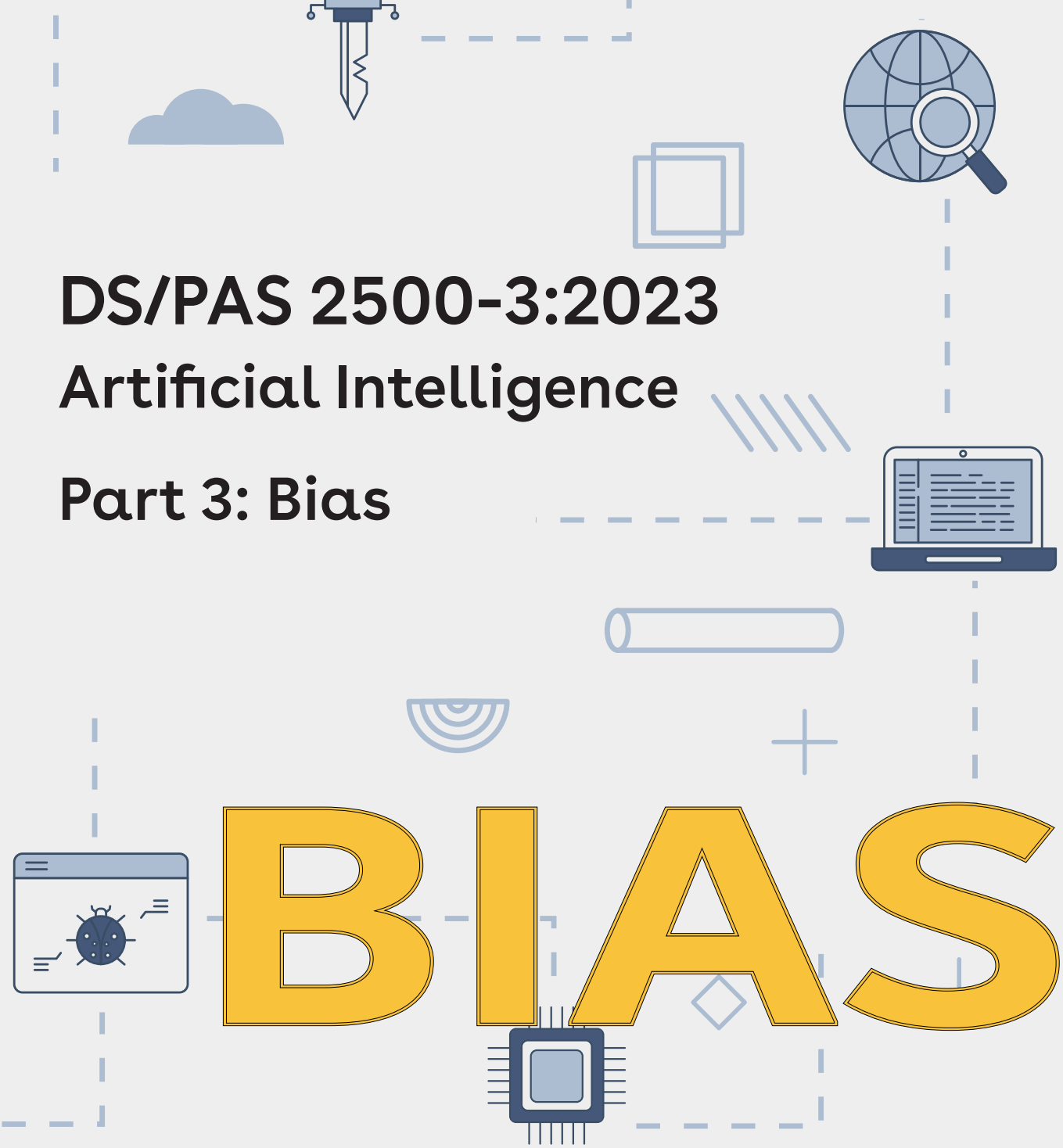
DANISH STANDARDS

DS/PAS 2500-3:2023

Artificial Intelligence

Part 3: Bias



DS/PAS 2500-3:2023

Artificial Intelligence

Part 3:
Bias

DANISH STANDARDS

DS/PAS 2500-3:2023,
Artificial Intelligence – Part 3: Bias

DS/PAS 2500-3:2023
Project number M393732
Graphic design: Jenni Søndergaard
Editor: Jenni Søndergaard
Printed by: Danish Standards, Published 2025
First edition

Published by Danish Standards,
Göteborg Plads 1
DK-2150 Nordhavn

Phone: 39 96 61 01
ds@ds.dk www.ds.dk

This is a POD publication, printed in Denmark

Images: Colourbox and IStockphoto.

Contents

Foreword

DS publication type

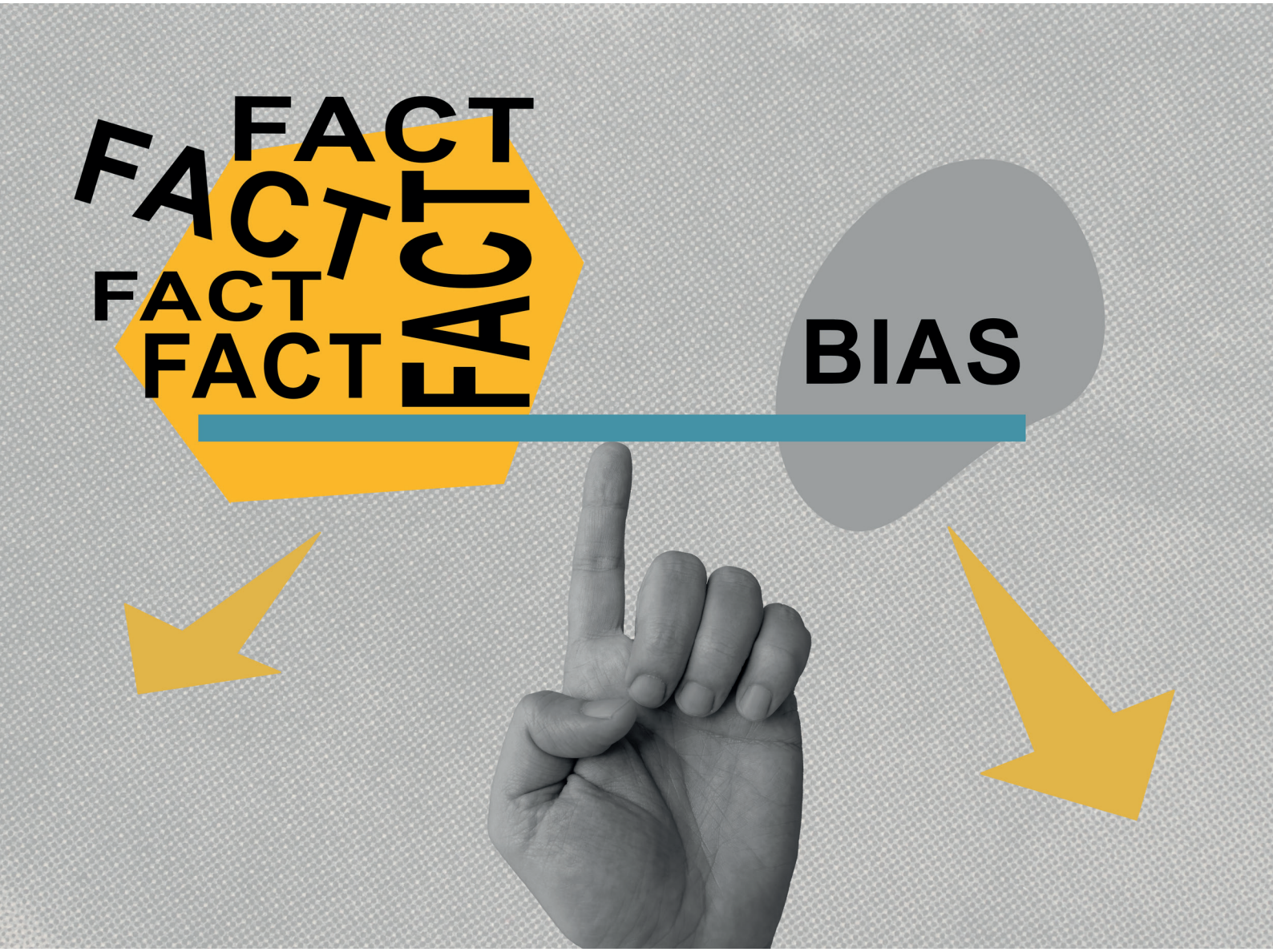
All designations for publications published by Danish Standards begin with DS followed by one or more prefixes and a number, e.g. DS 282, DS/EN 5414 etc. If the publication is part of a series, this will be indicated with a hyphen followed by the number in the series, e.g. DS/PAS 2500-1 and DS/PAS 2500-2. A series is a non-prioritised sequence of publications, each of which can be read individually but addresses the same topic.

