

INSTITUT FOR  
MENNESKE  
RETTIGHEDER

# DET OFFENTLIGES BRUG AF KUNSTIG INTELLIGENS – ET MENNESKERETLIGT PERSPEKTIV

Marya Akhtar, Juridisk chef

# AUTOMATISEREDE BESLUTNINGER

---

- Beslutninger truffet af algoritmiske systemer, foretager statistisk avanceret analyse baseret på store datamængder
- Algoritmiske systemer er i stand til at finde frem til uigennemskuelige – men statistisk pålidelige – mønstre.
- På baggrund af disse kan systemerne danne algoritmiske profiler af borgeres adfærd, præferencer, handlemønstre mv.

# ANVENDELSE

---

- Algoritmiske profileringer kan anvendes af offentlige og private aktører. Også når det alene vedrører det offentlige, spiller Big Tech en rolle i forvaltningen...
- Fuldautomatiseret vs. beslutningsstøttende anvendelse (human in the loop), men...

# RETTIGHEDER OG PRINCIPPER

---

Berørte rettigheder omfatter bl.a. privatliv, databeskyttelse, diskriminationsforbuddet

Berørte principper omfatter centrale retssikkerheds- og retsstatsprincipper som:

- Forudsigelighed og begrundelse – herunder kendskab til sin retsstilling
- Transparens og offentlighed for at afbalancere magt- og informationsasymmetri (sunshine is the best disinfectant...)

# PRINCIPPER - FORTSAT

---

Det omfatter også principper som:

- Kontrol – fx rekurs og domstolsprøvelse
- Saglighed og lighed vs. statistisk pålidelighed
- Det forvaltningsretlige skøn – som udtryk for retssikkerhed

# UIGENNEMSIGTIGHED (I)

---

Profileringsmodeller kan være uigennemsigtige på forskellig vis:

- Statistisk pålidelige, men ulogiske/ikke-intuitive sammenhænge (*the why*)
- Høj grad af kompleksitet i måden, de opererer på (*the what*)
- Explainable AI kan hjælpe med *the what*, men...

## UIGENNEMSIGTIGHED (II)

---

... det kan være på bekostning af øvrige egenskaber i modellen (designvalg/dilemmaer):

- Explainability vs. accuracy
- Hvad skal forklares/hvilken slags transparens? (holistisk, modulær, global og lokal transparens)
- Hvilket/hvilke formål tjener forklaringen? (fidelitet, stabilitet, anvendelighed)

# DISKRIMINATION

---

Risikoen for diskrimination kan opstå på forskellig vis:

- Videreførelse af eksisterende diskrimination
- Algoritmisk forstærket diskrimination (negative feedback-loops)
- Diskriminationsrisici i designvalg/dilemmaer:
  - Direkte vs. indirekte diskrimination
  - Samtlige beskyttede grupper og intersektionalitet
  - Ikke-diskrimination i afgørelsen – men hvad med fejlrate og fejltype?



# DERFOR VIGTIGT MED...

---

Periodiske konsekvensanalyser rettet mod risici (følger af såvel GDPR, AI-forordningen og almindelige forvaltningsretlige forpligtelser)

*Systemisk* transparens, der kan opveje manglende *algoritmisk* transparens. Systemisk transparens kan bl.a. vedrøre:

- Modellens træning: Informationer om det datasæt, der er brugt til udvikling af modellen
- Modellens kvalitet og risici: Informationer om modellens fejlrate og ydeevne, når den anvendes på nye sager og risici i forhold til rettigheder og retssikkerhed
- Modellens anvendelse: Information om, hvad modellen skal bruges til og om den er egnet til formålet samt overordnede og letforståelige beskrivelser af modellens funktion, type og design.
- Kontrol og tilsyn med modellen: Information om rammer for tilsyn og oplysninger om, hvordan modellen forklares for borgeren

Effektivt tilsyn med fornøden (menneske)retlig og teknologisk ekspertise

# TAK FOR I DAG

---

Rapport: <https://menneskeret.dk/udgivelser/naar-algoritmer-sagsbehandler-rettigheder-retssikkerhed-offentlige-myndigheders-brug>