DANSK STANDARD

DS/PAS 2500-3:2023

Kunstig intelligens
Del 3: Bias

© Danish Standards Foundation

DS/PAS 2500-3:2023, *Kunstig intelligens – Del 3: Bias*

© Dansk Standard 2023
Kopiering ikke tilladt uden særlig tilladelse

DS/PAS 2500-3:2023
Projektnummer: M360243
Grafisk tilrettelæggelse: Maja Bille – grafisk design
Tryk: Dansk Standard
Udgivet 2023
1. udgave, 1. oplag

Udgivet af Dansk Standard
Göteborg Plads 1
2150 Nordhavn
Telefon: 39 96 61 01
ds@ds.dk
www.ds.dk

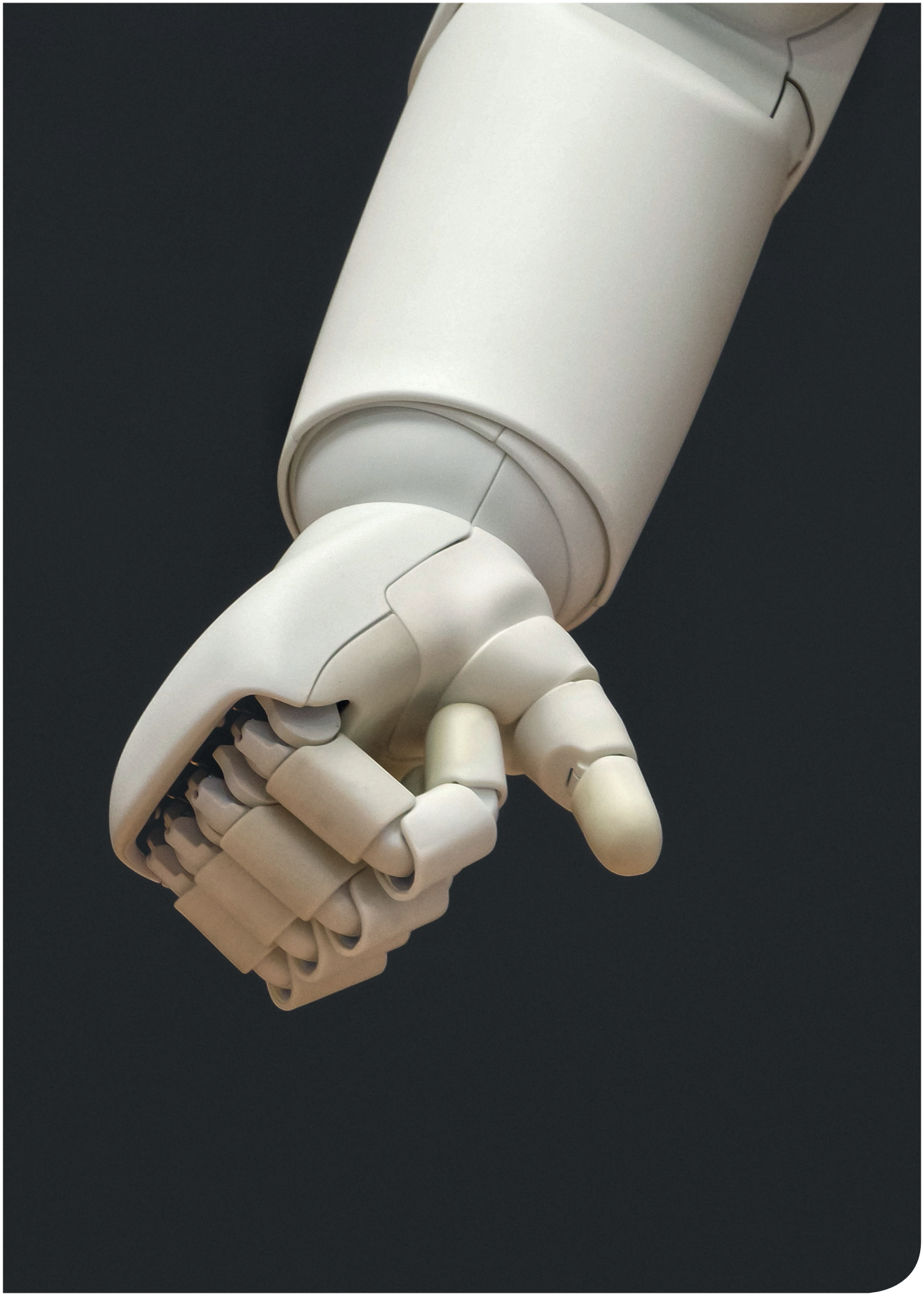
Dette er en POD-publikation
Printet i Danmark

© Danish Standards Foundation

Indholdsfortegnelse

Forord 5
Introduktion 7
1 Anvendelsesområde 8
2 Normative referencer 8
3 Termer og definitioner 9
4 Bias 12
4.1 Introduktion til bias 12
4.2 Menneskelig bias 13
4.3 Databias 14
5 Håndtering af bias i forbindelse med anvendelsen af kunstig intelligens 15
5.1 Overvejelser 15
5.2 Livscyklusmodel 15
5.3 Interessentgrupper og fokusområder 18
5.4 Overvejelser i forbindelse med de forskellige faser i IT-projektet 20
Anneks A 34
Anneks B 36
Anneks C 40
Anneks D 41
Anneks E 42
Bibliografi 44

© Danish Standards Foundation



© Danish Standards Foundation

Forord

DS-publikationstype

Alle betegnelser for publikationer udgivet af Dansk Standard begynder med DS efterfulgt af et eller flere præfikser og et nr., fx DS 282, DS/EN 5414 osv. Hvis publikationen er en del af en serie, vil dette blive benævnt med en bindestreg efterfulgt af nummeret i serien, fx DS/PAS 2500-1 og DS/PAS 2500-2. En serie er en ikke-prioriteret række af publikationer, der kan læses uafhængigt af hinanden, men som omhandler samme tema.

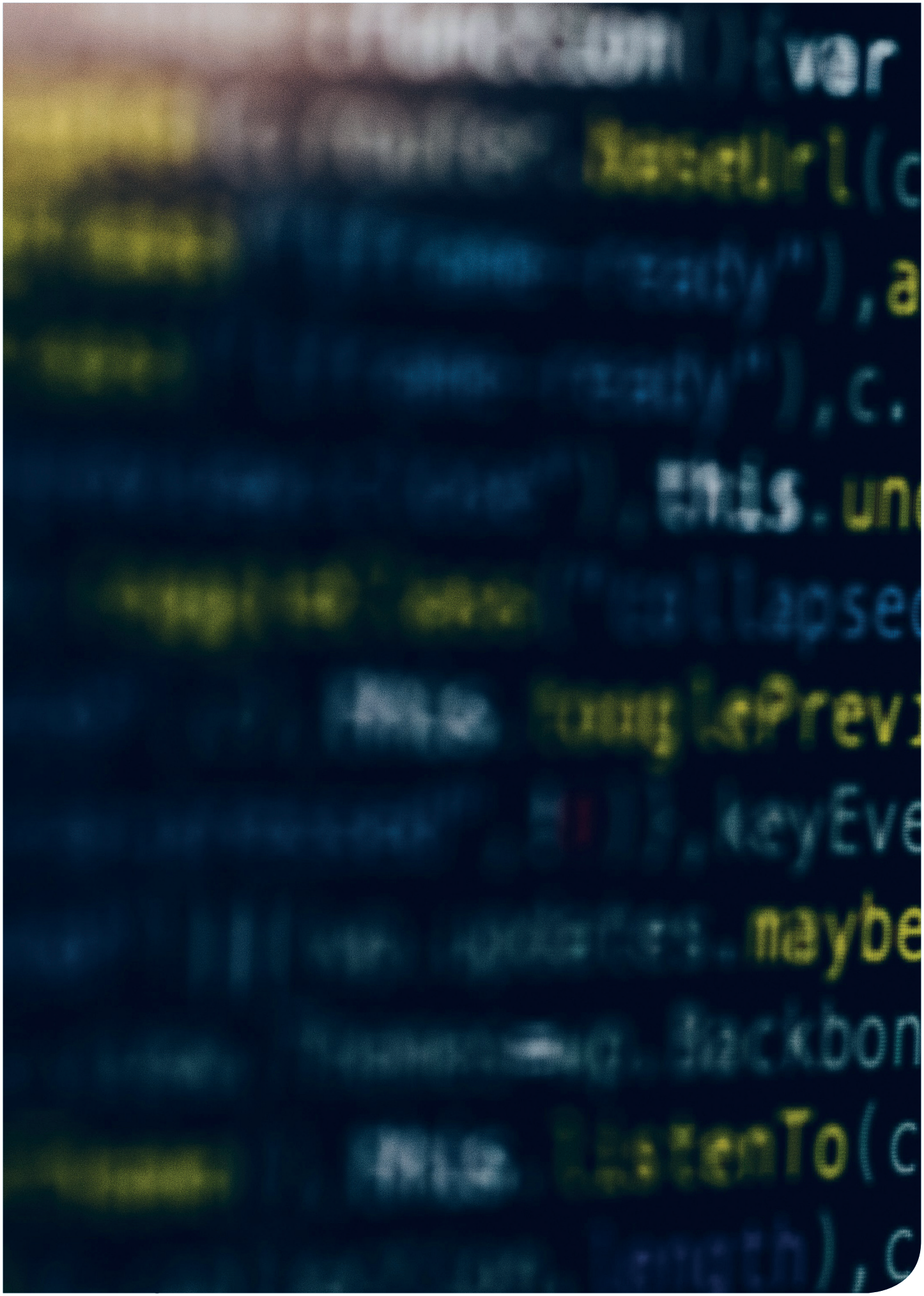
DS/PAS

Dette dokument er en DS/PAS. PAS er en forkortelse for 'Publicly Available Specification', som er en publikation, der er udarbejdet på nationalt niveau og ikke har samme status som en international standard. En PAS adskiller sig fra en international standard ved fx ikke at have de samme krav til graden af interessentinvolvering eller opsætning. En DS/PAS specificerer heller ikke krav, der skal overholdes, men indeholder i stedet anbefalinger, information og gode råd.