DS/PAS

This document is a DS/PAS. PAS is an abbreviation for 'Publicly Available Specification', which is a publication developed at the national level, that does not have the same status as an international standard. A PAS differs from an international standard, for example, by not having the same requirements for the level of stakeholder involvement or layout. A DS/PAS also does not specify requirements that must be complied with, but instead offers recommendations, information, and advice.

Preparation of the publication

This publication has been developed together with the Department of Computer Science at the University of Copenhagen (DIKU), Piedboeuf and Bureau M, and with input from Danish stakeholders through workshops, interviews, and written comments in accordance with the procedure for developing PAS publications at Danish Standards.


FACT
FACT
FACT
FACT
FACT
BIAS

Introduction

As artificial intelligence continues to develop and spread, attention has been drawn to the unwanted biases that can find their way into IT systems. These biases, when applied in specific IT systems, can impact various population groups based on factors such as gender, colour of skin, age, disability, social class, and more.

The presence of these unwanted biases in IT systems results from the complex intersection of IT, human involvement, and processes – both organizational and technical – during the development of these systems. This specification focuses on IT systems that are involved in or contribute to decision-making processes affecting groups of people, particularly AI-based systems. The development of IT systems also includes a range of human decisions: at each stage of the development process, choices are made, and biases can emerge during these decision-making processes.

The specification aims to provide guidance on how to create transparency about these biases and their potential impacts. From there it is the responsibility of the reader to determine which biases are considered wanted, unwanted, or neutral.

The specification does not provide guidance on determining what is fair, but instead helps the reader establish a basis for making informed decisions, considering applicable legislation, as well as personal views, values, and perceptions of fairness.

The goal is for the specification to be widely adopted and reach a diverse audience, including IT developers, legal professionals, HR departments, internal decision-makers, as well as purchasers, public authorities, and citizens. By doing so, the specification aims to raise general awareness that artificial intelligence will almost inevitably reflect biases that result from human bias in the development of IT systems. As a result, the focus should be on identifying and managing unwanted biases rather than attempting to eliminate bias entirely.

1. Scope

This document addresses the management of biases that occur in the interaction between IT systems, processes, and people. This includes:

- The design and development of IT systems and processes
- A general approach to identifying biases in this context and
- A method for defining design and development requirements that support the responsible management of bias.

This document provides recommendations for a method to identify biases in the design and development of IT systems. The approach outlined can be applied to both systems and processes used for automated decision-making or decision support, including artificial intelligence, as well as those that do not.

The primary target group includes requesters, designers, developers, and users of IT systems who are either interested in or required to identify and specify requirements regarding bias. A secondary target group includes stakeholders who would benefit from a greater understanding and awareness of bias in systems and processes, such as public authorities, purchasers, companies, consumers, and citizens.

The objective is to foster transparency and awareness of bias, rather than making decisions on how to address these biases based on normative views on e.g. fairness.

2. Normative references

There are no normative references.

3. Terms and definitions

In this document, the following terms and definitions apply.

3.1
algorithm
description of a process to solve a general problem

3.2
artificial intelligence
a system's ability to acquire, process, and apply *knowledge* (3.3) and/or skills to find solutions to problems, make or support decisions

Note 1 to term: Artificial intelligence is generally compared to human intelligence, although this is not an exhaustive definition.

Note 2 to term: Machine learning is a subcategory of artificial intelligence.

3.3
knowledge
data, information, and skills acquired through experience or training

3.4
decision
choice made between several options

3.5
decision domain
subject area, including options, for a *decision* (3.4), e.g. allocation of a social benefit by a public authority or bank credit

3.6

decision support

automating a *decision* (3.4) or providing information that a human can use to decide

Note 1 to term: There are three degrees of human involvement in decisions.

3.6.1

human-centered decision support

decision support (3.6) that requires human involvement, e.g. by providing parts of the data basis for the decision or being responsible for executing the decision

3.6.2

human-monitored decision support

decision support (3.6) that operates independently of humans but under human monitoring with intervention when necessary

3.6.3

autonomous decision support

decision support (3.6) that operates without human involvement

3.7

bias

systematic difference or error in data processing

Note 1 to term: Processing is any kind of action, including perception, observation, representation, prediction, or decision.

3.8

machine learning

statistical method or algorithm that, based on sample data, can find an algorithm capable of solving a given decision problem with a certain accuracy

3.9

interested party
stakeholder

individual, group, or organization that can affect, is affected by, or perceives itself to be affected by a decision or activity

3.10

system

combination of interacting elements organized to achieve one or more specified purposes

3.11

IT system

system (3.10) based on information technology

3.12

technical expert

person who designs, develops, tests, verifies, or audits the technical function of an *IT system* (3.11) or elements thereof

3.13

supervisor

person, public authority, or company that oversees that an IT project or the use of an *IT system* (3.11) complies with specified requirements

3.14

data bias

data characteristics that, if not addressed, can lead to a system with *bias* (3.7) towards different groups

3.15

group

subset of people in a domain that share common characteristics

3.16

characteristics

defining characteristics of objects in a domain

3.17

human bias

bias (3.7) that occurs when people process and interpret information

3.18

bias effect

the result of *bias* (3.7) in a system that may lead to differences in treatment between specific individuals and groups compared to others

Note 1 to term: A bias effect can be either positive/wanted, neutral or negative/unwanted.

3.19

data drift

result of a development in data that invalidates the data model in relation to *the decision domain* (3.5)

3.20

concept drift

change over time in an unforeseen way of a variable that the system is trying to predict

4. Bias

4.1 Introduction to bias

In this specification bias is understood as a systematic difference in data processing, which can have the effect of creating a difference in the treatment of specific objects, individuals, or groups compared to others. This could be between people of different genders and ages or between people of different ethnic origins. Treatment can be any form of action, including perception, observation, classification, representation, prediction, or decision-making. Any form of non-trivial treatment will be biased, as the treatment would otherwise be independent of the object being treated. It is therefore relevant to discuss the effect of bias in question and whether this effect is positive (wanted), negative (unwanted), or neutral.

