

**Diarienummer:** 2018/387  
**Handling:** 37  
**Handläggare:** Christina Jönsson och Mira Hjelm Östh  
**Status:** Fastställd

## Promemoria om automatiserad databehandling med algoritmer och risker för diskriminering: inom rekrytering och kreditgivning<sup>1</sup>

---

<sup>1</sup> Redovisning av modul 2 i projektet avseende utveckling av kunskap om automatiserad databehandling med algoritmer (artificiell intelligens) och stordata och diskriminering eller risker för diskriminering

## Innehåll

Sammanfattning .....	5
Slutsatser.....	6
DL:s tillämplighet .....	6
Aspekter av betydelse för DO:s tillsyn .....	6
Utvecklingen och användningen av ADB och AI .....	7
Risker för diskriminering.....	8
Inledning.....	10
Syfte.....	10
Avgränsningar .....	10
Problem med begreppet diskriminering i en AI-kontext.....	11
Begrepp .....	11
Bestämmelser om diskriminering.....	12
Diskrimineringslagen .....	12
Arbetslivsområdet – rekrytering .....	12
Tillhandahållande av tjänster.....	12
EU-rätten .....	14
Europakonventionen .....	15
Bestämmelser om ADB och AI .....	15
Regler på nationell nivå.....	15
Regler på EU-nivå .....	16
Deklarationer och riktlinjer med mera om ADB och AI.....	16
Nationell nivå.....	16
EU-nivå.....	17
EU-kommissionen .....	17
Europeiska Ekonomiska och Sociala Kommittén .....	19
CEPEJ:s etiska stadga för användning av AI i rättssystem.....	19
DL:s tillämplighet.....	20
Arbetslivsområdet – rekrytering .....	21
Diskrimineringsförbudet för arbetsgivare .....	21
Vem som kan likställas med arbetsgivare.....	21
Instruktioner att diskriminera .....	22
Ansvar vid diskriminering .....	22
Tillhandahållande av tjänster .....	23
Ansvar vid diskriminering .....	24
Utvecklingen av ADB och AI.....	24
Från deskriptiv till prediktiv analys och modellering .....	25

Prediktiv analys och prediktiv modell .....	26
Konsumentprofilering .....	26
Risker för diskriminering .....	28
Generella risker och felkällor .....	28
Omedveten diskriminering.....	29
Medveten diskriminering .....	31
Avslutande reflektioner.....	31
Exempel på användning av AI och ADB.....	32
Rekrytering .....	32
Konkreta exempel.....	33
Krediter .....	37
Riskbedömning .....	37
Handel med personuppgifter .....	38
Riskbedömning i USA – exempel.....	39
Riskbedömning under utveckling .....	40
Kreditgivares skyldighet att efterfråga information enligt lagen (2004:297) om bank- och finansieringsrörelse .....	41
DO:s tillsynsbeslut .....	42
Finska Diskriminerings- och jämställdhetsnämndens beslut.....	42
Aspekter av betydelse för DO:s tillsyn.....	43
Slutsatser .....	44
DL:s tillämplighet.....	44
Aspekter av betydelse för DO:s tillsyn.....	45
Utvecklingen och användningen av ADB och AI .....	45
Risker för diskriminering.....	47
Avslutande diskussion .....	47
Ordlista för vissa centrala begrepp för kartläggningen.....	50
Referenser.....	52
DO-dokument .....	52
Offentligt tryck.....	52
EU-rättsakter .....	52
EU-publikationer.....	53
Vetenskapliga artiklar .....	53
Övriga artiklar .....	53
Europarådet.....	54
Litteratur .....	54
Myndighetspublikationer .....	54

Websidor .....54

## Sammanfattning

Utvecklingen av automatiserad databehandling (ADB) och artificiell intelligens (AI) pågår i snabb takt inom flera samhällsområden. AI har potential att bidra med betydande nytta genom att bland annat bidra till lösningar på ekonomiska, miljömässiga och sociala samhällsutmaningar. Redan i dag finns exempel på att AI kan bidra till att diagnosticera sjukdomar, minska energianvändningen, reducera trafikolyckor, skapa nya tjänster, effektivisera industriell produktion, utveckla nya läkemedel och förkorta handläggningstider. I takt med den ökande utvecklingen och användningen av AI har bland annat regeringen, olika EU-institutioner samt internationella organisationer uppmärksammat behovet av att utveckla regler, standarder, normer och etiska principer, i syfte att vägleda utveckling och användning av etisk och hållbar AI. Inom EU och olika internationella organisationer pågår just nu arbete med att ta fram riktlinjer och liknande inom området.<sup>2</sup>

Denna promemoria har tagits fram inom ramen för Diskrimineringsombudsmannens (DO:s) projekt för utveckling av kunskap om diskriminering eller risker för diskriminering i samband med automatiserad databehandling med algoritmer (artificiell intelligens, nedan ”AI”) och stordata (eng: big data), nedan AI-projektet. I AI-projektets direktiv<sup>3</sup> samt beslut om fördjupad kartläggning inom ramen för AI-projektet<sup>4</sup> framgår att AI-projektet med utgångspunkt i en fördjupad kartläggning ska genomföra en rättsutredning avseende **utvecklingen och användningen av automatiserad databehandling med algoritmer och stordata och eventuell koppling till diskriminering och/eller risker för diskriminering**. Under arbetets gång har det visat sig att det material som funnits tillgängligt inte kunnat ligga till grund för en rättsutredning. I stället har slutsatserna sammanfattats i denna promemoria. Promemorian behandlar de två områdena rekrytering inom arbetslivet samt ansökan om kredit.

---

<sup>2</sup> En av de fyra etiska principerna som ingår i de etiska riktlinjer som tagits fram av EU:s expertgrupp för AI är **fairness**, som bland annat innebär att säkerställa att individer och grupper är fria från fördomar, diskriminering och stigmatisering, samt respekt för proportionalitetsprincipen i förhållandet mellan medel och mål. Se vidare avsnitt EU-kommissionen på sidan 16.

<sup>3</sup> 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.

<sup>4</sup> Beslut om fördjupad kartläggning inom ramen för uppdraget AI och diskriminering, LED 2018/387, handling 8.

## Slutsatser

### **DL:s tillämplighet**

DL aktualiseras om en arbetssökande diskrimineras av en presumtiv arbetsgivare inom ramen för ett rekryteringsförfarande.

Diskrimineringslagen omfattar i princip alla situationer i förhållandet mellan en arbetsgivare och en arbetstagare. Det följer av att diskrimineringsförbudet i lagen anger en personkrets som skyddas, och inte specifika situationer. Vidare anges vem som förbudet att diskriminera riktar sig mot och vem som ansvarar – nämligen arbetsgivaren (eller den som är att likställa med arbetsgivaren). Diskrimineringsförbudet på arbetslivsområdet är avsett att vara heltäckande, vilket framgår av lagens förarbeten. Om arbetsgivaren använder sig av 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.

Diskrimineringsförbudet i DL vid tillhandahållande av tjänster, exempelvis kreditgivning, till allmänheten träffar den fysiska eller juridiska person som, i enlighet med de rekvisit som uppställs i bestämmelsen, tillhandahåller en tjänst. Utifrån ordalydelsen i och syftet med bestämmelsen i 2 kap. 12 § DL kan man dra slutsatsen att ansvaret för diskriminering som sker vid tillhandahållande av en tjänst ligger 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.

### **Aspekter av betydelse för DO:s tillsyn**

Inom ramen för en tillsyn har DO möjlighet att från tillsynsobjektet efterfråga alla uppgifter som kan vara av betydelse för ombudsmannens tillsyn enligt DL, och ett tillsynsobjekt har motsvarande skyldighet att till DO lämna efterfrågade uppgifter<sup>5</sup>, exempelvis material för att kunna ta del av och granska uppbyggnaden av en algoritm. Ett tillsynsobjekt kan dock åberopa undantaget från uppgiftsskyldigheten i DL, om särskilda skäl talar emot. Undantaget kan komma i fråga exempelvis om de uppgifter som begärs ut av DO gäller försvaret, utrikestjänsten, affärs- eller

---

<sup>5</sup> Prop. 2007/08:95, s. 545.

driftsförhållanden eller andra uppgifter där motstående intressen är så starka att uppgiften inte bör lämnas ut.<sup>6</sup>

### Utvecklingen och användningen av ADB och AI

Utvecklingen av ADB och AI pågår ständigt och användningsområdena kommer sannolikt att förändras och utvecklas.

Inom området **rekrytering i arbetslivet** kan sammanfattningsvis sägas att produktutvecklare anser att de digitala rekryteringsverktyg (inbegripet robotar) som utvecklas ska bidra till att bland annat effektivisera rekryteringsprocesser, genomföra och analysera jobbintervjuer utan att påverkas av fördomar, magkänsla och andra subjektiva faktorer, samt vara tidsbesparande. Det pågår projekt i Sverige och i andra länder för att utveckla AI-baserade rekryteringsverktyg och robotar.