Udarbejdelse af publikationen

Denne publikation er blevet til med hjælp fra Datalogisk Institut på Københavns Universitet (DIKU), Piedboeuf og Bureau M, og input fra danske interessenter gennem workshops, interviews og skriftlige kommentarer i overensstemmelse med proceduren for udvikling af PAS-publikationer i Dansk Standard.

© Danish Standards Foundation



© Danish Standards Foundation

Introduktion

I takt med udviklingen og udbredelsen af kunstig intelligens er der kommet en opmærksomhed på de uønskede bias, der kan finde vej ind i IT-systemer. Det er bias, der i anvendelsen af det konkrete IT-system kan have konsekvenser for forskellige befolkningsgrupper alt efter køn, hudfarve, alder, funktionsdygtighed, social klasse, m.m.

De uønskede bias, der kan findes i IT-systemer, er resultatet af det komplekse krydsfelt mellem IT, mennesker og processer - organisatoriske så vel som tekniske - som IT-systemer udvikles i. Denne specifikation omhandler IT-systemer, der udfører eller bidrager til beslutninger vedrørende grupper af mennesker, herunder specifikt IT-systemer baseret på kunstig intelligens. Tilblivelsen af IT-systemer indbefatter også en række menneskelige beslutninger: i alle faser af udviklingsprocessen bliver der taget beslutninger, og i disse menneskelige beslutningsprocesser kan bias komme til udtryk.

Specifikationen har til formål at vejlede i, hvordan man kan skabe gennemsigtighed om disse bias og deres mulige konsekvenser. Herefter er det op til læseren selv at vurdere, hvilke bias der er ønskede, uønskede og neutrale.

Specifikationen vejleder ikke i, hvordan man tager stilling til, hvad der er retfærdigt, men vil i stedet hjælpe læseren med at skabe et grundlag, som der herefter kan tages oplyst stilling til ud fra gældende lovgivning samt egne holdninger, værdier og ønsker om retfærdighed.

Det er håbet, at specifikationen vil få bred anvendelse og finde vej til en bred målgruppe af IT-udviklere, jurister, HR-afdelinger og interne beslutningstagere, såvel som indkøbere, myndigheder og borgere. På den måde kan specifikationen være med til at skabe almen bred bevidsthed om, at en kunstig intelligens næsten uundgåeligt vil indeholde bias, der er et resultat af menneskelig bias i udviklingen af IT-systemet, og det derfor er vigtigere at kunne identificere og håndtere uønsket bias fremfor at prøve helt at undgå bias.

© Danish Standards Foundation

1 Anvendelsesområde

Dette dokument beskriver håndtering af de bias, der findes i samspillet mellem IT-systemer, processer og mennesker. Dette inkluderer

- design og udvikling af IT-systemer og processer
- en generel tilgang til, hvordan bias i denne sammenhæng kan identificeres, og
- en metode til at specificere krav til design og udvikling, der kan bidrage til ansvarlig håndtering af bias.

Dokumentet specificerer anbefalinger til en fremgangsmåde til at identificere bias i forbindelse med design og udvikling af IT-systemer. Tilgangen og metoden kan anvendes både i systemer og processer, som anvendes til automatiske beslutninger eller beslutningsstøtte – herunder kunstig intelligens – og de, som ikke gør.

Målgruppen er kravstillere, designere, udviklere og brugere af IT-systemer, som enten ønsker eller er forpligtet til at identificere og specificere krav vedrørende bias. En sekundær målgruppe er de interessenter, der har gavn af et øget kendskab og opmærksomhed på bias i systemer og processer, fx myndigheder, indkøbere, virksomheder, forbrugere og borgere.

Målet er, at der skabes gennemsigtighed og bevidsthed om bias, ikke at tage stilling til, hvordan disse bias skal håndteres ud fra normative betragtninger om fx retfærdighed.

2 Normative referencer

Der er ingen normative referencer.

© Danish Standards Foundation

3 Termer og definitioner

I dette dokument gælder følgende termer og definitioner.

3.1
algoritme
beskrivelse af en proces til at løse et generelt problem

3.2
kunstig intelligens
systems evne til at tilegne, behandle og anvende *viden* (3.3), og/eller færdigheder til at finde løsninger på problemer, træffe eller støtte beslutninger

Note 1 til term: Kunstig intelligens sammenlignes som regel med menneskelig intelligens, hvilket dog ikke er udtømmende for begrebet.

Note 2 til term: Maskinlæring er en underkategori af kunstig intelligens.

3.3
viden
data, information og færdigheder tilegnet via erfaring eller læring

3.4
beslutning
valg truffet imellem flere muligheder

3.5
beslutningsdomæne
genstandsområde, herunder valgmuligheder, for *beslutning* (3.4), fx tildeling af en social ydelse af offentlig myndighed eller en kredit i banken

3.6
beslutningsstøtte
automatisering af en *beslutning* (3.4) eller levering af information, som et menneske kan benytte til at træffe en *beslutning* (3.4)

Note 1 til term: Der er tre grader af menneskelig involvering i beslutninger.

3.6.1
menneskeinddragende beslutningsstøtte
beslutningsstøtte (3.6), der kræver menneskelig inddragelse, fx ved at levere dele af datagrundlaget for beslutningen eller som ansvarlig for at udføre beslutningen

© Danish Standards Foundation

3.6.2
menneskemonitoreret beslutningsstøtte
beslutningsstøtte (3.6), der opererer uafhængigt af mennesker, men under menneskelig overvågning, med intervention når det er nødvendigt

3.6.3
menneskeuafhængig beslutningsstøtte
beslutningsstøtte (3.6), der operer uden menneskelig inddragelse

3.7
bias
systematisk forskel eller fejl i behandling af data

Note 1 til term: Behandling er enhver form for handling, herunder opfattelse, observation, repræsentation, forudsigelse eller beslutning.

3.8
maskinlæring
statistisk metode eller algoritme, der, på baggrund af eksempeldata, kan finde en model, der kan løse et givet beslutningsproblem til en vis grad

3.9
interessent
individ, gruppe, eller organisation, der kan påvirke, bliver påvirket af, eller opfatter sig selv som påvirket af en beslutning eller en aktivitet

3.10
system
kombination af interagerende elementer, der er organiseret med henblik på at opnå et eller flere angivne formål

3.11
IT-system
system (3.10) baseret på informationsteknologi

3.12
teknisk ekspert
person, der designer, udvikler, tester, verificerer eller auditerer den tekniske funktion af et *IT-system* (3.11) eller elementer der indgår i et *IT-system* (3.11)

3.13
tilsynsførende
person, myndighed eller virksomhed, der fører tilsyn med, at et IT-projekt eller anvendelsen af et *IT-system* (3.11) lever op til givne krav

© Danish Standards Foundation

3.14
databias

dataegenskaber, der, hvis de ikke adresseres, kan føre til et system, der har *bias* (3.7) i forhold til forskellige grupper

3.15
gruppe

delmængde af personer i et domæne, der deler fælles karakteristika

3.16
karakteristika

definerende kendetegn for objekter i et domæne

3.17
menneskelig bias

bias (3.7), der opstår, når mennesker behandler og fortolker information

3.18
biaseffekt

resultatet af *bias* (3.7) i et system, der kan føre til forskel i behandling mellem specifikke personer og grupper i forhold til andre

Note 1 til term: En biaseffekt kan være enten positiv/ønsket, neutral eller negativ/uønsket.

3.19
dataskred

resultat af en udvikling i data, der ugyldiggør datamodellen i forhold til *beslutningsdomænet* (3.5)

3.20
konceptskred

ændring over tid på en uforudset måde af en variabel, som systemet prøver at forudsige

© Danish Standards Foundation

4 Bias

4.1 Introduktion til bias

Denne specifikation forstår bias som systematisk forskel i behandling af data, som kan have den effekt, at der sker en forskel i behandlingen af specifikke objekter, personer eller grupper i forhold til andre. Det kan fx være mellem personer af forskellige køn og alder eller mellem personer med forskellig herkomst. En behandling kan være enhver form for handling, herunder opfattelse, observation, klassifikation, repræsentation, forudsigelse eller beslutning. Enhver form for ikke-triviel behandling vil indeholde bias, da behandlingen ellers vil være uafhængig af det objekt, der behandles. Det er derfor relevant at tale om, hvilken effekt den pågældende bias har, og om denne effekt er positiv (ønsket), om den er negativ (uønsket) eller om den er neutral.

