



# Innebygd diskrimineringsvern

En veileder for å avdekke og  
forebygge diskriminering i utvikling  
og bruk av kunstig intelligens

# Likestillings- og diskrimineringsombudet

- Juridisk veiledning
- Pådriver for likestilling
- Konvensjonstilsyn
  
- Strategisk satsing på ny teknologi



# På agendaen

- Likestillings- og diskrimineringslovens relevans i møte med KI
- Hvordan bruke veilederen i et konkret case
- Innspill og diskusjon



# DISKRIMINERINGSNEMNDA

Diskrimineringsnemnda er et nøytralt organ som avgjør saker om diskriminering og trakassering.

Rettferdighet

Bias

Ansvarlig KI

Etikk

Diskriminering

**Målgruppe:** Ansvarlige for utvikling, innkjøp og bruk av KI-systemer som kan få betydning for personers rettigheter og plikter.

**Fundament:** Veilederen tar utgangspunkt i likestillings- og diskrimineringsloven og de menneskerettslige prinsippene om likeverd, autonomi og rettferdighet loven bygger på.

**Innhold:** Skisserer relevante problemstillinger som bør drøftes for å avdekke diskrimineringsrisiko gjennom alle utviklingsfaser i et KI-system.

**Bidrag:** Et diskusjons- og beslutningsverktøy som legger til rette for en tverrfaglig dialog mellom involverte aktører, med mål om å fremme likestilling og hindre diskriminering.

# Norsk lov forbyr diskriminering

*– men hva er det?*



**Diskriminering**

=

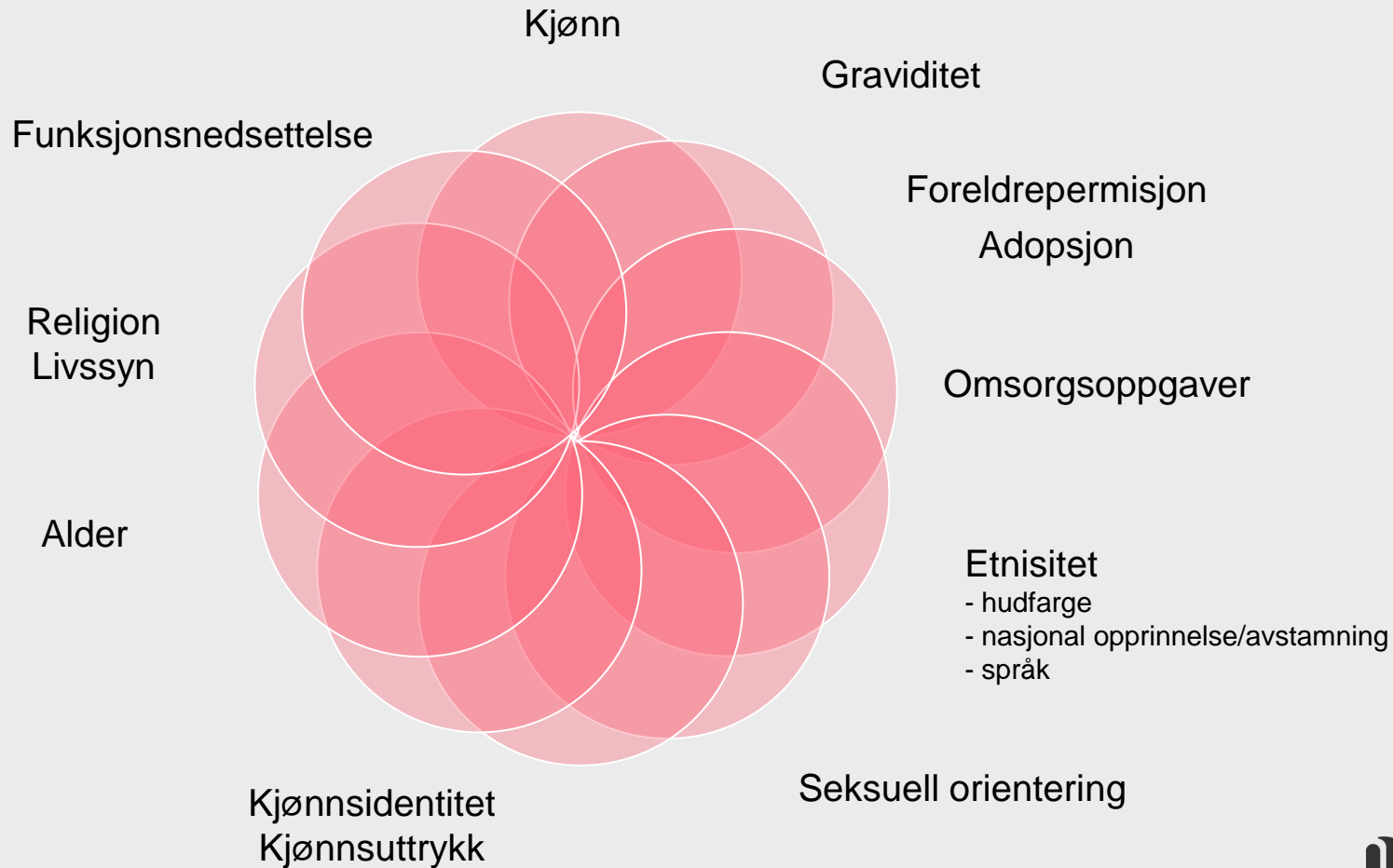
**Usaklig forskjellsbehandling  
knyttet til et eller flere  
diskrimineringsgrunnlag**