ADB eller AI används, eller kan förväntas komma att användas, för bland annat följande ändamål:

- Bakgrundskontroller via sökningar eller sökrobotar på internet.
- Ansökningar via webbförmulär som sedan behandlas av robotar.
- Granskning av CV och personliga brev i en första sällning av kandidater innan en människa tar vid i nästa skede i rekryteringsprocessen.
- Digitala eller webbaserade tester av exempelvis intelligenskvot.
- Numerisk analys, läsförståelse och personlighet.
- Intervjuer med robotar.
- Teknik för analys av ansiktsuttryck och ansiktsrörelser på videofilmade intervjuer.
- Så kallade chatbotar på rekryteringsföretags webbplatser.
- Automatiska sökningar i kandidatdatabaser med efterföljande automat-genererade mail till potentiella kandidater.

Inom området **krediter** har banker och kreditinstitut börjat intressera sig för den nya tekniken och den stora mängd information om individer som i dag kan inhämtas i så kallad stordata, som genereras av bland annat användning av smarta telefoner och internet. Den information man kan hämta i stordata kan bland annat ge underlag för att göra modeller och

---

<sup>6</sup> Prop. 2007/08:95, s. 546.

analyser som används i syfte att försöka förutspå framtida scenarier samt beteenden hos individer. Teknikutvecklingen har också medfört att en marknad för informations- eller datamäklare (eng: data brokers) har växt fram. Dessa aktörer samlar in information om konsumenter från många källor och säljer profiler till näringsidkare.

I dag finns modeller och algoritmer som utgår från andra källor än de traditionella för kreditvärdering, för att identifiera kreditvärdiga personer och företag där bankerna inte gör det. Det öppnar för nya kundgrupper och anses göra produkterna mer konkurrenskraftiga. Information hämtas från alternativa datakällor, som kan vara mobilanvändning, hur ofta man skickar pengar till andra, vad man köper på nätet och i handeln eller om man pendlar regelbundet. En persons sociala nätverk och beteende kan också vara relevant. Användning av alternativa datakällor är enligt dess förespråkare en förutsättning för att fler ska få tillgång till finansiella tjänster, men det är också en balansakt där personlig integritet och diskriminering är risker som måste hanteras.<sup>7</sup> Förekomsten av kreditgivning från svenska kreditgivare med stöd av information från alternativa datakällor är okänt.

### **Risker för diskriminering**

Användningen av ADB och AI kan komma att medföra en risk för diskriminering, beroende på hur tekniken och verktygen eller systemen programmeras, tränas, utvecklas och används samt vilken möjlighet till insyn i programvara och algoritmer som kommer att finnas. Därtill kommer eventuell framtida reglering på området i form av till exempel lagar, riktlinjer, rekommendationer, branschetiska regler eller vad som kan komma att anses vara lämpligt.

Vid utvecklingen och användningen av AI är det viktigt dels att beakta de etiska aspekterna i form av att respektera grundläggande rättigheter, tillämplig lagstiftning samt grundläggande principer och värderingar, dels göra tekniken robust och tillförlitlig/pålitlig (eng: trustworthy) för att undvika avsiktlig och oavsiktlig skada. Möjligheten att kontrollera, följa upp och utvärdera uppbyggnaden, utvecklingen och användningen av algoritmer är också betydelsefullt – det som inbegrips i det som kallas transparens (eng: transparency). Detta för att kunna identifiera exempelvis felkällor i de

---

<sup>7</sup> Benjamin Ståhl, FinTech - Den digitala transformeringen av finansmarknaden, Blue Institute, Stockholm, 2017, s. 34f; Kevin Petrasic, Benjamin Saul, James Greig, Matthew Bornfreund och Katherine Lamberth, Algorithms and bias: What lenders need to know, White & Case LLP, 2017, s. 3.



resultat en algoritm producerar eller felaktigheter (eng: bias) i den data som en algoritm tränats i eller programmerats med.<sup>8</sup>

Sammanfattningsvis kan konstateras att det är viktigt med kunskap om hur tekniken fungerar, i vilka sammanhang tekniken används samt vilka konsekvenser användningen kan få. I DO:s tillsyn kommer det att vara viktigt att få tillgång till information om hur en algoritm har fattat ett beslut eller på annat sätt utgjort en del i ett beslutsunderlag, samt vilka data som använts. Det kan också vara viktigt att göra allmänheten uppmärksam på problematiken.

---

<sup>8</sup> Se till exempel EU-kommissionens High-Level Expert Group on Artificial Intelligence, Draft Ethics Guidelines for Trustworthy AI, Working document for stakeholders' consultation, Bryssel, 18 december 2018, s. i; Nationell inriktning för artificiell intelligens, artikelnummer N2018.14, Stockholm 2018.

## Inledning

Denna promemoria fokuserar på automatiserade urvalsprocesser som äger rum när individer rekryteras till anställning samt ansöker om krediter.

## Syfte

Syftet med promemorian är att inom de två områdena rekrytering till anställning samt ansökan om kredit eller lån redogöra för utvecklingen och användandet av automatiserad databehandling med algoritmer och stordata samt AI och eventuell koppling till diskriminering och/eller risker för diskriminering.<sup>9</sup> Promemorian avser därutöver att besvara följande frågor:

- Omfattas användningen av ADB och AI av diskrimineringslagen (2008:567), DL?
- Om användningen av ADB och AI omfattas av DL, vem kan hållas ansvarig för potentiell diskriminering som ägt rum?
- Vilka möjligheter eller svårigheter har DO att i sin tillsyn utreda potentiell diskriminering genom användning av ADB och AI?

## Avgränsningar

Denna promemoria behandlar endast de bestämmelser om förbud mot diskriminering i DL och annan lagstiftning som är av relevans för de två samhällsområdena rekrytering (arbetsliv) samt ansökan om krediter (tjänster). Området för användningen av ADB och AI är under utveckling och förändras snabbt, varför slutsatserna i denna promemoria kan bli inaktuella.

Promemorian redogör inte för ADB och AI i förhållande till DL:s bestämmelser om aktiva åtgärder i 3 kap. DL.

Avsikten med promemorian är att ge en övergripande bild av utvecklingen och användningen av ADB och AI, var de största riskerna i diskrimineringshänseende kan sägas finnas, samt ge exempel på användning av ADB och AI som identifieras inom ramen för AI-projektet. Promemorian gör inte anspråk på att vara uttömmande i fråga om utvecklingen och användandet av ADB och AI inom de två områden som promemorian avser.

---

<sup>9</sup> Se projektdirektivet, LED 2018/387, handling 1, s. 3, dokumentet Underlag för prioritering av undersökningsområden för AI-projektet, LED 2018/387, handling 7, s. 3-4, samt Beslut om fördjupad kartläggning inom ramen för uppdraget AI och diskriminering, LED 2018/387, handling 8.

## Problem med begreppet diskriminering i en AI-kontext

Många exempel i det material som vi inom ramen för AI-projektet och denna promemoria har tagit del av har handlat om att personer särbehandlats utifrån ekonomiska förhållanden och social status, vilket ju faller utanför tillämpningsområdet för DL och de olika EU-direktiven på diskriminerings- och likabehandlingsområdet. En svaghet som lyfts av flera, bland annat professor Frederik Zuiderveen Borgesius<sup>10</sup>, är att diskrimineringslagstiftning typiskt sett tar sikte på skyddade grunder. Många av de nya typerna av AI-driven särbehandling framstår som orättvisa och problematiska, men faller utanför tillämpningsområdet för de flesta icke-diskrimineringsförfattningar. Även forskaren jur. dr. Stanley Greenstein framhåller i sin doktorsavhandling att beslut som fattas med hjälp av prediktiv modellering, utförd av AI, kan vara diskriminerande i ordets mer generella betydelse, utan att strida mot diskrimineringslagstiftningen, enligt vilken det krävs att särbehandlingen kan kopplas till en specifik diskrimineringsgrund.<sup>11</sup>

En viktig utgångspunkt för förståelsen av problematiken som behandlas i denna promemoria är alltså att definitionen av begreppet diskriminering kan variera i en internationell kontext, vilket därför kan påverka hur man ser på till exempel omfattningen av riskerna för diskriminering kopplat till AI. Det kan även påverka val av angreppssätt för att adressera problematiken. Vidare kan det i framtiden finnas anledning att reflektera över de diskrimineringsgrunder som skyddas av DL, om det kan finnas anledning att diskutera eller lyfta frågan om DL:s tillämpningsområde borde vidgas och i så fall om DO borde förespråka en så kallad öppen lista.

## Begrepp

I denna promemoria används begreppet ”krediter” som ett samlingsbegrepp för olika former av lån.

