

# BIAS

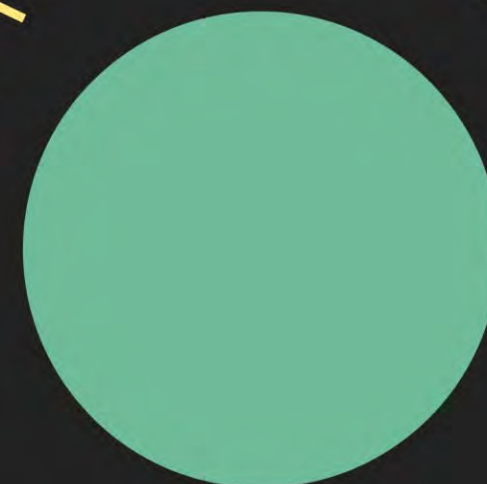
Mitigating biases  
of AI in the  
labour market

# Hvilken rolle kan KI spille når mangfold er målet?

Siri Øyslebø Sørensen

Professor, Institutt for tverrfaglige kulturstudier, NTNU

Frokostseminar: Bedre ansettelser med KI?  
Digitaliserings- og forvaltningsdepartementet  
23. Oktober 2024



# BIAS


## BIAS-prosjektet

- Finansiert € 4.2 million av EUs *Horizon Europe* 2022-2026.
- Ledet av professor Roger A. Søråa ved Institutt for tverrfaglige kulturstudier, NTNU
- 9 partnerorganisasjoner over hele Europa, med over 40 medarbeidere.

BIAS  
CONSORTIUM MAP

- NORWAY – NTNU
- SWITZERLAND – BFH
- ICELAND – HI
- PORTUGAL – LOBA
- IRELAND – CHX
- ITALY – SVEN
- NETHERLANDS – LEID
- ESTONIA – DIGI
- TURKEY – FARPL



 Funded by  
the European Union

[www.biasproject.eu](http://www.biasproject.eu)





## Utgangspunkt: Ulikhet i arbeidslivet

- Strukturelle forhold
- Kulturelle forestillinger
  
- → Diskriminering og forskjellsbehandling
  
- Rekruttering
- Karriereutvikling



## KI brukt i HR arbeid

**Organisering av arbeid:** skiftordninger, produktivitet, monitorering etc.

**Rekruttering og ansettelsesprosesser:**

- Generere utlysningstekster
- Nå ut til potensielle søkere
- Prosessere søknader - sortere og/ eller rangere
- Trekke ut informasjon fra søknadsbrev

→ Automatiserte prosesser?



Hvordan unngår vi å reprodusere diskriminering?

Hvordan forhindre nye former for diskriminering?



Kan kunstig intelligens bidra til å skape  
likere muligheter for alle i arbeidslivet?

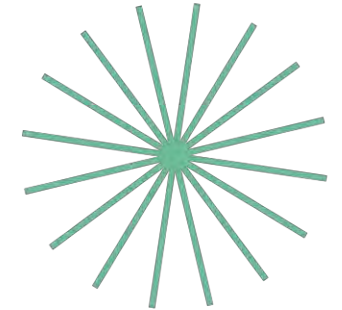


Når noen søker på en jobb, hvordan kan vi sørge for at «rett person» får jobben?

Er mennesker eller kunstig intelligens best til å vurdere hvem som er «rett» for en jobb?