# Innebygd diskrimineringsvern

En veileder for å avdekke og  
forebygge diskriminering i utvikling  
og bruk av kunstig intelligens



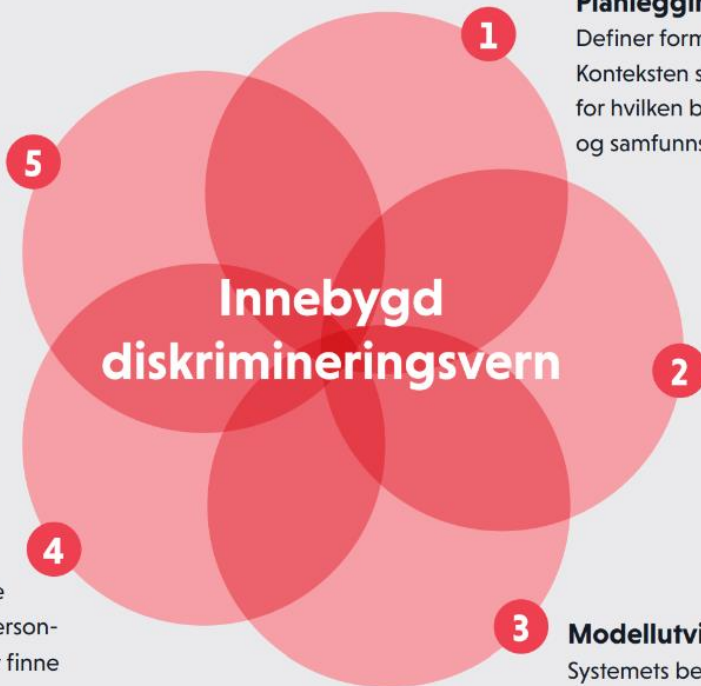
## Implementering og tilsyn

Undersøk om skjevheter i systemet kan utjevnes ved kompenserende tiltak. Menneskelig kontroll med systemet må sikres, og kunnskap om systemet gjøres kjent for de berørte, slik at de kan ivareta sine rettigheter.



## Testing

Kartlegg om systemet fungerer like presist for alle grupper. Selv om personopplysninger fjernes, kan systemet finne mønstre i dataene slik at opplysninger som sammenfaller med diskrimineringsgrunnlagene ulovlig vektlegges.



## Planlegging

Definer formål og tiltenkt bruk av KI-systemet. Konteksten systemet skal operere i er avgjørende for hvilken betydning det kan ha på individ- og samfunnsnivå.



## Treningsdata

Datagrunnlaget modellen er trent på må være representativt ut fra modellens formål, og historiske skjevheter må kartlegges og korrigeres.

## Modellutvikling

Systemets beregning må sammenfalle med det overordnede formålet. De variablene systemets beregninger baserer seg på må være relevante. Påvis saklig formål, nødvendighet og forholdsmessighet for variabler som kan knyttes til et diskrimineringsgrunnlag.