In everyday language and understanding, discrimination is a normative concept, as it is most often used to refer to a discriminatory practice that puts someone at a disadvantage compared to others. In this specification, however, the term "difference in treatment" is used as a neutral concept that refers to whether an IT system is designed to – intentionally or unintentionally – treat various groups differently. An interpretation of difference in treatment as discriminatory thus only applies in cases where this is neither intentional nor purposeful. An example of this is a system that uses a predictive model where group affiliation appears as a parameter in the function, for instance so that the prediction is mathematically adjusted depending on whether the individual is male or female.

Artificial intelligence in this specification is characterised by a system's abilities to process and apply knowledge and/or skills to find solutions to problems, make, or support decisions. Different types of artificial intelligence methods, such as the use of machine learning, rule-based systems, and statistical models, will use different types of knowledge and data, and will also have different ways of processing and applying this knowledge, thus also having different ways in which bias is introduced and identified. Machine learning methods are typically based on examples of solutions to a problem or decision. Rule-based systems may also be based on underlying learning from examples but can also be based on expert knowledge, which may have been obtained in various ways.

Finally, statistical models typically rely on observations and assumptions about a mathematical model that adequately explains these observations. It is beyond the scope of this specification to go into detail on method-specific sources of bias, but this is relevant for how the individual questions presented later in the document are answered. For example, there will be different types of knowledge or types of

data underlying machine learning, rule-based, and statistical models, respectively.

However, it is relevant to distinguish between human (cognitive) bias and data bias, including statistical bias.

4.2 Human bias

4.2.1 General

Human (cognitive) bias is the bias associated with human data processing, including selection or decision-making. Human bias comes in many forms and can be introduced at any stage of the development and use of an IT system. Examples of human bias include automation bias, group attribution bias, confirmation bias, time period bias, in-group bias, and social contextual bias. Specifically in relation to the development of IT systems, several choices are made that contain bias, such as choice of parameters, choice of method, and, depending on the method, choice of algorithm, inference method, or statistical model.

4.2.2 Automation bias

Automation bias is the type of bias where, in the context of decision support with human involvement or monitoring, an action, choice, or decision made or suggested by an automated system is uncritically accepted.

4.2.3 Group attribution bias

Group attribution bias occurs when a person assumes that a characteristic of an individual or element within a group applies to the entire group. For example, it could be a teenager assuming that all parents are stupid, or a parent assuming that all teenagers are lazy.

4.2.4 Confirmation bias

Confirmation bias refers to when people prefer choices, information, or decisions that confirm their own perception of what is right.

4.2.5 Time period bias

Time period bias occurs because of the system being based on knowledge or data from a limited time period that is not representative of the time at which the system is used.

4.2.6 In-group bias

An example of in-group bias occurs when a person treats the group they belong to differently from other groups, often by assigning greater authority to that group. Another type of in-group bias occurs when a person perceives a group (often a group they do not belong to) as more homogeneous than it is, thus risking treating members of the group uniformly in cases where differentiation should be made.

4.2.7 Social contextual bias

Social contextual bias refers to human bias that is particularly prevalent in society and thus may become embedded in data, organizations/institutions, organizational rules, e.g. stereotypes. This can be an assumption that unemployment is largely self-inflicted, which can be reflected in legislation designed to monitor and activate the unemployed in specific ways.

4.3 Data bias

4.3.1 General

Data bias refers to characteristics of data used in systems that will lead to bias in the system's treatment of specific groups compared to others if these characteristics are not considered. Data bias is often a result of human bias in the selection and processing of data and knowledge used to develop the system. Biased in connection with selection generally occurs when data is not selected in a way that is representative of the true distribution of data. Examples of such biases can include selection bias, coverage bias, and participation bias. Treatment bias can occur in the classification (labeling) of the data that forms the basis for system development into categories that do not reflect the variation in society, for example, by dividing people into men and women or into politically "blue" or "red."

4.3.2 Selection bias

Selection bias occurs when data is not collected randomly, which can happen, for example, when survey respondents are selected among personal acquaintances or users of a specific social media platform.

4.3.3 Coverage bias

Coverage bias occurs when the population in the dataset does not correspond to the population you want to make decisions for. For example, an IT system for facial recognition that has only been trained on white faces being used on all types of faces.

4.3.4 Participation bias

Participation bias occurs when the respondents in a survey are not representative of the group to be studied. For example, an online survey may not include participants who have chosen not to use a computer.

5. Bias management and the use of artificial intelligence

5.1 Considerations

It is necessary to consider when it is ethically responsible to use an IT system that applies artificial intelligence, or when it is ethically responsible for humans to solve the task. These considerations should be included in all phases of IT system development and involve all relevant stakeholders. The considerations must be documented to a degree that allows those responsible for the project and the individual phases and activities to be held accountable for carrying out activities and initiating subsequent phases.

5.2 Lifecycle model

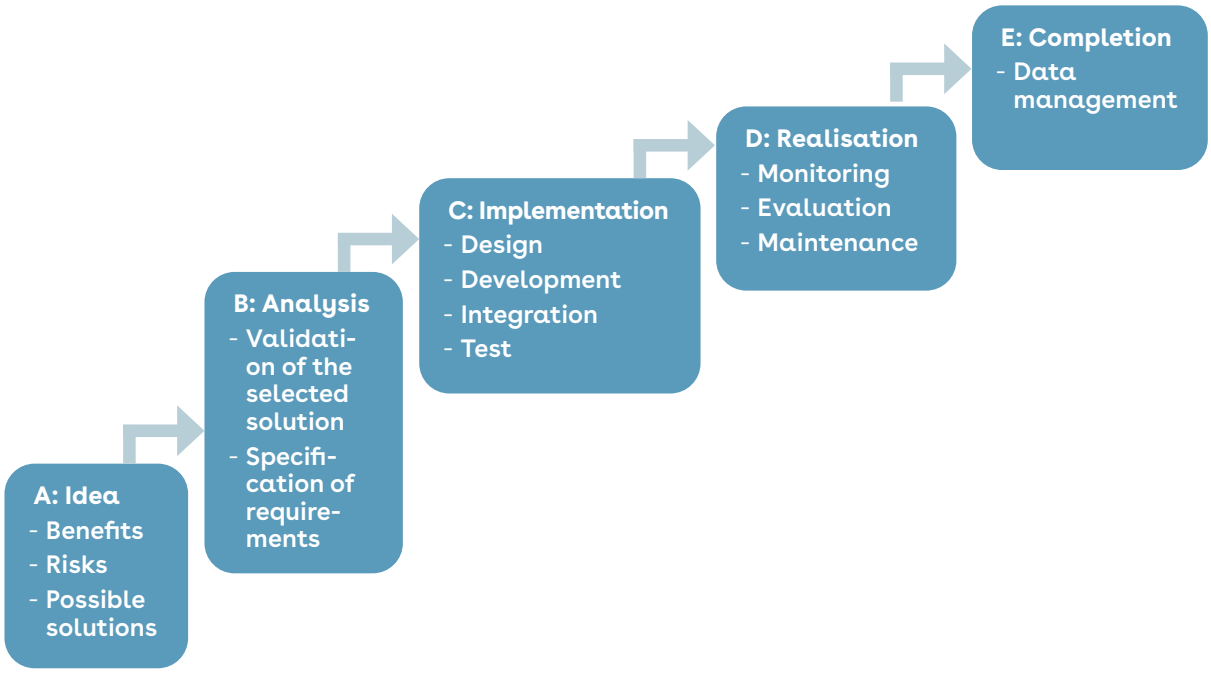
5.2.1 General

In this document, the overall phases in the Danish State IT project model[1] are used as a starting point: Idea, analysis, implementation, and realisation. In this document, the "completion" phase has also been added. Each phase contains many activities, as shown in Figure 1.