I daglig tale og forståelse er forskelsbehandling et normativt begreb, idet det oftest bruges til at betegne en diskriminerende praksis, der stiller nogen eller nogle dårligere end andre. I denne specifikation anvendes i stedet "forskel i behandling" som et neutralt begreb, der betegner, om et IT-systems er designet til - intenderet eller ikke-intenderet - at behandle forskellige grupper forskelligt. En tolkning af forskellen i behandling som værende diskriminerende finder således kun anvendelse i de tilfælde, dette ikke er intenderet eller formålstjenstligt. Et eksempel herpå er et system, der anvender en prædiktiv model, hvor gruppetilhørsforhold optræder som parameter i funktionen, fx således at forudsigelsen matematisk justeres, afhængigt af om den pågældende person er mand eller kvinde.

Kunstig intelligens er i denne specifikation karakteriseret ved et systems evner til at behandle og anvende viden, og/eller færdigheder til at finde løsninger på problemer, træffe eller støtte beslutninger. Forskellige typer af metoder til kunstig intelligens, såsom brug af maskinlæring, regel-baserede systemer og statistiske modeller, vil anvende forskellige typer viden og data, og også have forskellige måder at behandle og anvende denne viden på, og dermed også forskellige måder hvorpå bias introduceres og identificeres. Maskinlæringsmetoder baseres typisk på eksempler på løsninger på et problem eller beslutning. Regelbaserede systemer kan også være baseret på en bagvedlæggende læring fra eksempler, men kan også være baseret på ekspertviden, der kan være opnået på mange forskellige måder.

Endelig baserer statistiske modeller sig typisk på observationer og antagelser om en matematisk model, der i tilstrækkelig grad forklarer disse observationer. Det er uden for anvendelsesområdet for denne specifikation at gå i detaljer med metodespecifikke kilder til bias, men det har betydning for, hvordan de enkelte spørgsmål, som præsenteres senere i dokumentet, besvares. Fx vil der være tale om forskellige former for viden eller typer af data, der ligger til grund for henholdsvis maskinlærings-, regelbaserede og statistiske modeller.

Det er dog relevant at skelne imellem menneskelig (kognitiv) bias og databias, herunder statistisk bias.

© Danish Standards Foundation

4.2 Menneskelig bias

4.2.1 Generelt

Menneskelig (kognitiv) bias er den bias, der er forbundet med den menneskelige behandling af data, herunder udvælgelse eller beslutningstagen. Der findes mange former for menneskelig bias, og de kan introduceres i alle faser af udviklingen og brugen af et IT-system. Eksempler på menneskelig bias kan være automatiseringsbias, gruppetilskrivningsbias, bekræftelsesbias, tidsperiodebias, gruppebias og samfundskontekstuel bias. Specifikt i forbindelse med udviklingen af IT-systemer træffes der en række valg, der indeholder bias, såsom valg af parametre, valg af metode, og afhængigt af metode, valg af algoritme, inferensmetode eller statistisk model.

4.2.2 Automatiseringsbias

Automatiseringsbias er den type bias, hvor man i forbindelse med beslutningsstøtte med menneskelig inddragelse eller overvågning ukritisk accepterer en handling, et valg eller en beslutning foretaget eller foreslået af et automatisk system.

4.2.3 Gruppetilskrivningsbias

Gruppetilskrivningsbias finder sted, når et menneske antager, at en egenskab ved et enkelt individ eller element i en gruppe gælder for hele gruppen. Det kunne fx være en teenager, der antager, at alle forældre er dumme, eller en forælder, der antager, at alle teenagere er dovne.

4.2.4 Bekræftelsesbias

Bekræftelsesbias referer til, når mennesker foretrækker valg, information eller beslutninger, der bekræfter deres opfattelse af, hvad der er rigtigt.

4.2.5 Tidsperiodebias

Tidsperiodebias optræder som resultat af, at systemet er baseret på viden eller data fra en begrænset tidsperiode, der ikke er repræsentativ for tidspunktet, hvor systemet anvendes.

4.2.6 Gruppebias

Et eksempel på gruppebias optræder, når et menneske behandler den gruppe, som vedkommende tilhører, anderledes end andre grupper, ofte ved at tildele gruppen større autoritet. En anden type gruppebias optræder, når et menneske opfatter en gruppe (ofte en gruppe, som vedkommende ikke tilhører) som mere homogen, end den i virkeligheden er, og dermed risikerer at behandle medlemmer af gruppen ensartet i tilfælde, hvor der burde gøres forskel.

4.2.7 Samfundskontekstuel bias

Samfundskontekstuel bias betegner menneskelig bias, som er særlig udbredt i et samfund og som derfor kan blive indlejret i data, organisationer/institutioner, virksomhedsregler og lovgivning; eksempelvis stereotyper. Det kan fx være en forestilling om, at arbejdsløshed i høj grad er selvforskyldt og give sig udtryk i lovgivning, der tjener til at overvåge og aktivere arbejdsløse på særlige måder.

© Danish Standards Foundation

4.3 Databias

4.3.1 Generelt

Databias er egenskaber ved data anvendt i systemer, der vil føre til bias i systemernes behandling af specifikke grupper i forhold til andre, hvis ikke der tages højde for disse egenskaber. Databias er ofte et resultat af menneskelig bias i forbindelse med udvælgelse og behandling af den data og viden, der ligger til grund for udviklingen af systemet.
Bias i udvælgelse finder generelt sted, når data ikke er udvalgt repræsentativt for den reelle fordeling af data. Eksempler på sådanne bias kan være selektionsbias, dækningsbias og deltagelsesbias. Bias i behandling kan finde sted i forbindelse med inddeling (labelling) af den data, der danner grundlag for udvikling af systemet i kategorier der ikke er dækkende for variationen i samfundet, fx ved inddeling af personer i mænd og kvinder eller i politisk blå eller røde.

4.3.2 Selektionsbias

Selektionsbias optræder, når data ikke er indsamlet tilfældigt, hvilket for eksempel kan ske ved at respondenter på en undersøgelse er udvalgt blandt egne bekendte eller brugere af et specifikt socialt medie.

4.3.3 Dækningsbias

Dækningsbias optræder, når populationen i datasættet ikke svarer til den population, man ønsker at træffe beslutninger for. Det kunne fx være, at et IT-system til ansigtsgenkendelse, der kun er trænet på hvide ansigter, skal anvendes på alle typer af ansigter.

4.3.4 Deltagelsesbias

Deltagelsesbias opstår, når respondenterne i en undersøgelse ikke er repræsentative i forhold til den gruppe, man gerne vil undersøge. Det kan fx være en onlineundersøgelse, hvor man ikke får deltagere med, der har fravalgt at bruge en pc.

© Danish Standards Foundation

5 Håndtering af bias i forbindelse med anvendelsen af kunstig intelligens

5.1 Overvejelser

Det er nødvendigt at overveje, hvornår det er etisk ansvarligt at anvende et IT-system, der benytter kunstig intelligens, eller hvornår det er etisk ansvarligt, at mennesker løser opgaven. Disse overvejelser bør indgå i alle faser af udviklingen af IT-systemet og processer og inddrage alle relevante interessenter. Overvejelserne skal dokumenteres i en grad, der muliggør, at ansvarlige for projektet og de enkelte faser og aktiviteter kan holdes til ansvar for udførsel af aktiviteter og igangsættelse af efterfølgende faser.

5.2 Livscyklusmodel

5.2.1 Generelt

I dette dokument anvendes som udgangspunkt de overordnede faser i statens IT-projektmodel[1]: Idé, Analyse, Gennemførelse, Realisering. I dette dokument er desuden tilføjet fasen "Afslutning". De enkelte faser indeholder mange aktiviteter, som angivet i figur 1.

Faserne kan udføres strengt sekventielt som i en såkaldt vandfaldsmodel, der følger rækkefølgen angivet i figuren, eller overlappe og gentages som i såkaldte iterative eller agile modeller. Organisationen, der er ansvarlig for projektet (eller organisationerne, der er ansvarlige for de enkelte faser), bør udpege en eller flere personer med ansvar for at sikre, at overvejelserne i denne specifikations tjekliste (5.4) indgår i beslutningen om at igangsætte aktiviteter og efterfølgende faser.

Figur 1. Faser i statens IT-projektmodel og aktiviteter, de som minimum bør indeholde, tilføjet er fasen "Afslutning".

[1] https://digst.dk/styring/projektstyring/statens-it-projektmodel/

© Danish Standards Foundation

5.2.2 A: Idéfasen

Denne fase bør som minimum indeholde identifikation af ønskede gevinster, mulige løsninger og mulige risici samt håndtering af risici.

5.2.3 B: Analysefasen

Denne fase bør som minimum indeholde validering af løsninger i forhold til ønskede gevinster og risici og specifikation af krav.

5.2.4 C: Gennemførelsesfasen

Denne fase indeholder design, udvikling, integration og test.

5.2.5 D: Realiseringsfasen

Denne fase indeholder overvågning, evaluering, vedligehold af systemet eller dele af systemet, herunder personhenførbare data, der ikke længere er nødvendige.

5.2.6 E: Afslutningsfasen

Ved afskaffelse af et system er der nogle bestemte krav, som bringes i spil. Disse krav drejer sig primært om, hvorvidt eksisterende data håndteres korrekt (typisk i forhold til lovgivning).

