

Hvad er algoritmisk bias

Algoritmisk bias er et vigtigt teknisk fænomen ved algoritmer og kunstig intelligens (AI), men kan forstås på flere forskellige måder.

I en **neutral** forstand betyder algoritmisk bias, at en algoritme eller AI systematisk behandler data på en bestemt måde. Denne form for bias findes i alle algoritmer og AI.

I en **kritisk** forstand betyder algoritmisk bias, at en algoritme eller AI systematisk behandler data på en måde, vi opfatter som problematisk. Typisk skyldes det, at den systematisk behandler nogle **personer** anderledes end andre.

Fire former for algoritmisk bias

En AI eller algoritme kan systematisk behandle personer anderledes på flere måder:

- **Brug af data om grupper** – den bruger data om for eksempel køn, alder, eller etnicitet til at lave en vurdering.
- **Forskelle i vurderinger** – den vurderer oftere personer fra en bestemt gruppe positivt eller negativt.
- **Forskelle i præcision** – den rammer oftere rigtigt, når den vurderer personer fra en bestemt gruppe.
- **Forskelle i fejltipe** – den begår forskellige slags fejl, for eksempel ved ofte at overvurdere en bestemt gruppe og undervurdere en anden.

Kan man fjerne algoritmisk bias?

En AI eller algoritme kan få bias ved:

- at bruge **data om medlemskab** af en gruppe,
- at bruge en **matematisk model, der virker forskelligt** for forskellige grupper, og
- at bruge **data, som viser forskelle** på grupperne, eller repræsenterer grupper forskelligt.

Det er ikke muligt at undgå alle slags algoritmisk bias, men der findes **metoder** til at fjerne nogle bias ved at:

- **Ændre træningsdata**, for eksempel ved at fjerne data om medlemskab af en gruppe, tilføje kopier af datapunkter, eller ændre værdien for visse data.
- **Justere træningsprocessen**, for eksempel ved at få modellen til at lægge vægt på et bestemt mål for lighed.
- **Filtrere resultaterne**, for eksempel ved at bruge forskellige tærskler til at vurdere, eller omgøre bestemte resultater.

Kunstig intelligens er kun så klog som vi gør den. Det kræver dataetisk eftertænksomhed, at udvikle og bruge teknologien ansvarligt.