OKTOBER 2021

NÅR ALGORITMER SAGSBEHANDLER: RISIKO FOR DISKRIMINATION I DET OFFENTLIGES BRUG AF PROFILERINGSMODELLER

Staten er menneskeretligt forpligtet til at sikre, at ingen borgere diskrimineres. Det gælder også, når offentlige myndigheder benytter sig af kunstig intelligens til at træffe afgørelser om borgere. Dog skaber det offentliges brug af såkaldte profileringsmodeller en risiko for diskrimination. Dette notat belyser de menneskeretlige udfordringer med særligt fokus på diskrimination ved brugen af algoritmer i offentlig sagsbehandling.

INSTITUT FOR MENNESKERETTIGHEDER ANBEFALER, AT:

- Justitsministeriet med inddragelse af Datatilsynet og Digitaliseringsstyrelsen udsteder en vejledning om offentlige myndigheders brug af profileringsmodeller med fokus på de rettigheds- og retssikkerhedsmæssige udfordringer ved modellerne.
- At der i vejledningen stilles krav om konsekvensanalyser for kunstig intelligens (AI-konsekvensanalyser) tidligt i beslutningsfasen og periodisk herefter under modellens udvikling og anvendelse til både beslutningsstøtte og fuldautomatisering.
- At konsekvensanalyserne bl.a. bør omfatte en vurdering af, hvilke diskriminationsrisici modellen rejser, samt hvorledes samtlige risici vil blive imødegået.

Danmark er blandt de førende i verden, når det gælder digitalisering af borgernes kontakt med stat, region og kommune.[1] Digitaliseringen omfatter i stigende grad også brug af kunstig intelligens, hvor store mængder data selv "lærer" at analysere opgaver og finde løsninger. Ansigtsgenkendelse, maskinoversættelser og rutevejledninger er alle eksempler på brug af kunstig intelligens, som de fleste kender, men teknologien kan også bruges i offentlige myndigheders sagsbehandling, når de skal træffe afgørelser om for eksempel ydelser, indsatser eller sanktioner overfor borgere.

I 2019 udgav regeringen "National strategi for kunstig intelligens", hvori man understreger vigtigheden af fortsat udvikling på området.[2] Strategiens vision er, at "Danmark skal gå forrest med ansvarlig udvikling og anvendelse af kunstig intelligens", og visionen er allerede nu udmøntet i en række signaturprojekter[3] og forskningsprogrammer på både nationalt og EU-plan.

Trods det politiske fokus og de mange tiltag mangler der imidlertid et overblik over, hvordan offentlige myndigheders brug af kunstig intelligens griber ind i menneskerettighederne. Spørgsmålet er kun sporadisk blevet adresseret i Danmark, og det offentliges forskellige digitaliseringsstrategier kommer ikke nærmere ind på, hvilke følger det kan få for borgernes retssikkerhed, at dele af sagsbehandlingen udføres af algoritmiske profileringsmodeller.

Både i EU, Europarådet og FN[4] er der udtrykt bekymring for, at kunstig intelligens kan underminere beskyttelsen af den enkeltes rettigheder. Brugen af profileringsmodeller kan fx skabe risici for diskrimination på en række forskellige måder.

Institut for Menneskerettigheder anser offentlige myndigheders brug af kunstig intelligens for at have stor betydning for borgernes rettigheder, og instituttets arbejde tager netop udgangspunkt i en rettighedsbaseret tilgang. Dette har som forudsætning, at det skal være muligt at få indblik i myndighedernes brug af profileringsmodeller og til en vis grad også indblik i, hvordan de virker.