## Hjerterom for etisk AI

Sluttrapport fra sandkasseprosjektet til Ahus (EKG AI)

Tema: Algoritmeskjevhet og rettferdige algoritmer

Februar 2023





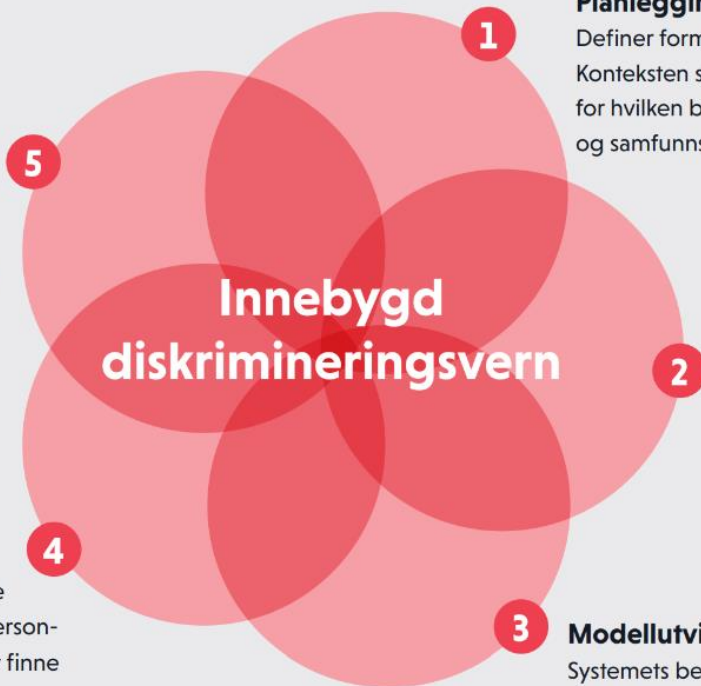
## Implementering og tilsyn

Undersøk om skjevheter i systemet kan utjevnes ved kompenserende tiltak. Menneskelig kontroll med systemet må sikres, og kunnskap om systemet gjøres kjent for de berørte, slik at de kan ivareta sine rettigheter.



## Testing

Kartlegg om systemet fungerer like presist for alle grupper. Selv om personopplysninger fjernes, kan systemet finne mønstre i dataene slik at opplysninger som sammenfaller med diskrimineringsgrunnlagene ulovlig vektlegges.



## Planlegging

Definer formål og tiltenkt bruk av KI-systemet. Konteksten systemet skal operere i er avgjørende for hvilken betydning det kan ha på individ- og samfunnsnivå.