The phases can be executed strictly sequentially, as in a so-called waterfall model that follows the order shown in the figure or overlap and be repeated as in so-called iterative or agile models. The organization responsible for the project (or the organizations responsible for each phase) should appoint one or more people responsible for ensuring that the considerations in the checklist (5.4) are included in the decision to initiate activities and subsequent phases.

1 https://oes.dk/it-og-oekonomistyring/it-projektstyring/dokumenter-og-vejledninger/ (in Danish)

Figure 1: Phases in the Danish State IT project model and the activities they should include as a minimum.



5.2.2 A: The idea phase

As a minimum, this phase should include identifying desired benefits, possible solutions, and possible risks, as well as risk management.

5.2.3 B: The analysis phase

As a minimum, this phase should include validation of solutions against desired benefits and risks, and specification of requirements.

5.2.4 C: The implementation phase

This phase includes design, development, integration, and test.

5.2.5 D: The realisation phase

This phase includes monitoring, evaluation, and maintenance of the system or parts of the system, including personally identifiable information (PII) that is no longer needed.

5.2.6 E: The completion phase

When phasing out a system, there are certain requirements to be aware of. These requirements concern, for example, whether data is handled correctly (typically in relation to legislation).

5.3 Stakeholder groups and focus areas

Different groups and stakeholders have vastly different needs and expectations for an IT system. Therefore, this specification is based on a number of different groups and their immediate needs. This inherently introduces a bias, but it is a wanted bias intended to help structure the work.

The checklists are based on the following stakeholder groups:

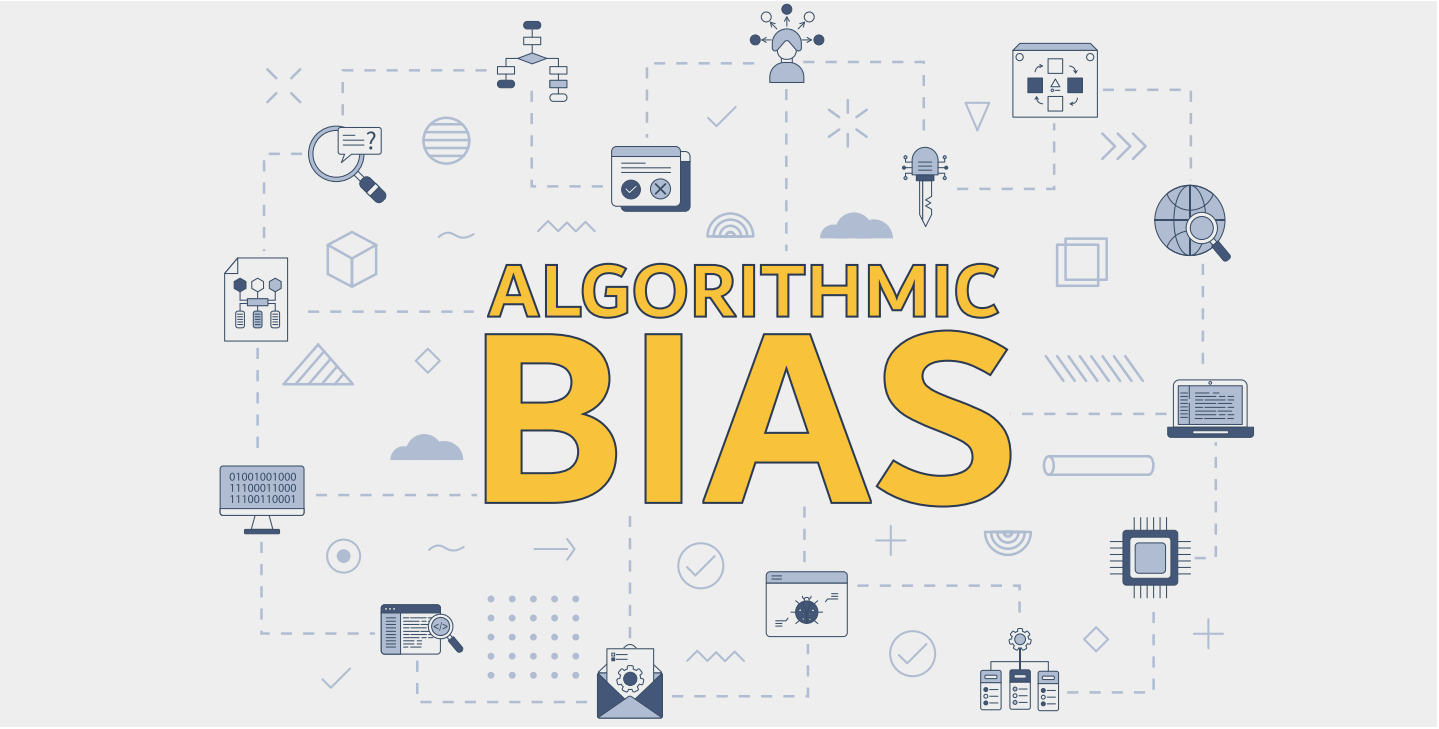
- Project owner/requester
- End users
- Citizens (indirectly and directly affected)
- Suppliers
- Developers
- Legal staff/supervisors/unbiased technical experts.

On the next page, there is an overview of stakeholder groups and their typical focus areas (Table 1). In section 5.4, there are checklists with questions that are relevant to the stakeholder groups and to the documentation of each phase in the development of a system that uses artificial intelligence.

STAKEHOLDER GROUPS	DESCRIPTION	FOCUS AREA FOR THE STAKEHOLDER GROUP
Project owner/ requester	The organization, department, or person responsible for developing and operating an IT system. *For example, a technical or organizational project manager.*	Identification of bias related to idea, analysis, design, requirements specification, and the resulting effect of using the system.
End users	A user who has direct interaction with the system. *An example could be an insurance advisor handling a case using the system, or a citizen using an online self-help service.*	Identification of bias related to the idea/design phase (involvement), use of the IT system, and specific outcomes of using the system.
Citizens - directly affected	People who do not interact directly with the system but are affected or perceive themselves to be affected by the effects of using the system. *An example could be customers whose insurance terms are changed based on other customers' cases settled by an IT system.*	Identification of bias in the resulting effect of using the system.
Citizens - indirectly affected	People who may be indirectly affected by the effects of the system or if the system is used in a context of general public interest. *An example could be that the general insurance terms across the industry change based on the use of the system.*	Identification and general precision of bias related to e.g. population groups.
Suppliers	Companies or individuals who act as suppliers of parts of an IT system or the whole system. This also includes data suppliers. Suppliers can be involved in all phases of a project (see Figure 1). *This could be, for example, a supplier of analysis components for estimating insurance risks or a supplier of data for a construction project management system.*	Identification of bias in connection with the development and maintenance of the system.