# Kan vi i stedet utvikle en KI som bidrar til å fjerne negative effekter av fordommer og stereotypier?

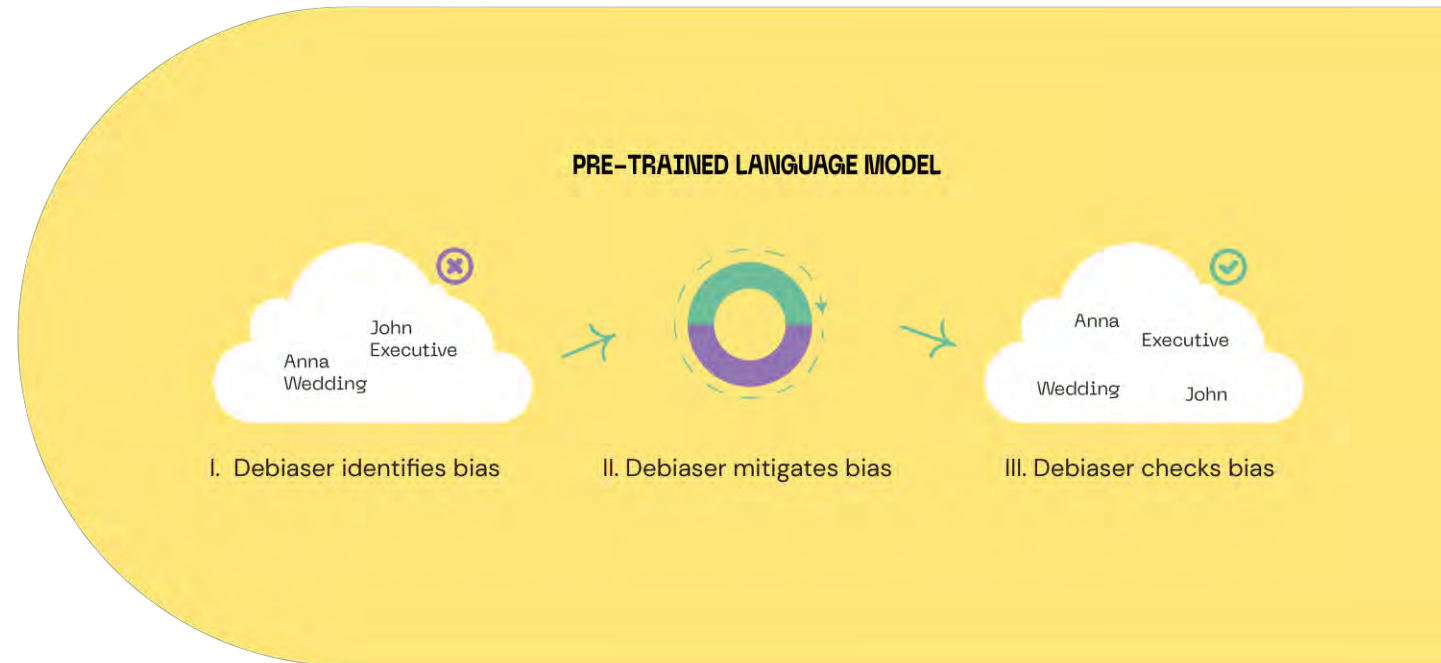


NLP –naturlig språkprosessering (Natural language processing)

KI som trenes på store tekstkorpus for å håndtere menneskelig språk

Fare: svakheter i treningsdata og fordommer (ofte ubevisste) hos de som utvikler koden kan bidra til skjevheter

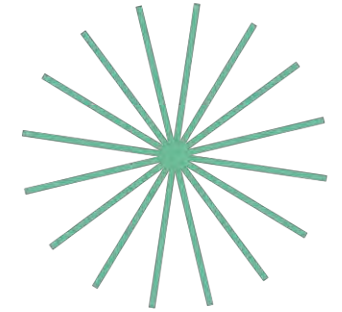
→ behov for å utvikle verktøy som kan identifisere og utligne skjevheter (bias) når KI skal brukes i HR arbeid