Instituttet vurderer overordnet, at der er behov for regler om, at offentlige myndigheder skal påvise, hvordan de sikrer samtlige rettigheder og retsgarantier for borgerne, når de bruger algoritmiske profileringsmodeller. Dette gøres efter instituttets vurdering bedst ved gennemførelsen af såkaldte 'artificial intelligence'-konsekvensanalyser (AI-konsekvensanalyser), som vi redegør for i et beslægtet notat.[5]

FOKUS FOR NOTATET

Instituttet fokuserer i dette notat på den helt centrale menneskeretlige udfordring ved profileringsmodeller, at profileringsmodeller skaber risiko for diskrimination.

Notatet er tæt beslægtet men en række øvrige notater om det offentliges brug af profileringsmodeller, som der løbende henvises til nedenfor. Det er nemlig en central pointe, at risikoen for diskrimination i modellerne skal imødegås gennem mange forskellige tiltag, der vedrører styring, transparens og datakvalitet.

HVAD ER EN ALGORITMISK PROFILERINGSMODEL?

En algoritmisk profileringsmodel er en computerdrevet matematisk formel, som på baggrund af statistiske data eller variabler vurderer, om et objekt – eksempelvis en borger – har en bestemt egenskab, for eksempel om borgeren er berettiget til et tilskud.

Profileringsmodeller udvikles og trænes på baggrund af store mængder data om allerede afgjorte sager på området. Modellerne kan analysere og behandle data og finde sammenhænge og mønstre, der er langt mere komplekse, end hvad mennesker er i stand til, og de kan derfor være en stor ressource i sagsbehandlingen.

Profileringsmodeller kan både bruges til at træffe forvaltningsretlige afgørelser (fuldautomatisering) eller til at oplyse eller støtte myndighedens afgørelse (automatiseret beslutningsstøtte). Bruges modellen til beslutningsstøtte, kan dette enten ske for at afklare faktuelle omstændigheder som led i sagsoplysning, eller som støtte i den juridiske afvejning af en sag. Udfordringerne med transparens er til stede, uanset om modellen bruges fuldautomatiseret eller til automatiseret beslutningsstøtte.

Bruges modellen til beslutningsstøtte, kan dette enten ske for at afklare faktuelle omstændigheder som led i sagsoplysning eller som støtte i den juridiske afvejning af en sag. Udfordringerne med uigennemsigtighed er til stede, uanset om modellen bruges fuldautomatiseret eller til automatiseret beslutningsstøtte.

Profileringsmodeller er allerede under udvikling og i brug i det offentlige Danmark. Blandt andet har Styrelsen for Arbejdsmarked og Rekruttering udviklet og implementeret en model til forudsigelse af langtidsledighed blandt nyledige, og Gladsaxe Kommune udviklede i 2018 – men implementerede ikke – en model til forudsigelse af mistrivsel i børnefamilier. I begge tilfælde behandler modellen udvalgte data helholdvis om nyledige og om børnefamilier, og modellens resultater kan få betydning for den indsats, der igangsættes over for de identificerede ledige og børnefamilier.

DIREKTE OG INDIREKTE DISKRIMINATION I PROFILERINGSMODELLER

Risikoen for diskrimination eller ulovlig forskelsbehandling opstår, når en myndighed stiller borgeren ringere end andre på grund af beskyttede kendetegn som for eksempel køn, etnicitet eller seksuel orientering.

HVAD ER DISKRIMINATION?

Forbuddet mod diskrimination følger af en lang række internationale konventioner, Den Europæiske Menneskerettighedskonvention, EU-retten og dansk lovgivning. I lovgivningen skelnes der mellem direkte og indirekte diskrimination.

Direkte diskrimination indebærer, at en person på grund af et beskyttet kriterium som køn, etnicitet, religion eller tro, handicap, alder eller seksuel orientering behandles ringere end en anden i en sammenlignelig situation. Når det fremgår, at grunden til den ringere behandling skyldes et beskyttet kriterium, er der tale om direkte diskrimination eller ulovlig forskelsbehandling.