Developers	People responsible for the development of an IT system or its components. This includes both internal and external developers. *For example, it could be the project owner's own IT department, temporary consultants, or developers from a subcontractor.*	Identification of bias associated with data and algorithms, including data weighting.
Legal staff/ supervisors/ unbiased technical experts	People who ensure compliance with standards, oversee that case management is carried out correctly, or determine quality. *This could include certification of insurance systems, monitoring of public case management systems, or quality audits.*	Bias that influences the ability to monitor and approve the system. Bias management in connection with audits and documentation.

Table 1: Stakeholder groups and focus areas



5.4 Considerations for the different phases of the IT project

5.4.1 General

The checklists for the different phases are a systematic way to ensure that the various stakeholder groups (Table 1) are considered in the IT project and that the considerations are documented. The checklists are therefore an internal tool that can be used by the project team.

In the "Question" column, a question is posed, and the considerations are noted in the table. In most cases, it is relevant to involve the stakeholder in the considerations, but it is not the intention that stakeholders fill in the answers in the checklist. Note that for a specific system, there may be stakeholders who are not relevant and stakeholders that may not have been included. Similarly, organizations representing stakeholders, such as interest groups, may be involved rather than individual members of that group.

In the "Action" column, the necessary action that the considerations in the "Question" column give rise to is noted. In this context, models can be applied to assess risks and organize actions, such as the risk model (see Annex C) and the Plan-Do-Check-Act (PDCA) method (see Annex D). These methods will help minimise "putting out fires", which is when problem-solving occurs through quick fixes, with a focus on symptom treatment.

Please note that the checklists do not have to be completed in the given order and therefore some questions are included in multiple phases. The checklists are also not based on specific industries or specific types of systems that use artificial intelligence.

5.4.2 Phase A: The idea phase

The organization responsible for the idea phase should at a minimum consider and document the answers to the questions in A1-A4 for each intended use, involving relevant stakeholders or representatives of these.

ID	QUESTIONS	ACTIONS
A1	**Which groups in the decision domain of the system are relevant in terms of bias?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
A1-1: Project owner/requester		
A1-2: End users		
A1-3: Citizens (indirectly and directly affected)		
A1-4: Suppliers		
A1-5: Developers		
A1-6: Legal staff/supervisors/ unbiased technical experts		

Table 2 continues on the following pages

A2	**What legislation and which human rights (see Annex A) are relevant in this context?** **Have the groups (defined in A1) expressed values, opinions, or wishes that should be considered?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
A2-1: Project owner/requester		
A2-2: End users		
A2-3: Citizens (indirectly and directly affected)		
A2-4: Suppliers		
A2-5: Developers		
A2-6: Legal staff/supervisors/ unbiased technical experts		
A3	**Is the IT system accessible for those with special needs or disabilities? Are there other groups that risk being excluded?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
A3-1: Project owner/ requester		
A3-2: End users		
A3-3: Citizens (indirectly and directly affected)		
A3-4: Suppliers		
A3-5: Developers		
A3-6: Legal staff/supervisors/ unbiased technical experts		

A4	**How is the IT system secured against cyberattacks that could affect data quality?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
A4-1: Project owner/ requester		
A4-2: End users	(Probably not applicable)	
A4-3: Citizens (indirectly and directly affected)	(Probably not applicable)	
A4-4: Suppliers		
A4-5: Developers		
A4-6: Legal staff/supervisors/ unbiased technical experts		

Table 2: Checklist for the idea phase

5.4.3 Phase B: The analysis phase

The organization responsible for the analysis phase should at a minimum consider and document the answers to the questions in B1-B6 for each intended use, involving relevant stakeholders or representatives of these.

ID	QUESTIONS	ACTIONS
B1	**Which groups in the decision domain of the system are relevant in terms of bias? (May have been answered in A1)**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
B1-1: Project owner/requester		
B1-2: End users		
B1-3: Citizens (indirectly and directly affected)		
B1-4: Suppliers		
B1-5: Developers		
B1-6: Legal staff/supervisors/ unbiased technical experts		
B2	**What human cognitive biases may be present in the system, and do they have a negative/ unwanted, neutral, or positive/ wanted effect?** **In the case of negative/ unwanted biases, has a risk analysis and impact assessment been conducted?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
B2-1: Project owner/requester		
B2-2: End users		
B2-3: Citizens (indirectly and directly affected)		
B2-4: Suppliers		
B2-5: Developers		
B2-6: Legal staff/supervisors/ unbiased technical experts		

Table 3 continues on the following pages

B3	**What data biases may be present in the system, and do they have a negative/ unwanted, neutral, or positive/ wanted effect?** **Are the data representative of the relevant groups, such as demographic parameters like gender, income, origin, and age?** **In the case of negative/ unwanted biases, has a risk analysis and impact assessment been conducted?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
B3-1: Project owner/ requester		
B3-2: End users		
B3-3: Citizens (indirectly and directly affected)		
B3-4: Suppliers		
B3-5: Developers		
B3-6: Legal staff/supervisors/ unbiased technical experts		
B4	**How stable are the assumptions and data the system is based on, and what is the risk of data drift?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
B4-1: Project owner/ requester		
B4-2: End users	(Probably not applicable)	
B4-3: Citizens (indirectly and directly affected)	(Probably not applicable)	
B4-4: Suppliers		
B4-5: Developers		
B4-6: Legal staff/supervisors/ unbiased technical experts		

B5	**Can the system be developed to be robust, reliable, and secure, and can it be monitored and audited?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
B5-1: Project owner/ requester		
B5-2: End users	(Probably not applicable)	
B5-3: Citizens (indirectly and directly affected)	(Probably not applicable)	
B5-4: Suppliers		
B5-5: Developers		
B5-6: Legal staff/supervisors/ unbiased technical experts		
B6	**Can the system be developed to be robust, reliable, and secure, and can it be monitored and audited?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
B6-1: Project owner/ requester		
B6-2: End users	(Probably not applicable)	
B6-3: Citizens (indirectly and directly affected)	(Probably not applicable)	
B6-4: Suppliers		
B6-5: Developers		
B6-6: Legal staff/supervisors/ unbiased technical experts		

Table 3: Checklist for the analysis phase

5.4.4 Phase C: The implementation phase

The organization responsible for the implementation phase should at a minimum consider and document the answers to the questions in C1-C4 for each intended use, involving relevant stakeholders or representatives of these.