I framställningen används förkortningen ”ADB och AI” i betydelsen alla processer med automatiserade eller algoritmiska inslag, inklusive automatiserat beslutsstöd, och omfattar allt från automatiserad databehandling (ADB) till AI. För definition av begrepp, se Ordlista för vissa centrala begrepp för kartläggningen, sidan 50.

---

<sup>10</sup> Discrimination, artificial intelligence and algorithmic decision-making, Frederik Zuiderveen Borgesius, Europarådet, s. 20.

<sup>11</sup> Greenstein, S., *Our Humanity Exposed, Predictive Modelling in a legal context*, 2017, s. 204f.

## Bestämmelser om diskriminering

I detta avsnitt redogörs för relevanta bestämmelser inom diskrimineringsområdet.

### Diskrimineringslagen

#### Arbetslivsområdet – rekrytering

I 2 kap. 1 § DL framgår att arbetsgivare<sup>12</sup> inte får diskriminera arbetstagare, arbetssökande eller den som gör en förfrågan om arbete, är praktikant eller är inhyrd eller inlånad arbetskraft. Förbudet mot diskriminering i form av bristande tillgänglighet gäller inte i fråga om den som gör en förfrågan om arbete. Enligt 2 kap. 1 § andra stycket DL ska den som i arbetsgivarens ställe har rätt att besluta i frågor som rör någon som avses i bestämmelsens första stycke likställas med arbetsgivaren.

Det finns i 2 kap. 2 § DL undantag från huvudregeln om diskrimineringsförbud i 2 kap. 1 § DL. Undantagsregeln anger att förbudet mot diskriminering i 2 kap. 1 § DL inte hindrar (1) särbehandling som föranleds av en egenskap som har samband med någon av diskrimineringsgrunderna om, vid beslut om anställning, befordran eller utbildning för befordran, egenskapen på grund av arbetets natur eller det sammanhang där arbetet utförs utgör ett verkligt och avgörande yrkeskrav som har ett berättigat syfte och kravet är lämpligt och nödvändigt för att uppnå syftet, (2) åtgärder som är ett led i strävanden att främja jämställdhet mellan kvinnor och män och som avser annat än löne- eller andra anställningsvillkor, (3) tillämpning av åldersgränser för rätt till pensions-, efterlevande- eller invaliditetsförmåner i individuella avtal eller kollektivavtal, eller (4) särbehandling på grund av ålder, om den har ett berättigat syfte och de medel som används är lämpliga och nödvändiga för att uppnå syftet.

#### Tillhandahållande av tjänster

Enligt 2 kap. 12 § 1 st. DL är diskriminering förbjuden för den som utanför privat- och familjelivet tillhandahåller bland annat tjänster till allmänheten.

---

<sup>12</sup> Begreppet "arbetsgivare" kan ha olika betydelse i olika situationer och det är inte ovanligt att frågan om vem som rättsligt sett ska betraktas som arbetsgivare i förhållande till en anställd är en fråga som måste prövas rättsligt. En utgångspunkt är emellertid att "arbetsgivaren" är den som är motpart till arbetstagaren i anställningsavtalet. Både juridiska och fysiska personer kan vara arbetsgivare (Lunning/Toijer, Anställningsskydd, Wolters Kluwer, 11 uppl. 2016, s. 43, 46, 51.) Inom diskrimineringslagstiftningen på arbetslivsområdet har utgångspunkten varit att arbetsgivarbegreppet ska vara detsamma som inom arbetsrätten i övrigt (se till exempel prop. 1978/79:175, s.111, och prop. 1990/91:113, s. 181 f.).

Den som i förhållande till allmänheten företräder den som avses i bestämmelsens första stycke ska likställas med denne.

Enligt förarbetena till DL avses med ”tjänst” något som någon gör eller utför åt någon annan mot ersättning. Som exempel nämns hantverks- och arbetsprestationer, lån och försäkringar samt kurser för enskilda. Med tillhandahållande avses köp, byte, upplåtelse och gåva, det vill säga den civilrättsliga transaktionen. Regeln innebär att det i princip inte är tillåtet att tillämpa olika priser eller andra villkor på något sätt som har samband med någon av diskrimineringsgrunderna.<sup>13</sup>

I kravet på tillhandahållande till allmänheten ligger att tjänsten ska erbjudas till en obestämd krets av personer. Både transaktioner som sker yrkesmässigt eller annars i näringsverksamhet och privatpersoners transaktioner faller inom förbudets tillämpningsområde.<sup>14</sup>

Begreppet ”tjänst” är ett EU-rättsligt begrepp och ska tolkas i enlighet med EU-rätten. Tjänster definieras i artikel 57 FEUF<sup>15</sup> som prestationer som normalt utförs mot ersättning, i den utsträckning de inte faller under bestämmelserna om fri rörlighet för varor, kapital och personer. Med tjänster ska särskilt avses verksamhet av industriell natur, av kommersiell natur, inom hantverk samt inom fria yrken.

#### Undantag från diskrimineringsförbudet

Det finns i 2 kap. 12 a - c §§ DL ett antal undantag från diskrimineringsförbudet i 2 kap. 12 § DL. Undantagen gäller i vissa situationer, för diskrimineringsgrunderna kön, ålder och bristande tillgänglighet.

#### Undantag för diskrimineringsgrunden kön

Enligt 2 kap. 12 a § DL hindrar inte förbudet mot diskriminering som har samband med kön i 2 kap. 12 § DL att kvinnor och män behandlas olika i fråga om bland annat tjänster, om det har ett berättigat syfte och de medel som används är lämpliga och nödvändiga för att uppnå syftet. Enligt lagens förarbeten<sup>16</sup> och fast praxis från EU-domstolen ska undantag tillämpas restriktivt.

---

<sup>13</sup> Prop. 2007/08:95, s. 518.

<sup>14</sup> Prop. 2007/08:95, s. 518f.

<sup>15</sup> Fördraget om Europeiska Unionens funktionssätt.

<sup>16</sup> Prop. 2007/08:95, s. 520.

Undantag för diskrimineringsgrunden ålder

I 2 kap. 12 b § DL framgår att förbudet mot diskriminering som har samband med ålder i 2 kap. 12 § DL inte hindrar särbehandling på grund av ålder bland annat vid tillämpning av bestämmelser i lag som föreskriver viss ålder, vid tillhandahållande av försäkringstjänster, eller om särbehandlingen har ett berättigat syfte och de medel som används är lämpliga och nödvändiga för att uppnå syftet.

## EU-rätten

Inom EU-rätten finns flera direktiv på arbetslivets område och området tillhandahållande av varor och tjänster, exempelvis det så kallade Recast-direktivet<sup>17</sup> och Rådets direktiv 2004/113/EG av den 13 december 2004 om genomförande av principen av likabehandling av kvinnor och män när det gäller tillgång till och tillhandahållande av varor och tjänster. Där till ska beaktas praxis från EU-domstolen vid tolkning och tillämpning av direktiv och lagstiftning som införlivar direktiven i nationell rätt.

I EU:s grundfördrag finns bestämmelser om icke-diskriminering. I EU-stadgan om de grundläggande rättigheterna (EU-stadgan)<sup>18</sup> framgår i artikel 21 att all diskriminering på grund av bland annat kön, ras, hudfärg, etniskt eller socialt ursprung, genetiska särdrag, språk, religion eller övertygelse, politisk eller annan åskådning, tillhörighet till nationell minoritet, förmögenhet, börd, funktionshinder, ålder eller sexuell läggning ska vara förbjuden.

I artikel 2 i Fördraget om Europeiska unionen (EU-fördraget)<sup>19</sup> slås bland annat fast att unionen ska bygga på värdena respekt för människans värdighet, frihet, demokrati, jämlikhet, rättsstaten och respekt för de mänskliga rättigheterna och att dessa värden ska vara gemensamma för medlemsstaterna i ett samhälle som kännetecknas av mångfald, icke-diskriminering, tolerans, rättvisa, solidaritet och principen om jämställdhet mellan kvinnor och män. I artikel 18 i Fördraget om Europeiska unionens funktionssätt (FEUF-fördraget)<sup>20</sup> framgår också att inom fördragens tillämpningsområde och utan att det påverkar tillämpningen av någon

---

<sup>17</sup> Europaparlamentets och rådets direktiv 2006/54/EG av den 5 juli 2006 om genomförandet av principen om lika möjligheter och likabehandling av kvinnor och män i arbetslivet.

<sup>18</sup> Europeiska Unionens stadga om de grundläggande rättigheterna (2010/C 83/02), C 83/392.

<sup>19</sup> Fördraget om Europeiska unionen (konsoliderad version 2016), EGT C 202, 7.6.2016.

<sup>20</sup> Fördraget om Europeiska unionens funktionssätt (konsoliderad version 2016), EGT C 202, 7.6.2016

särskild bestämmelse i fördragen, ska all diskriminering på grund av nationalitet vara förbjuden.