© Danish Standards Foundation

© Danish Standards Foundation

5.3 Interessentgrupper og fokusområder

Forskellige grupper og interessenter har vidt forskellige behov og forventninger til et IT-system. Derfor tager denne specifikation udgangspunkt i en række forskellige grupper og deres umiddelbare behov. Heri er der allerede indbygget en bias. Men det er en ønsket bias, der skal hjælpe med at strukturere arbejdet.

Tjeklisterne tager udgangspunkt i følgende interessentgrupper:

- Projektejer/kravstiller
- Slutbrugere
- Borgere (både indirekte og direkte berørte)
- Leverandører
- Udviklere
- Retsligt personale/tilsynsførende/uvildige tekniske eksperter.

På næste side findes en oversigt over interessentgrupper og deres typiske fokusområder (tabel 1). I afsnit 5.4 findes tjeklister med spørgsmål, der er relevante i forhold til interessentgrupperne og i forhold til dokumentationen af hver fase i udviklingen af et system, der anvender kunstig intelligens.

© Danish Standards Foundation

INTERESSENT-GRUPPER	BESKRIVELSE	FOKUSOMRÅDE FOR INTERESSENTGRUPPEN
Projektejer/ kravstiller	Den organisation, afdeling eller person, som har ansvaret for at udvikle og drifte et IT-system. *Eksempelvis teknisk eller organisatorisk projektleder.*	Identifikation af bias knyttet til idé, analyse, design, kravspecifikation og den afledte effekt af systemets anvendelse.
Slutbrugere	En bruger, der har direkte interaktion med systemet. *Et eksempel kunne være en forsikringsrådgiver, der behandler en sag ved hjælp af systemet, eller en borger, som benytter sig af en online selvhjælpstjeneste.*	Identifikation af bias knyttet til idefase/ design (involvering), brug af IT-systemet og specifikke udfald af systemets anvendelse.
Borgere – direkte berørte	Personer, som ikke har direkte interaktion med systemet, men som bliver berørt eller opfatter sig som berørt af effekter af systemets anvendelse. *Et eksempel kunne være kunder, hvis forsikringsvilkår ændres på baggrund af andre kunders sager, afgjort ved hjælp af et IT-system.*	Identifikation af bias i den afledte effekt af systemets anvendelse.
Borgere – indirekte berørte	Personer, som kan blive indirekte berørt af systemets effekter, eller hvis systemet bliver anvendt i en kontekst, som har befolkningens generelle interesse. *Et eksempel kunne være, at de generelle forsikringsvilkår i hele branchen ændres på baggrund af brugen af systemet.*	Identifikation og en generel præcision af bias knyttet til eksempelvis befolkningsgrupper.
Leverandører	Virksomheder eller personer, som fungerer som leverandører af dele af et IT-system eller hele systemet. Dette drejer sig også om dataleverandører. Leverandørerne kan optræde i alle faser af et projekt (se figur 1). *Dette kan eksempelvis være en leverandør af analysekomponenter til estimering af forsikringsrisici eller en leverandør af data til et byggesagssystem.*	Identifikation af bias i forbindelse med udvikling og vedligeholdelse af systemet.

Fortsættes på næste side

© Danish Standards Foundation

Udviklere	Personer, som står for udviklingen af et IT-system eller dettes komponenter. Dette drejer sig om både interne og eksterne udviklere. *Eksempelvis kan det være den projektejende virksomheds egen IT-afdeling, midlertidige konsulenter eller udviklere hos en underleverandør.*	Identifikation af bias knyttet til data og algoritmer, herunder vægtning af data.
Retsligt personale/ tilsynsførende/ uvildige tekniske eksperter	Personer, som kontrollerer overensstemmelse i forhold til standarder, fører tilsyn med, at sagsbehandling udmøntes korrekt, eller afgør kvalitet. *Dette kan eksempelvis være certificering af forsikringssystemer, overvågning af offentlige sagsbehandlingssystemer eller efterkontrol i forhold til kvalitet.*	Bias, der har indflydelse på mulighed for monitorering og godkendelse af systemet. Håndtering af bias i forbindelse med audit og dokumentation.

Tabel 1 – Interessentgrupper og fokusområder

5.4 Overvejelser i forbindelse med de forskellige faser i IT-projektet

5.4.1 Generelt

Tjeklisterne for de forskellige faser er en systematisk måde at sikre, at de forskellige interessentgrupper (tabel 1) tænkes ind i IT-projektet, og at overvejelserne dokumenteres. Tjeklisterne er således et internt værktøj, som kan anvendes af projektgruppen.

I Spørgsmål-kolonnen stilles et spørgsmål, og overvejelserne noteres i skemaet. I de fleste tilfælde er det relevant at inddrage interessenten i overvejelserne, men det er ikke hensigten, at interessenterne selv skal skrive svaret ind i tjeklisten. Bemærk, at der for et konkret system kan være interessenter, der ikke er relevante, samt interessenter, der ikke er medtaget nedenfor. Tilsvarende kan organisationer, der repræsenterer interessenter, som fx interesseorganisationer, inddrages, i stedet for medlemmer af denne gruppe.

I Handling-kolonnen noteres den nødvendige handling, som overvejelserne i Svar-kolonnen giver anledning til. I den forbindelse kan man med fordel anvende modeller til at vurdere risici og organisere handlinger, fx risicimodellen (se anneks C) samt Plan-Do-Check-Act (PDCA)-metoden (se anneks D). Disse metoder vil medvirke til at minimere "brandslukning". Brandslukning er, når problemløsning sker via hurtige løsninger og fokus er på symptombehandling.

Bemærk, at tjeklisterne ikke behøver at blive udfyldt i den givne rækkefølge, og derfor medtages enkelte spørgsmål i flere faser. Tjeklisterne tager heller ikke udgangspunkt i specifikke brancher eller bestemte typer af systemer, der anvender kunstig intelligens.

© Danish Standards Foundation

5.4.2 Fase A: Idéfasen

Organisationen med ansvar for idéfasen bør som minimum overveje og dokumentere svarene på spørgsmålene i A1-A4 for hver tænkt anvendelse, med inddragelse af relevante interessenter eller repræsentanter for disse.

ID	SPØRGSMÅL	HANDLING
A1	**Hvilke grupper i beslutningsdomænet for systemet er relevante i forhold til bias?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
A1-1: Projektejer/ kravstiller		
A1-2: Slutbrugere		
A1-3: Borgere (både indirekte og direkte berørte)		
A1-4: Leverandører		
A1-5: Udviklere		
A1-6: Retsligt personale/tilsynsførende/uvildige tekniske eksperter		

Fortsættes på næste side

© Danish Standards Foundation

A2	**Hvilke love og menneskerettigheder (se anneks A) er relevante i denne sammenhæng?** **Har grupperne (defineret i A1) givet udtryk for særlige værdier, holdninger eller ønsker, der bør tages hensyn til?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
A2-1: Projektejer/ kravstiller		
A2-2: Slutbrugere		
A2-3: Borgere (både indirekte og direkte berørte)		
A2-4: Leverandører		
A2-5: Udviklere		
A2-6: Retsligt personale/tilsynsførende/uvildige tekniske eksperter		
A3	**Er dit IT-system brugbart for dem med særlige behov eller handicap? Er der andre grupper, der risikerer at blive udelukket?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
A3-1: Projektejer/ kravstiller		
A3-2: Slutbrugere		
A3-3: Borgere (både indirekte og direkte berørte)		
A3-4: Leverandører		
A3-5: Udviklere		
A3-6: Retsligt personale/tilsynsførende/uvildige tekniske eksperter		

© Danish Standards Foundation

A4	**Hvordan er dit IT-system sikret mod cyberangreb, som kan påvirke datakvaliteten?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
A4-1: Projektejer/ kravstiller		
A4-2: Slutbrugere	(sandsynligvis ikke relevant)	
A4-3: Borgere (både indirekte og direkte berørte)	(sandsynligvis ikke relevant)	
A4-4: Leverandører		
A4-5: Udviklere		
A4-6: Retsligt personale/tilsyns-førende/uvildige tekniske eksperter		

Tabel 2 – Tjekliste for idéfasen

© Danish Standards Foundation

5.4.3 Fase B: Analysefasen

Organisationen med ansvar for analysefasen bør som minimum overveje og dokumentere svarene på spørgsmålene B1-B6, med inddragelse af relevante interessenter eller repræsentanter for disse.