Indirekte diskrimination indebærer, at en tilsyneladende neutral bestemmelse, betingelse eller praksis stiller personer med en bestemt alder, handicap, køn, race eller etnicitet, religion, tro eller seksuel orientering ringere end andre personer. Der er dog ikke tale om indirekte diskrimination, hvis bestemmelsen er objektivt begrundet i et sagligt formål, og midlerne til at opfylde formålet er hensigtsmæssige og nødvendige.

En profileringsmodel kan føre til direkte diskrimination af en gruppe, hvis modellen bruger et beskyttet kendetegn i sin vurdering. Den kan for eksempel lægge vægt på personers køn, så det at være mand øger sandsynligheden for at blive klassificeret positivt, mens det at være kvinde øger sandsynligheden for at blive klassificeret negativt – eller omvendt.

Den enkleste metode til at forhindre direkte algoritmisk diskrimination er at "blinde" modellen ved at fjerne de beskyttede kendetegn, så de ikke optræder i det datasæt, modellen bliver trænet på. Man kan også benytte såkaldte "særskilte læringsprocesser"[6], hvor oplysningerne ikke fjernes, men modellen ikke bruger de variabler, der repræsenterer beskyttede kendetegn, selvom de er i datasættet. Begge løsninger fjerner risikoen for direkte diskrimination, men kan føre til, at modellen får en højere fejlrate og derfor bliver dårligere til at udføre det arbejde, som den var tiltænkt.

En profileringsmodel kan også føre til indirekte diskrimination. Det sker ved, at modellen i praksis behandler en gruppe anderledes end andre, selvom modellen ikke direkte lægger vægt på de forhold, der kendetegner den beskyttede gruppe. Indirekte diskrimination af en gruppe kan forekomme, selvom modellen er gjort blind over for gruppen for at beskytte mod direkte diskrimination. Det skyldes, at der i datasættet kan være systematiske sammenhænge, der fører tilbage til det beskyttede kendetegn – for eksempel en sammenhæng mellem bopæl og etnicitet eller mellem forældremyndighed og køn. Hvis modellen forskelsbehandler på

baggrund af bopæl eller forældremyndighed, som ikke i sig selv er beskyttede kendetegn, kan den dermed indirekte forskelsbehandle på baggrund af etnicitet eller køn, som er beskyttede kendetegn.

Det kan være svært overhovedet at opdage, at en profileringsmodel diskriminerer indirekte, da den kan gøre det på grundlag af sammenhænge, som er svære eller umulige for mennesker at få indsigt i. Dette skaber særlige bevismæssige udfordringer, som ikke behandles her.[7]

DISKRIMINATIONSRISIKO PÅ TRE NIVEAUER

Enhver borger har ret til en fair og uvildig behandling af sin sag i forvaltningen, og menneskeretten kræver, at borgeren ikke diskrimineres. Der opstår nye risici for diskrimination, når dele af afgørelsen overlades til en profileringsmodel, der analyserer borgernes oplysninger på måder, som de færreste kan forstå.

Hertil kommer, at når myndighederne gør brug af profileringsmodeller som led i vidtgående beføjelser – ofte over for borgere i en udsat position – har algoritmernes diskriminationsrisici en øget betydning.

Risikoen for diskrimination i profileringsmodeller har været erkendt og behandlet i forskningen i over et årti, og der findes i dag en omfattende litteratur, som udforsker tekniske løsninger, der forhindrer eller reducerer diskrimination i algoritmisk profilering.[8] Problemet, der rummer nogle principielle menneskeretlige dilemmaer, er altså på ingen måde ukendt.

Imidlertid er emnet ikke i særlig grad behandlet i en dansk kontekst og har ikke fået den fornødne bevågenhed i arbejdet med den øgede digitalisering af offentlige myndigheder. Dette er efter instituttets vurdering problematisk, da profileringsmodeller kan rejse alvorlige udfordringer i forhold til diskriminationsforbuddet.