## Europakonventionen

Europeiska konventionen om skydd för de mänskliga rättigheterna (EKMR) är inkorporerad som lag i Sverige genom lag (1994:1219) om den europeiska konventionen angående skydd för de mänskliga rättigheterna och de grundläggande friheterna. Av relevans för denna promemoria är bland annat artikel 8 och artikel 1 i tilläggsprotokoll nr 12 som beskriver rätten till skydd för privat- och familjeliv samt artikel 14 som föreskriver att åtnjutandet av de fri- och rättigheter som anges i konventionen ska säkerställas utan någon åtskillnad såsom på grund av kön, ras, hudfärg, språk, religion, politisk eller annan åskådning, nationellt eller socialt ursprung, tillhörighet till nationell minoritet, förmögenhet, börd eller ställning i övrigt.

## Bestämmelser om ADB och AI

I detta avsnitt redogörs för relevanta bestämmelser eller andra dokument i fråga om ADB och AI, som vi har kunnat identifiera inom ramen för arbetet med denna promemoria.

### Regler på nationell nivå

Den 25 maj 2018 trädde lag (2018:218) med kompletterande bestämmelser till EU:s dataskyddsförordning (dataskyddslagen) samt förordning (2018:219) med kompletterande bestämmelser till EU:s dataskyddsförordning ikraft. Utgångspunkten för dataskyddslagen och tillhörande förordning var EU:s nya dataskyddsförordning (dataskyddsförordningen)<sup>21</sup> som också den trädde i kraft den 28 maj 2018. De nationella reglerna är kompletterande lagar inom vissa av dataskyddsförordningens områden, och dataskyddslagen är subsidiär i förhållande till avvikande bestämmelser i annan lag eller förordning som reglerar behandling av personuppgifter.

---

<sup>21</sup> Europaparlamentets och Rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning).

## Regler på EU-nivå

I dataskyddsförordningen fastställs bland annat bestämmelser om skydd för fysiska personer med avseende på behandlingen av personuppgifter och om det fria flödet av personuppgifter. Förordningen skyddar fysiska personers grundläggande rättigheter och friheter, särskilt deras rätt till skydd av personuppgifter. Dataskyddsförordningen ska som huvudregel tillämpas på behandling av personuppgifter som helt eller delvis företas på automatisk väg samt på annan behandling än automatisk av personuppgifter som ingår i eller kommer att ingå i ett register. Dataskyddsförordningen berör bland annat profilering, i artikel 4.1 definierat som varje form av automatisk behandling av personuppgifter som används för att bedöma vissa personliga egenskaper hos en fysisk person, i synnerhet för att analysera eller förutsäga denna fysiska persons arbetsprestationer, ekonomiska situation, hälsa, personliga preferenser, intressen, pålitlighet, beteende, vistelseort eller förflyttningar. I artikel 22 framgår i fråga om automatiserat individuellt beslutsfattande, inbegripet profilering att den registrerade som huvudregel ska ha rätt att inte bli föremål för ett beslut som enbart grundas på automatiserad behandling, inbegripet profilering, som har rättsliga följder för eller på liknande sätt i betydande grad påverkar honom eller henne.

I EU:s stadga om de grundläggande rättigheterna finns bestämmelser som behandlar de människans rätt till integritet (artikel 3), respekt för privatlivet och familjelivet (artikel 7), skydd av personuppgifter (artikel 8), förbud mot diskriminering (artikel 21), och de garanterade rättigheternas räckvidd (artikel 52).

## Deklarationer och riktlinjer med mera om ADB och AI

I detta avsnitt redogörs för relevanta deklarerationer och riktlinjer i fråga om ADB och AI.

### Nationell nivå

Regeringen har antagit en nationell inriktning för artificiell intelligens<sup>22</sup>. Regeringen uttalar bland annat att det är regeringens målsättning att Sverige ska vara ledande i att ta tillvara möjligheterna som användning av AI kan ge, med syftet att stärka både den svenska välfärden och den svenska konkurrenskraften. Vidare anförs att Sverige behöver utveckla regler,

---

<sup>22</sup> Nationell inriktning för artificiell intelligens, artikelnummer N2018.14, Näringsdepartementet, Regeringskansliet, Stockholm 2018.



standarder, normer och etiska principer i syfte att vägleda etisk och hållbar AI och användning av AI.

## EU-nivå

Inom EU och olika internationella organisationer pågår arbete med att ta fram olika riktlinjer och liknande inom området.

### EU-kommissionen

Den 10 april 2018 undertecknade 25 EU-medlemsstater en deklARATION om artificiell intelligens. EU-kommissionen har tagit fram en strategi för artificiell intelligens Artificial Intelligence for Europe (COM [2018] 237 final), där tre pelare utgör kärnan i den föreslagna strategin: EU ska (1) vara i framkant när det gäller teknologisk utveckling och industriell kapacitet och stärka offentliga och privata aktörer, (2) förbereda för socioekonomiska förändringar som AI för med sig, genom att modernisera utbildningssystem och stötta system för övergångar på arbetsmarknaden och sociala skyddsnet och (3) säkerställa ett lämpligt etiskt och juridiskt ramverk för AI, baserat på EU:s värderingar och i linje med EU-stadgan för grundläggande rättigheter.

EU-kommissionen har också i en kommunikation, Coordinated Plan on Artificial Intelligence (COM [2018] 795 final), den 7 december 2018 lagt fram en koordinerad plan för AI. Syftet är att maximera effekten av investeringar i EU och på nationell nivå, främja synergier och samarbete över Europa, utbyta "best practices" och definiera en väg framåt för att säkerställa att EU kan konkurrera globalt. Kommissionen uppmanar också medlemsstaterna att genomföra den koordinerade planen och utveckla nationella AI-strategier till mitten av 2019, samt uppmanar Rådet och Europaparlamentet att anta återstående lagstiftningsinitiativ på området. Kommissionen har också utsett en självständig expertgrupp<sup>23</sup>, för att ta fram etiska riktlinjer för AI.

Expertgruppen presenterade den 8 april 2019 slutversionen av sina etiska riktlinjer för tillförlitlig AI (Ethics Guidelines for Trustworthy AI)<sup>24</sup>. Tillförlitlig AI består av tre komponenter som överlappar och samverkar med varandra; AI bör för det första vara lagenlig (eng: lawful) och förenlig med alla tillämpliga lagar och regler. AI bör för det andra vara etisk (eng: ethical), säkerställa anslutning till etiska principer och värderingar, och för det tredje vara robust (eng: robust), både från ett tekniskt och ett socialt

---

<sup>23</sup> EU-kommissionens Independent High-Level Expert Group on Artificial Intelligence.

<sup>24</sup> <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>.

perspektiv eftersom AI kan orsaka oavsiktlig skada även om intentionerna är goda. Expertgruppen presenterar i sina etiska riktlinjer fyra etiska principer och sammanhängande värderingar som måste följas för att garantera att AI utvecklas på ett människocentrerat sätt (eng: "human-centric manner"). Dessa principer bygger på principer som tagits fram av The European Group on Ethics in Science and new technologies (EGE) och AI4Peoples projekt "AI4People – An Ethical Framework for a Good AI society: Opportunities, Risks, Principles and Recommendations".<sup>25</sup> De fyra etiska principerna är följande:

**Respect for human autonomy**, vilket bland annat innebär mänsklig översyn över arbetsprocesser och rätt till mänskligt självbestämmande.

**Prevention of harm**, vilket innefattar bland annat att skydda mänsklig värdighet liksom mental och kroppslig integritet, att säkerställa att AI-system inte är öppna för ondskefull användning och att hänsyn tas till naturen och alla levande varelser.

**Fairness** innebär bland annat att säkerställa jämlik och rättvis fördelning av fördelar och kostnader. Det innebär också att säkerställa att individer och grupper är fria från fördomar, diskriminering och stigmatisering, samt respekt för proportionalitetsprincipen i förhållandet mellan medel och mål. Det handlar också om möjligheten att överklaga beslut som fattats av AI-system, vilket förutsätter att den enhet som är ansvarig för ett beslut kan identifieras och att beslutsprocesser kan förklaras.

**Explicability** betyder att processer måste vara transparenta, med öppen kommunikation kring förmågor och syfte med AI-system och beslut måste (i den utsträckning det är möjligt) vara möjliga att förklara för dem som påverkas direkt eller indirekt. Om en förklaring inte är möjlig krävs i stället andra förklaringsåtgärder såsom spårbarhet, möjlighet till granskning och transparent kommunikation om systemets förmågor (eng: capabilities).

I riktlinjerna anges också att det bör fästas särskild uppmärksamhet på situationer som rör mer sårbara grupper som barn, personer med funktionsnedsättning eller andra som historiskt har missgynnats eller riskerar uteslutning, eller på situationer med ojämlika förhållanden avseende makt eller information.