ID	SPØRGSMÅL	HANDLING
B1	**Hvilke grupper i beslutnings-domænet for systemet er relevante i forhold til bias? (Er evt. tidligere besvaret i A1)**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
B1-1: Projektejer/ kravstille		
B1-2: Slutbrugere		
B1-3: Borgere (både indirekte og direkte berørte)		
B1-4: Leverandører		
B1-5: Udviklere		
B1-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		
B2	**Hvilke menneskelige, kognitive bias kan der være i systemet, og har de negativ/ uønsket, neutral eller positive/ ønsket effekt?** **I tilfælde af negative/ uønskede bias, er der lavet en risikoanalyse og konsekvensanalyse?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
B2-1: Projektejer/ kravstiller		

© Danish Standards Foundation

B2-2: Slutbrugere		
B2-3: Borgere (både indirekte og direkte berørte)		
B2-4: Leverandører		
B2-5: Udviklere		
B2-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		
B3	**Hvilke databias kan der være i systemet, og har de negativ/uønsket, neutral eller positiv/ønsket effekt?** **Er data repræsentative for de relevante grupper, som fx de-mografiske parametre som køn, indkomst, herkomst og alder?** **I tilfælde af negative/uønskede bias, er der lavet en risikoana-lyse og konsekvensanalyse?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
B3-1: Projektejer/kravstiller		
B3-2: Slutbrugere		
B3-3: Borgere (både indirekte og direkte berørte)		
B3-4: Leverandører		
B3-5: Udviklere		
B3-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		

Fortsættes på næste side

© Danish Standards Foundation

B4	**Hvor stabile er de antagelser og data, systemet bygger på, og hvor stor er risikoen for dataskred?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
B4-1: Projektejer/ kravstiller		
B4-2: Slutbrugere	(sandsynligvis ikke relevant)	
B4-3: Borgere (både indirekte og direkte berørte)	(sandsynligvis ikke relevant)	
B4-4: Leverandører		
B4-5: Udviklere		
B4-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		
B5	**Er der udviklet en plan for håndtering af bias?** **Indeholder denne plan opdatering af auditlog for alle relevante dele af systemets livscyklus?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
B5-1: Projektejer/ kravstiller		
B5-2: Slutbrugere	(sandsynligvis ikke relevant)	
B5-3: Borgere (både indirekte og direkte berørte)	(sandsynligvis ikke relevant)	
B5-4: Leverandører		
B5-5: Udviklere		
B5-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		

© Danish Standards Foundation

B6	**Kan systemet udvikles, så det bliver robust, pålideligt og sikkert, og kan det overvåges og auditeres?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
B6-1: Projektejer/ kravstiller		
B6-2: Slutbrugere	(sandsynligvis ikke relevant)	
B6-3: Borgere (både indirekte og direkte berørte)	(sandsynligvis ikke relevant)	
B6-4: Leverandører		
B6-5: Udviklere		
B6-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		

Tabel 3 – Tjekliste for analysefasen

© Danish Standards Foundation

5.4.4 Fase C: Gennemførelsesfasen

Organisationen med ansvar for gennemførelsesfasen bør som minimum overveje og dokumentere svarene på spørgsmålene C1-C4, med inddragelse af relevante interessenter eller repræsentanter for disse.

C1	**Er der tilstrækkelig gennemsigtighed[2] i designet af systemet, så bias kan identificeres, auditeres og overvåges?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
C1-1: Projektejer/ kravstiller		
C1-2: Slutbrugere		
C1-3: Borgere (både indirekte og direkte berørte)		
C1-4: Leverandører		
C1-5: Udviklere		
C1-6: Retsligt personale/tilsynsførende/uvildige tekniske eksperter		
C2	**Er der tilstrækkelig oplæring og bevidsthed om bias i de relevante grupper (se A1)?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
C2-1: Projektejer/ kravstiller		
C2-2: Slutbrugere		
C2-3: Borgere (både indirekte og direkte berørte)		
C2-4: Leverandører		
C2-5: Udviklere		
C2-6: Retsligt personale/tilsynsførende/uvildige tekniske eksperter		

[2] Se evt. DS/PAS 2500-1:2020 *Kunstig intelligens – Del 1: Gennemsigtighed*

© Danish Standards Foundation

C3	**Er der etableret en procedure for at fjerne potentiel uønsket bias?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
C3-1: Projektejer/ kravstiller		
C3-2: Slutbrugere	(sandsynligvis ikke relevant)	
C3-3: Borgere (både indirekte og direkte berørte)	(sandsynligvis ikke relevant)	
C3-4: Leverandører		
C3-5: Udviklere		
C3-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		
C4	**Er dokumentationen mht. bias tilstrækkelig gennemsigtig?** **Kan eventuelle forpligtigelser i forhold til benyttelse af data dokumenteres?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	Handling (udfyld felter)
C4-1: Projektejer/ kravstiller		
C4-2: Slutbrugere		
C4-3: Borgere (både indirekte og direkte berørte)		
C4-4: Leverandører		
C4-5: Udviklere		
C4-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		

Tabel 4 – Tjekliste for gennemførelsesfasen

© Danish Standards Foundation

5.4.5 Fase D: Realiseringsfasen

Organisationen med ansvar for realiseringsfasen bør som minimum overveje og dokumentere svarene på spørgsmålene D1-D5, med inddragelse af relevante interessenter eller repræsentanter for disse.

ID	SPØRGSMÅL	HANDLING
D1	**Er der etableret en procedure for overvågning og identificering af uønsket bias?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
D1-1: Projektejer/ kravstiller		
D1-2: Slutbrugere		
D1-3: Borgere (både indirekte og direkte berørte)		
D1-4: Leverandører		
D1-5: Udviklere		
D1-6: Retsligt personale/tilsynsførende/uvildige tekniske eksperter		

© Danish Standards Foundation

D2	**Er der etableret en procedure for håndtering af uønsket bias?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
D2-1: Projektejer/ kravstiller		
D2-2: Slutbrugere		
D2-3: Borgere (både indirekte og direkte berørte)		
D2-4: Leverandører		
D2-5: Udviklere		
D2-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		
D3	**Er der monitorering og løbende kontrol af systemets overholdelse af relevante love (se anneks A), menneskerettig-heder, demokratiske værdier og diversitetsspørgsmål?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
D3-1: Projektejer/ kravstiller		
D3-2: Slutbrugere		
D3-3: Borgere (både indirekte og direkte berørte)		
D3-4: Leverandører		
D3-5: Udviklere		
D3-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		

Fortsættes på næste side

© Danish Standards Foundation

D4	**Er der indbygget automatiske mekanismer til håndtering af eventuelle ændringer i bias (fx konceptskred)?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
D4-1: Projektejer/ kravstiller		
D4-2: Slutbrugere	(sandsynligvis ikke relevant)	
D4-3: Borgere (både indirekte og direkte berørte)	(sandsynligvis ikke relevant)	
D4-4: Leverandører		
D4-5: Udviklere		
D4-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		
D5	**Er auditloggen vedligeholdt korrekt (også for runtime-ændringer)?**	[Anvend gerne risicimodel og/eller PDCA-metoden beskrevet i anneks C og D]
	Svar (udfyld felter)	**Handling** (udfyld felter)
D5-1: Projektejer/ kravstiller		
D5-2: Slutbrugere	(sandsynligvis ikke relevant)	
D5-3: Borgere (både indirekte og direkte berørte)	(sandsynligvis ikke relevant)	
D5-4: Leverandører		
D5-5: Udviklere		
D5-6: Retsligt personale/tilsynsfø-rende/uvildige tekniske eksperter		

Tabel 5 – Tjekliste for realiseringsfasen

© Danish Standards Foundation

5.4.6 Fase E: Afskaffelsesfasen

Ved afskaffelse af et system er der nogle bestemte krav, som bringes i spil. Disse krav drejer sig primært om, hvorvidt eksisterende data afskaffes korrekt (typisk i forhold til lovgivning). Dette punkt indeholder ikke en eksplicit tjekliste, da denne specifikation ikke forholder sig til den til enhver tid gældende lovgivning. Aftager og modtager opfordres til at konstruere formålstjenlige tjeklister allerede i idéfasen.

© Danish Standards Foundation

Anneks A

(informativt)

Overblik over relevant lovgivning

A.1 Relevant lovgivning

A.1.1 Generelt

Når man behandler bias i forhold til demografiske parametre som fx køn, etnicitet, og alder, skal der være en opmærksomhed på den relevante lovgivning, blandt andet EU-retten samt international menneskeret.

A.1.2 International lovgivning

Den Europæiske Menneskerettighedskonvention sikrer et forbud mod diskrimination på baggrund af køn, race, farve, sprog, religion, politisk eller anden overbevisning, national eller social oprindelse, tilhørighed til et nationalt mindretal, formueforhold, fødsel eller ethvert andet forhold. Den Europæiske Menneskerettighedskonvention er inkorporeret i dansk lovgivning, hvilket betyder, at den er en del af dansk ret.

EU-retten indeholder i traktaterne og EU's Charter om Grundlæggende Rettigheder et ligebehandlingsprincip og et diskriminationsforbud. Ud over de generelle regler og principper er der vedtaget en række direktiver, der er implementeret i dansk ret, der beskytter imod diskrimination i særlige situationer. Direktiverne beskytter imod diskrimination på baggrund af race og etnisk oprindelse, religion eller tro, handicap, alder, seksuel orientering og køn (1).

A.1.3 National lovgivning

Den relevante nationale lovgivning vil i denne sammenhæng ofte – men ikke kun - være:

- Ligebehandlingsloven
- Forskelsbehandlingsloven
- Ligestillingsloven
- Lov om etnisk ligebehandling
- Lov om forbud mod diskrimination på grund af handicap.

Er der tale om et IT-system, hvis beslutningsdomæne relaterer sig til adgang til og/eller muligheder på arbejdsmarkedet, er det ligebehandlingsloven og forskelsbehandlingsloven, man med fordel kan orientere sig i. Det kunne fx være rekrutterings- eller performance management-systemer.