Overordnet kan man tale om risiko for diskrimination på tre niveauer. I det nederste og enkleste niveau viderefører profileringsmodellen diskrimination fra det datasæt, den trænes på. I det næste niveau forstærker modellen risikoen for diskrimination ved at finde nye sammenhænge mellem data – og endelig finder vi øverst et niveau, hvor modellen i sig selv kan skabe en række nye risici for diskrimination.

Når myndighederne skal sikre overholdelsen af diskriminationsforbuddet, er det derfor afgørende, at de forstår – og afhjælper – de risici, der eksisterer på de tre niveauer. På samme måde er det vigtigt, at tilsyns- og kontrolmyndigheder efterprøver, om en profileringsmodel strider mod forbuddet mod diskrimination ud fra alle tre niveauer.

DISKRIMINATIONSRISICI I PROFILERINGSMODELLER

KORREKTE DATA AFGØRENDE FOR RETSSIKKERHEDEN

Den simpleste form for diskriminationsrisiko opstår, når profileringsmodellen trænes på baggrund af afgjorte sager, som på den ene eller anden måde diskriminerer.

Hvis der er diskriminerende praksis i datasættet, som modellen trænes på, vil modellen profilere ud fra denne diskriminerende praksis. Det behøver ikke dreje sig om egentlig diskrimination i afgørelserne – hvilket heldigvis meget sjældent forekommer – men kan skyldes, at der i datasættet foreligger færre eller ringere data om nogle grupper borgere end om andre. Sådan en forskel i datakvalitet kan føre til, at modellen i højere grad træffer forkerte afgørelser for disse borgere.

De fejl og mangler, der findes i de data, som modellen trænes på, videreføres dermed af modellen – et fænomen, der populært går under betegnelsen "garbage in, garbage out" – på dansk "affald ind, affald ud".

De afgørelser, som udvælges til modellens træning, vil udgøre en "grundsandhed" for modellen, og enhver usaglighed, forskelsbehandling eller misvisende udvælgelse af afgørelser vil afspejle sig i modellens resultater, da den vil bruge skævvredne data som eksempel på en korrekt – og lovlig – afgørelse.[9]

Vi anbefaler derfor, at myndighederne tidligt i forløbet og løbende under modellens brug forholder sig til datakvalitet i AI-konsekvensanalyserne, således at bl.a. udvælgelsen af datasæt ikke fører til diskrimination.

FORSTÆRKET FORSKELSBEHANDLING OG NEGATIVE FEEDBACKLOOPS

Profileringsmodellen kan "lære" en diskriminerende praksis, hvis den udvikles på baggrund af diskriminerende eller fejlbehæftet data – men den kan også, gennem analyser og behandling af data, selv forstærke den diskriminerende praksis.

Dette fænomen kan illustreres gennem et kendt forsøg, hvor forskere fra Cornell University i USA trænede en model til at identificere kønnet på personer i billeder. De brugte med vilje et datasæt, der overrepræsenterede kvinder i køkkener: af de billeder, der forestillede mennesker i et køkken, var to tredjedele kvinder, mens en tredjedel var mænd. Fordi køkkenet bag personen udgjorde en letgenkendelig og signifikant variabel, tillagde modellen det en betydning, der førte til, at den efterfølgende identificerende 84 pct. af alle personer i køkkener som kvinder – hvilket var langt mere end de 67 pct., der var i det datasæt, modellen blev trænet på.[10] Selv relativt små forskelle mellem grupper i et datasæt kan med andre ord føre til en model, som behandler grupperne forskelligt.