I syfte att erbjuda vägledning kring genomförandet och förverkligandet av tillförlitlig AI innehåller riktlinjerna också en icke uttömmande lista med sju

---

<sup>25</sup> Minds and Machines, 28(4): 689–707.

kriterier som bygger på de fyra principerna; Human agency and oversight, Technical robustness and safety, Privacy and data governance, Transparency, Diversity, Non-discrimination and fairness, Societal and environmental wellbeing och Accountability.

### **Europeiska Ekonomiska och Sociala Kommittén**

Europeiska Ekonomiska och Sociala Kommittén (EESK) har på eget initiativ lämnat ett yttrande<sup>26</sup> om konsekvenserna av AI för arbetslivet. Syftet med yttrandet är att föregripa konsekvenserna av AI för arbetslivet för att säkerställa en rättvis övergång. I yttrandet förespråkar EESK bland annat att etiska riktlinjer ska inbegripa öppenhetsprinciper. Dessa öppenhetsprinciper gäller användning av AI-system för anställning, bedömning och ledningsmässig kontroll av arbetstagare samt principer för hälsa och säkerhet och förbättring av arbetsvillkoren, samt garantera skyddet av rättigheter och friheter när det gäller behandling av arbetstagarnas personuppgifter, i enlighet med principerna om icke-diskriminering. EESK anger också att effekterna av AI på sysselsättning och arbete kräver särskild uppmärksamhet på politisk nivå, eftersom institutionerna i synnerhet har till uppgift att se till att de ekonomiska omvandlingsprocesserna blir socialt hållbara.

### **CEPEJ:s etiska stadga för användning av AI i rättssystem**

Europarådets European Commission for the Efficiency of Justice (CEPEJ) har utarbetat det första rättsliga instrumentet på europainivå som ställer upp fem väsentliga och metodologiska principer som är tillämpliga för det automatiska processandet av juridiska beslut, domstolsbeslut och data, baserade på AI-teknik. CEPEJ:s etiska stadga har tagits med i promemorian för att belysa vilka regleringar som finns på europainivå avseende AI, men stadgan har inte någon direkt tillämpning på de områden som behandlas i promemorian.

De fem principerna beskrivs i dokumentet "The CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment".<sup>27</sup> Principerna antogs av CEPEJ i Strasbourg den 3 december 2018. De fem principerna är följande, och återges på engelska för att undvika eventuell förvanskning:

---

<sup>26</sup> Yttrande från Europeiska ekonomiska och sociala kommittén om Artificiell intelligens, yttrande på eget initiativ (2018/C 440/01), vid dess 537:e plenarsession den 19–20 september 2018, publicerat i Official Journal den 6 december 2018.

<sup>27</sup> <https://www.coe.int/en/web/cepej>.

1. **Principle of respect for fundamental rights:** Ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights.
2. **Principle of non-discrimination:** Specifically prevent the development or intensification of any discrimination between individuals or groups of individuals.
3. **Principle of quality and security:** With regard to the processing of judicial decisions and data, use certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment.
4. **Principle of transparency, impartiality and fairness:** Make data processing methods accessible and understandable, authorise external audits.
5. **Principle “under user control“:** Preclude a prescriptive approach and ensure that users are informed actors and in control of their choices.

CEPEJ skriver i sin presentation<sup>28</sup> av den etiska stadgan avseende användningen av AI i rättssystem att bland de fem principerna är respekten för mänskliga rättigheter och icke-diskriminering av principiell betydelse. Syftet är att säkerställa, från idé till praktisk tillämpning, att lösningarna garanterar respekt för de rättigheter som garanteras i EKMR och Europarådets konvention Nr. 108. Principen om icke-diskriminering anges uttryckligen med anledning av förmågan hos vissa bearbetningsoperationer med automatiska medel – särskilt i brottmål – att avslöja befintlig diskriminering genom att aggregera eller klassificera data relaterad till personer eller grupper av personer. Offentliga och privata aktörer måste därför säkerställa att dessa tillämpningar inte reproducerar eller förvärrar denna diskriminering och inte leder till deterministiska analyser eller praxis.

## DL:s tillämplighet

I detta avsnitt redogörs för rättsfrågorna om DL:s tillämplighet vid användning av ADB och AI samt vem som enligt DL kan hållas ansvarig för eventuell diskriminering.

---

<sup>28</sup> <https://rm.coe.int/presentation-note-en-for-publication-4-december-2018/16808f699d>, hämtad 2019 – 04 – 02.

## Arbetslivsområdet – rekrytering

### Diskrimineringsförbudet för arbetsgivare

Av 2 kap. 1 § DL framgår att arbetsgivare inte får diskriminera arbetstagare, arbetssökande eller den som gör en förfrågan om arbete, är praktikant eller är inhyrd eller inlånad arbetskraft<sup>29</sup>. Lagtexten anger den personkrets som omfattas av diskrimineringsförbudet<sup>30</sup>, det vill säga vem som skyddas av bestämmelsen och vem som kan bryta mot diskrimineringsförbudet. Diskrimineringsförbudet gäller således utan begränsningar till vissa angivna situationer. Alla situationer som kan relateras till arbetet och som gäller förhållandet mellan arbetsgivare och arbetstagare omfattas av förbudet.<sup>31</sup> Skyddet för arbetssökande gäller under hela rekryteringsprocessen: själva anställningsbeslutet och alla åtgärder och underlåtenheter som föregår detta.<sup>32</sup> Den som ”tydligt gett till känna att han eller hon söker en anställning” räknas som arbetssökande. Även den som enbart har gjort en förfrågan om arbete omfattas av förbudet mot diskriminering.<sup>33</sup>

Om en arbetssökande diskrimineras av en potentiell arbetsgivare inom ramen för ett rekryteringsförfarande bör man kunna utgå från att DL är tillämplig, utifrån vad som anförts ovan om att diskrimineringsförbudet i lagen anger en personkrets som skyddas, och inte specifika situationer. Om arbetsgivaren använder sig av ett digitalt verktyg som behandlar information eller tar fram beslutsunderlag med stöd av AI eller ADB så är det fortfarande arbetsgivarens ansvar.

### Vem som kan likställas med arbetsgivare

Av 2 kap. 1 § tredje stycket DL följer att den som i arbetsgivarens ställe har rätt att besluta i frågor som rör någon som avses i första stycket<sup>34</sup> ska likställas med arbetsgivaren. I förarbetena till bestämmelsen uttalas att: ”Om en arbetsgivare på något sätt delegerat sin beslutanderätt eller arbetsgivarfunktion till någon är det klart att denne ”är” arbetsgivaren på så sätt att hans eller hennes ageranden är att likställa med arbetsgivarens

---

<sup>29</sup> I 2 kap. 2 § DL finns en regel som i vissa fall medger undantag från huvudregeln i 2 kap. 1 § DL.

<sup>30</sup> I lagtexten i tidigare diskrimineringslagar tillämpliga inom arbetslivets område framgick specifika typsituationer på vilka diskrimineringsförbudet var tillämpligt, exempelvis vid anställning, befordran och uppsägning.

<sup>31</sup> Prop. 2007/08:95, s. 134.

<sup>32</sup> Fransson/Stüber, Diskrimineringslagen – en kommentar, 2 uppl., Norstedts, 2015, s. 129.

<sup>33</sup> Prop. 2007/08:95, s. 135.

<sup>34</sup> Det vill säga arbetstagare, arbetssökande, praktikanter, inhyrda et cetera.

handlande.”<sup>35</sup> När det gäller rekryteringsföretag uttalas att ”[o]m en arbetsgivare anlitar så kallade headhunters eller rekryteringsföretag för att sköta eller hjälpa till i ett anställningsförfarande likställs även handlande av dessa personer med arbetsgivarens handlande”.<sup>36</sup> Vidare uttalas följande: ”Det får sedan ankomma på arbetsgivaren att i sin tur vända sig till sin avtalspart och regressvis söka ersättning. En sådan ordning innebär till exempel att en arbetsgivare inte kan kringgå diskrimineringsförbudet genom att till exempel uppdra åt ett rekryteringsföretag att sköta en rekrytering och därvid ge direktiv att sälla bort arbetssökande till exempel av ena könet eller av viss etnisk tillhörighet.”