Ligebehandlingsloven påbyder, at arbejdsgivere skal behandle alle uanset køn, kønsidentitet, kønsudtryk og kønskarakteristika lige ved ansættelse, forfremmelse og forflyttelse. Det er også med ligebehandlingsloven man sikrer, at kvinder behandles lige på arbejdsmarkedet under graviditet, fødsel og barsel.

© Danish Standards Foundation

Forskelsbehandlingsloven forbyder direkte eller indirekte forskelsbehandling på arbejdsmarkedet på baggrund af race, hudfarve, religion eller tro, politisk anskuelse, seksuel orientering, alder, handicap eller national, social eller etnisk oprindelse. Det kan være ved ansættelse, afskedigelse, forflyttelse, forfremmelse eller med hensyn til løn- og arbejdsvilkår.

Til områder, der ikke er direkte eller indirekte knyttet til arbejdsmarkedet, er det **Ligestillingsloven**, der sikrer ligestilling mellem kvinder og mænd, mellem seksuelle minoriteter og majoriteter, og forbyder forskelsbehandling på grund af kønsidentitet, kønsudtryk og kønskarakteristika. Ifølge Ligestillingsloven skal disse grupper behandles lige i offentlig, almen og erhvervsmæssig virksomhed. Fx må den samme vare ikke prissættes forskelligt til mænd og kvinder.

Lov om etnisk ligebehandling og Lov om forbud mod diskrimination på grund af handicap beskytter også mod diskrimination uden for arbejdsmarkedet. Det kunne fx være i forhold til adgang til boliger eller myndigheders tildeling af sociale ydelser.

(1): https://menneskeret.dk/sites/menneskeret.dk/files/media/document/Algoritmer_8.K.pdf

© Danish Standards Foundation

Anneks B

(informativt)

Oversigt over relevante standarder

B.1 Internationale standarder

DS/ISO/IEC TR 24027

Informationsteknologi – Kunstig intelligens (AI) – Bias i AI-systemer og AI-assisteret beslutningstagning

Dette dokument adresserer bias i forhold til AI-systemer, især med hensyn til AI-støttet beslutningstagning. Måleteknikker og metoder til vurdering af bias beskrives med det formål at adressere og behandle biasrelaterede sårbarheder. Alle AI-systemets livscyklusfaser er inkluderet, herunder dataindsamling, træning, løbende læring, design, test, evaluering og brug.

DS/ISO/IEC 23894

Informationsteknologi – Kunstig intelligens – Risikoledelse

Dette dokument giver retningslinjer for styring af risici, som organisationer står over for under udvikling og anvendelse af kunstig intelligens (AI) teknikker og systemer. Retningslinjerne har også til formål at hjælpe organisationer med at integrere risikostyring i deres AI-relaterede aktiviteter og funktioner. Den beskriver desuden processer til effektiv implementering og integration af AI-risikostyring. Anvendelsen af disse retningslinjer kan tilpasses til enhver organisation og dens kontekst.

DS/ISO/IEC TR 24368

Informationsteknologi – Kunstig intelligens – Overblik over etiske og samfundsmæssige hensyn

Dette dokument giver et overblik over etiske og samfundsmæssige hensyn ved anvendelsen af kunstig intelligens (AI). Derudover indeholder dette dokument information i relation til principper, processer og metoder på dette område. Dokumentet er beregnet til teknikere, myndigheder, interessegrupper og samfundet som helhed. Dokumentet er ikke beregnet til at advokere for et bestemt sæt værdier (værdisystemer). Dette dokument indeholder også en oversigt over internationale standarder, der behandler spørgsmål vedrørende etiske og samfundsmæssige hensyn.

© Danish Standards Foundation

DS/ISO/IEC 5338

Informationsteknologi – Kunstig intelligens – AI-systemers livscyklusprocesser

Dette dokument definerer et sæt processer og tilknyttede koncepter til beskrivelse af AI-systemers livscyklus baseret på maskinlæring og heuristiske systemer. Den er baseret på ISO/IEC/IEEE 15288:2015 og ISO/IEC/IEEE 12207:2017 med ændringer og tilføjelser af AI-specifikke processer fra ISO/IEC 22989:– og ISO/IEC 23053:2022. Dokumentet beskriver processer, der understøtter definition, kontrol, styring, udførelse og forbedring af AI-systemet i dets livscyklusstadier. Disse processer kan også bruges i en organisation eller et projekt, når man udvikler eller anskaffer AI-systemer.

B.2 Nationale standarder

DS 5001

Ledelsessystemer for ligestilling og diversitet – Krav og vejledning

Denne standard fastlægger krav til et ledelsessystem for ligestilling og diversitet i de tilfælde, hvor en organisation har behov for at a) demonstrere sin evne til vedvarende at skabe større ligestilling og diversitet i organisationen, b) dokumentere at organisationen arbejder systematisk med ligestilling og diversitet og c) styrke sin troværdighed via transparente indsatser samt dokumentation herfor. Alle krav i denne standard er generiske og er beregnet til at kunne anvendes af enhver organisation, uanset type eller størrelse, eller hvilke produkter og ydelser den leverer. Omfanget af den dokumenterede information for ledelsessystemet skal afstemmes til organisationens størrelse, kompleksitet, lokationer samt tilpasses de enkelte indsatsområder.

DS/PAS 2500-1:2020

Kunstig intelligens – Del 1: Gennemsigtighed

Dette dokument beskriver gennemsigtighed, en generel tilgang til, hvordan det kan opnås, og en metode til at selvevaluere niveauet af gennemsigtighed. Denne metode kan også anvendes til at specificere krav til gennemsigtighed. Tilgangen og metoden kan anvendes både i systemer, som anvender kunstig intelligens, og de, som ikke gør. Målgruppen er designere, udviklere og brugere, som enten ønsker eller er forpligtet til at dokumentere og/eller øge gennemsigtigheden i deres systemer. En sekundær målgruppe er de interessenter, der har gavn af en øget gennemsigtighed, fx myndigheder, indkøbere, virksomheder og forbrugere. Dette dokument specificerer anbefalinger til en fremgangsmåde til at opnå gennemsigtighed i systemer, som anvendes til automatiske beslutninger eller beslutningsstøtte, herunder slutninger (inferens) og brug af data. Ofte vil sådanne systemer anvende kunstig intelligens, men det er ikke et krav. I tillæg kan fremgangsmåden anvendes til at vurdere om den gennemsigtighed, der er opnået, har et tilstrækkeligt niveau. Dokumentet omfatter ikke reproducerbarhed, forklaringer i en bredere forstand og vedligehold (fx konceptskred).

© Danish Standards Foundation

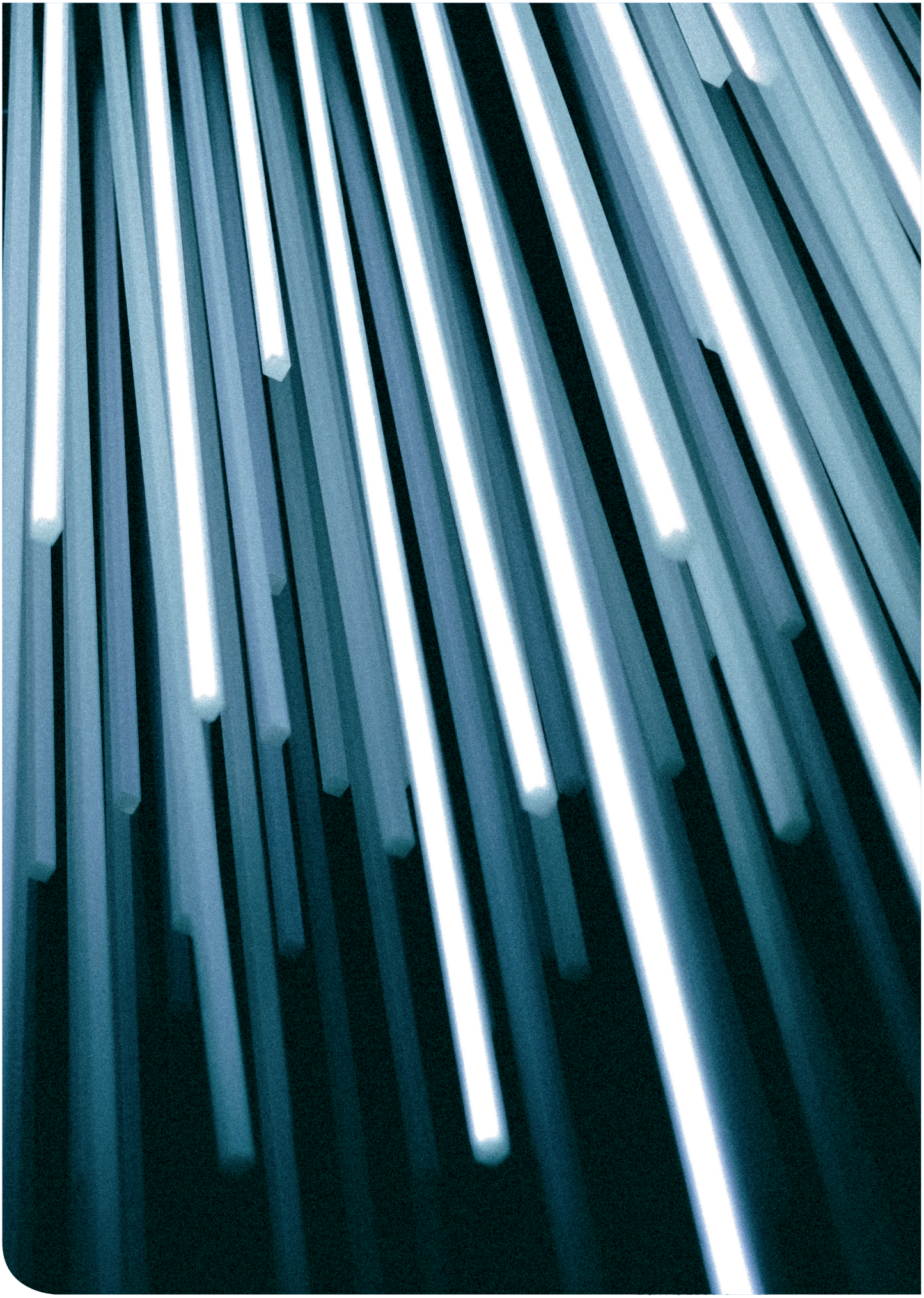
DS/PAS 2500-2:2020

Kunstig intelligens – Del 2: Beslutningsstøttende anvendelse i offentlig sagsbehandling