Forskelsbehandlingen kan blive forstærket yderligere, når modellens vurderinger efterfølgende bruges i myndighedens praksis. Her kan myndighederne uforvarende komme til at diskriminere på baggrund af modellens resultater, og når data om denne nye praksis inkluderes i modellens datasæt, vil det yderligere forstærke skævvridningen. Denne proces kaldes en negativ feedbacksløjfe, og den beskæftiger vi os mere indgående med i et notat om datakvalitet gennem hele modellens levetid.[11]

Et af de mest omdiskuterede eksempler på negative feedbacksløjfer handler om politiarbejde i USA. Her bruger politiet i nogle stater en algoritmisk model til at vurdere, hvordan de skal fokusere deres ressourcer. For eksempel udpeger algoritmer, hvilke områder politiet skal patruljere mere intensivt i.[12] Forskning peger imidlertid på, at de data, som modellerne trænes på, i mange tilfælde afspejler politiets langvarige diskrimination af sorte borgere. For eksempel er politipatruljer tilbøjelige til i højere grad at undersøge og arrestere sorte borgere. Hvis det er tilfældet, vil datasættet afspejle denne diskrimination, og man vil derfor træne en model med tendens til at vurdere, at politiets ressourcer bør fokuseres i områder med mange sorte borgere. Følger politiet herefter modellens vurdering, vil det føre til endnu flere interaktioner mellem politi og borgere i disse områder, og når data om denne praksis bruges til at opdatere modellens datasæt, vil det igen forstærke modellens tendens til at forskelsbehandle.[13]

Eksempler som disse viser, at offentlige myndigheder ikke blot skal sikre datakvalitet under udviklingen af en model, men også sikre sig mod diskriminerende praksis, når modellen er taget i brug. Fordi modellen til en vis grad bliver overladt til sig selv, hvad der jo netop er formålet med maskinlæring, kan den komme til at forstærke eksisterende diskriminationsrisici og skabe negative feedbacksløjfer. Disse mekanismer kan kun forebygges ved grundige og gentagne tests af modellerne.

Som nævnt foroven anbefaler vi derfor, at myndighederne periodisk forholder sig til risikoen for diskrimination i AI-konsekvensanalyserne, herunder den risiko der er for, at diskrimination kan blive forstærket ved modellens brug.

ALGORITMISKE DILEMMAER

De diskriminationsrisici, vi indtil nu har berørt, er alle sammenlignelige med risici, der også kunne være til stede uden brugen af profileringsmodeller. Her aktualiserer brugen af profileringsmodeller en risiko for diskrimination, hvor vi stadig er på kendt grund rent juridisk. Med den tredje og sidste type diskriminationsrisiko berører vi et område, hvor beskyttelsen mod diskrimination efter instituttets vurdering bliver udfordret på en ny måde.

En profileringsmodel har brug for utvetydige anvisninger for at kunne danne sine resultater og træffe en afgørelse. Hvis der i disse anvisninger tilføjes elementer, som har til hensigt at forebygge diskrimination, bliver modellen mere teknisk kompleks at udvikle og vedligeholde. Det betyder, at den kan komme til at få en højere fejlrate og dermed bliver dårligere til at udføre sit arbejde

Er modellen tænkt til at opspore mistrivsel i børnefamilier, kan man risikere, at den bliver dårligere til dette, hvis den for eksempel (også) skal undgå indirekte diskrimination. Dette skyldes, lidt forenklet, at de tekniske metoder til at forhindre indirekte diskrimination og metoderne til at sikre en effektiv og retvisende profileringsmodel ikke altid er forenelige.

Det kan også medføre, at modellen ikke er i stand til at sikre mod diskrimination af alle beskyttede befolkningsgrupper på samme tid, uden væsentligt tab af kvalitet i modellen.

Hertil kommer, at modellens behov for utvetydige anvisninger kan føre til, at der skal træffes et valg i modellens design mellem for eksempel beskyttelsen mod direkte versus indirekte diskrimination. Vælger man eksempelvis at blinde en model (se ovenfor), hvor man undlader beskyttede kendetegn som køn eller etnicitet, begrænser det muligheden for at kontrollere for indirekte diskrimination på baggrund af andre variabler, der hænger systematisk sammen med disse kendetegn. Det kan for eksempel være en sammenhæng mellem bopæl og etnicitet.