### **Instruktioner att diskriminera**

Enligt 1 kap. 4 § 6 DL räknas också som diskriminering de fall där någon instruerar någon annan att diskriminera. Instruktionen eller ordern att diskriminera ska lämnas till någon som står i beroende- eller lydnadsförhållande till den som uppmanar till att diskriminera. Förbudet gäller också när instruktioner ges till någon som åtagit sig att fullgöra ett uppdrag. Det kan handla om den som står i ett uppdragsförhållande till den som ger instruktionen och exempelvis åtagit sig att välja ut personer i olika typer av urvalssituationer. Det kan till exempel vara ett bemanningsföretag, arbetsförmedlingen, en antagningsenhet eller en fastighetsmäklare.<sup>37</sup>

### **Ansvar vid diskriminering**

I enlighet med vad som anförts ovan är DL tillämplig om en arbetssökande diskrimineras av en presumtiv arbetsgivare inom ramen för ett rekryteringsförfarande. Det följer av att diskrimineringsförbudet i lagen anger en personkrets som skyddas, och inte specifika situationer. Vidare anges vem som förbudet att diskriminera riktar sig mot och vem som ansvarar – nämligen arbetsgivaren (eller den som är att likställa med arbetsgivaren). Om arbetsgivaren använder sig av ett digitalt verktyg som behandlar information eller tar fram beslutsunderlag med stöd av AI eller ADB så är det arbetsgivarens ansvar om en arbetssökande diskrimineras.

Det torde därmed vara arbetsgivarens ansvar om användningen av en algoritm leder till att någon diskrimineras i ett rekryteringsförfarande, oavsett vilken insyn arbetsgivaren har i hur algoritmen fungerar.

---

<sup>35</sup> Prop. 2007/08:95, s. 137.

<sup>36</sup> Prop. 2007/08:95, s. 501.

<sup>37</sup> Fransson/Stüber, Diskrimineringslagen – en kommentar, 2 uppl., Norstedts, 2015, s. 89 och prop. 2007/08:95, s. 494.

Diskrimineringsförbudet på arbetslivsområdet är avsett att vara heltäckande, vilket framgår av följande uttalande i förarbetena: ”Avsikten är att diskrimineringsförbudet ska täcka alla situationer som kan uppkomma mellan en arbetsgivare och en arbetssökande eller arbetstagare på en arbetsplats eller över huvud taget i samband med anställningen.”<sup>38</sup>

Även om till exempel en arbetsgivare kan tänkas invända att denne saknar kunskap om innehållet i koden, och att programmeringen gjorts av och köpts in från ett externt företag är det svårt att se hur den drabbade enskilde (till exempel en arbetssökande) skulle kunna rikta ersättningsanspråk på grund av diskriminering mot någon annan än arbetsgivaren. Av DL framgår tydligt att (bland annat) en arbetsgivare är ansvarig, då det i 5 kap. 1 § 1 st. DL föreskrivs att den som bryter mot förbudet mot diskriminering enligt denna lag ska betala diskrimineringsersättning för den kränkning som överträdelsen innebär. Mellan företaget som använder ett till exempel ett AI-verktyg i en rekryteringsprocess och dataföretaget som utvecklat verktyget finns ett kontraktsförhållande, men däremot finns inte något sådant förhållande mellan den arbetssökande och dataföretaget.

## Tillhandahållande av tjänster

I 2 kap. 12 § DL definieras kretsen som träffas av diskrimineringsförbudet så som ”den som [...] tillhandahåller varor, tjänster eller bostäder till allmänheten [...]”. Diskrimineringsförbudet träffar alltså den fysiska eller juridiska person som, i enlighet med de rekvisit som uppställs i bestämmelsen, tillhandahåller en tjänst.<sup>39</sup>

Enligt 2 kap. 12 § andra stycket DL ska den som i förhållande till allmänheten företräder den som avses i bestämmelsens första stycke likställas med denne. I förarbetena ges exempel på vad detta innebär för den som äger och driver en restaurang eller annat nöjesställe: Det är den som äger och driver restaurangen, butiken et cetera som har ansvaret för hela verksamheten, i den meningen att ingen skillnad görs mellan hur ägarens anställda agerar och hur dennes samarbetspartners handlar gentemot gäster och kunder när diskrimineringsförbudet tillämpas. Bestämmelsen omfattar anställda, och är ett förtydligande av rättsläget och arbetsgivares principalansvar. Formuleringen omfattar också ombud som i förhållande till kunder agerar på någons uppdrag. Syftet med bestämmelsens andra stycke är att tjäna som incitament för verksamhetsutövare att vara noggranna i

---

<sup>38</sup> Prop. 2007/08:95, s. 139.

<sup>39</sup> Prop. 2007/08:95, s. 242.

sina val av medarbetare eller samarbetspartners. Det är också ett sätt att utvidga diskrimineringsförbudets räckvidd på så sätt att även den som inte står i avtalsrelation till gäster eller kunder ska avhållas från att diskriminera. Ersättningsanspråk eller talan i ett diskrimineringsmål kan därmed riktas mot den fysiska eller juridiska person som, i enlighet med de rekvisit som uppställs i bestämmelsen, tillhandahåller en tjänst (i exemplen i förarbetena ägaren till en restaurang eller butik). Om den som tillhandahåller en tjänst med tillämpning av bestämmelsen i 2 kap. 12 § andra stycket DL tvingas betala diskrimineringsersättning för diskriminering som en annan person orsakat så bör det vara möjligt för tillhandahållaren av tjänsten att regressvis söka få sin kostnad ersatt av det företag eller den person som vållat skadan.<sup>40</sup>

### Ansvar vid diskriminering

Utifrån ordalydelsen i och syftet med bestämmelsen i 2 kap. 12 § DL, vilket redogjorts för i föregående stycke, bör man kunna dra slutsatsen att ansvaret för förekomst av diskriminering vid tillhandahållande av en tjänst ligger hos den fysiska eller juridiska person som, i enlighet med de rekvisit som uppställs i bestämmelsen, tillhandahåller tjänsten. Detta oavsett vilken insyn tillhandahållaren av tjänsten har i hur den algoritm eller det program som algoritmen förekommer i fungerar.

## Utvecklingen av ADB och AI

I takt med den tekniska och digitala utvecklingen och den ökande mängden av tillgång till data, exempelvis genom användningen av smarta telefoner och internet, har datainsamlingen ökat. Den stora mängd data som produceras bidrar till det som ibland kallas stordata. Stordata karaktäriseras bland annat av brist på struktur, vilket gör att man söker efter samband och mönster i data. Stordata relaterar också till begrepp som maskinlärande, algoritmer och prediktiva modeller.<sup>41</sup> Nedan beskrivs kort något om utvecklingen av den nya tekniken.

Inom **rekryteringsbranschen** har användningen av AI och ADB ökat stort det senaste decenniet. Det blir allt vanligare att lediga tjänster söks via webformulär och att olika screeningverktyg som digitala begåvnings-, lämplighets- och personlighetstester används i rekryteringsprocesser. Det är

---

<sup>40</sup> Prop. 2007/08:95, s. 242 f.

<sup>41</sup> Stanley Greenstein, *Our Humanity Exposed – Predictive Modelling in a Legal Context*, Stockholm 2017, s. 61 och 77.



även vanligt att automatiska svar på mottagen ansökan skickas till kandidater.

På **kreditmarknaden** har den ökande mängden tillgång till data medfört bland annat att nya metoder för konsumentprofilering utvecklats, eller i vart fall skapat förutsättningar för att utvecklas, från mer traditionella kreditvärderingsmodeller. På området finns också, i förhållande till banker och andra kreditinstitut, nya aktörer som vid kreditprövning använder, eller förespråkar användning av, alternativa datakällor som exempelvis surfhistorik och köpvanor på internet.

## Från deskriptiv till prediktiv analys och modellering

Ny teknologi utvecklas för att utnyttja den ökade datamängden, exempelvis i stordata, för att få fram information som tidigare inte var tillgänglig. Det finns problematiska aspekter i fråga om dels att själva insamlingen ofta sker utan individens insyn, medvetande och kontroll, dels att insamlingen är relevant också i förhållande till vad analysen av informationen kan leda till. Den nya teknologin kan också införlivas i beslutsfattande processer, vilket kan medföra risker om informationen används som en del i beslutsunderlaget, processerna automatiseras i hög utsträckning och man i alltför hög grad litar på den automatiserade processen.<sup>42</sup>

Den utvecklade tekniken har bidragit till att marknadsaktörer har kunnat gå från att vara **deskriptiva** till att vara **prediktiva**. Det innebär att man kunnat gå från att analysera information så att man förstår hur saker och ting är, till att göra beräkningar av sannolikheter.<sup>43</sup> Inte sällan antas att resultat av dataprocesser är objektiva, vilket kan leda till att man förlitar sig på att resultaten alltid är korrekta och återspeglar den sanna verkligheten. Man bör dock ha i åtanke att den rådata som utnyttjas eller analyseras i en dataprocess inte nödvändigtvis behöver vara objektiv, eftersom rådata för det mesta är en produkt av mänsklig aktivitet (jämför avsnitt Risker för diskriminering, sidan 28). Varken data eller tekniken som använder data är alltså nödvändigtvis objektiv.<sup>44</sup>

---

<sup>42</sup> Stanley Greenstein, *Our Humanity Exposed – Predictive Modelling in a Legal Context*, Stockholm 2017, s. 21f.