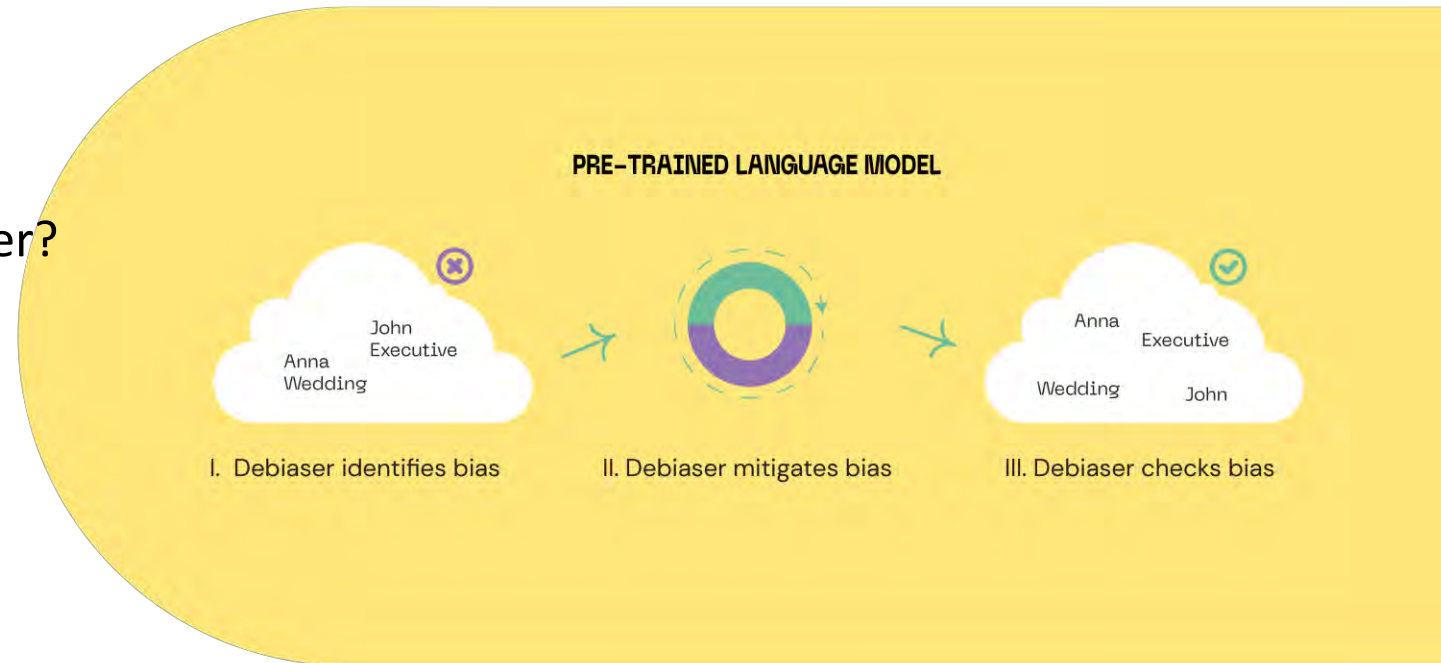




# Hvordan bestemme hva som er «rett»? (fairness), skal vi ta hensyn til...



- Gruppe eller individ?
- Prosedyrelighet eller påvirkningslikhet?
- Skjevheter vi allerede kjenner til eller potensielt nye og foranderlige skjevheter?



\*\*\*

Rigotti, Carlotta, Alexandre Puttick, Eduard Fosch-Villaronga and Mascha Kurpicz-Briki (2023) **Mitigating Diversity Biases of AI in the Labor Market**. *EWAF'23: European Workshop on Algorithmic Fairness, CEUR Workshop Proceedings*.



## Hvordan bruke KI på etisk forsvarlige måter?

- Dagens likestillings- og diskrimineringslovverk har i liten grad tatt høyde for algoritmisk diskriminering
- Perspektiver fra ulike aktører (stakeholders) blir dermed viktig, både for teknologiutvikling og for regulering av KI - teknologien og bruken av den fremover

*«...we assume that fairness constitutes situated knowledge and practices, and acknowledge that it is not a one-size-fits all concept but is contingent on specific contexts.» (Rigotti and Fosch-Villaronga 2024: 2)*

\*\*\*

Rigotti, Carlotta and Eduard Fosch-Villaronga (2024) **Fairness, AI & recruitment**, *Computer Law & Security Review*, Vol 53.  
<https://doi.org/10.1016/j.clsr.2024.105966>



# Aktiviteter i BIAS prosjektet

- **Teknologiutvikling** – verktøy for å identifisere og eliminere bias i AI (NLP) systemer som brukes i HR arbeid (The Debiaser)
- **Involvering av interessenter** – aktører i arbeidslivet (National Labs)
- **Kunnskaps- og erfaringsutveksling** på tvers av fagmiljøer (Capacity building)
- **Kvalitativ forskning:**
  - intervjuer og spørreundersøkelser med HR- og KI-eksperter, ledere og medarbeidere.
  - Etnografisk feltarbeid for dybdeforståelse for bruk, oppfatninger, erfaringer, behov etc.





## Innledende undersøkelser: ekspertintervjuer og 'labs'

- 71 intervjuer med eksperter på KI- (35) og HR (36) fra Estland, Frankrike, Island, Italia, Norge, Sveits, Nederland og Tyrkia.
- National labs – workshops med aktører i arbeidslivet (HR, KI, ledere, medarbeidere/fagforeninger)
- Temaer:
  - KI og bias/ skjevheter
  - KI i personaloppfølging og rekruttering
  - Strategier for å håndtere bias/ potensielle skjevheter

Hovedfunn fra innledende undersøkelser →



## KI og skjevheter

- Under/overrepresentasjon av grupper i treningsdata leder til skjevhet.
- Samfunnsmessige skjevheter gjenspeiles i treningsdata og leder til forsterkede skjevheter.
- Skjevheter rapporteres relatert til kategoriene kjønn, rase, etnisitet, nasjonalitet, geografisk opprinnelse, hudfarge, navn, alder og funksjonsevne
- Skjevheter oppstår også som konsekvens av svake KI modeller (f.eks feilmarginer og inputparametre)
- Det er en betydelig risiko for at ubevisste fordommer påvirker beslutninger både i utvikling av KI og i beslutningstaking basert på output fra algoritmene.