ID	QUESTIONS	ACTIONS
C1	**Is there sufficient transparency[2] in the design of the system so that bias can be identified, audited, and monitored?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
C1-1: Project owner/requester		
C1-2: End users		
C1-3: Citizens (indirectly and directly affected)		
C1-4: Suppliers		
C1-5: Developers		
C1-6: Legal staff/supervisors/ unbiased technical experts		
C2	**Is there sufficient training and awareness of bias in the relevant groups (see A1)?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
C2-1: Project owner/requester		
C2-2: End users		
C2-3: Citizens (indirectly and directly affected)		
C2-4: Suppliers		
C2-5: Developers		
C2-6: Legal staff/supervisors/ unbiased technical experts		

2 See DS/PAS 2500-1:2020, *Artificial Intelligence - Part 1: Transparency*

Table 4 continues on the following page

C3	**Has a procedure been established to remove potential unwanted bias?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
C3-1: Project owner/ requester		
C3-2: End users	(Probably not applicable)	
C3-3: Citizens (indirectly and directly affected)	(Probably not applicable)	
C3-4: Suppliers		
C3-5: Developers		
C3-6: Legal staff/supervisors/ unbiased technical experts		
C4	**Is the documentation regarding bias sufficiently transparent?** **Can any obligations related to the use of data be documented?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
C4-1: Project owner/ requester		
C4-2: End users		
C4-3: Citizens (indirectly and directly affected)		
C4-4: Suppliers		
C4-5: Developers		
C4-6: Legal staff/supervisors/ unbiased technical experts		

Table 4: Checklist for the implementation phase

5.4.5 Phase D: The realisation phase

The organization responsible for the realisation phase should at a minimum consider and document the answers to the questions in D1-D5 for each intended use, involving relevant stakeholders or representatives of these.

ID	QUESTIONS	ACTIONS
D1	**Has a procedure been established for monitoring and identifying unwanted bias?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
D1-1: Project owner/requester		
D1-2: End users		
D1-3: Citizens (indirectly and directly affected)		
D1-4: Suppliers		
D1-5: Developers		
D1-6: Legal staff/supervisors/ unbiased technical experts		
D2	**Has a procedure been established for managing unwanted bias?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
D2-1: Project owner/requester		
D2-2: End users		
D2-3: Citizens (indirectly and directly affected)		
D2-4: Suppliers		
D2-5: Developers		
D2-6: Legal staff/supervisors/ unbiased technical experts		

Table 5 continues on the following page

D3	**Is there monitoring and ongoing control of the system's compliance with relevant legislation (see Annex A), human rights, democratic values, and diversity issues?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
D3-1: Project owner/ requester		
D3-2: End users		
D3-3: Citizens (indirectly and directly affected)		
D3-4: Suppliers		
D3-5: Developers		
D3-6: Legal staff/supervisors/ unbiased technical experts		
D4	**Are there automatic mechanisms integrated to handle potential changes in bias (e.g. concept drift)?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
D4-1: Project owner/ requester		
D4-2: End users	(Probably not applicable)	
D4-3: Citizens (indirectly and directly affected)	(Probably not applicable)	
D4-4: Suppliers		
D4-5: Developers		
D4-6: Legal staff/supervisors/ unbiased technical experts		

D5	**Is the audit log maintained correctly (including for runtime changes)?**	[The risk model and/or PDCA method described in Annex C and D may be used]
	Answer (to be filled in)	**Action** (to be filled in)
D5-1: Project owner/ requester		
D5-2: End users	(Probably not applicable)	
D5-3: Citizens (indirectly and directly affected)	(Probably not applicable)	
D5-4: Suppliers		
D5-5: Developers		
D5-6: Legal staff/supervisors/ unbiased technical experts		

Table 5: Checklist for the realisation phase

5.4.6 Phase E: The completion phase

When a system is decommissioned, there are specific requirements that come into play. These requirements primarily concern whether existing data is properly disposed of (typically in relation to legislation). This section does not include an explicit checklist, as this specification does not relate to current legislation. Consumers and recipients are encouraged to create suitable checklists already in the idea phase.

Annex A

A

(informative)

Overview of relevant legislation

A.1 Relevant legislation

A.1.1 General

When addressing bias in relation to demographic parameters such as gender, ethnicity, and age, attention must be paid to the relevant legislation, including EU law and international human rights.

A.1.2 International legislation

The European Convention on Human Rights prohibits discrimination based on gender, race, colour of skin, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth, or any other status. The European Convention on Human Rights is incorporated into Danish law, which means that it forms part of the Danish legal framework.

EU law includes a principle of equal treatment and a prohibition of discrimination **in the Treaties and the EU Charter of Fundamental Rights**. In addition to the general rules and principles, a number of directives have been adopted and implemented in Danish law, which protect against discrimination in specific situations. The directives protect against discrimination based on racial and ethnic origin, religion or belief, disability, age, sexual orientation, and gender[3].

A.1.3 National legislation

The relevant national legislation in this context will often – but not exclusively – include:

- The Equal Treatment Act
- The Act on the Prohibition of Differences of Treatment
- The Gender Equality Act
- The Act on Ethnic Equal Treatment
- The Act on Prohibition of Differences of Treatment on the grounds of Disability.

If the IT system's decision domain relates to access to and/or opportunities in the labour market, it is advisable to refer to the Equal Treatment Act and the Act on the Prohibition of Differences of Treatment. This could e.g. include systems for recruitment or performance management.

The Equal Treatment Act requires employers to treat everyone equally regardless of gender, gender identity, gender expression, and gender characteristics, during hiring, promotion, and transfer. The Equal Treatment Act also ensures that women are treated equally in the labor market during pregnancy, childbirth, and maternity leave.

3 https://menneskeret.dk/sites/menneskeret.dk/files/media/document/Algoritmer_8.K.pdf (in Danish)

B

Annex B

(informative)

Overview of relevant standards

B.1 International standards

The Act on the Prohibition of Differences of Treatment prohibits direct or indirect discrimination in the labour market based on race, colour of skin, religion or belief, political opinion, sexual orientation, age, disability, or national, social or ethnic origin. This may be in relation to hiring, dismissal, transfer, promotion, or regarding salary and working conditions.

For areas that are not directly or indirectly related to the labour market, **the Gender Equality Act** ensures equality between women and men, between sexual minorities and majorities, and prohibits discrimination based on gender identity, gender expression, and gender characteristics. According to the Gender Equality Act, these groups must be treated equally in public, general, and commercial enterprises. For example, the same product cannot be priced differently for men and women.

The Act on Ethnic Equal Treatment and The Act on Prohibition of Differences of Treatment on the grounds of Disability also protect against discrimination outside the labour market. This could, for instance, relate to access to housing or the allocation of social benefits by the public authorities.

DS/ISO/IEC TR 24027 Information technology – Artificial intelligence (AI) – Bias in AI systems and AI aided decision making

This document addresses bias in relation to AI systems, especially with regards to AI aided decision making. Measurement techniques and methods for assessing bias are described with the aim of addressing and managing bias-related vulnerabilities. All stages of the AI system's lifecycle are included, such as data collection, training, ongoing learning, design, testing, evaluation, and use.

DS/ISO/IEC 23894 Information technology – Artificial intelligence – Guidance on risk management

This document provides guidelines for managing risks that organizations face during the development and use of artificial intelligence techniques and systems. The guidelines also aim to help organizations integrate risk management into their AI-related activities and functions. Additionally, it outlines processes for the effective implementation and integration of AI risk management. The application of these guidelines can be adapted to any organization and its context.

DS/ISO/IEC TR 24368 Information technology – Artificial intelligence – Overview of ethical and societal concerns

This document provides an overview of ethical and societal concerns regarding the use of artificial intelligence. In addition, this document contains information related to principles, processes, and methods in this area. The document is intended for technicians, public authorities, interest groups, and society in general. The document is not intended to advocate a specific set of values (value systems). This document also contains an overview of international standards that address issues related to ethical and societal concerns.