I dette dokument angives en tjekliste for relevante overvejelser i de forskellige faser i IT-projekter, hvor der anvendes kunstig intelligens til beslutningsstøtte i offentlig sagsbehandling. Det er vigtigt, at disse overvejelser medtages i grundlaget for krav og løsninger, samt i beslutningen om at udføre aktiviteter og gå videre til næste fase i IT-systemets livscyklus. Det er brugerens ansvar at afgøre, hvorvidt tjeklisten er anvendelig til et givet formål. Målgruppen er designere, udviklere, udbydere, indkøbere, tilsynsførere og brugere af kunstig intelligens til beslutningsstøtte i offentlig sagsbehandling, som ønsker eller er forpligtede til at evaluere systemets anvendelse. En sekundær målgruppe er de interessenter, der er berørt af systemet, fx myndigheder og borgere generelt. Dokumentet omfatter ikke lovgivning og særlige krav, der kan være ved anvendelse af beslutningsstøtte indenfor specifikke domæner. Dokumentet skal anvendes sammen med organisationens eksisterende retningslinjer for sikring af overholdelse af regler, herunder, men ikke begrænset til, håndtering af privatliv, cybersikkerhed, robusthed af IT-løsninger, datagovernance, klage- og appelhåndtering samt lovgivning. Dette dokument beskæftiger sig ikke eksplicit med transparens i kunstig intelligens, men kan med fordel bruges sammen med DS/PAS 2500-1: 2020, *Kunstig intelligens – Del 1: Gennemsigtighed*, der specificerer anbefalinger til en fremgangsmåde til at opnå gennemsigtighed i systemer, som anvendes til automatiske beslutninger eller beslutningsstøtte.

Alle standarder kan købes i Dansk Standards webshop: https://webshop.ds.dk/
DS/PAS 2500-serien er gratis.

© Danish Standards Foundation



© Danish Standards Foundation

Anneks C

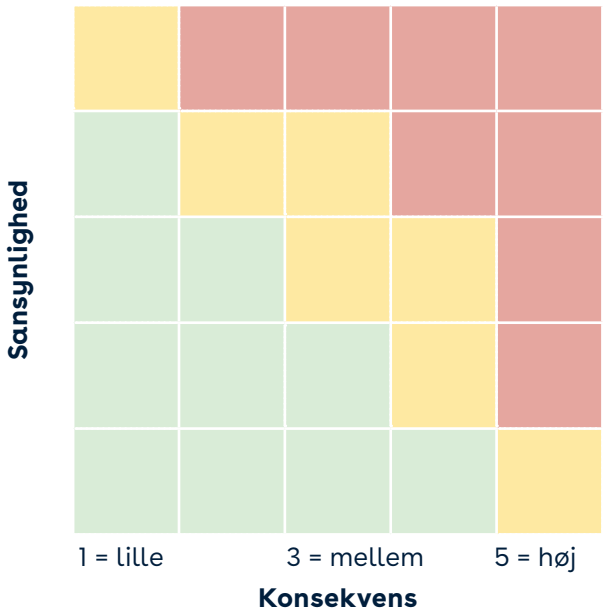
(informativt)

C.1 Risicimodel lavet på baggrund af DS/EN IEC 31010:2019, *Risikoledelse – Teknikker til risikovurdering*

Følgende model kan anvendes til at vurdere risikoen i de enkelte faser i projektet.

Risiko	**Kort beskrivelse af risici**	**Konsekvens** (1 = lille; 3 = mellem; 5 = høj)	**Sandsynlighed** (1 = lille; 3 = mellem; 5 = høj)	**Risikofaktor** (konsekvens x sandsynlighed)	**Hvordan kan vi mitigere?**
#1					
#2					
#3					

Tabel C.1 - Risikoskema til udfyldelse



Figur C.2 - Risikomatrice

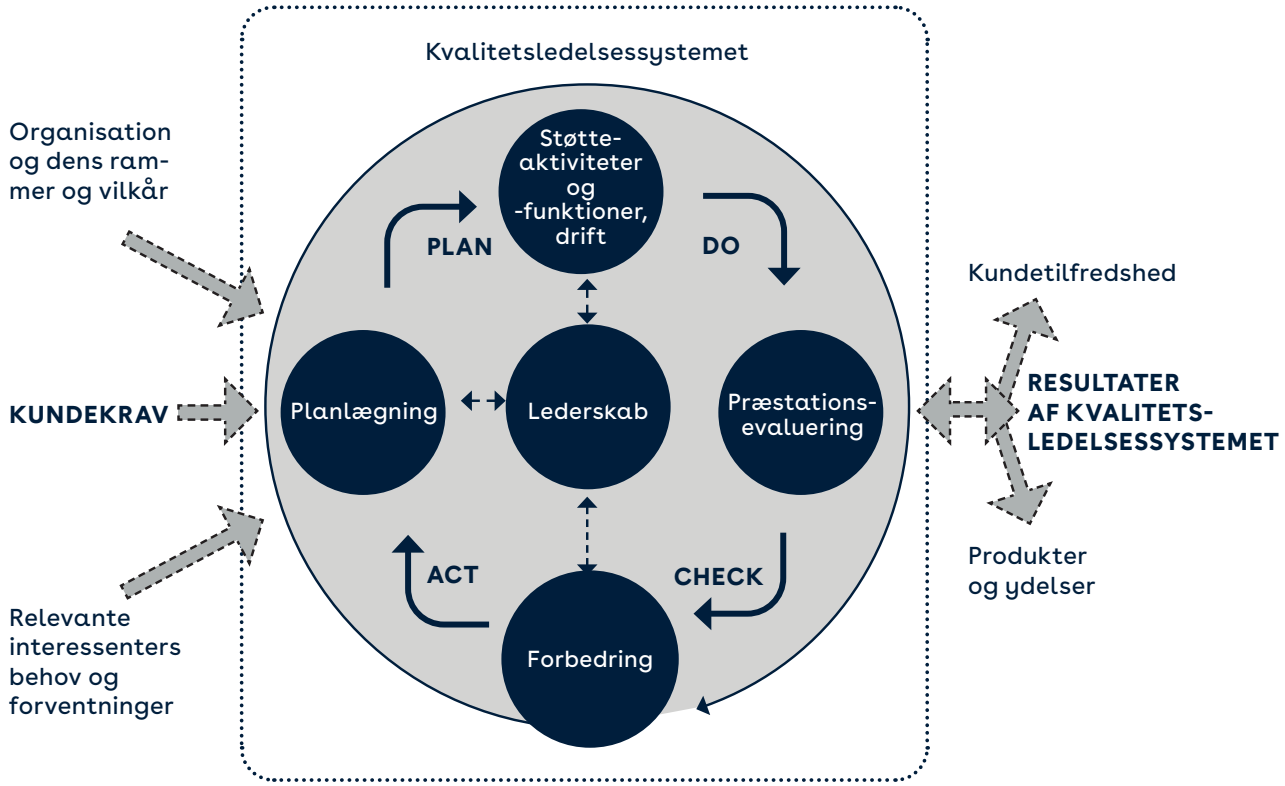
© Danish Standards Foundation

Anneks D

(informativt)

D.1 Plan-Do-Check-Act (PDCA)-metoden fra DS/EN ISO 9001:2015, *Kvalitetsledelsessystemer - Krav*

PDCA-cyklussen kan anvendes på alle processer og på et kvalitetsledelsessystem som helhed. Figuren illustrerer, hvordan de forskellige aktiviteter grupperes i forhold til PDCA-cyklussen.



Figur D.1 - Plan-Do-Check-Act-cyklus (Figur 2 i DS/EN ISO 9001:2015 - Kvalitetsledelsessystemer - Krav)

PDCA-cyklussen kan kort beskrives på følgende måde:

- **Plan:** Opstille målene for systemet og dets processer
- **Do:** Implementere det planlagte
- **Check:** Overvåge og (hvor det er relevant) måle og evaluere processer, produkter og ydelser i forhold til mål og krav.
- **Act:** Iværksætte handlinger efter behov for at forbedre præstationen.