<sup>43</sup> 2017:1 Kunskapsöversikter inom det konsumentpolitiska området, dnr 2017/34, Konsumentverket, 2017, s. 30; 2016:12 Digitalisering och konsumentintresset – En litteraturöversikt, Konsumentverket 2016, s. 39; Stanley Greenstein, *Our Humanity Exposed – Predictive Modelling in a Legal Context*, Stockholm 2017, s. 21f och 61.

<sup>44</sup> Stanley Greenstein, *Our Humanity Exposed – Predictive Modelling in a Legal Context*, Stockholm 2017, s. 78.

## Prediktiv analys och prediktiv modell

Användningen av **prediktiv analys** (eng: predictive analysis) växer och används i fler och fler sammanhang, exempelvis i bankers arbete mot bedrägerier (när kortaktiviteter avviker från prediktionen kan kortet vara stulet), försäkringsbranschens nytta av att på individnivå kunna prediktera (förutsäga) skada, sjukdom och även levnadslängd, och detaljister och reklambranschen kan dra nytta av att förutsäga vem som handlar vad.<sup>45</sup> Den information man kan få genom prediktiv analys kan användas för att göra **prediktiva modeller** (eng: predictive modelling): Den information man får fram genom prediktiv analys kan införlivas i en statistisk modell som kan tillämpas på framtida scenarier.<sup>46</sup>

I sammanhanget ska nämnas något om skillnaden mellan profilering och prediktiva modeller. Vid **profilering** kategoriseras data för att få insikt i dessa kategorier, utan att för den skull sätta någon etikett på grupperna som sådana. Profilering betonar det förflutna och nutid. Den data som används vid profilering är traditionellt sådan data som är lagrad i databaser för att kategorisera individer och placera dem i fördefinierade grupper. Profilering kan vara variera i komplexitet från enkel till mycket komplex. I dess komplexa form kan profilering likna prediktiva modeller. **Prediktiva modeller** identifierar exempel på beteenden (det vill säga data grupperas), med etiketter för respektive grupp med målet att identifiera dessa grupperingar i nutid och framtid. Prediktiva modeller är framtidsorienterade, på så sätt att man söker efter mönster i mänskligt beteende och försöker förutspå dessa mönster. Den data som används i prediktiva modeller förknippas med data på ett mer generellt plan, oavsett från vilken källa eller struktur data kommer ifrån. Det kan vara sociala medier, bloggar eller andra digitala källor. Fokus är att extrahera trender från sådana generella data.<sup>47</sup>

## Konsumentprofilering

Ett exempel på utvecklingen och användningen av prediktiv analys är så kallad konsumentprofilering. Konsumentprofilering innebär att marknadsaktörer skaffar sig förståelse för konsumenter i syfte att ha informationen

---

<sup>45</sup> Konsumentverket 2016, 2016:12 Digitalisering och konsumentintresset – En litteraturöversikt, s. 39f.

<sup>46</sup> Stanley Greenstein, Our Humanity Exposed – Predictive Modelling in a Legal Context, Stockholm 2017, s.104.

<sup>47</sup> Stanley Greenstein, Our Humanity Exposed – Predictive Modelling in a Legal Context, Stockholm 2017, s. 100–102.

som beslutsunderlag för strategiska beslut och använda den för till exempel marknadsföring eller produktinformation. Företeelsen är inte ny, men metoderna för insamling av data har förändrats och informationsunderlaget för dagens konsumentprofilering inkluderar en stor mängd och volym av informationskällor där bland annat dagens smarta telefoner och den data som kan erhållas från individens användning av sin smarta telefon kan ge mycket information som marknadsaktörerna är intresserade av att få tillgång till. Informationen används ofta för att till exempel kunna skilja ut de mest lönsamma respektive olönsamma kunderna från andra, eller för att förutsäga sannolikheten för vilka kunder som kommer att sköta avbetalningar.<sup>48</sup>

Inhämtandet av information från stordata sker ofta genom automatiserade processer genom exempelvis algoritmer som programmerats att reglera exempelvis hur inhämtad information ska hanteras, tolkas och hur automatiserade beslut ska tas, vilket gör att det är svårt att observera de inblandade processerna. Insynen i processerna försvåras alltså, vilket brukar kallas brist på transparens. Brist på transparens i fråga om exempelvis algoritmisk design och automatiserade utfall är relevant i relation till diskriminering.<sup>49</sup>

Med ökad digital konsumentprofilering följer frågor kring diskriminering, och dessa frågor behöver adresseras. Forskare har i ett experiment byggt ett mjukvaruverktyg för att spåra hur beteende och användarinställning hängde ihop. Experimentet visade bland annat hur man fick färre reklam erbjudanden om höginkomsttjänster om användaren angav att den var kvinna i sina inställningar jämfört med om den angav att den var man. Detta enkla exempel visar att behovet av att kunna studera automatiserade utfall och algoritmisk design i relation till bland annat diskriminering. Forskningen är bara i sin linda på området för diskriminerande utfall av digital konsumentprofilering, men knyter an till ett växande fält kring ansvar i relation till algoritmer och automatiserade verktyg.<sup>50</sup>

---

<sup>48</sup> 2017:1 Kunskapsöversikter inom det konsumentpolitiska området, dnr 2017/34, Konsumentverket, 2017, s. 30f.

<sup>49</sup> 2017:1 Kunskapsöversikter inom det konsumentpolitiska området, dnr 2017/34, Konsumentverket, 2017, s. 30–32 samt 39 och 41.

<sup>50</sup> 2017:1 Kunskapsöversikter inom det konsumentpolitiska området, dnr 2017/34, Konsumentverket, 2017, s. 41f.

## Risker för diskriminering

Användningen och utvecklingen av ADB, maskinlärande och AI motiveras ofta av bland annat möjligheten att nå bättre resultat än vad människor kunnat göra och att nå dem snabbare och utifrån mer data än en människa kan bearbeta. Ytterligare motiv för användning är förhoppningen om att minska, eller helt utesluta, subjektiva och fördomsfulla värderingar som kan förekomma vid beslutsfattande av en människa. Men användningen av algoritmer innebär dock inte att de diskriminerande effekter som kan förekomma vid mänskligt beslutsfattande försvinner eller minskar. Det finns ett tydligt samspel mellan en algoritms effekter och den data den förses och tränats med. Om så kallad biased (snedvridna, skeva, fördomsfulla; nedan används det engelska begreppet "biased") data används kan även en logiskt välfungerande algoritm ge snedvridna resultat. Om en algoritm tränas med biased data kan resultatet medföra diskriminerande effekter. Ju mer avancerad en algoritm är, desto större självständighet får algoritmen att välja ut och analysera data. Det bidrar till minskad transparens i fråga om att förstå och få insyn i dels de parametrar som algoritmen tagit i beaktande och analyserat, dels det resultat algoritmen producerat.<sup>51</sup> Det kan inte heller uteslutas att ADB och AI avsiktligt programmeras på ett sätt som får diskriminerande effekter.

## Generella risker och felkällor

I en studie publicerad av Europarådet redogör professor Fredrik Zuiderveen Borgesius för hur AI kan leda till risker för diskriminering, genom följande exempel. AI-system är ofta så kallade "svarta lådor", det vill säga att man inte har någon insyn i hur algoritmen i dataprogrammen fungerar. Det är ofta oklart varför systemet har fattat ett visst beslut om en person. Det är därför också svårt för människor att bedöma om de har blivit diskriminerade. AI-drivet beslutsfattande kan leda till diskriminering på flera olika sätt.<sup>52</sup> I studien refereras till forskarna Barocas och Selbst<sup>53</sup> som pekar ut fem sätt. Dessa sätt, genom vilka beslutsfattande med hjälp av AI

---

<sup>51</sup> Markus Naarttijärvi, Rättsstatlighet och algoritmiska svarta lådor, Jubileumsskrift till Juridiska Institutionen 40 år, Umeå Universitet, Juridiska institutionen, Umeå 2017, s. 250 – 253; Kevin Petrasic, Benjamin Saul, James Greig, Matthew Bornfreund och Katherine Lamberth, Algorithms and bias: What lenders need to know, White & Case LLP, 2017, s. 2.

<sup>52</sup> Discrimination, artificial intelligence and algorithmic decision-making, prof. Frederik Zuiderveen Borgesius, studie publicerad av Europarådet 2018, s. 10.

<sup>53</sup> Barocas S. och Selbst A.D., "Big Data's disparate impact (2016) Calif Law Review, vol. 104, s. 671–732.

omedvetet kan leda till diskriminering, redogörs för nedan. Vidare redogörs även kort för hur medveten diskriminering kan ske.