DS/ISO/IEC 5338 Information technology – Artificial intelligence – AI system lifecycle processes

This document defines a set of processes and associated concepts for describing the lifecycle of AI systems based on machine learning and heuristic systems. It is based on ISO/IEC/IEEE 15288:2015 and ISO/IEC/IEEE 12207:2017, with modifications and additions of AI-specific processes from ISO/IEC 22989:– and ISO/IEC 23053:2022. The document outlines processes that support the definition, control, management, execution, and improvement of the AI system throughout its lifecycle stages. These processes can also be used in an organization or project when developing or acquiring AI systems.

B.1 National standards

DS 5001 Diversity and equality management systems – Requirements and guidelines

This standard establishes the requirements for diversity and equality management systems in cases where an organization needs to: a) demonstrate its ability to continuously promote greater diversity and equality within the organization, b) document that the organization is systematically working with diversity and equality, and c) strengthen its credibility through transparent efforts and documentation. All requirements in this standard are generic and intended to be applicable to any organization, regardless of type or size, or the products and services it provides. The scope of the documented information for the management system should be aligned with the organization's size, complexity, locations, and tailored to the individual focus areas.

DS/PAS 2500-1:2020 Artificial intelligence – Part 1: Transparency

This document describes transparency, a general approach to achieving it, and a method for self-assessing the level of transparency. This method can also be used to specify requirements for transparency. The approach and method can be applied both to systems that use artificial intelligence and those that do not. The target group includes designers, developers, and users who either wish to or are required to document and/or increase transparency in their systems. A secondary group is the stakeholders who benefit from increased transparency, such as public authorities, purchasers, companies, and consumers. This document specifies recommendations for an approach to achieve transparency in systems used for automated decision-making or decision support, including conclusions (inference) and data usage. Such systems will often use artificial intelligence, but this is not a requirement. Additionally, the approach can be used to assess whether the transparency achieved meets an adequate level. The document does not cover reproducibility, explanations in a broader sense, or maintenance (e.g. concept drift).

DS/PAS 2500-2:2020 Artificial intelligence – Part 2: Decision support application in public case management

This document provides a checklist for relevant considerations in the various phases of IT projects using artificial intelligence for decision support in public case management. It is important that these considerations are included as the basis for requirements and solutions, as well as in the decision to carry out activities and proceed to the next phase in the IT system's lifecycle. It is the responsibility of the user to determine whether the checklist is applicable for a specific purpose. The target group includes designers, developers, providers, purchasers, auditors, and users of artificial intelligence for decision support in public case management who wish or are required to evaluate the system's use. A secondary group includes stakeholders affected by the system, such as public authorities and citizens in general. The document does not cover legislation and specific requirements that may apply when using decision support within specific domains. The document must be used in conjunction with the organization's existing guidelines for ensuring compliance with regulations, including, but not limited to, privacy management, cybersecurity, robustness of IT solutions, data governance, complaint and appeal management, as well as legislation. This document does not explicitly address transparency in artificial intelligence but can be used alongside DS/PAS 2500-1:2020, Artificial Intelligence – Part 1: Transparency, which specifies recommendations for an approach to achieving transparency in systems used for automated decision-making or decision support.

All standards can be purchased in the Danish Standards webshop: https://webshop.ds.dk/
The DS/PAS 2500 series is free of charge.

Annex C

(informative)

C.1 Risk model based on DS/EN IEC 31010:2019, *Risk management – Risk assessment techniques*

The following model can be used to assess the risk in each phase of the project.

RISK	Brief description of risks	Consistency (1 = low; 3 = medium; 5 = high)	Probability (1 = low; 3 = medium; 5 = high)	Risk factor (consequence x probability)	How can we mitigate?
#1					
#2					
#3					