## Treningsdata

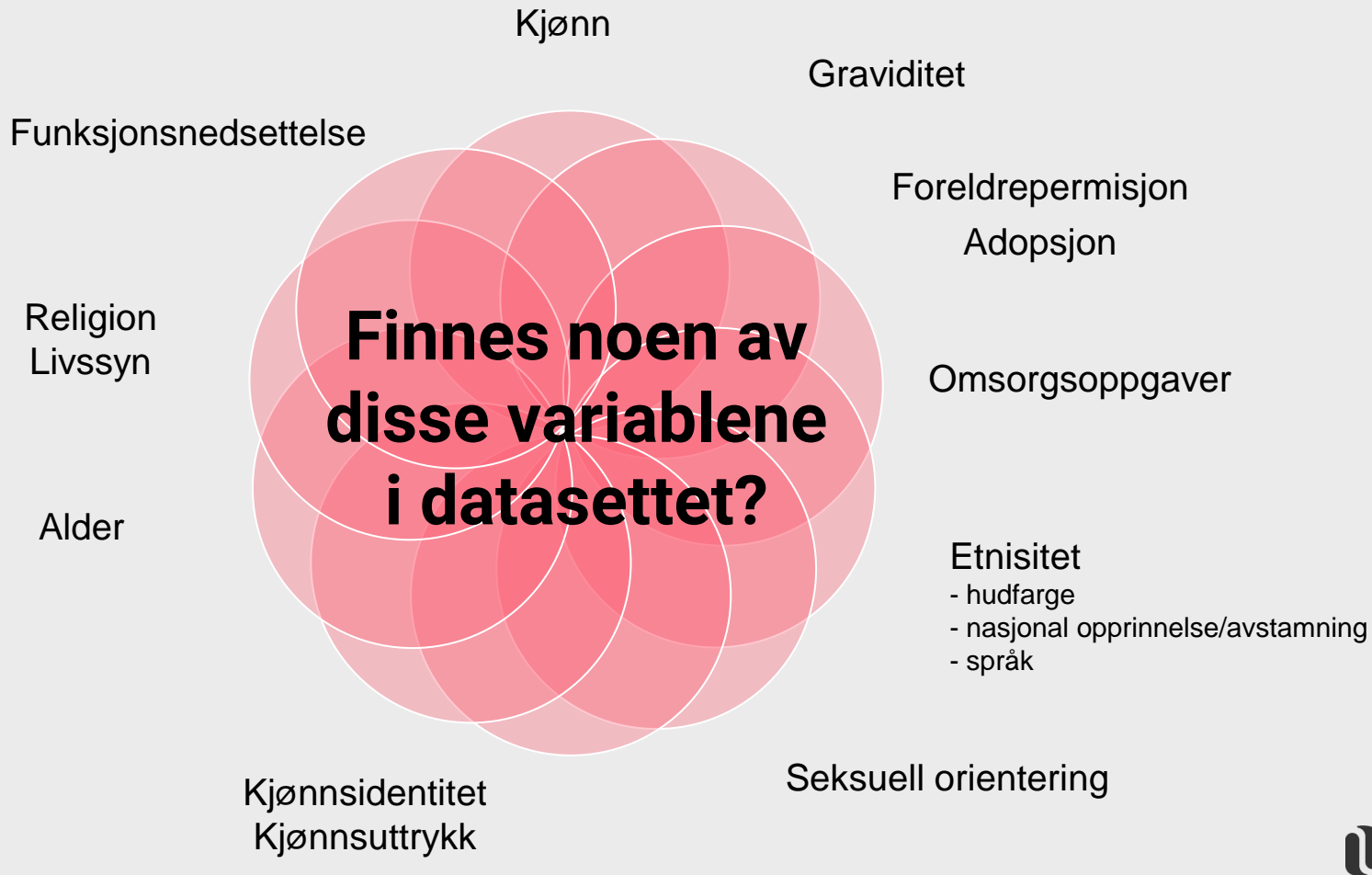
Datagrunnlaget modellen er trent på må være representativt ut fra modellens formål, og historiske skjevheter må kartlegges og korrigeres.

## Modellutvikling

Systemets beregning må sammenfalle med det overordnede formålet. De variablene systemets beregninger baserer seg på må være relevante. Påvis saklig formål, nødvendighet og forholdsmessighet for variabler som kan knyttes til et diskrimineringsgrunnlag.









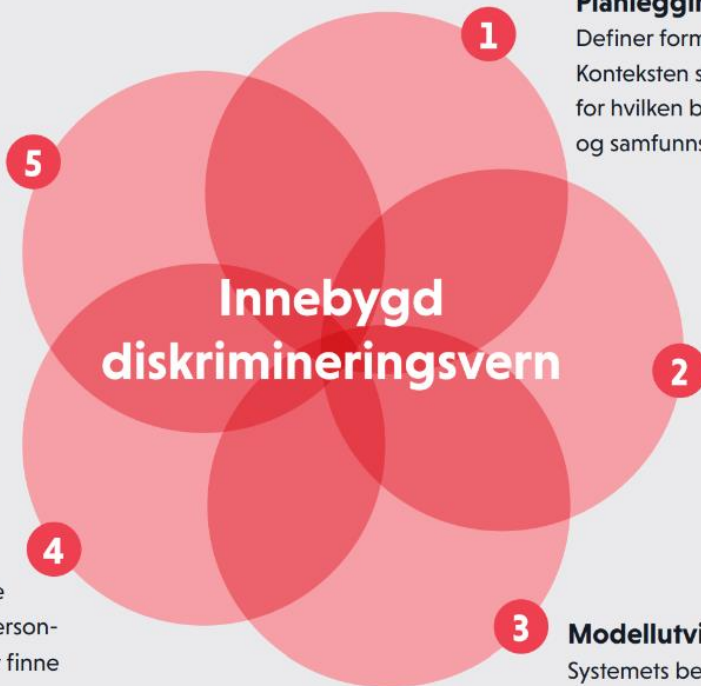
## Implementering og tilsyn

Undersøk om skjevheter i systemet kan utjevnes ved kompenserende tiltak. Menneskelig kontroll med systemet må sikres, og kunnskap om systemet gjøres kjent for de berørte, slik at de kan ivareta sine rettigheter.