Modellen er baseret på en risikobaseret tankegang. Begrebet risikobaseret tankegang handler om at gennemføre forbyggende handling for at fjerne potentielle afvigelser, at analysere afvigelser, og handle for at forhindre gentagelse på en hensigtsmæssig måde i forhold til afvigelsens effekt.

© Danish Standards Foundation

Anneks E

(informativt)

E.1 Oversigt over tjeklistespørgsmål

FASE (ID)	SPØRGSMÅL (ID)
Idéfasen (A)	Hvilke grupper i beslutningsdomænet for systemet er relevante i forhold til bias? (A1) Hvilke love og menneskerettigheder (se anneks A) er relevante i denne sammenhæng? (A2) Har grupperne (defineret i A1) givet udtryk for særlige værdier, holdninger eller ønsker, der bør tages hensyn til? (A2) Er dit IT-system brugbart for dem med særlige behov eller handicap? (A3) Er der andre grupper, der risikerer at blive udelukket? (A3) Hvordan er dit IT-system sikret mod cyberangreb, som kan påvirke datakvaliteten? (A4)
Analysefasen (B)	Hvilke grupper i beslutningsdomænet for systemet er relevante i forhold til bias? (Er evt. tidligere besvaret i A1) (B1) Hvilke menneskelige, kognitive bias kan der være i systemet, og har de negativ/uønsket, neutral eller positiv/ønsket effekt? (B2) I tilfælde af negative/uønskede bias, er der lavet en risikoanalyse og konsekvensanalyse? (B2) Hvilke databias kan der være i systemet, og har de negativ/uønsket, neutral eller positiv/ønsket effekt? (B3) Er data repræsentative for de relevante grupper, som fx demografiske parametre som køn, indkomst, herkomst og alder? (B3) I tilfælde af negative/uønskede bias, er der lavet en risikoanalyse og konsekvensanalyse? (B3) Hvor stabile er de antagelser, datasystemet bygger på, og hvor stor er risikoen for dataskred? (B4) Er der udviklet en plan for håndtering af bias? (B5) Indeholder denne plan opdatering af auditlog for alle relevante dele af systemets livscyklus? (B5) Kan systemet udvikles, så det bliver robust, pålideligt og sikkert, og kan det overvåges og auditeres? (B6)

© Danish Standards Foundation

Gennemførsels-fasen (C)	Er der tilstrækkelig gennemsigtighed i designet af systemet, så bias kan identificeres, auditeres og overvåges? (C1) Er der tilstrækkelig oplæring og bevidsthed om bias i de relevante grupper (se A1)? (C2) Er der etableret en procedure til at fjerne potentiel uønsket bias? (C3) Er dokumentationen mht. bias tilstrækkelig gennemsigtig? (C4) Kan eventuelle forpligtigelser i forhold til benyttelse af data dokumenteres? (C4)
Realiserings-fasen (D)	Er der etableret en procedure for overvågning og identificering af uønsket bias? (D1) Er der etableret en procedure for håndtering af uønsket bias? (D2) Er der monitorering og løbende kontrol af systemets overholdelse af relevante love (se anneks A), menneskerettigheder, demokratiske værdier og diversitetsspørgsmål? (D3) Er der indbygget automatiske mekanismer for at håndtere eventuelle ændringer i bias (fx konceptskred)? (D4) Er auditloggen vedligeholdt korrekt (også for runtime-ændringer)? (D5)
Afskaffelses-fasen (E)	(ingen spørgsmål)

© Danish Standards Foundation

Bibliografi

[1] Den Europæiske Menneskerettighedskonvention, https://www.retsinformation.dk/eli/lta/1996/423

[2] Den Europæiske Unions Charter om Grundlæggende Rettigheder, https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:da:PDF

[3] Bekendtgørelse af lov om ligebehandling af mænd og kvinder med hensyn til beskæftigelse m.v. (Ligebehandlingsloven), https://www.retsinformation.dk/eli/lta/2011/645

[4] Bekendtgørelse af lov om forbud mod forskelsbehandling på arbejdsmarkedet m.v. (Forskelsbehandlingsloven), https://www.retsinformation.dk/eli/lta/2017/1001

[5] Bekendtgørelse af lov om etnisk ligebehandling, https://www.retsinformation.dk/eli/lta/2012/438

[6] Lov om forbud mod diskrimination på grund af handicap, https://www.retsinformation.dk/eli/lta/2018/688

[7] DS/ISO/IEC TR 24027, *Informationsteknologi – Kunstig intelligens (AI) – Bias i AI-systemer og AI-assisteret beslutningstagning*

[8] DS/ISO/IEC 23894, *Informationsteknologi – Kunstig intelligens – Risikoledelse*

[9] DS/ISO/IEC TR 24368, *Informationsteknologi – Kunstig intelligens – Overblik over etiske og samfundsmæssige hensyn*

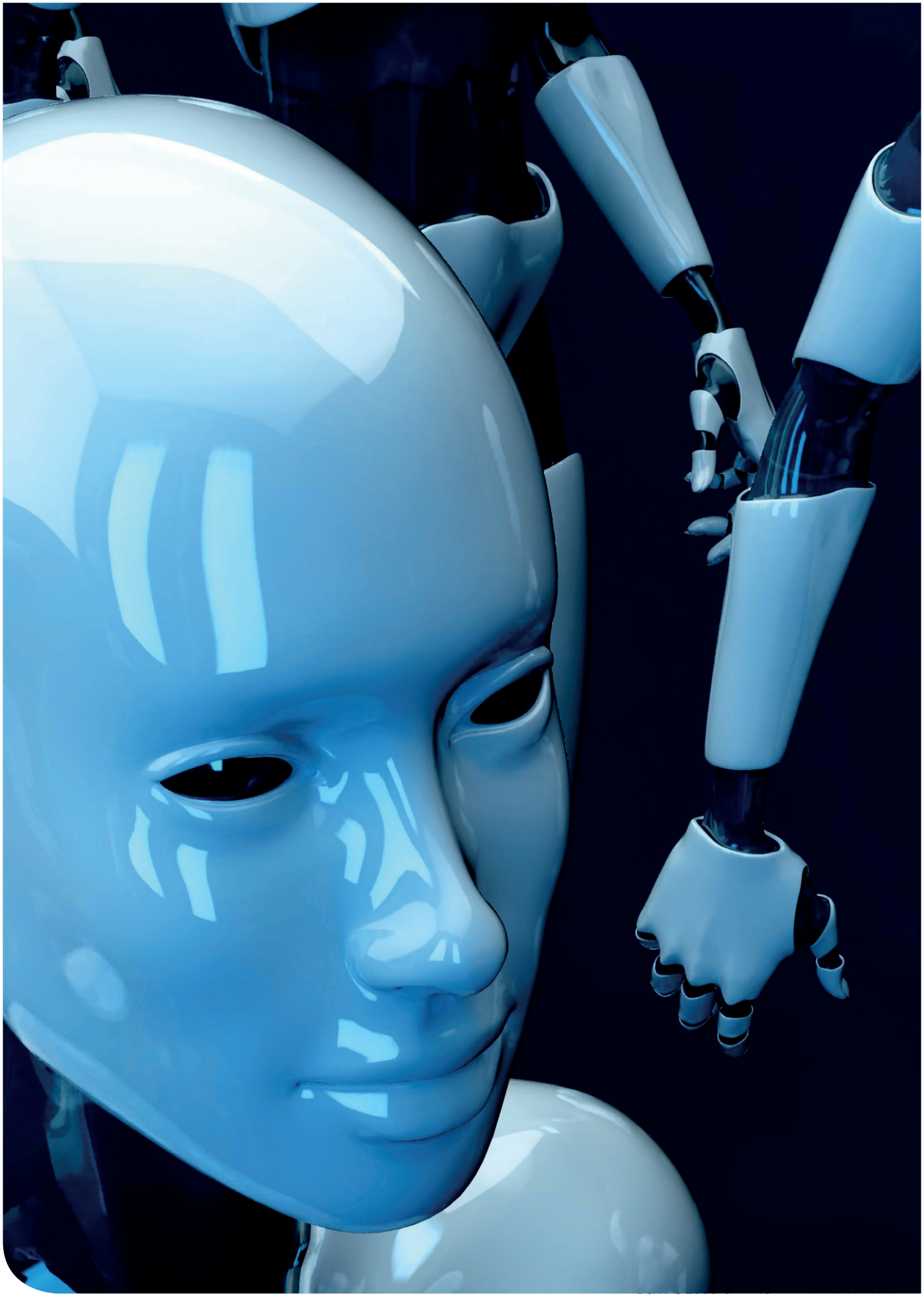
[10] DS/ISO/IEC 5338, *Informationsteknologi – Kunstig intelligens – AI-systemers livscyklusprocesser*

[11] DS 5001, *Ledelsessystemer for ligestilling og diversitet – Krav og vejledning*

[12] DS/PAS 2500-1:2020, *Kunstig intelligens – Del 1: Gennemsigtighed*

[13] DS/PAS 2500-2:2020, *Kunstig intelligens – Del 2: Beslutningsstøttende anvendelse i offentlig sagsbehandling*

[14] DS/EN IEC 31010:2019, *Risikoledelse – Teknikker til risikovurdering*

[15] DS/EN ISO 9001:2015, *Kvalitetsledelsessystemer – Krav*

© Danish Standards Foundation