Table C.1: Risk form to be filled in

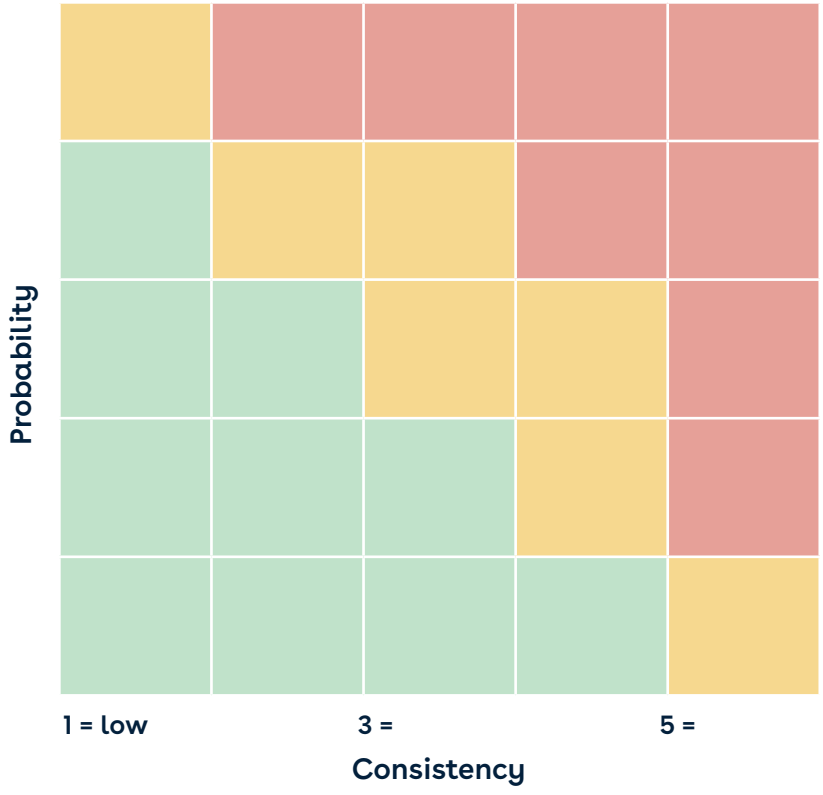


Figure C.2: Risk matrix

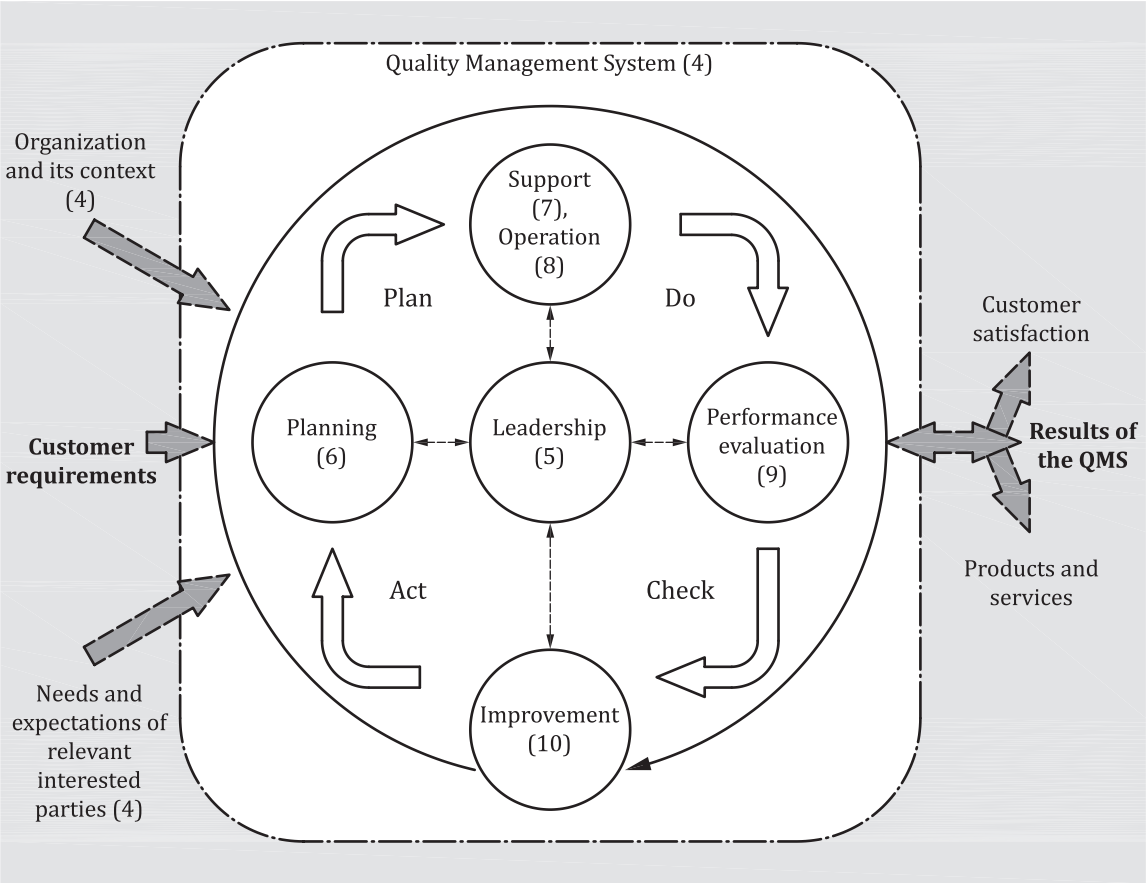
Annex D

(informative)

D.1 The Plan-Do-Check-Act (PDCA) method from DS/EN ISO 9001:2015, *Quality management systems – Requirements*

The PDCA cycle can be applied to all processes and to a quality management system. The figure illustrates how the different activities can be grouped in relation to the PDCA cycle.

Figure D.1 – The Plan-Do-Check-Act cycle (Figure 2 in DS/EN ISO 9001:2015 Quality management systems – Requirements)



The PDCA cycle can be briefly described as follows:

- **Plan**: Establish the objectives of the system and its processes
- **Do**: Implement what was planned
- **Check**: Monitor and (where applicable) measure processes and the resulting products and services against objectives and requirements
- **Act**: Take action to improve performance, as necessary.

The model is based on risk-based thinking. The concept of risk-based thinking is based on carrying out preventive action to eliminate potential nonconformities, analysing nonconformities, and taking action to prevent recurrence that is appropriate for the effects of the nonconformity.

Annex E

(informative)

E.1 Overview of checklist questions

PHASE (ID)	QUESTION (ID)
Idea phase (A)	• Which groups in the decision domain of the system are relevant in terms of bias? (A1) • What legislation and which human rights (see Annex A) are relevant in this context? (A2) • Have the groups (defined in A1) expressed values, opinions, or wishes that should be considered? (A2) • Is the IT system accessible for those with special needs or disabilities? (A3) • Are there other groups that risk being excluded? (A3) • How is the IT system secured against cyberattacks that could affect data quality? (A4)
Analysis phase (B)	• Which groups in the decision domain of the system are relevant in relation to bias? (May have been answered in A1) (B1) • What human cognitive biases may be present in the system, and do they have a negative/unwanted, neutral, or positive/wanted effect? (B2) • In the case of negative/unwanted biases, has a risk analysis and impact assessment been conducted? (B2) • What data biases may be present in the system, and do they have a negative/unwanted, neutral, or positive/wanted effect? (B3) • Are the data representative of the relevant groups, such as demographic parameters like gender, income, origin, and age? (B3) • In the case of negative/unwanted biases, has a risk analysis and impact assessment been conducted? (B3) • How stable are the assumptions and data the system is based on, and what is the risk of data drift? (B4) • Has a plan been developed for managing bias? (B5) • Does this plan include updating the audit log for all relevant phases of the system's lifecycle? (B5) • Can the system be developed to be robust, reliable, and secure, and can it be monitored and audited? (B6)

Implementation phase (C)	• Is there sufficient transparency in the design of the system so that bias can be identified, audited, and monitored? (C1) • Is there sufficient training and awareness of bias in the relevant groups (see A1)? (C2) • Has a procedure been established to remove potential unwanted bias? (C3) • Is the documentation regarding bias sufficiently transparent? (C4) • Can any obligations related to the use of data be documented? (C4)
Realisation phase (D)	• Has a procedure been established for monitoring and identifying unwanted bias? (D1) • Has a procedure been established for managing unwanted bias? (D2) • Is there monitoring and ongoing control of the system's compliance with relevant legislation (see Annex A), human rights, democratic values, and diversity issues? (D3) • Are there automatic mechanisms integrated to handle potential changes in bias (e.g. concept drift)? (D4) • Is the audit log maintained correctly (including for runtime changes)? (D5)
Completion phase (E)	(No questions)

Bibliography

European Convention on Human Rights, www.retsinformation.dk/eli/lta/1996/423

The Charter of Fundamental Rights of the European Union, https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:da:PDF

The Gender Equality Act www.retsinformation.dk/eli/lta/2011/645

The Act on the Prohibition of Differences of Treatment in the Labour Market, www.retsinformation.dk/eli/lta/2017/1001

The Act on Ethnic Equal Treatment www.retsinformation.dk/eli/lta/2012/438

The Act on Prohibition of Differences of Treatment on the grounds of Disability www.retsinformation.dk/eli/lta/2018/688

DS/ISO/IEC TR 24027, *Information technology – Artificial intelligence (AI) – Bias in AI systems and AI-aided decision making*

DS/ISO/IEC 23894, *Information technology – Artificial intelligence – Guidance on risk management*

DS/ISO/IEC TR 24368, *Information technology – Artificial intelligence – Overview of ethical and societal concerns*

DS/ISO/IEC 5338, *Information technology – Artificial intelligence – AI system lifecycle processes*

DS 5001, *Diversity and equality management systems – Requirements and guidelines*

DS/PAS 2500-1:2020, *Artificial intelligence – Part 1: Transparency*

DS/PAS 2500-2:2020, *Artificial intelligence – Part 2: Decision support application in public case management*

DS/EN IEC 31010:2019, *Risk management – Risk assessment techniques*

DS/EN ISO 9001:2015, *Quality management systems – Requirements*

As artificial intelligence continues to develop and spread, attention has been drawn to the unwanted biases that can find their way into IT systems. These biases, when applied in specific IT systems, can impact various population groups based on factors such as gender, colour of skin, age, disability, social class, and more.

The specification provides guidance on how to create transparency about these biases and their potential impacts. From there it is the responsibility of the reader to determine which biases are considered wanted, unwanted, or neutral.

The goal is for the specification to be widely adopted and reach a diverse audience, including IT developers, legal professionals, HR departments, internal decision-makers, as well as purchasers, public authorities, and citizens to raise general awareness that artificial intelligence will almost inevitably reflect biases that result from human bias in the development of IT systems. As a result, the focus should be on identifying and managing unwanted biases rather than attempting to eliminate bias entirely.

NORDIC SWAN ECOLABEL
Printed matter
5041 0968