## Bruk av KI i personaloppfølging og rekruttering

- KI brukes oftere i større selskaper, og sjeldnere mindre.
- KI oppfattes som et verktøy som bør brukes til å bistå ansatte, ikke til å kontrollere dem.
- Menneskelig intervensjon forstås som viktig og nødvendig.
- Akseptabel bruk av KI beskrives som transparent, underlagt menneskelig kontroll, forklarbarhet, og i tråd med sikkerhetsregler og personvern.
- Grunner til at KI *ikke* tas i bruk: personvernreguleringer, kostnader, motstand fra fagforeninger, ønske om 'fairness' og usikkerhet omkring dette



## Strategier for å unngå skjevheter ved bruk av KI

- Sørge for mangfold både i KI utvikler- og HR team som bruker KI.
- Mangfoldskvotering i rekruttering (uten å gå bort fra prinsipper om kvalifikasjoner).
- Øke bevissthet og kunnskap om bias (systemskjevheter, fordommer og stereotypier).
- Sjekke treningsdata
- Anonymisere søkerdata
- Transparens og menneskelig involvering
- Utvikle KI systemer for å avsløre skjevheter



# Innledende studie i Norsk kontekst (fase 1)

- PhD prosjekt, Shan Wang, NTNU
- 37 intervjuer (HR arbeidere, KI utviklere og medarbeidere)
- **Individuell bruk av KI i HR arbeid**
  - Støtte i egne arbeidsoppgaver (ideer)
  - Nye krav til digital kompetanse
- **Utviklere diskuterer hva som er god bruk av KI**
  - KI (GPT) brukes som add-on i eksisterende plattformer, innrettet mot første fase i en ansettelsesprosess (stillingsbeskrivelser, kvalifikasjonsoppsummeringer, første seleksjon)
  - Bekymring knyttet til hvorvidt vi kan stole på KI i tilstrekkelig grad for å brukes i HR arbeid
- **Usikkerhet knyttet til KI fra jobbsøkerperspektivet**
  - Forbedring av søknadsbrev
  - Forberedelse til intervjuer





## Fase 2 (i oppstart nå): case studier

- Norge: Fokus på jobbsøkerperspektivet (ulike faser), og særlig de med innvandringsbakgrunn og minoritetsstatus
- Italia: Fokus på HR perspektivet: intuisjon, erfaring (generasjon), fremtidsforestillinger
- Nederland: Fokus på teknologiutvikling og operasjonalisering av forståelser av 'fairness' i praksis
- Tyrkia: Fokus på hvordan ALTAI prinsippene tas i bruk (fra flere perspektiv)
- Island: *KI i annet HR arbeid* - KI brukt i lønnsbonusordning (fiskeribedrifter)

# Oppsummert (så langt): Hvilken rolle kan KI spille når mangfold er målet?



- Fremtidig bruk av KI krever aktiv, menneskelig deltakelse med kunnskap om- og forståelse for historiske skjevheter, fordommer og stereotypier knyttet til sosiale kategorier (kjønn, seksualitet, etnisitet, rase, funksjonsevne, alder osv.)
- Framtidig bruk av KI krever *også* økt allmenn forståelse for teknologien og at folk med forskjellige identiteter og bakgrunner er involvert i utviklingen
- *Hvis* vi får til dette *kan* KI bli med på å fremme mangfold og forhindre diskriminering i arbeidslivet



Takk for oppmerksomheten!

<https://www.biasproject.eu/>

# BIAS

Mitigating biases  
of AI in the  
labour market

# Consortium



NTNU



UNIVERSITY  
OF ICELAND

LOBA®



CrowdHelix



SMARTVENICE<sup>spa</sup>



Universiteit  
Leiden  
eLaw



DigiTouch

farplas



Bern University  
of Applied Sciences



Funded by  
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the Swiss State Secretariat for Education, Research and Innovation can be held responsible for them.



**email**

[info@biasproject.eu](mailto:info@biasproject.eu)

**website**

[www.biasproject.eu](http://www.biasproject.eu)

**social media**

[@BIASProjectEU](https://twitter.com/BIASProjectEU)

The Associated Partner Bern University of Applied Sciences has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI)