

## Kunskapsöversikt om användningen och utvecklingen av automatiserad databehandling med algoritmer (artificiell intelligens) och stordata och diskriminering eller risker för diskriminering

## Innehåll

Sammanfattning .....	3
Inledning.....	6
Bakgrund .....	6
Syfte.....	9
Genomförande .....	10
Avgränsningar .....	11
Begrepp och definitioner .....	11
Resultat .....	12
Faktorer som kan leda till diskriminering eller risker för diskriminering vid automatiserat beslutsfattande.....	12
Att pröva eventuell diskriminering förutsätter transparens och möjlighet till förståelse .....	14
Rekrytering bygger allt mer på automatiserat beslutsfattande.....	16
Kreditgivning bygger allt mer på automatiserat beslutsfattande .....	20
Projektets iakttagelser.....	23
Avslutande diskussion.....	25
Möjliga åtgärder .....	25
Ordlista för centrala begrepp som används i denna översikt .....	27
Referenser.....	29

# Sammanfattning

## Bakgrund

Bakgrunden till denna kunskapsöversikt är att Diskrimineringsombudsmannen, DO, vill utveckla kunskap om hur automatiserad databehandling med algoritmer (artificiell intelligens) och stordata används och utvecklas inom samhällsområden som täcks av diskrimineringsförbudet enligt diskrimineringslagen.<sup>1</sup>

Standardiserade formler används alltmer i olika former av automatiserad databehandling med algoritmer (artificiell intelligens) för att välja ut eller bort människor i olika sammanhang. Samtidigt ökar användningen av en mängd data från olika källor – så kallad stordata – såväl nationellt, inom EU, som globalt. Debatten om hur stordata och användandet av algoritmer kan missgynna respektive gynna individer eller grupper av människor ökar i motsvarande takt liksom debatten kring riskerna för diskriminering. DO erfar också att företeelser som har koppling till att ansvariga aktörer använt sig av algoritmer, som kan leda till eller har lett till att individer utsatts för diskriminering eller risker för diskriminering, förekommer allt oftare i genomförandet av myndighetens uppdrag.

## Syfte

Användande av automatiserad databehandling med algoritmer (artificiell intelligens) och stordata kan leda till diskriminering av individer inom en rad områden som regleras i diskrimineringslagen. DO vill därför utreda sambandet mellan automatiserad databehandling och risker för diskriminering. Syftet är också att påverka utvecklingen av användande av automatiserad databehandling med algoritmer och stordata på sådant sätt att dessa praktiker inte leder till att individer som skyddas enligt diskrimineringslagen utsätts för diskriminering eller risker för diskriminering.

## Genomförande

Denna kunskapsöversikt bygger på en genomgång av forskning, debatt i olika tidskrifter och andra medier, DO:s rättspromemorior, intervjuer, samtal med enskilda aktörer och intresseorganisationer samt information från utbildningar, föreläsningar, konferenser, seminarier och workshops.

---

<sup>1</sup> Se ärende, LED 2018/387. Begreppet automatiserad databehandling med algoritmer används inledningsvis i direktiv och planering. För vår undersökning har vi valt att huvudsakligen använda begreppet automatiserat beslutsfattande. Bägge begreppen används dock här och ses som i stort överlappande varandra.

Fokus i översikten är användningen av automatiserat beslutsfattande vid rekrytering och kreditgivning.

## Resultat

Projektet har arbetat med att identifiera och föreslå områden där utvecklingen är av särskilt intresse ur ett diskrimineringsperspektiv. Denna kunskapsöversikt ger en bild av utvecklingen av användningen av automatiserat beslutsfattande med en fördjupning inom områdena rekrytering och kreditgivning.

Projektet ser att automatiserat beslutsfattande förekommer i olika former och omfattning inom rekrytering och kreditgivning liksom inom övriga samhällsområden som omfattas av diskrimineringslagen. Mer avancerade former av artificiell intelligens är än så länge relativt ovanliga i Sverige. Däremot är andra former av enklare automatiserat beslutsfattande redan vardag inom många områden och antalet nya tillämpningarna växer snabbt, något som kan påverka DO:s arbete med anmälningar, vägledning, tillsyn, processer och kunskapsspridning.

Projektet konstaterar att forskning och erfarenhet visar att olika inslag i automatiserat beslutsfattande kan riskera att leda till diskriminering. Det kan röra sig om felaktiga dataunderlag, snedvridna dataprogram som leder till missgynnande, bristande analys av resultaten av datakörningar, med mera. I en tilltagande debatt uppmärksammas dessa risker men det kommer också förslag på hur dessa brister kan övervinnas.

Projektets rättspromemoria om automatiserad databehandling med algoritmer och risker för diskriminering inom rekrytering och kreditgivning, har kommit fram till att om arbetsgivaren vid rekrytering använder ett digitalt verktyg eller en algoritm som behandlar information eller tar fram beslutsunderlag med stöd av automatisk databehandling eller artificiell intelligens, är det arbetsgivarens ansvar om en arbetssökande diskrimineras. Detta oavsett vilken insyn arbetsgivaren har i hur algoritmen fungerar. Vid tillhandahållande av tjänsten kreditgivning ligger ansvaret hos den fysiska eller juridiska person som tillhandahåller tjänsten. Detta oavsett vilken insyn tillhandahållaren av tjänsten har i hur den algoritm eller det program som algoritmen förekommer i fungerar.

Projektet har identifierat att det kan vara svårt för tillsynsmyndigheter att få tillräcklig insyn där automatiserat beslutsfattande används. Det kan i sin tur göra det svårt att identifiera risker för diskriminering liksom möjligheten att pröva om en individ har utsatts för diskriminering. Bristen på transparens kan bero på ovilja att redovisa grunderna bakom ett automatiserat beslut av

exempelvis affärsmässiga skäl. Det kan också finnas tekniska och juridiska hinder. Ett annat hinder kan vara bristande kompetens att tolka och förstå information om programmering och funktionssätt.

Projektets rättspromemorior visar dock att när det gäller rekrytering och kreditgivning finns det långgående rättigheter för DO att utifrån diskrimineringslagen begära ut information.<sup>2</sup> För enskilda finns det rättigheter till information i Dataskyddsförordningen (GDPR). Exakt vad dessa rättigheter innebär i praktiken är dock i vissa fall oklart.

#### Avslutande diskussion

Sammanfattningsvis ser projektet att DO bör ta hänsyn till följande för fortsatt problemformulering och åtgärder:

- Artificiell intelligens och automatiserat beslutsfattande förekommer i olika hög grad inom alla diskrimineringslagens utpekade samhällsområden.
- Användningen av artificiell intelligens liksom annat automatiserat beslutsfattande kan innebära en risk för diskriminering enligt diskrimineringslagen beroende på hur tekniken, verktygen och systemen programmeras, tränas, utvecklas och används.
- När artificiell intelligens och automatiskt beslutsfattande används inom rekrytering och kreditgivningen omfattas detta av diskrimineringslagen, DL. Ingenting tyder på att något annat skulle gälla för användning av artificiell intelligens och automatiskt beslutsfattande inom alla samhällsområden som omfattas av DL
- Det pågår utveckling av standarder, guider, manualer, praktiker på branschnivå, nationell, internationell och EU-nivå och (eventuell lagstiftning) med möjlighet för DO att försöka påverka utvecklingen.
- Tillsynsmyndigheter och individer behöver få tillgång till begriplig information för att kunna bedöma förekomsten av diskriminering. Möjliga svårigheter är hänvisningar till särskilda skäl, som exempelvis företagshemligheter samt den så kallade svarta lådan-problematiken.

---

<sup>2</sup> Se LED 2018/387, handling 37, Diskrimineringsombudsmannen (2019a) och handling 44 Diskrimineringsombudsmannen (2019b).

# Inledning

## Bakgrund

Under de senaste åren har tekniken för automatiserat beslutsfattande utvecklats och används allt mer inom alla de samhällsområden som omfattas av diskrimineringslagen.<sup>3</sup> Möjligheter att kombinera stora mängder av datakällor med förmågan att göra dataanalyser kopplas till stora förväntningar om framsteg. Både privata och offentliga organisationer, såväl i Sverige som inom EU, satsar allt mer resurser på att utveckla automatiserat beslutsfattande.<sup>4</sup> I Sverige förekommer än så länge jämfört med länder som USA, Kina och Storbritannien, mindre av avancerad tillämpning av artificiell intelligens.<sup>5</sup> Men andra, mindre avancerade former av automatiserat beslutsfattande sprids i allt större utsträckning. Inom flertalet av de granskade samhällsområdena är automatiserat beslutsfattande inte längre en framtidsvision utan vardag. En del av dessa nya tekniker anses av debattörer och forskare kunna minska diskrimineringen i samhället. Detta genom att ersätta mänskligt beslutsfattande med dess risk för fördomar och begränsad uppfattningsförmåga med datoriserad analys. Men DO har uppmärksammat att forskare och debattörer också ser att användningen av automatiserat beslutsfattande riskerar att leda till diskriminering.<sup>6</sup> Risker som tar sig nya uttryck och återfinns inom nya områden.

Ett fåtal anmälningar om misstänkt diskriminering, där automatiserat beslutsfattande är ett inslag, har kommit in till

---

<sup>3</sup> Se Underlag för beslut om prioritering av samhällsområden för fördjupade studier, LED 2018/387, handling 7.

<sup>4</sup> Exempel på sådana satsningar är etablerandet av Myndigheten för digital förvaltning, programsatsningar inom Vinnova, se Vinnova (2013, 2018), och Knut och Alice Wallenbergs stiftelses flermiljarderssatsning, Wallenberg Artificial Intelligence, Autonomous Systems, Software Programs, WASP, samt Europeiska kommissionens strategi för artificiell intelligens, Europeiska Unionen (2018a). Se även regeringens inriktning för artificiell intelligens, Näringsdepartementet (2018), sid 10, och Sveriges kommuner och landsting (2018).

<sup>5</sup> Intervju med forskargruppen Transparenta algoritmer i Försäkringsbranschen, RISE, KTH, 2019-02-26; möte med forskargrupp för intervjurobotar, Gabriel Skantze, Daniel Karlsson, KTH, 2019-03-22 samt Institutet för socialförsäkringar (2018a).

<sup>6</sup> Se exempelvis debattartiklar av Anna Felländer, med flera, där risker för olika behandling av män och kvinnor med artificiell intelligens lyfts fram, Dagens Industri (2017a, 2017b); Ny Teknik (2018), Sydsvenskan (2018). Många av de kritiska debattörerna och forskarna ser både fördelar och nackdelar med den nya tekniken men vill öka medvetenheten om möjliga risker.

Diskrimineringsombudsmannen i Sverige, liksom till motsvarande myndighet i Finland.<sup>7</sup>

Det är inte bara Diskrimineringsombudsmannen som uppmärksammat att artificiell intelligens och automatiserat beslutsfattande riskerar att leda till diskriminering. Även likabehandlingsmyndigheter i andra EU-länder liksom olika EU-organ som Europeiska kommissionen, ministerrådet och Europaparlamentet, uppmärksammar möjliga risker för diskriminering genom bland annat felaktig programmering, olämplig data, och oreflekterad analys och användning.<sup>8</sup> Ett exempel på en sådan växande medvetenhet är att likabehandlingsmyndigheternas samarbetsorganisation Equinet har påbörjat ett arbete med dessa frågor.<sup>9</sup> Ett annat exempel är FRA, EU:s organ för grundläggande rättigheter, som konstaterar att:

When algorithms are used for decision making, there is a potential for discrimination against individuals. The principle of non-discrimination, as enshrined in Article 21 of the Charter of Fundamental Rights of the European Union (EU) needs to be taken into account when applying algorithms to everyday life.<sup>10</sup>

Den svenska regeringen uppmärksammar också risker med automatiserat beslutsfattande och artificiell intelligens som kan leda till systematiska (medvetna såväl som omedvetna) fel vid insamlande av data, val av källor eller märkning av data.<sup>11</sup> Vidare uppmärksammar olika internationella organ såsom FN, OECD, World Economic Forum, med flera, frågor som rör etisk och hållbar artificiell intelligens och risker för diskriminering.<sup>12</sup> Allt fler etablerade såväl som nystartade NGO:s och tankesmedjor liksom olika människorättsorganisationer engagerar sig också i debatten.<sup>13</sup>

---

<sup>7</sup> Se Diskriminerings- och jämställdhetsnämnden (2018a); Diskrimineringsombudsmannens tillsynsrapport om SSAB, Diskrimineringsombudsmannen (GRA 2017/80).

<sup>8</sup> Se exempelvis FRA (2018a, b; 2019); Europaparlamentet (2017); Europeiska kommissionen (2019 och 2018); Council of Europe, Council of Europe Committee of Ministers (2019).

<sup>9</sup> Så har Equinet startat ett projekt kring AI, Equinet (2019a). Artificial intelligens och digitalisering är också prioriterat i organisationens framtidsplanering, se Equinet (2019b).

<sup>10</sup> FRA (2018a).

<sup>11</sup> Näringsdepartementet (2018); Europeiska kommissionen (2018).

<sup>12</sup> World Economic Forum, (2018).

<sup>13</sup> Se exempelvis Amnesty International, Access Now (2018); AI Now Institute (2018), AlgorithmWatch (2019a; 2019b); FRA (2018a, b); Europeiska kommissionen (2019); och Europaparlamentet (2017).

Inom forskningen uppmärksammas allt mer att automatiserat beslutsfattande kan leda till risker för diskriminering.<sup>14</sup>

Det är således många aktörer som diskuterar risker med ett automatiserat beslutsfattande. Flera av dessa aktörer arbetar samtidigt med att ta fram och föreslå lösningar. Exempel på lösningar som diskuteras är ökad transparens kring programmering och använd data, ökad användning av kunskap om diskriminering, uppföljning och kontroll samt certifiering. Ett exempel är de etiska riktlinjer som 2019 togs fram av Europeiska kommissionens expertgrupp för AI-frågor (High Level Expert Group on Artificial Intelligence (AI HLEG)).<sup>15</sup> I AI HLEG:s riktlinjer konstateras att området inte är oreglerat utan det finns en rad lagar som gäller. Riktlinjerna överensstämmer i mycket med vad som föreslås i andra initiativ. Europeiska kommissionen har uttalat sig om att de ser riktlinjerna som en viktig grund för fortsatt uppdateringen av medlemsländernas gemensamma plan för investeringar på artificiell intelligens.<sup>16</sup> Flera större aktörer som Microsoft, Google och andra företag liksom forskare från EU:s medlemsländer, har medverkat i framtagandet av riktlinjerna. De etiska riktlinjerna har kompletterats med en AI HLEG rapport på begäran från kommissionen som underlag för utveckling av EU:s kommande ramverk och policy – inklusive reglering och samordning av styrning.<sup>17</sup>

I AI HLEG:s etiska riktlinjer identifieras risker för sårbara grupper samt ojämlika makt- eller informationsövertag, som mellan arbetsgivare och anställda eller näringsliv och konsumenter. Riktlinjerna präglas av ett uttryckligt systemperspektiv, där både teknik, sociala, juridiska och samhällseliga aspekter omfattas. De etiska åtagandena ska grundas på grundläggande mänskliga rättigheter. För att förverkliga riktlinjerna behövs enligt AI HLEG både tekniska metoder som icke-tekniska åtgärder som exempelvis utbildning och forskning. AI HLEG menar att det inte räcker med allmänna etiska riktlinjer. Dessa måste anpassas till de konkreta förutsättningarna inom olika branscher. De etiska riktlinjerna föreslås därför konkretiseras i branschspecifika checklistor. Arbetet med sådana har startat.

---

<sup>14</sup> Se översikten från AlgorithmWatch Institute, AlgorithmWatch (2019a) och Mittelstand, med flera (2016), Raso, med flera (2018). I Sverige finns verksamheter vid universiteten i Lund, Umeå, Linköping och Stockholm samt KTH.

<sup>15</sup> För mer information, se webbplatsen för Europeiska kommissionen: <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence> samt rapporten Europeiska kommissionen (2019a).

<sup>16</sup> Se Europeiska kommissionen (2019a).

<sup>17</sup> Se Europeiska kommissionen (2019b).



Olika debattörer menar att flera offentliga och kommersiella satsningar på artificiell intelligens och automatiserat beslutsfattande brister när det gäller att reflektera kring och motverka risker med artificiell intelligens, algoritmer och stordata, inte minst diskrimineringsrisker.<sup>18</sup> Därför diskuteras nödvändigheten av att etablera kontroll- och styrmekanismer för att hindra oetiska och diskriminerande användningar av automatiserat beslutsfattande. Det handlar om frivilliga riktlinjer från branschorganisationer men det förekommer också diskussioner såväl i Sverige som inom EU om att det kan finnas behov av ny lagstiftning kring artificiell intelligens och dess användningar.<sup>19</sup> En del debattörer menar dock att det finns goda möjligheter att nyttja existerande lagstiftning – som exempelvis diskrimineringslagar.<sup>20</sup>

Oavsett om det blir ny lagstiftning eller ej, menar professor Zuiderveen Borgesius, författare till en rapport till Europarådet, att i olika EU-länder finns behov av att stärka existerande tillsynsmyndigheters kompetenser och resurser att övervaka automatiserat beslutsfattande och eventuellt även inrätta helt nya tillsynsmyndigheter för att få tillräcklig kompetens för att utöva tillsyn.<sup>21</sup>

Mot denna bakgrund har DO tagit fram denna promemoria för att övergripande beskriva utvecklingen av användningen av automatiserat beslutsfattande, med ett fördjupande fokus inom områdena rekrytering och kreditgivning.

## Syfte

Användande av automatiserad databehandling med algoritmer (artificiell intelligens) och stordata riskerar att leda till diskriminering av individer inom en rad områden som regleras i diskrimineringslagen. DO vill därför utreda hur automatiserad databehandling med algoritmer (artificiell intelligens) och stordata används och utvecklas inom dessa områden samt sambandet mellan automatiserad databehandling och risker för diskriminering. Detta för att bidra till att användningen av automatiserad databehandling med algoritmer och stordata inte leder till att individer

---

<sup>18</sup> Dagens Industri (2017a, b); Ny Teknik (2018), Sydsvenskan (2018), samtal med Anna Felländer 2018-11-21, föreläsning med Jen Gennai, chef för ansvarig innovation vid Google, 2019-05-23.

<sup>19</sup> Zuiderveen Borgesius (2018), Greenstein (2017), SOU (2018:25).

<sup>20</sup> I Allen och Masters (2019) konstateras att det saknas särskild lagstiftning kring artificiell intelligens och diskriminering i europeisk lagstiftning. Men författarna menar att "the existing equality framework" väl klarar av att hantera diskriminerande teknologi. Det finns dock behov av att säkra en rätt till meningsfull insyn i funktionen hos algoritmer.

<sup>21</sup> Zuiderveen Borgesius, (2018), FRA (2019) och Wahlgren (2018).

utsätts för diskriminering eller risker för diskriminering. Av särskilt intresse är att utvidga Diskrimineringsombudsmannens stöd till arbetsgivare i sitt arbete med aktiva åtgärder.<sup>22</sup>

## Genomförande

Inom ramen för projektet har det i en modul, modul 1, tagits fram denna översiktspromemoria samt rättspromemorian Tillgång till information som förutsättning för att bedöma förekomsten av diskriminering i samband med automatiserat beslutsfattande.<sup>23</sup> Detta utgör slutredovisningen av denna modul. I ytterligare en modul, projektets modul 2, har en rättspromemoria om diskriminering och automatiserat beslutsfattande inom områdena rekrytering och kreditgivning tagits fram.<sup>24</sup>

Projektet har tagit del av information från webbplatser, medier, forskningsartiklar, rapporter, böcker, utredningar, promemorior, nätverksträffar samt underlag från tankesmedjor och olika bransch- och andra intresseorganisationer. Dessutom har vi intervjuat forskare och företrädare för intresseorganisationer, myndigheter samt debattörer när det gäller artificiell intelligens, digitalisering, automatiserat beslutsfattande och förhållanden inom de samhällsområden och branscher, främst rekrytering, och kreditgivning, som vi har undersökt. Vi har även tagit del av utbildningar, studiebesök, seminarier, konferenser och mässor. Metodmässigt har vi arbetat med en explorativ ansats där intervjuade har tipsat om intressanta andra källor och miljöer. Vi har undersökt förekomsten av automatiserat beslutsfattande inom olika samhällsområden, men fokus har legat på företeelserna rekrytering och kreditgivning. Huvuddelen av arbetet har bestått i att samla in existerande kunskap för att systematisera och analysera dessa utifrån våra behov.

Projektgruppen har bestått av Ragnar Andersson, projektledare ANA, Carl Ekström, ANA, Arja Lehto, ANA, Mira Hjelm Östh, Rättsenheten arbetsliv, Julia DeMarinis Giddings, Processenheten, Christina Jönsson, Rättsenheten samhällsliv och utbildningsväsende, som ersatte tjänstlediga Patricia Isaksson Rivas, samma enhet, Ammar Makboul, Utvecklingssektionen samt Anna Gustavsson, Kommunikation.

---

<sup>22</sup> Se direktiv LED 2018/387, Diskrimineringsombudsmannen (2018).

<sup>23</sup> Rättspromemoria, Tillgång till information som förutsättning för att bedöma förekomsten av diskriminering i samband med automatiserat beslutsfattande.

(Diskrimineringsombudsmannen (2019b) Se LED 2018/387, handling 44. För information om projektets projektplan, se LED 2018/387, handling 3.

<sup>24</sup> Se Diskrimineringsombudsmannen, 2019b och LED 2018/387, handling 37.

## Avgränsningar

Översikten beskriver utvecklingen i Sverige och tekniker som används eller har ansetts nära att tas i bruk. Där vi bedömer att det är relevant beskriver vi även utvecklingen i andra länder. Fokus ligger på användning av automatiserat beslutsfattande som kan riskera att leda till diskriminering enligt diskrimineringslagen, i synnerhet inom områdena rekrytering och kreditgivning.

I en första fas av detta arbete genomförde vi en översiktlig kartläggning av olika samhällsområden utifrån diskrimineringslagen för att göra ett urval av områden för fördjupade studier. Förslaget till urval gjordes utifrån kriterier om förekomsten av automatiserat beslutsfattande, hur många som berörs inom samhällsområdena, risker för allvarlig skada vid missgynnande samt hur sannolikt det är med diskriminering. Detta resulterade i att samhällsområden som hälso- och sjukvården samt socialtjänst blev högt prioriterade. Trots detta valdes dessa områden bort från vår undersökning. Detta då dessa områden kännetecknas av komplicerade styrningsförhållanden som projektet inte bedömdes ha tid att kartlägga.<sup>25</sup> Utifrån resultatet fattades beslutet att vår studie skulle fokusera på vår fördjupade studie till:<sup>26</sup>

- Samhällsområdet Arbetsliv, delområde: Rekrytering.
- Samhällsområdet Varor, tjänster och bostäder; delområde: Kreditgivning.
- Samhällsområdet Varor, tjänster och bostäder; delområde: Försäkringsbranschen.

Av resursskäl gick projektet inte vidare med Försäkringsbranschen.

## Begrepp och definitioner

I vår översikt fokuserar vi på det vi har valt att benämna automatiserat beslutsfattande vilket för oss innebär urval och sortering av data med hjälp av algoritmer och datorstöd. När vi använder begreppet automatiserat beslutsfattande omfattar det processer där beslut fattas helt automatiserat och såväl som sådana där människor fattar de slutliga besluten med hjälp av automatiserade beslutsstöd.<sup>27</sup> När det uttalat handlar om helt automatiserat

---

<sup>25</sup> För en diskussion kring medicin och artificiell intelligens, se exempelvis kunskapsöversikten AI Sustainability Center (2019). Se även Modlitba (2018) och Sjöberg (2016). Försäkringsbranschen är också intressant.

<sup>26</sup> Se LED 2018/387, Underlag för beslut, handling 7 och beslut om fördjupningsområden, LED 2018/387, handling 8.

<sup>27</sup> Denna definition överensstämmer med praxis inom stora delar av EU, exempelvis i Europeiska kommissionens rapporter samt även med formuleringar i GDPR.

beslutsfattande eller processer för beslutstöd, så skrivs detta ut. (För mer definitioner, se begreppslista i slutet av studien).

Det förekommer många begrepp som kan kopplas till automatiserat beslutsfattande. Ett är artificiell intelligens. Andra är begrepp som stordata, digitalisering, robotisering, RPA, maskinlärande, med flera.<sup>28</sup> Dessa olika begrepp står för dataprogram med olika grader av komplexitet. Det handlar om allt från relativt enkla program med förprogrammerade och spårbara arbetssätt såväl som mer avancerade självlärande mjukvaruprogram.<sup>29</sup> De enklare formerna innehåller automatiserat beslutsfattande där datorprogram med algoritmer läser in beslutsunderlag och gör enkla urval och sorteringar. Programmering, underlag och funktionssättet hos inblandade algoritmer är relativt lätta att beskriva och förstå för lekmän och externa bedömare. De mer avancerade formerna omfattar vad som brukar sammanfattas med beteckningen artificiell intelligens, sådant som stordata, databrytning, maskinläring, djuplärande eller liknande. Det handlar här ofta om datorprogram som självständigt gör kopplingar och ser mönster och samband mellan olika data från en rad större datakällor.<sup>30</sup> Det kan vara svårt även för datorprogrammets programmerare att beskriva hur och varför dessa program drar vissa slutsatser. Detta brukar kallas Svarta lådan-problemet, vilket försvårar för eventuella tillsynsmyndigheter att pröva om någon utsatts för diskriminering i juridisk mening.<sup>31</sup>

## Resultat

### Faktorer som kan leda till diskriminering eller risker för diskriminering vid automatiserat beslutsfattande

Det finns en rad olika faktorer som riskerar att leda till diskriminering vid automatiserat beslutsfattande. Nedan räknas några av dessa upp.

En sådan faktor är användandet av icke-relevant dataunderlag. Kvaliteteten och rättssäkerheten vid en automatiserat beslutsfattande blir inte större än vad processens underlagsdata ger möjlighet till, risker som ökar i och med en delvis oreglerad växande marknad för information. När information säljs

---

<sup>28</sup> För en närmare förklaring av olika begrepp, se ordlistan i slutet av rapporten.

<sup>29</sup> Se exempelvis diskussionen i AlgorithmWatch (2019b).

<sup>30</sup> Se beskrivning av maskinlärande i ordlista i rapportens slut.

<sup>31</sup> Se ordlista samt Zuiderveen Borgesius, (2018); Greenstein (2017, 2018); Naarttjärvi (2017).

och byts finns risker för att det blir svårare att garantera kvaliteten på informationen.<sup>32</sup>

Icke-relevant data kan också påverka den så kallade träningen av maskinlärande datorprogram. De datamodeller som skapas vid maskinlärandet kan riskera att leda till diskriminering när de används. Även sådant som hur man bestämmer sig för att vikta betydelsen av olika variabler samt val av klassificeringar och bedömningskriterier kan påverka.

Olämplig (biased) data kan uppstå vid datainsamlingar om sökkategorier och kriterier för urval formuleras på ett sådant sätt att de bidrar till diskriminering.

Ytterligare faktorer är medvetet eller omedvetet vinklade val av data, eller av andra skäl bristande kvalitet på data.<sup>33</sup> Risker som ökar i och med en delvis växande och oreglerad informationsmarknad. När information säljs och byts kan det bli svårare att garantera kvaliteten på informationen.<sup>34</sup>

En annan faktor är att det vid träning av självlärande dataprogram används i sig korrekt data som dock baserar sig på samhällliga strukturer som leder till ojämlikhet och ojämställdhet – exempelvis traditionella yrkesval och karriärer för kvinnor och män. Ett självlärande datorprogram som använder sådan data kan resultera i att förhållandena reproduceras.

Det kan också konstateras att det finns brister som beror på att programmerare inte har kunskaper om vad som eventuellt kan riskera att leda till diskriminering. Vidare kan programmerare bära på omedvetna fördomar eller medvetet vinkla programmeringen med diskriminerande resultat. Detta diskuteras i en forskningsrapport från det tvärvetenskapliga forskningscentret, AI Now Institute, vid New York University.<sup>35</sup>

En ökad standardisering av olika program för automatiserat beslutsfattande innebär att olika delprogram från skilda underleverantörer byggs samman till paketlösningar. I många fall kan då även den som säljer den sammansatta paketlösningen ha svårt att fullt ut ha insyn i underleverantörernas programmeringar och om dessa riskerar att leda till diskriminering.

Ett förhållande som kan öka risken för att diskriminering vid automatiserat beslutsfattande inte upptäcks är bristande uppföljning och kontroll av

---

<sup>32</sup> Se Rieke, med flera (2016).

<sup>33</sup> Ibid; Barocas och Selbst (2016).

<sup>34</sup> Myers, med flera (2019).

<sup>35</sup> Ibid.

beslutens utfall. När automatiserat beslutsfattande används i vetenskapliga sammanhang görs oftast olika former av uppföljning av resultaten. Exempelvis så kontrolleras inom sjukvård hur effektiv automatiserad diagnostisering är genom att kontrollera denna mot manuell diagnostisering med kontrollgrupper. Vid användningar av automatiserat beslutsfattande som vid kreditbedömning och rekrytering förekommer sällan motsvarande uppföljningar av exempelvis diskriminerande effekter.<sup>36</sup>

Användningen av underlagen från automatiserat beslutsstöd kan om människor oreflekterat litar på informationen från datorkörningar riskera att leda till diskriminering. Detta brukar kallas automatiseringsbias. (automation bias).

Det kan också behövas samhällsvetenskapliga kunskaper eller juridisk expertis för att fullt ut kunna analysera resultaten från automatiserat beslutsfattande.

Tillgången till stordata gör det lätt att ta fram olika jämförelsegrupper, eller så kallade proxygrupper.<sup>37</sup> Sådana statistiska dubbelgångargrupper delar olika egenskaper och beteenden och används för att förutsäga personers beteenden. Det kan exempelvis handla om att jämföra en ung man när det gäller eventuell kreditrisk med andra unga män som delar en rad egenskaper som ålder, kön, utbildning, aktiviteter på sociala medier och tidigare återbetalningar. Jämförelsegruppen kanske som helhet bedöms innebära en hög kreditrisk. Men oavsett om en person delar flera egenskaper med en sådan statistisk jämförelsegrupp så förutsätter diskrimineringslagen, med vissa undantag, en individuell bedömning. Det kan finnas individer som faktiskt innebär en låg kreditrisk. Genom att utan individuella hänsyn då bedömas enbart utifrån jämförelsegruppens resultat kan innebära ett missgynnande.<sup>38</sup>

## Att pröva eventuell diskriminering förutsätter transparens och möjlighet till förståelse

För att upptäcka och motverka eventuell diskriminering i samband med automatiserat beslutsfattande krävs transparens när det gäller kodning, data och tolkning. Detta understryks av en rad debattörer. Såväl enskilda individer som tillsynsmyndigheter har behov av insyn. Det finns flera

---

<sup>36</sup> O'Neill, C (2017); Europeiska kommissionen (2019). Se här även Institutet för socialförsäkringars rapporter (2018a) och (2018b). I rapporten 2018a redovisas olika risker för diskriminering i samband med profilering.

<sup>37</sup> Prince och Schwarcz, D (2020).

<sup>38</sup> Detta ansågs ligga bakom utslaget kring konsumentkreditfallet i Finland som diskuteras på sidan 22 och framåt i denna kunskapsöversikt.

möjliga tekniska, juridiska såväl som resursmässiga hinder för transparens.<sup>39</sup> Ett resursmässigt hinder är brist på kompetens för att tolka algoritmer och programkod. Detta gäller såväl för enskilda som för tillsynsmyndigheter. För en del självlärande algoritmer finns också svårigheter att klargöra hur dessa kommer fram till sina resultat. Detta gäller även den eller de som programmerat det självlärande programmet.<sup>40</sup> Men även där det finns kompetens nog att förstå och tolka algoritmer och programkod kan detta var otillräckligt. Det kan också behövas kunskap om att diskriminering kan uppstå samt ett juridiskt kunnande för att avgöra om en algoritm kan leda till diskriminering i juridisk mening.

För att öka möjligheten till insyn behövs det kompetens i att förklara artificiell intelligens så att även icke-expertter kan förstå.<sup>41</sup> Dessutom behövs tydliga lagar och regler. Här kan Dataskyddsförordningen få en viktig roll bland annat när det gäller olika former av profilering.<sup>42</sup> Det finns dock frågetecken kring användningen av Dataskyddsförordningen (GDPR) och om vad dess uppgiftsskyldigheter egentligen innebär i praktiken.<sup>43</sup>

En rättspromemoria som tagits fram av Diskrimineringsombudsmannen visar att baserat på diskrimineringslagen har ett tillsynsobjekt skyldighet – med vissa undantag – att lämna ut efterfrågade uppgifter om algoritmers eller programs uppbyggnad. Men hur långt dessa skyldigheter sträcker sig och vem som bestämmer om informationen är begriplig, är ännu oklart.<sup>44</sup>

---

<sup>39</sup> Patent och registreringsverket (2019); New Economics Foundation (2018); Larsson, (2018); Watcher, med flera (2017).

<sup>40</sup> Zuiderveen Borgesius (2018); Naartijärvi (2017); Greenstein (2018); Intervjuer med forskargruppen Transparenta algoritmer i Försäkringsbranschen, RISE, KTH, 2019-02-26, Ulrik Franke och Jacob Dexe, samt med Gabriel Skantze, KTH, 2019-03-22.

<sup>41</sup> Förståbar artificiell intelligens är ett växande forskningsområde. Bland annat forskas kring hur man kan förstå och visualisera algoritmers arbete och resultat. Se också Allen och Masters (2019) om behovet av meningsfull information om hur algoritmer fungerar.

<sup>42</sup> Se Dataskyddsförordningen (2019). Profilering står här för varje form av automatisk behandling av personuppgifter då uppgifterna används för att bedöma vissa personliga egenskaper, i synnerhet för att analysera eller förutsäga personens arbetsprestationer, ekonomiska situation, hälsa, personliga preferenser, intressen, pålitlighet, beteende, vistelseort eller förflyttningar.

<sup>43</sup> För en redovisning av legala aspekter av transparens, se den rättsliga undersökningen, Diskrimineringsombudsmannen, (2019b). Dataskyddsförordningens (GDPR:s) term automatiserat beslutsfattande verkar överensstämma med vår användning.

<sup>44</sup> Rättspromemoriorna om automatiserad databehandling med algoritmer och risker för diskriminering, Diskrimineringsombudsmannen (2019 a; b). I en artikel, Allen och Masters (2019) argumenterar författarna för att en kändare kan få stöd för en stämning om eventuell diskriminering på grund av brist på transparens. Detta utifrån rättsfallet, Case - 109/88, Handels och Kontorsfunktionäerernes Forbund i Danmark v Dansk Arbejdsgiverforening Exp. Danfoss A/S EU:1989: 383. Danfoss EU:C1989:383, där ett icke-transparent lönesystem innebär att arbetsgivaren fick ansvaret för att bevisa att detta inte diskriminerade. Samma princip bör enligt Allen och Masters gälla även för annan eventuellt diskriminerande teknik, som artificiell intelligens och automatiserat beslutsfattande.

Vid tillsyn kan det eventuellt räcka med en innehållsdeklaration över använda datakällor och en översiktlig beskrivning av hur programmet arbetar för att kunna göra en bedömning. Finansinspektionen har vid en tillsyn prövat att få ut information om ett automatiserat beslutsfattande. Myndigheten fick först ut information i kodad form som den inte kunde tolka. Men i en dialog med tillsynsobjektet kom Finansinspektionen så småningom fram till en beskrivningsnivå som var tillräcklig för att kunna genomföra tillsynen.<sup>45</sup> Genom olika tillsyner från olika myndigheter kan framöver utvecklas en praxis för meningsfull och förståelig information från tillsynsobjekt kring använda automatiska beslutsfattanden som inbegriper algoritmer. Något som kan komma att påverka även DO tillsynsmöjligheter.

## Rekrytering bygger allt mer på automatiserat beslutsfattande

Ovan har allmänt behandlats risker som kan leda till diskriminering i samband med automatiserat beslutsfattande. Här redovisas en fördjupad studie av risker i samband med rekrytering. Flera av de tidigare nämnda faktorerna återfinns även här.

Rekryteringsprocesser omfattar flera moment där automatiserat beslutsfattande kan förekomma. Rekrytering omfattas av diskrimineringslagens förbudsdel, kapitel 2, såväl som dess föreskrifter om aktiva åtgärder, kapitel 3. Det innebär att vid rekrytering – och då även med automatiserat beslutsfattande – är en presumtiv arbetsgivare ansvarig för att undvika eventuell diskriminering oavsett vilka digitala verktyg eller algoritmer som använts. Likaså ska arbetsgivaren med aktiva åtgärder löpande och systematiskt arbeta förebyggande för att motverka diskriminering och för att främja lika villkor. Ansvar gällande även för diskriminering som orsakas vid användningen av automatiserat beslutsfattande av exempelvis underleverantörer i flera led eller andra utförare. Detta oavsett om den presumtiva arbetsgivaren vet om eller har insyn i eller förstår hur de verktyg som använts fungerar.<sup>46</sup> Arbetsgivaren är vid en tillsyn enligt diskrimineringslagen skyldig att redovisa hur eventuella verktyg för automatiserat beslutsfattande fungerar – till exempel uppbyggnaden av en algoritm. Det finns dock vissa undantag från uppgiftsskyldigheten i diskrimineringslagen.<sup>47</sup>

---

<sup>45</sup> Intervju med Finansinspektionens Innovationscenter 2019–04–09.

<sup>46</sup> Se Rättspromemorian, Diskrimineringsombudsmannen (2019a). Det har också klargjorts att så kallade headhunters eller rekryteringsföretag som anlitas för att hjälpa till vid en anställning likställs med att det är arbetsgivaren som handlar. Detta anses också gälla vid anlitaandet av en robot eller ett datorprogram.

<sup>47</sup> Det finns dock vissa undantag från den informationskyldigheten. Se Diskrimineringsombudsmannen (2019b) för ett resonemang om Prop. 2007/08:95: Om det i



Användningen av olika former av automatiserat beslutsfattande skiljer sig således inte från användningen av andra verktyg eller tjänster vid rekrytering när det gäller ansvar och förpliktelser.

En översikt över rekryteringsprocesser innefattar ofta moment som: kartläggning av anställningsbehov; framtagande av kravprofiler; annonsering/ matchning/sourcing; urval av sökande genom sällning/skanning av ansökningar; intervjuer; tester och eventuellt granskningar på sociala medier; slutliga intervjuer och beslut om anställning.

En del bedömare anser att det går att se en förskjutning när det gäller urvalet av sökande där ett fokus på tidigare meriter ersätts av olika framtidsinriktade bedömningar av sökandes potentiella arbetsprestationer.<sup>48</sup> Flera av de faktorer som kan leda till diskriminering som nämns ovan återfinns också vid rekrytering med automatiserat beslutsfattande, artificiell intelligens och stordata.

Automatiserat beslutsfattande används i större eller mindre omfattning inom i stort sett alla moment som kan förekomma i en rekryteringsprocess. Och antalet tillämpningar ökar. Nedan följer exempel på hur automatiserat beslutsfattande används inom några sådana moment.

#### Kravprofiler och matchning

Kravprofilering inför en rekrytering spelar en avgörande roll i rekryteringsprocessen och kan numera genomföras med stöd av automatiserat beslutsfattande.<sup>49</sup> Data för kravprofileringen samlas in från exempelvis uppföljningar av anställdas prestationer, chefsutvärderingar, olika personlighetstest, löneutveckling och karriär. Sådan data används för att skapa kravprofiler som avsiktligt eller oavsiktligt kan leda till diskriminering.

Ett omtalat exempel på detta är Amazons försök med en rekryteringsrobot som visade sig gynna män. Roboten arbetade med maskinlärning och tränades framförallt på cv:n från i branschen historiskt sett framgångsrika tekniker, vilket visade sig vara män. Försök gjordes att korrigera

---

det enskilda fallet finns särskilda skäl som talar emot, exempelvis om de uppgifter som begärts ut av Diskrimineringsombudsmannen gäller utrike-tjänsten, försvaret, affärs- eller driftsförhållanden eller andra uppgifter där motstående intressen är så starka att uppgiften inte bör lämnas ut.

<sup>48</sup> Intervju med vd, Manne Didehvar, Sveriges HR-förening, 2019-04-25; Svenskt Näringsliv (2018).

<sup>49</sup> Intervju med vd, Manne Didehvar, Sveriges HR-förening, 2019-04-25 samt inlägg i olika branschbloggar och tidskrifter.

snedrekryteringen genom att ta bort hänvisningar till kön. Men det självlärande, maskinlärande, datorprogrammet fortsatte att i sin analys identifiera samma grupp av anställda som framgångsrika. Men denna gång kopplade programmet framgången till mer indirekta faktorer som exempelvis tekniska manligt dominerade utbildningsinstitutioner. Trots flera korrigeringsförsök lyckades inte Amazons utvecklare komma till ett läge där de kunde garantera att det inte förekom en snedrekrytering utifrån kön. Projektet lades ner.<sup>50</sup>

Detta och liknande resultat har öppnat för en diskussion om att risker för diskriminering kan befastas om kravprofiler bygger på existerande förhållanden och att rekryteringsverktyg i stället bör bygga på normativa modeller baserade på hur det borde se ut i arbetslivet.<sup>51</sup>

Automatiserat beslutsfattande kan användas för att ge en snabb och nischad marknadsföring samt för att välja målgrupper, kanaler och för att utforma annonser. Beroende på hur detta görs kan det öka riskerna för exkludering av olika grupper liksom minska möjligheterna till transparens.<sup>52</sup> Kunskapen om eventuell omfattning av exkludering verkar vara mycket begränsad.

En annan möjlighet vid rekrytering är att matcha arbetssökandes profiler med arbetsgivares kravprofiler i olika databaser och jobbplattformar. En ny företeelse är så kallad sourcing där rekryterare proaktivt förbereder för matchning genom att bygga upp databaser med uppgifter om lämpliga kandidater för framtida rekryteringar.<sup>53</sup> I båda fallen är automatiserat beslutsfattande en förutsättning.<sup>54</sup> Både när det gäller matchning och sourcing finns risk för att förfarandet leder till diskriminering och att arbetssökande riskerar att missgynnas genom att väljas bort av osakliga skäl.

---

<sup>50</sup> Intervju med forskargruppen för Transparenta algoritmer i Försäkringsbranschen, RISE, KTH, 2019-02-26, Se även rättspromemoria om automatiserad databehandling med algoritmer och risker för diskriminering, Diskrimineringsombudsmannen (2019a), SVT Nyheter (2018) och The Guardian (2018).

<sup>51</sup> Se exempelvis Rieke, med flera (2018); Myers West, med flera (2019); samt Person (2016).

<sup>52</sup> Detta omfattas inte av den svenska diskrimineringslagen. I exempelvis USA så finns dock reglering gällande detta.

<sup>53</sup> Sourcing är ett begrepp från inköpsbranschen som betyder att hitta leverantörer. Har i rekryteringsbranschen kommit att beteckna att hitta, kontakta och engagera möjliga kandidater för framtida rekryteringar. Alternativt svenskt begrepp verkar saknas. Se gärna utredningen om Arbetsförmedlingens stöd till arbetssökande som inte använder digitala tjänster, Inspektionen för arbetslöshetsförsäkringen (2018).

<sup>54</sup> Exempel här är LinkedIn, Monster, Blocket, Facebook Jobs, Jobsafari. Men det finns många fler. Dataskyddsförordningen påverkar här möjligheterna till att spara information och använda i nya sammanhang. Se Diskrimineringsombudsmannen (2019b).

Urval av arbetssökande

#### **Urval utifrån arbetssökandes ansökningar**

Automatiserat beslutsfattande kan användas för en första sällning utifrån formella kriterier och skullkrav när det gäller ansökningar och cv: n. Detta moment anses vara tids- och resurskrävande och stöd i form av automatiserat beslutsfattande ses som resursbesparande. Utformningen av kravprofiler blir viktig.

#### **Intervjuer**

Intervjuer är vanliga inslag vid rekrytering och görs allt mer med hjälp av standardiserade frågeverktyg. Flera marknadsaktörer menar att mer av standardisering och strukturerade intervjuer kan minska riskerna för diskriminering.<sup>55</sup> Standardiserade intervjumallar möjliggör utvecklingen av automatiserat beslutsfattande i form av så kallade intervjurobotar.<sup>56</sup> Intervjuerna kan ske såväl skriftligt som muntligt, med direkt närvaro eller via exempelvis videoinspelningar eller internet. Användning av intervjurobotar gör det möjligt att genomföra ett stort antal intervjuer med olika kandidater. Sådana intervjuer genomförs allt oftare tidigare i rekryteringsprocessen och i en större omfattning jämfört med tiden före automatiserat beslutsfattande.

Även arbetssökandes ansiktsuttryck, gester och språkliga mönster kan analyseras, påstår olika företag, för att ge information om huruvida de intervjuade talar sanning, deras sociala förmåga samt deras personlighets- och karaktärstyper.<sup>57</sup> Det är oklart om detta över huvud taget förekommer i Sverige. I USA och England förekommer dock sådana intervjurobotar. Än så länge används intervjurobotar i Sverige framförallt för beslutsstöd och tolkningen av intervjuerna sker manuellt av människor. Men utomlands – och det finns sådana ambitioner även i Sverige – förekommer att de slutliga besluten om uteslutning eller att gå vidare, sker automatiserat med exempelvis maskinlärande.<sup>58</sup>

#### **Tester**

Personlighets-, kognitiva-, karaktärs-, motivations-, social förmåga, arbetsförmåga och färdighetstester är inte något nytt vid rekrytering. Men användningen av tester kan öka med hjälp av automatiserat beslutsfattande

---

<sup>55</sup> Intervju med vd, Manne Didehvar, Sveriges HR-förening, 2019-04-25, samt se Svenskt Näringsliv (2018), Diskrimineringsombudsmannen (2016).

<sup>56</sup> Intervju med Gabriel Skantze, KTH, 2019-03-22.

<sup>57</sup> Se kritisk forskning kring ansiktigenkänning, Buolamwini (2016, 2018-06-21).

<sup>58</sup> Intervju med forskargrupp intervjurobotar, Gabriel Skantze, Daniel Karlsson, KTH, 2019-03-22, samt intervju med vd, Manne Didehvar, Sveriges HR-förening, 2019-04-25.

och ske tidigare i rekryteringsprocessen.<sup>59</sup> En ny form av tester är i form av digitala spel där den sökande snabbt utvärderas med automatiserat beslutsfattande.<sup>60</sup>

Det finns forskning såväl som debatt om fördelar och risker med tester ur ett diskrimineringsperspektiv. Men det finns lite kunskap om hur eventuella risker för diskriminering vid användandet av tester påverkas av automatiserat beslutsfattande.

#### **Kontroll av arbetssökande på nätet**

Det blir allt vanligare att rekryterare använder sig av kartläggningar av kandidater via olika former av jobbplattformar och sociala medier som Facebook, Instagram, LinkedIn och sökmotorer som Google, och liknande, något som möjliggörs av artificiell intelligens.<sup>61</sup> I Sverige verkar sådana kontroller göras manuellt och i relativt stor omfattning.<sup>62</sup> Men det finns datorprogram på marknaden som automatiserat söker upp och analyserar jobbkandidaters data i olika sociala medier, inkluderat uppgifter om de jobbsökandes vänners profiler

#### **Kreditgivning bygger allt mer på automatiserat beslutsfattande**

Tillgången till allt mer datoriserad analyskraft och ett växande antal datakällor, så kallad stordata, anses leda till att kreditgivare istället för att utgå från historiska analyser av låntagares tidigare betalningar och ekonomi, allt mer fokuserar på så kallade prediktiva analyser av låntagares beteenden: Kommer kredittagarna att betala räntor och amorteringar? Finns det risk för bedrägerier? Har kredittagare riskfyllda beteenden? Via så kallad konsumentprofilering försöker man på ett djupare sätt förstå konsumenters beteenden.<sup>63</sup> Genom sådana profileringar möjliggörs snabba beslut – ibland i realtid. En pådrivande kraft i utvecklingen är uppfattningen att konsumenter förväntar sig snabbare och mer personanpassade

---

<sup>59</sup> Det finns tester för färdigheter; personlighets- och karaktärsdrag; kognitiva förmågor; arbetsmotivation och emotionell intelligens; social förmåga; grader av självförtroende, narcissism, perfektionism, passiv aggression, kontrollbehov och på senare tid om uthållighet över tid. Se artikel om ökat antal tester, Dagens Juridik (2015). Se även Wilson, med flera (2018).

<sup>60</sup> Evry (2019).

<sup>61</sup> Begrepp för detta är bland annat cybervetting eller on-line vetting. Stockholms Handelskammare (2017); Backman och Hedenus (2018); Hedenus, med flera (2019); Hedenus och Backman (2018).

<sup>62</sup> Stockholms Handelskammare (2017).

<sup>63</sup> Se Rättspromemoria om automatiserad databehandling med algoritmer och risker för diskriminering, Diskrimineringsombudsmannen (2019a) samt Ståhl (2017); Petrasic, med flera (2017); Konsumentverket (2017a, b); Greenstein (2017, 2018).

kredittjänster.<sup>64</sup> Kreditgivaren ser sig därför beroende av automatiserat beslutsfattande för snabba beslut.

Den fysiska eller juridiska person som tillhandahåller kredittjänster eller lån är ansvarig för eventuell diskriminering som uppstår i sin egen verksamhet eller hos eventuella underleverantörer eller samarbetspartners, detta oavsett vilken insyn tillhandahållaren av tjänsten har i hur den algoritm eller det program som algoritmen förekommer i, fungerar.<sup>65</sup> Vid en tillsyn utifrån diskrimineringslagen är en kreditgivare skyldig att på ett förståeligt sätt ge information av betydelse, vilket omfattar använda datorprogram och algoritmer.<sup>66</sup>

Framväxten av lagar och system för kreditvärdighet som innebar fokus på verkliga inkomster, tillgångar och betalningshistoria innebar minskade risker för diskriminering. Men nu frestar möjligheten att med automatiserat beslutsfattande kombinera traditionella personuppgifter om betalningshistoria och inkomst med information från sociala medier.<sup>67</sup> Exempel på information som kan användas för kreditvärdering är vilka nätverk låntagaren är del av, låntagarens sökningar på internet, köpvanor, köpkraft och brottslighet i låntagarens grannskap, typ av grannskap, typ av boende, och utbildningsnivå. Många av dessa datakällor kan i sig eller ihopkopplade till kluster av sammanhängande mönster i praktiken direkt eller indirekt leda till en individ missgynnas på ett sätt som har samband med någon av diskrimineringsgrunderna.<sup>68</sup> Historiskt skapades uttrycket Redlining för fenomenet att hela grannskap utestängdes från kredit genom inramning med rödpenna på en karta. Nu har termen Weblining börjat dyka upp för att beskriva när olika grupper på ett liknande sätt utestängs digitalt via internet.

Kreditupplysningar kan användas i andra sammanhang än vid kreditgivning. Det innebär att eventuella risker för diskriminering kan sprida sig till dessa sammanhang – som exempelvis rekrytering. I USA uppger forskarna O´Brien och Kiviat att cirka hälften av arbetsgivarna begär

---

<sup>64</sup> Samtal med Finansinspektionens Innovationscenter, 2019-04-09 samt Grevelius (2018).

<sup>65</sup> Se 2 kap. 12§ samt 12a-c §§ i diskrimineringslagen och dess förarbeten samt för en utförlig redovisning, DO:s rättspromemoria, Diskrimineringsombudsmannen, 2019a. Se LED 2018/387, handling 37.

<sup>66</sup> Ibid. Det finns dock undantag från uppgiftsskyldigheten i diskrimineringslagen om det i det enskilda fallet finns särskilda skäl som talar emot, exempelvis om de uppgifter som begärts ut av Diskrimineringsombudsmannen gäller utrikestjänsten, försvaret, affärs- eller driftsförhållanden eller andra uppgifter där motstående intressen är så starka att uppgiften inte bör lämnas ut.

<sup>67</sup> Hurley och Adebayo (2017).

<sup>68</sup> Ibid.

tillstånd av jobbsökare för göra kreditgranskning av dem.<sup>69</sup> En studie visar att en kombination av en kreditupplysning som visar på betalningsproblem och att vara kvinna jämfört med att vara man eller med ”svart” bakgrund jämfört med ”vit”, leder till sämre chanser till anställning.<sup>70</sup> Effekterna av kreditupplysningar riskerar således att sträcka sig långt. Dessutom finns en växande handel med kreditupplysningar via så kallade datamäklare vilken försvårar kontroll och överblick.<sup>71</sup> Data från kreditupplysningar kan kombineras med en rad andra data av olika kvalitet och bakgrund.

I ett fall som illustrerar en del av det som beskrivs ovan fann den finska Diskriminerings- och jämställdhetsnämnden att ett svenskt kreditinstitut verksamt i Finland överträtt den finska diskrimineringslagen.<sup>72</sup> Vid en ansökan om en konsumentkredit i samband med näthandel jämfördes en person med hjälp av automatiserat beslutsfattande utifrån kön, ålder, bostadsort och språk genom statistisk metod med de finansiella beteendena hos olika jämförelsegrupper. Slutsatsen av jämförelsen innebar att den sökande fick för låga kreditpoäng för att få kredit. Ingen kontroll gjordes av den sökandes verkliga finansiella situation. Kreditinstitutet hävdade att förfaringssättet är standard i branschen och en förutsättning för ett snabbt kreditavgörande.<sup>73</sup> Men Diskriminerings- och jämställdhetsnämnden konstaterar att finansiella skäl, det vill säga kostnader för kreditgivaren, inte är en godkänd ursäkt för en missgynnande särbehandling. Nämnden förbjöd kreditinstitutet att tillämpa denna metod.

Under utredningen framkom att andra kreditgivare troligen agerar på ett liknande sätt. Exemplet visar på hur ett relativt enkelt automatiserat beslutsfattande kan leda till diskriminering genom användning av olika data som i sig kan ses som relevanta.

Ytterligare ett exempel på automatiserat beslutsfattande och diskriminering visade sig vid en tillsyn genomförd av Diskrimineringsombudsmannen i Sverige. Den statliga banken SBAB Bank AB nekade en bankkund ett utökat bolån med hänvisning till att dennes ålder närmade sig pensionsåldern. Kundens aktuella inkomst skrevs därför automatiskt ner av bankens datasystem med 30 procent eftersom banken antog att inkomsten skulle minska vid pensionering. Diskrimineringsombudsmannen kom fram till att den automatiserade och schabloniserade kalkylmetod som användes av

---

<sup>69</sup> O’Brien och Kiviat (2018).

<sup>70</sup> Ibid. Det finns dock ett antal stater i USA där det är olagligt att begära kreditupplysningar vid jobbsökande just på grund av risker för att det ska leda till diskriminering.

<sup>71</sup> Robinson och Yu, (2014).

<sup>72</sup> Diskriminering och jämställdhetsnämnden (2018).

<sup>73</sup> Se rättspromemorian, Diskrimineringsombudsmannen (2019a).

banken vid bedömningar av låneansökningar från personer som fyllt, eller under det året bankkrediten söktes skulle fylla 60 år, utgjorde en överträdelse av diskrimineringsförbudet som hade samband med ålder.<sup>74</sup>

## Projektets iakttagelser

Projektet har arbetat med att identifiera områden där utvecklingen av användandet av automatiserat beslutsfattande är av särskilt intresse ur ett diskrimineringsperspektiv och omfattas av diskrimineringslagen. Utifrån vår studie gör vi följande iakttagelser:

- Automatiserat beslutsfattande förekommer i olika former och omfattning inom rekrytering och kreditgivning liksom i övriga samhällsområden som omfattas av diskrimineringslagen.
- Mer avancerade former av artificiell intelligens som självlärande program, exempelvis maskinlärande, är än så länge relativt ovanliga i Sverige. Däremot är andra former av enklare automatiserat beslutsfattande inte längre avlägsna framtidsvisioner utan redan vardag inom många områden och antalet tillämpningar ökar snabbt. Detta blir tydligt i vår fördjupade granskning av områdena rekrytering och kreditgivning. Dessa tillämpningar kan påverka många av de verksamheter som DO genomför i sitt arbete med vägledning, tillsyner, processer och kunskapsspridning.

Forskning och erfarenheter visar att olika inslag i automatiserat beslutsfattande riskerar att leda till diskriminering. Det kan röra sig om missgynnande dataprogram, olämplig data, bristande analys, med mera. Allt mer uppmärksammas sådana risker. Men det förs även fram förslag på hur dessa brister kan överkommas.

- Projektets rättspromemoria om automatiserad databehandling med algoritmer och risker för diskriminering inom rekrytering och kreditgivning, har kommit fram till att om arbetsgivaren vid rekrytering använder ett ”digitalt verktyg eller en algoritm som behandlar information eller tar fram beslutsunderlag med stöd av ADB eller AI så är det arbetsgivarens ansvar om en arbetssökande diskrimineras, oavsett vilken insyn arbetsgivaren har i hur algoritmen fungerar.” Vid kreditgivning gäller att vid tillhandahållande av tjänsten kreditgivning ”kan man dra slutsatsen att ansvaret ligger hos den fysiska eller juridiska person som tillhandahåller tjänsten. Detta oavsett vilken insyn

---

<sup>74</sup> Se Diskrimineringsombudsmannen (2017), GRA 2017/80.

tillhandahållaren av tjänsten har i hur den algoritm eller det program som algoritmen förekommer i fungerar.”<sup>75</sup>

- Utifrån vår genomgång av områdena rekrytering och kreditgivning menar vi att det finns ett stort behov av detaljerad kunskap om arbetssätt, regler och rutiner inom olika samhälls- och verksamhetsområden samt branscher för att kunna bedöma risker för diskriminering i samband med automatiserat beslutsfattande. Det räcker inte med en generell kunskap om automatiserat beslutsfattande. Här kan fortsatt omvärldsbevakning bidra med kunskap liksom olika former av undersökningar. Men även tillsyner samt rättsliga processer kan bidra med kunskap.<sup>76</sup>
- Det kan vara svårt för tillsynsmyndigheter att få tillräcklig insyn där automatiserat beslutsfattande används för att kunna pröva om en individ har utsatts för diskriminering. Brist på transparens kan bero på ovilja att redovisa grunderna för processen bakom ett automatiserat beslut av exempelvis affärsmässiga skäl. Det kan också finnas tekniska och juridiska hinder. Ett annat hinder kan vara bristande kompetens att tolka och förstå information om programmering och funktionssätt. Det finns dock försök att ta fram metoder för förståelig och meningsfull information om automatiserat beslutsfattande.
- Projektets rättspromemorior (LED 2018/387, handling 37 och handling 44) visar att när det gäller rekrytering och kreditgivning finns det långgående rättigheter för tillsynsmyndigheter att utifrån diskrimineringslagen begära ut information. För individer finns det vissa rättigheter till information utifrån Dataskyddsförordningen (GDPR). Exakt vad dessa rättigheter innebär i praktiken är dock i vissa fall oklart. Här menar projektet att det vore intressant med en större klarhet kring tillämpningen av dessa rättigheter. Förhållandet mellan diskrimineringslagens och dataskyddsförordningens krav på insyn och rätten till immaterialrättsliga skydd och eventuella andra begränsningar kan behövas utredas. Det kan också finnas oklarheter kring diskrimineringslagens krav på att uppgifter sparas för att kunna se om diskriminering förekommit och dataskyddsförordningens krav på att personuppgifter bör tas bort när de inte används.

---

<sup>75</sup> Promemoria om automatiserad databehandling med algoritmer och risker för diskriminering inom rekrytering och kreditgivning, LED 2018/387, handling 37, sid 6.

<sup>76</sup> I samband med den tillsyn den finska Diskriminerings- och jämställdhetsnämnden gjorde av ett kreditinstitut föreslogs en tillsyn av hela kreditbranschen, se pressmeddelande, Diskriminerings- och jämställdhetsnämnden (2018b). Samtal med Sara Markstedt, 2019-03-27 om behovet av fördjupad analys om risker och förutsättningar utifrån olika samhällsområden och myndigheter.



## Avslutande diskussion

Projektet har avslutningsvis några tankar kring DO:s bidrag till att påverka utvecklingen av användandet av automatiserat beslutsfattande utifrån myndighetens uppdrag och vision. Sammanfattningsvis ser projektet att DO bör ta hänsyn till följande för fortsatt problemformulering och åtgärder:

- Artificiell intelligens och automatiserat beslutsfattande förekommer i olika hög grad inom alla diskrimineringslagens utpekade samhällsområden.
- Användningen av artificiell intelligens liksom annat automatiserat beslutsfattande kan innebära en risk för diskriminering enligt diskrimineringslagen beroende på hur tekniken, verktygen och systemen programmeras, tränas, utvecklas och används.
- När artificiell intelligens och automatiskt beslutsfattande används inom rekrytering och kreditgivningen omfattas detta av diskrimineringslagen, DL. Ingenting tyder på att något annat skulle gälla för användning av artificiell intelligens och automatiskt beslutsfattande inom alla samhällsområden som omfattas av DL
- Det pågår utveckling av standarder, guider, manualer, praktiker på branschnivå, nationell, internationell och EU-nivå och (eventuell lagstiftning) med möjlighet för DO att försöka påverka utvecklingen.
- Tillsynsmyndigheter och individer behöver få tillgång till begriplig information för att kunna bedöma förekomsten av diskriminering. Möjliga svårigheter är hänvisningar till särskilda skäl, som exempelvis företagshemligheter samt den så kallade svarta lådan-problematiken.

### Möjliga åtgärder

Projektet ser att DO kan påverka aktörer – såväl privata som offentliga, nationella som internationella – med kunskap om risker för diskriminering med automatiserat beslutsfattande. DO kan också tillföra aktörer inom området expertis om juridiska aspekter utifrån diskrimineringslagen i en utveckling som ofta präglas av en allmän icke-juridisk förståelse av begreppet diskriminering. DO kan också i sin vägledning kring aktiva åtgärder lyfta fram att dessa omfattar även automatiserat beslutsfattande.

DO kan bidra med ökad kunskap om konkreta tillämpningar av automatiserat beslutsfattande och deras koppling till diskriminering genom kunskapsbyggande och genom att vid anmälningar, tillsyner, processer och kontakter med arbetsmarknadens parter samt civilsamhället, liksom vid andra aktiviteter, uppmärksamma aspekter av automatiserat beslutsfattande. Andra aktörer kan behöva den kunskap om diskriminering som DO har. En möjlighet att lära mer om den pågående utvecklingen i

Sverige är att uppmuntra till anmälningar kopplade till automatiserat beslutsfattande.<sup>77</sup>

Projektet menar att för att kunna genomföra de insatser som skissas ovan kan det behövas en fördjupad kompetens om problemområdet och att denna kompetens kan och bör användas i det egna arbetet på olika nivåer och delar av DO. En fortsatt omvärldsbevakning är nödvändig. En möjlighet här är att delta i Equinets satsningar på området för att få en inblick i utvecklingen inom EU och dess medlemsländer. Inte minst intressant är att följa utvecklingens olika juridiska aspekter – exempelvis genom att följa arbetet inom den av Europeiska kommissionen utsedda expertgruppen AI HLEG samt annan relevant debatt.

---

<sup>77</sup> I samband med avgörande av den tillsyn den finska Diskriminerings- och jämställdhetsnämnden gjorde av ett kreditinstitut kring användandet av automatiserat beslutsfattande, ansåg nämnden att en tillsyn av hela kreditbranschen skulle vara intressant, se pressmeddelande Diskriminerings- och jämställdhetsnämnden (2018b).

## Ordlista för centrala begrepp som används i denna översikt

**Automatiserat beslutsfattande:** Automatiska algoritmiska dataprocesser för att ta fram beslutsstöd för människor eller helt automatiserat beslutsfattande utan mänsklig kontroll. Motsvaras på engelska av Automated Decision-Making i olika policydokument liksom i GDPR.

**Algoritm:** En systematisk procedur eller formel för hur man genom ett begränsat antal steg löser ett problem och tar beslut, exempelvis för att välja ut eller bort något alternativ. I många fall vill man kunna använda detta för prediktiva analyser som att förutsäga om någon blir en bra arbetskraft eller god kund. Algoritmer styr mycket av vad som görs på internet och möjliggör hanterandet av mycket stora mängder data som de samlar in, filtrerar, sorterar, betygssätter och anpassar.

**Artificiell intelligens:** Samlingsbegrepp som omfattar många andra begrepp. Men innebär inblandning av datorer och program som försöker likna intelligent beteende.

**Bias:** engelskt begrepp, ursprungligen från statistik, står för data som är påverkad av fördomar, snedvriden, vinklad, avviker från riktiga förhållanden. Bias används ofta även i svenska sammanhang.

**Databrytning** (data mining, datautvinning): Innebär att med datorer, och ofta algoritmer, gå igenom stora mängder data för att finna mönster, samband eller trender. En del av det som nämns artificiell intelligens.

**Digitalisering:** Att gå från analog till digital representation av information. Är ett övergripande begrepp som omfattar informationsteknologi, webb, internet, men också artificiell intelligens och maskinlärande.

**Maskinlärande:** Sjävlärande dataprogram. Efter att det en gång programmerats så tränas maskinen till att själv hitta olika mönster och relationer. Dessa behöver inte nödvändigtvis motsvara verkliga orsakssamband utan kan vara korrelationer som i sig bara visar på samtidigt förekommande fenomen. Att klargöra varför ett självlärande program fattat ett visst beslut, exempelvis varför någon uteslutits från en ansökningsprocess, är ofta mycket svårt. (Se Svart låda).

**Prediktiv analys:** Databaserad analys för att förutsäga individers och andra aktörers beteende och framgång, exempelvis vid nyanställning och kreditgivning. Oftast baserade på olika statistiska relationer mellan olika iakttagelser.

**Profilerings:** Olika data används för att bygga upp beskrivningar och antaganden om beteenden, för exempelvis olika individer, så kallade profiler. Det kan röra konsumentbeteenden, bedömningar av arbetsförmåga och bedömningar av riskbeteenden vid kreditgivning. I det senare fallet talas

ofta om kreditvärdighetspoäng eller e-scores. Begreppet återfinns även i GDPR där de som utsätts för profilering har rätt att få veta detta.

**Proxygrupper** eller **jämförelsegrupper**: Används allt mer för så kallade prediktiva analyser av arbetssökande, låntagare, konsumenter, försäkringstagare, bidragstagare och kriminella. Individer jämförs med det kartlagda beteendet hos olika grupper utifrån antagandet att de ska bete sig på ett liknande sätt. Man letar dubbelgångare med så många gemensamma egenskaper med den man vill utvärdera som möjligt. Kan leda till diskriminering då på avgörande punkter den enskilde individen kan skilja sig åt från jämförelsegruppen.

**Sourcing** är ett begrepp som används även i svenska sammanhang och som står för att bygga upp databanker med upplysningar om människor som kan vara intressanta för framtida erbjudanden från arbetsgivare.

**Stordata (Big Data)**: När ett stort antal källor används som underlag för automatiserat beslutsfattande och behandlas med algoritmer och maskinlärande. Det kan handla om att bearbeta miljoner eller fler informationsbitar från en rad källor i form av såväl ord, siffror som ljud och bilder. Även klick på internet, sökningar, vad du skriver på sociala medier, vem du känner och umgås med eller vilka låtar du väljer på Spotify kan användas som data för olika analyser.

**Svart låda eller Black Box**: Syftar i dessa sammanhang på att när det gäller automatiserat beslutsfattande och då särskilt mer avancerade former som maskinlärande det ofta är svårt att visa eller rekonstruera vad som händer i den Svarta lådan.<sup>78</sup>

**Svarta lådan-testande eller Black Box Testing**: Effekten av ett automatiserat beslutsfattande system, dess utfall, följs upp utan att man har tillgång till använd träningsdata, programmering eller liknande. Kan genomföras med olika kvantitativa eller kvalitativa metoder. Exempelvis genom att testa systemet med olika in-data och jämföra med olika utfall.

**Träningsdata**: Historiska data används för att upptäcka möjliga förutsägande mönster och relationer genom algoritmiska modeller. Processen att tillföra ett algoritmiskt program data brukar kallas att träna ett program. Så kan för att ett program ska lära sig känna igen tomater det visas ett mycket stort antal bilder av tomater såväl som icke-tomater, för att tränas att känna igen faktorer som förknippas med just bilder av tomater.

---

<sup>78</sup> Pasquale (2015).

## Referenser

AI Now Institute (2018). Litigating Algorithms. Challenging Government Use and Algorithmic Decisions Systems. New York: New York University, AI Now Institute. Publicerad på: <https://ainowinstitute.org/litigatingalgorithms.pdf> .

AI Sustainability Center (2019). Hållbar AI. Rapport. Stockholm: AI Sustainability Center.

AlgorithmWatch (2019a). AI Ethics Guidelines Global Inventory, A list of Ethics Guidelines for ADM systems worldwide. Berlin: AlgorithmWatch.

AlgorithmWatch (2019b). Automating Society: Taking Stock of Automated Decision-Making in the EU. Rapport. Berlin: AlgorithmWatch in cooperation with Bertelsmann Stiftung, Open Society Foundations.

Allen, Robin, Masters, Dee (2019). Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making: ERA Forum.

Amnesty International, Access Now (2018). The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems: Toronto.

Backman, C, Hedenus, A, (2018). Online privacy in job recruitment processes? Boundary work among cybervetting recruiters. Konferensbidrag. Århus8<sup>th</sup> Biannual Surveillance and Society Conference.

Backman, C, Hedenus, A (2019). Beskrivning av forskarprogrammet Arbetsgivares användning av informationssökning via internet i samband med rekrytering, Vid Göteborgs Universitet. Se <https://socav.gu.se/forskning/aktuella-projekt/arbetsgivares-anvandning-av-informationssokning-pa-internet-i-samband-med-rekrytering./projektbeskrivning>. Hämtad 2019-05-12.

Barocas, S, Selbst, D, A (2016). Big Data´s Disparate Impact, ur California Law Review, Vol 104, Nr 671, 2016. Kalifornien, USA.

Buolamwini, J (2016). How I´m fighting bias in algorithms, TED talks, TEDxBeacon Street. November 2016. [https://www.ted.com/talks/joy\\_buolamwini\\_how\\_i\\_m\\_fighting\\_bias\\_in\\_algorithms](https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms)

Buolamwini, J (2018-06-21). When the Robot Doesn´t See Dark Skin, New York Times, Opinion. <https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html>

Council of Europe, Committee of Ministers (2019). Intervention of the Commissioner, High-Level Conference Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the rule of law. 26-27 February 2019, Helsinki, Finland. Konferensens webbplats. <https://www.coe.int/en/web/freedom-expression/aiconference2019>

Dagens Industri (2017a). Felländer, A, Lind, F, Sommansson, P, ”Därför är hållbar AI det nya CSR”. Debattartikel, Dagens Industri, Digital, publicerad 2017-10-24, se <https://digital.di.se/artikel/debatt-darfor-ar-hallbar-ai-det-nya-csr> .

Dagens Industri (2017b). Öhrwall, S. Diskriminerande algoritmer, i Dagens Industri, publicerad 2017-11-06. Stockholm: Dagens Industri.

Dagens Juridik (2015). Personlighetstester vid rekrytering ökar. Publicerad 2015-06-01. Stockholm: Dagens Juridik.

Dataskyddsförordningen (2019). Hämtad 2019-06-20. Se <https://eur-lex.europa.eu/legal-content/SV/TXT/PDF/?uri=CELEX:32016R0679&rid=1>.

Diskrimineringsombudsmannen (DO) (2019a). Promemoria om automatiserad databehandling med algoritmer och risker för diskriminering. Rättspromemoria. LED 2018/387, handling 37. Solna: DO

Diskrimineringsombudsmannen (DO) (2019b). Tillgång till information som förutsättning för att bedöma förekomsten av diskriminering i samband med automatiserat beslutsfattande. Rättspromemoria. LED 2018/387, handling 44. Solna: DO.

Diskrimineringsombudsmannen (DO) (2018). Projektdirektiv avseende utveckling av kunskap om automatiserad databehandling med algoritmer (artificiell intelligens) och stordata och diskriminering eller risker för diskriminering. LED 2018/387, handling 1. Solna: DO.

Diskrimineringsombudsmannen (DO) (2017). Tillsyn SSAB, GRA 2017/80.

Diskrimineringsombudsmannen (DO) (2016). Hundra möjligheter att rekrytera utan att diskriminera – det här visar forskningen. Stockholm: Diskrimineringsombudsmannen. Finns nedladdningsbar på [www.do.se](http://www.do.se)

Diskriminerings- och jämställdhetsnämnden (2018a) Diskriminerings- och jämställdhetsnämnden förbjöd ett kreditinstitut att använda statistiska metoder som diskriminerar kreditsökanden vid konsumentkrediter. Publicerad 2018-05-18. Helsingfors. Se <https://www.yvtltk.fi/sv/index/tiedotteet/2018/05/diskriminerings-ochjamstalldhetsnamndenforbjodettkreditinstitutattanvandastatistiskametoderpaet-tsattsomdiskriminerarkreditsokandenvidkonsumentkrediter.html>

Diskriminerings- och jämställdhetsnämnden (2018b). Kreditinstitut bör förnya sin praxis. Se <https://www.syrjinta.fi/web/sv/-/bedomning-av-kreditvardighet-endast-pa-basis-av-statistiska-uppgifter-ar-diskriminering-kreditinstitut-bor-fornya-sin-praxis>.

Europeiska kommissionen (2018). Algo:aware. Raising awareness on algorithms. Updated State of the Art Rapport. Brussel: Europeiska kommissionen.

Europeiska kommissionen (2018a). Communication from the Commission to the European Parliament, the European Council, The European Economic and Social Committee and the Committee of the Regions. Artificial Intelligence for Europe. {SWD (2018) 137 final}. Brussels: 25.4.2018. COM.

Europeiska kommissionen (2019). Meddelande från kommissionen till Europaparlamentet, Rådet, Den europeiska ekonomiska och sociala kommittén, samt Regionkommittén. Att skapa förtroende för människocentrerad artificiell intelligens. Bryssel den 8.4.2019 COM (2019) 168 final Se <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2019:0168:FIN:SV:PDF>

Europeiska kommissionen (2019a). Ethics Guidelines for Trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. Bryssel: Europeiska kommissionen.

Europeiska kommissionen (2019b). A Policy and investments recommendations for Trustworthy Artificial Intelligence. Independent High-Level Group on Artificial Intelligence set up by the European Commission. Bryssel: Europeiska kommissionen. Se <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>.

Europaparlamentet (2017). Betänkande om stordatas effekter på de grundläggande rättigheterna: integritet, uppgiftsskydd, icke-diskriminering, säkerhet och brottsbekämpning. 2016/2225 (INI), 2017-02 20. Bryssel: Europaparlamentet.

Evry (2019). Spela dataspel och bli anställd. Hämtad 2019-12-03. Webbplats: Se <https://www.evry.com/sv/om-evry/media/service-articles/spela-dataspel-och-bli-anstalld/>.

- Equinet (2019a). Call for proposal: Equality, artificial intelligence and algorithmic discrimination. Se <http://equineteurope.org/2019/04/24/call-for-proposals-equality-artificial-intelligence-and-algorithmic-discrimination/>.
- Equinet (2019b) Work and strategic plans. Equinet 2019. Se <http://equineteurope.org/equinet-at-a-glance/work-and-strategic-plans/>
- FRA (European union agency for fundamental rights) (2018a). #Big Data: Discrimination in data-supported decision-making, Wien: FRA Focus.
- FRA (European union agency for fundamental rights) (2018b). Artificial Intelligence, Big Data and Fundamental Rights. Tender/ Presentation av forskningsprojekt. Wien: FRA. Se <https://fra.europa.eu/en/call-for-tender/2019/artificial-intelligence-big-data-and-its-impact-fundamental-rights-eu>
- FRA (European union agency for fundamental rights) (2019). Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights Wien: FRA Focus. Se [https://fra.europa.eu/sites/default/files/fra\\_uploads/fra-2019-data-quality-and-ai\\_en.pdf](https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf)
- Greenstein, S (2017). Our Humanity Exposed: Predictive Modelling in a Legal Context. Avhandling. Stockholm: Stockholm University.
- Greenstein, S (2018), Artificial Intelligence (AI) and the law. Kurs vid Stockholm Universitet, 6 december, 2018.
- Grevelius, A. (2018). Fintech och robotrådgivning. Ur Livet med AI. SSF 29. Stockholm: SSF.
- The Guardian, (2018). Amazon ditched AI recruiting tool that favoured men for technical jobs. Artikel. Hämtad 2018-10-11., se <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>
- Hedenus, A, Backman, C, Håkansson, P (2019). Whom do you Know? Recruiters motives for assessing job-seekers online networks, ur The Journal of Human Resource Management, March 2019.
- Hedenus, A, Backman, C. (2018). Nätgranskning i samband med rekrytering. Hur, vad och varför (inte)? Arbetsliv i omvandling, nr. 1, 2018: 1–55.
- Hurley, M, Adebayo, J (2017). Credit Scoring in the Era of Big Data, ur Yale Journal of Law and Technology, Vol 18, nr, 1 2017.



Inspektionen för arbetslöshetsförsäkringen (2018). Arbetsförmedlingens stöd till arbetssökande som inte använder digitala tjänster. Rapport 2018:11. Katrineholm: IFA.

Institutet för socialförsäkring (ISF) (2018ab). Riskbaserade urvalsprofiler och likabehandling. Arbetsrapport 2018:1. Stockholm: ISF.

Institutet för socialförsäkring (ISF) (2018). Profilerings som arbetsmetod för riktade kontroller. Arbetsrapport 2018:5. Stockholm: ISF.

Konsumentverket (2017a). Kunskapsöversikter inom det konsumentpolitiska området. Rapport 2017:1. Stockholm, Karlstad: Konsumentverket.

Konsumentverket (2017b). Vår omvärld 2017. Rapport till regeringen 2017-11-30. Stockholm, Karlstad: Konsumentverket

Larsson, S (2018). Sju nyanser av transparens. Om artificiell intelligens och ansvaret för digitala plattformars påverkan. Särtryck. Tankesmedjan Fores. Stockholm: Fores.