Det rejser selvsagt alvorlige problemer, da menneskeretten ikke giver mulighed for at træffe den slags valg. Efter menneskeretten skal samtlige beskyttede befolkningsgrupper altid beskyttes mod samtlige former for diskrimination.

Alle disse dilemmaer opstår, fordi sagsbehandlingen overlades til algoritmer, der sjældent kan dække det fulde anvendelsesområde for diskriminationsforbuddet og derfor indebærer visse kompromiser, når modellen skal designes. Løsningen er naturligvis ikke at fravige borgernes rettigheder, hvorfor det fra et menneskeretligt perspektiv er bydende nødvendigt, at der sættes rammer for det offentliges brug af profileringsmodeller og krav til, hvorledes modellerne designes.

Et af de væsentligste måder, det kan gøres på er ved tidligst muligt i udviklingsfasen at identificere de risici, som modellen rejser. Det anbefaler vi sker ved gennemførelsen af AI-konsekvensanalyser.

Herudover er det afgørende, at der er transparens i og om modellerne og deres risici. Det anbefaler vi sker gennem systemisk og algoritmisk transparens, som vi har beskrevet i et andet notat.[14]

Sidst men ikke mindst er det nødvendigt at både myndigheden, der beslutter at bruge en model, sagsbehandleren, der træffer afgørelsen, og tilsynet, der skal kontrollere myndighedens arbejde, alle besidder den fornødne tekniske og rettighedsmæssige forståelse for området. Også det har vi udfoldet i et separat notat.[15]

ANBEFALINGER: HUSK RETTIGHEDERNE

Danmark er verdensmester i digital forvaltning. Offentlige myndigheder er ambitiøse, når det gælder udviklingen af digitale løsninger– men de har ikke taget stilling til, hvordan disse løsninger skal afspejle den fortsatte sikring af menneskerettigheder og retsprincipper.

En rettighedsbaseret tilgang til offentlige myndigheders brug af profileringsmodeller er særligt vigtig, da det gælder afgørelser om borgere. Der kan være tale om afgørelser, som har stor indvirkning på borgernes retsstilling som for eksempel afgørelser om udsatte borgeres behov for støtte. Offentlige myndigheder er forpligtet til at handle sagligt og til at sikre beskyttelsen af borgernes rettigheder og retssikkerhed, også når sagerne automatiseres.

Det er afgørende, at offentlige myndigheder ikke uforvarende forstærker eksisterende uligheder ved at implementere teknologi, der ikke kan tage højde for rettighedernes fulde anvendelsesområde.

Institut for Menneskerettigheder efterlyser et større fokus på de mange rettighedsmæssige udfordringer, som udviklingen af en digital forvaltning stiller os overfor. Det indebærer klarere rammer for myndighedernes anvendelse af profileringsmodeller, både når det gælder datakvalitet og test men også når det gælder risici og begrænsningerne i, hvad modellerne er i stand til.

LÆS MERE I VORES RAPPORT

I rapporten 'Når algoritmer sagsbehandler – Rettigheder og retssikkerhed i offentlige myndigheders brug af profileringsmodeller' udgivet i oktober 2021 kortlægger Institut for Menneskerettigheder, hvordan borgernes rettigheder og retssikkerhed påvirkes, når offentlige myndigheder benytter sig af kunstig intelligens. Rapporten er en af de første, der tager udgangspunkt i danske myndigheders praksis og dansk lovgivning.

Dette notat gengiver dele af rapportens konklusioner.

Læs hele rapporten på menneskeret.dk/algoritmer.