## Testing

Kartlegg om systemet fungerer like presist for alle grupper. Selv om personopplysninger fjernes, kan systemet finne mønstre i dataene slik at opplysninger som sammenfaller med diskrimineringsgrunnlagene ulovlig vektlegges.



## Planlegging

Definer formål og tiltenkt bruk av KI-systemet. Konteksten systemet skal operere i er avgjørende for hvilken betydning det kan ha på individ- og samfunnsnivå.



## Treningsdata

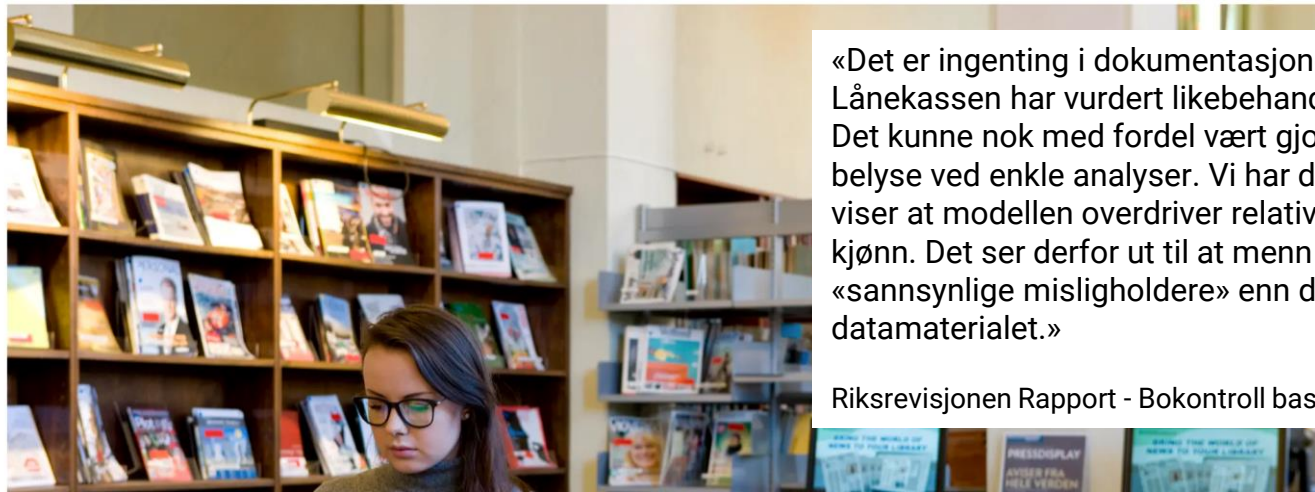
Datagrunnlaget modellen er trent på må være representativt ut fra modellens formål, og historiske skjevheter må kartlegges og korrigeres.

## Modellutvikling

Systemets beregning må sammenfalle med det overordnede formålet. De variablene systemets beregninger baserer seg på må være relevante. Påvis saklig formål, nødvendighet og forholdsmessighet for variabler som kan knyttes til et diskrimineringsgrunnlag.

# Kunstig intelligens avslørte Lånekassen-juks

Lånekassen har for første gang brukt kunstig intelligens for å velge ut studenter til bokkontroll. Det ga dobbelt så mange avsløringer av juks.



«Det er ingenting i dokumentasjonen vi har fått som tilsier at Lånekassen har vurdert likebehandling i utviklingen av ML-modellen. Det kunne nok med fordel vært gjort, særlig ettersom det er mulig å belyse ved enkle analyser. Vi har derfor gjort noen slike analyser, som viser at modellen overdriver relativt kraftig betydningen av nettopp kjønn. Det ser derfor ut til at menn i større grad blir plukket ut som «sannsynlige misligholdere» enn det det er grunnlag for i datamaterialet.»

Riksrevisjonen Rapport - Bokkontroll basert på maskinlæring (sak 2019/01316)

<https://www.aftenposten.no/norge/i/3jAqm0/kunstig-intelligens-avslørte-laanekassen-juks>

# Veien videre

- Utredning av likestillingslovverkets muligheter og begrensninger i møtet med KI
- Samarbeid med forsknings- og utviklermiljøer om konkrete verktøy