Manant, M, Pajak, S, Soulié, N (2018). Can social media lead to labor market discrimination? Evidence from a field experiment. Ur Journal of Economics and Management Strategy. Vol 28, nr 2, 2018.

Mittelstand, B, R, Allo, P, Taddeo, M, Wachter, S, Floridi, L (2016). The Ethics of algorithms: Mapping the debate, Big Data & Society, July-December 2016: 1-21. Oxford: Alan Turing Institute.

Modlitba, P (2018). Fyra förändrade drivkrafter för AI i vården, ur Livet med AI. Stockholm: Stiftelsen för strategisk forskning.

Myers West, S, Whittaker, M, Crawford (2019). Discriminating Systems. Gender, Race, and Power in AI. Rapport. New York: AI Now Institute. Se <https://ainowinstitute.org/discriminating-systems.html>.

Naarttijärvi, M (2017). Rättstatligheten och algoritmiska svarta lådor, ur Jubileumsskrift till Juridiska institutionen 40 år. (red): Edström, Ö, Lindholm, J, Mannelqvist, R, Umeå, Umeå Universitet, Juridiska institutionen. Se <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-141876>

New Economics Foundation. (2018). The Rise of the Data Oligarchs. Power and Accountability in the Digital Economy. Part 1: Data Collection. London: NEF.

- Ny Teknik (2018).”Algoritmer utan etik ger AI utan moral”. A Felländer. Debattartikel, publicerad 2018-03-08, se <https://www.nyteknik.se/opinion/algoritmer-utan-etik-ger-ai-utan-moral-6902790>
- O’Neill, C (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishing Group.
- Pasquale, F (2015). The Black Box Society. The Secrets Algorithms that Controls Money and Information. USA: Harvard University Press.
- Patent- och registreringsverket (2019). Mjukvara, appar och affärsmetoder. Programkod eller rena affärsmetoder går inte att patentera i Sverige. <https://www.prv.se/sv/patent/ansoka-om-patent/innan-ansokan/vad-kan-inte-patenteras/mjukvara-appar-och-affarsmetoder/> . Hämtad, 2019-04-25.
- Person, A (2016). Implicit Bias when using Data Profiling within Recruitments and Human- Resource Management, Uppsala: Uppsala University. Se <https://pdfs.semanticscholar.org/7fd1/6bf660bab7a5d398d19fc1200f3dee7099ce.pdf>
- Petrasic, K, Saul, B, Greig, J, Bornfreund, M, Lamberth, K (2017). Algorithms and bias: What Lenders need to know. Financial institutions must mind their fintech solutions to ensure they do not generate unintended consequences. Boston: White and Case. Se <https://www.whitecase.com/publications/insight/algorithms-and-bias-what-lenders-need-know>
- Prince, A, Schwarcz, D (2020). Proxy discrimination in the age of artificial intelligence and Big Data. (Forthcomming): Iowa: Iowa Law Review.
- Prop. 2007/08:95. Ett starkare skydd mot diskriminering. Se <https://www.regeringen.se/49bafd/contentassets/9992e1e8bedd4019aaa6a9e8565f778b/ett-starkare-skydd-mot-diskriminering-prop.-20070895>
- Raso, F, Hilligoss, H, Krishnamurthy, V, Bavitz, C, Kim, L (2018). Artificial Intelligence & Human Rights: Opportunities & Risks. Boston: Harvard University, The Berkman Klein Center for Internet & Society Research at Harvard University, 2018.
- Näringsdepartementet (2018). Nationell inriktning för artificiell intelligens. Stockholm: regeringen. se [https://www.regeringen.se/49a828/contentassets/844d30fb0d594d1b9d96e2f5d57ed14b/2018ai\\_webb.pdf](https://www.regeringen.se/49a828/contentassets/844d30fb0d594d1b9d96e2f5d57ed14b/2018ai_webb.pdf)

O'Brien, R, Kiviat, B (2018). Disparate Impact? 'Race, sex and Credit Reports in Hiring. Socius: Sociological Research for a Dynamic World, Nr 4: 1-20, Sage.

Rieke, A, Bogen, M, Robinson, D, G (2018). Public Scrutiny of Automated Decisions. Early lessons and Emerging Methods. Mumbai: Upturn and Omidyar Network Report.

Rieke, A, Yu, H, Robinson, D, von Hoboken, J (2016). Data Brokers in an Open Society. An Upturn Report, prepared for the Open Society Foundations <https://www.opensocietyfoundations.org/uploads/42d529c7-a351-412e-a065-53770cf1d35e/data-brokers-in-an-open-society-20161121.pdf>

Robinson, David, Yu, Harlan (2014). Knowing the Score: New Data, Underwriting, and Marketing in the Consumer Credit Marketplace. A Guide for Financial Inclusion Stakeholders. New York: The Ford Foundation. Version 1. October 2014. Se <https://perma.cc/29YH-J55K>

Sjöberg, P-O (2016) Scenario: Digitalisering för en hälsosammare framtid. Ur Temarapport 2016:2 från Digitaliseringskommissionen (N 2012:04) Stockholm.

SOU 2018:25 Juridik som stöd för förvaltningens digitalisering. Betänkande av Digitaliseringsrättsutredningen. Stockholm.

Sveriges Kommuner och Landsting (SKL) (2017). Artificiell intelligens. Möjligheter för välfärd. Stockholm: SKL.

Stockholms Handelskammare (2017). Sociala medier sållar bort allt fler vid rekrytering. Analysrapport. Stockholm.

Ståhl, B (2017). FinTech. Den digitala transformeringen av finansmarknaden. Rapport 18. Stockholm: Blue Institute.

Svenskt Näringsliv (2018). Jobbskaparna larmar! Kompetensbristen ökar. Rekryteringsenkäten. Rapport. Stockholm: Svenskt Näringsliv.

SVT Nyheter (2018). Amazon stängde ner rekryteringsrobot – diskriminerade kvinnor. Publicerad 2018-10-24, se <https://www.svt.se/nyheter/amazon-stangde-ner-rekryteringsrobot-diskriminerade-kvinnor>.

Sydsvenskan (2018). "I dag integreras artificiell intelligens i människors vardag utan att det finns tillräcklig kunskap om vad det innebär. Lagar och regler släpar efter". Anna Felländer, med flera. Debattartikel, publicerad 2018-05-07, se [www.sydsvenskan.se /2018-05-07/idag-integreras](http://www.sydsvenskan.se/2018-05-07/idag-integreras)

Wahlgren, P (2018). Reglering av artificiell intelligens. Ur Livet med AI. SSF rapport 29. Stockholm: SSF.

Watcher, S, Mittelstad, B, D, Floridi, L (2017). Why a right to explanation of automated decision-making does not exist in the General Protection Regulation. *International Data Privacy Law*, Volume 7, Issue 2, May 2017, Pages 76–99. Oxford: Oxford Internet Institute, Oxford University. Se <https://academic.oup.com/idpl/article/7/2/76/3860948>

Vertesi, J (2016). My experiment with Opting Out of Big Data Made Me Look Like a Criminal. *Time* 2016-10125. Se <https://time.com/83200/privacy-internet-big-data-opt-out/>

Vinnova (2013). Big Data Analytics. A research and Innovation Agenda for Sweden. Stockholm: Vinnova. Se <https://www.vinnova.se/globalassets/mikrosajter/strategiska-innovationsprogram/agendor/analys-av-stordata.pdf>

Vinnova (2018). Artificiell intelligens i svenskt näringsliv och samhälle. Analys av utveckling och potential. Vinnova Rapport VR 2018:08. Stockholm: Vinnova.

Wilson, M, Kurzweil, M, Alamuddin, R (2018). Mapping the Wild West of Pre-Hire Assessment. A landscape Of the Uncharted Technology-Facilitated Ecosystem, Report. New York: Ithaka SR.

World Economic Forum (2018). How to Prevent Discriminatory Outcomes in Machine Learning. White Paper Global Future Council, March 2018.

Zuiderveen Borgesius, F (2018). Discrimination, artificial intelligence and algorithmic decision-making. Study by Prof. of Law, Frederik Borgesius, Professor of Law, Institute for Computing and Information Sciences (iCIS), Radboud University Nijmegen, and Researcher at the Institute for Information Law, University of Amsterdam (the Netherlands), Strasbourg: Anti-discrimination department, Council of Europe. Se <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>.

Intervjuer, föreläsningar, workshops, seminarier, nätverksträffar:

2018-06-07 Om behovet av transparens och etik i svensk AI-utveckling. Sociala och rättsliga konsekvenser, seminarium, tankesmedjan Fores, Anna Felländer, Fredrik Heinz, Åsa Zetterberg, med flera. LED 2018/387: handling 9.

2018-06-11: Telefonmöte med Anna Felländer, forskare och debattör om AI, om deltagande vid DO personaldagar 2018. LED 2018/387: handling 10. Se även konversation med Åsa Zetterberg, Digital officer vid regeringskansliet, LED 2018/ 320.

2018-09-11 DO Personaldag om AI med Anna Felländer, med flera. Se uppdragsavtal LED 2018/347.

2018-10-10-11 E-förvaltningsdagarna. Samtal med GD för DIGG, med mera. Se LED 2018/387, handling 11 och 12.

2018-11-21. Telefonsamtal och e-postutbyte med Anna Felländer om AI Sustainability Center. Se LED 2018/387, handling 13.

2018-12-06 Utbildning: AI and the Law, Stockholms Universitet. Se 2018/387, handling 14.

2019-01-15 Det digitala plattformssamhället, boksläpp, presentation, diskussion, Fores. Se LED 2018/387, handling 16.

2019-01-20 Intervju med Lotta-Karin Hansson, Anna Claesson, HR, Diskrimineringsombudsmannen. Se LED 2018/387, handling 17.

2019-02-11 Intervju med Stanley Greenstein, Stockholms universitet. Se LED 2018/387, handling 18.

2019-02-26 Intervju med forskargrupp Transparenta algoritmer i försäkringsbranschen, RISE, KTH, Ulrik Franke, Jacob Dexe. Se LED 2018/387, handling 19.

2019-03-15 Telefonintervju Christel Backman, forskare, Göteborgs universitet. Se LED 2018/387, handling 20.

2019-03-22 Möte med forskargrupp intervjurobotar, Gabriel Skantze, Daniel Karlsson, KTH Se LED 2018/387, handling 22.

2019-03-27 Samtal med Sara Markstedt, föreläsare vid Myndighetsnätverket för AI, Se LED 2018/387, handling 23.

2019-04-08 Samtal med Konsumentverket. Se LED 2018/387, handling 26.

2019-04-09 Samtal med Innovationscenter, Finansinspektionen. Se LED 2018/387, handling 29.

2019-04-25, Intervju med vd Manne Didehvar, Svensk HR-förening. Se LED 2018/handling 32.

2019-05-16-17 Seminarium och konferens om AI, mänskliga rättigheter och lagen, ERA, Bryssel. Se LED 2019/36.

2019-05-17 Trendseminarium om AI och rekrytering, TNG, Arja Lehto. Se LED 2018/387, handling 30.

2019-05-23, AI Sustainability Center, Seminarium om ansvarsfull AI i praktiken, med bland annat Jen Gennai, chef för ansvarsfull innovation på Google. Se LED 2018/387, handling 31.