SLUTNOTER

[1] Digitalstyrelsen (2020), Ny FN-måling: Danmark er fortsat verdensmestre i offentlig digitalisering, tilgængelig på: https://digst.dk/nyheder/nyhedsarkiv/2020/juli/ny-fn-maaling-danmark-er-fortsat-verdensmestre-i-offentlig-digitalisering/

[2] Finansministeriet og Erhvervsministeriet (2019): National strategi for kunstig intelligens tilgængelig på: https://www.regeringen.dk/media/6537/ai-strategi_web.pdf

[3] Digitalstyrelsen (2019), Kommuner og regioner skal afprøve kunstig intelligens for at løfte kvaliteten i den offentlige service, tilgængelig på: https://digst.dk/nyheder/nyhedsarkiv/2019/oktober/kommuner-og-regioner-skal-afproeve-kunstig-intelligens-for-at-loefte-kvaliteten-i-den-offentlige-service/

[4] Kommissionens hvidbog (2020), "On Artificial Intelligence - A European approach to excellence and trust" Tilgængelig på: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf , Europarådets resolution, CM/Rec(2020)1, tilgængelig på: https://rm.coe.int/09000016809e1154 , FNs Højkommissariat for Menneskerettigheder (2019), "FN's Specialrapportør for ekstrem fattigdom", rapport A/74/48037.

[5] Institut for Menneskerettigheder, (2021) "Klare rammer for det offentliges brug af profileringsmodeller", https://menneskeret.dk/udgivelser/klare-rammer-offentliges-brug-profileringsmodeller

[6] Lipton, Z. C., A. Chouldechova and J. McAuley (2018), "Does mitigating ML's impact disparity require treatment disparity?" Tilgængelig på: https://arxiv.org/pdf/1711.07076.pdf

[7]Institut for Menneskerettigheder, (2021) "Når algoritmer sagsbehandler: Transparens i og om profileringsmodeller brugt i det offentlige", https://menneskeret.dk/udgivelser/transparens-profileringsmodeller-brugt-offentlige. Se desuden Barocas, Solon and Selbst, Andrew D., "Big Data's Disparate Impact" (2016). 104 California Law Review 671 (2016), tilgængelig på: http://dx.doi.org/10.2139/ssrn.2477899

[8] Et ofte citeret, banebrydende studie i denne henseende er Pedreschi, D., S. Ruggieri and F. Turini (2008). Discrimination-aware data mining. Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM

[9] Barocas, Solon and Selbst, Andrew D., "Big Data's Disparate Impact" (2016). 104 California Law Review 671 (2016), tilgængelig på: http://dx.doi.org/10.2139/ssrn.2477899

[10] Zhao, J., T. Wang, M. Yatskar, V. Ordonez and K.-W. Chang (2017). "Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints.", tilgængelig på: https://arxiv.org/pdf/1707.09457.pdf

[11] Institut for Menneskerettigheder, (2021) "Når algoritmer sagsbehandler: Krav til data i profileringsmodeller brugt i det offentlige", https://menneskeret.dk/udgivelser/krav-data-profileringsmodeller-brugt-offentlige

[12] Perry, W. L. (2013)."Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations", tilgængelig på: https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf Ferguson, A. G, (2017) "The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement", Richardson, R., J. Schultz and K. Crawford (2018). "Dirty Data, Bad Predictions: How

Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice" tilgængelig på: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333423

[13] Ensign, D., S. A. Friedler, S. Neville, C. Scheidegger and S. Venkatasubramanian (2017). "Runaway Feedback Loops in Predictive Policing." tilgængelig på: https://arxiv.org/pdf/1706.09847.pdf

[14] Institut for Menneskerettigheder, (2021) "Når algoritmer sagsbehandler: Transparens i og om profileringsmodeller brugt i det offentlige", https://menneskeret.dk/udgivelser/transparens-profileringsmodeller-brugt-offentlige

[15] Institut for Menneskerettigheder, (2021) "Når algoritmer sagsbehandler: Klare rammer for det offentliges brug af profileringsmodeller", https://menneskeret.dk/udgivelser/klare-rammer-offentliges-brug-profileringsmodeller