### Omedveten diskriminering

#### Målvariabler och gruppetiketter

AI kan innebära att datorer hittar korrelationer i en mängd data. Genom att definiera ”målvariabler” (eng: target variables) och ”gruppetiketter” (eng: class labels) kan man skapa träningsdata. En målvariabel definierar vilken data som man letar efter. Ett exempel är utvecklingen av spam-filter. Man matar då datorn med e-postmeddelanden som människor har klassat som ”spam”, respektive ”inte spam”. Datorn lär sig då hitta typiska kännetecken hos de e-postmeddelanden som klassats som ”spam”, till exempel att de ofta innehåller vissa fraser eller skickas från vissa IP-adresser. Genom att exponera maskinlärande algoritmer för exempel på intressanta fall, såsom e-postmeddelanden som tidigare definierats som spam, lär sig algoritmen vilka kännetecken som hör samman med vad man är intresserad av, den så kallade målvariabeln.<sup>54</sup>

Gruppetiketter delar in alla tänkbara värden i en målvariabel i olika kategorier med gemensamma nämnare. Ett exempel på målvariabel som ges är att ett företag i syfte att sortera jobbansökningar definierar en ”bra” anställd som någon som aldrig är sen till arbetet. Om det antas att människor med mindre god ekonomi sällan bor centralt och har längre resväg till arbetet och därför oftare kommer sent till arbetet än andra, på grund av trafikstockningar eller problem med kollektivtrafiken, skulle ett system där variabeln ”sällan sen till jobbet” ge personer med utländsk bakgrund en nackdel, eftersom de i genomsnitt har sämre ekonomi och längre resväg till arbetet. Det finns risk för att diskriminering uppstår i ett AI-system, men riskerna är beroende av hur målvariabler och gruppetiketter definieras.<sup>55</sup>

#### Träningsdata: Att sätta etiketter på exempel

Beslutsfattande med hjälp av AI kan få diskriminerande effekter om systemet tränas på så kallade biased data, eller om systemet lär sig på biased exempel. I dessa fall kommer AI-systemet att reproducera samma *bias*. De träningsdata som används kan vara behäftade med *bias* om de är resultatet av diskriminerande beslut som tidigare tagits av människor. Ett exempel på

---

<sup>54</sup> Discrimination, artificial intelligence and algorithmic decision-making, prof. Frederik Zuiderveen Borgesius, studie publicerad av Europarådet 2018, s. 11.

<sup>55</sup> Ibid s. 11.

detta är antagningen till en läkarutbildning i Storbritannien på 1980-talet, där man lät ett dataprogram behandla alla ansökningar. Som träningsdata användes ansökningarna för dem som antagits tidigare år. De gemensamma kännetecknen i ansökningarna hos dem som antagits tidigare år, användes av dataprogrammet för att dra slutsatser om vilka sökanden som borde antas till läkarutbildningen. Det visade sig sedan att dataprogrammet missgynnade kvinnor och personer med invandrarbakgrund, eftersom de var underrepresenterade i materialet som användes som träningsdata. Slutsatsen blir att om det träningsdata som används är biased, så riskerar AI-systemet att reproducera denna *bias*.<sup>56</sup>

#### Träningsdata: Insamling av data

Om insamlingen av data sker på ett selektivt (eng: biased) sätt så kan det också leda till diskriminerande effekt. Om till exempel Polisen fokuserar sina resurser på vissa etniska grupper eller områden där personer med en viss etnisk tillhörighet är i majoritet, så är det sannolikt att dessa etniska grupper eller områden kommer att vara överrepresenterade i Polisens register. Om ett AI-system tränas på sådan biased data, så kommer AI-systemet att fortsätta att reproducera samma *bias* genom att lära sig att personer med viss etnisk tillhörighet eller bakgrund har större sannolikhet att begå brott. Effekterna av ett sådant biased urval kan till och med förstärkas i så kallade prediktiva (förutsäggande) AI-processer. Om Polisen fokuserar på vissa etniska grupper och vissa bostadsområden, så kommer dessa etniska grupper och bostadsområden sannolikt bli överrepresenterade i statistiken. Det kan leda till att Polisen ytterligare förstärker sin närvaro i sådana områden, eftersom det statistiskt sett sker fler brott i dessa områden. Detta leder till en så kallad "feedback loop" (självförstärkande process).<sup>57</sup>

#### Urvalskategorier eller kriterier

De urvalskriterier eller kategorier som man väljer att använda i ett AI-system har betydelse. Om man vill använda AI för att automatiskt förutsäga någonting så kan det behöva göras en förenkling av verkligheten (till exempel att begåvade personer studerar vid elituniversitet) för att kunna fånga in relevant data. Man behöver alltså välja vilka urvalskategorier eller kriterier man observerar och sedan infoga det i sin analys. I en rekryteringsprocess är det ibland inte möjligt, eller åtminstone för kostsamt, att göra en fullständig bedömning av varje sökande. Arbetsgivaren kan då

---

<sup>56</sup> Ibid s. 12.

<sup>57</sup> Ibid s. 12.

välja att fokusera på vissa egenskaper som en sökande bör ha. Ett exempel på hur urvalskriterier kan få en diskriminerande effekt är om en arbetsgivare vid urval av kandidater använder examen från ett berömt universitet som urvalskriterium. Det kan då få diskriminerande effekter om exempelvis vissa etniska grupper är underrepresenterade vid dessa universitet.<sup>58</sup>

### Proxies

Ibland kan kriterier som är relevanta för att fatta rationella och välgrundade beslut också utgöra proxies (närliggande omständigheter eller data; nedan används det engelska begreppet "proxies") som har koppling till eller samband med skyddade diskrimineringsgrunder. Vissa data som används som träningsdata kan alltså ha samband med skyddade diskrimineringsgrunder. Ett exempel är om en bank använder ett AI-system som tränats på data avseende vilka låntagare som har haft svårt att betala sina lån, för att kunna förutsäga vilka lånesökande som kommer att få svårt att betala sina lån. Då kan systemet exempelvis lära sig att personer från ett visst postnummerområde ofta haft problem att betala av sina lån och använda det sambandet för att förutsäga problem med återbetalning av lån. Om man antar att postnummerområde har samband med etniskt ursprung så skulle människor med visst etniskt ursprung oavsett de egna ekonomiska förutsättningarna missgynnas om banken baserat på förutsägelsen nekar låneansökningar till människor bosatta i ett visst postnummerområde.<sup>59</sup>

### Medveten diskriminering

En verksamhet kan också medvetet diskriminera, till exempel genom att vid programmering av algoritmer använda proxies som har samband med skyddade diskrimineringsgrunder. Om man använder proxies så kan diskrimineringen vara svårare att upptäcka, än om man använder data som har direkt koppling till en skyddad diskrimineringsgrund.<sup>60</sup>

### Avslutande reflektioner

ADB och AI kan komma att medföra en ökad risk för diskriminering beroende på hur tekniken och verktygen eller systemen programmeras, tränas, utvecklas och används samt vilken möjlighet till insyn i programvara och algoritmer som kommer att finnas. Därtill kommer eventuell framtida

---

<sup>58</sup> Ibid s. 12.

<sup>59</sup> Ibid s. 13.

<sup>60</sup> Ibid s. 13

reglering på området i form av till exempel lagar, riktlinjer, rekommendationer, branschetiska regler eller vad som kan komma att anses vara lämpligt.

## Exempel på användning av AI och ADB

### Rekrytering

Det som kan vara av intresse ur ett diskrimineringsperspektiv gäller framförallt metoder för automatiserad databehandling som används vid rekrytering (avseende arbetssökande) och för olika typer av urval bland en arbetsgivares anställda för kompetensutveckling, befordran, etcetera.

I fråga om dagens teknik för rekrytering så finns ett antal exempel på situationer där ADB eller AI kan tänkas användas. Exempelvis kan tekniken användas för bakgrundskontroller via sökningar eller sökrobotar på internet, ansökningar via webbförmulär som sedan behandlas av robotar, granskning av CV och personliga brev i en första sällning av kandidater innan en människa tar vid i nästa skede i rekryteringsprocessen. Därtill kan det handla om digitala eller webbaserade tester av exempelvis intelligenskvot, numerisk analys, läsförståelse och personlighet. Det kan även handla om: intervjuer med robotar, teknik för analys av ansiktsuttryck och ansiktsrörelser på videofilmade intervjuer, så kallade chatbotar på rekryteringsföretags webbplatser eller automatiska sökningar i kandidatdatabaser med efterföljande automatgenererade mail till potentiella kandidater.

Ett stort internetpostorderföretag gjorde ett försök att utveckla ett rekryteringsverktyg i form av automatiserad hantering av ansökningar, se vidare nedan. Företaget kunde konstatera att programvaran, baserat på historiska data, gav ansökningar från kvinnor lägre poäng än ansökningar från män. Det handlade om att systemet hade tränats på data som speglade en verklighet där det råder obalans mellan könen. Försöket illustrerar att automatiserad behandling av ansökningar kan komma att sorteras efter vissa parametrar. Dessa kan vara direkt eller indirekt diskriminerande.

Olika typer av webbaserade begåvnings-, anlags- och personlighetstester kan innebära diskriminering av personer med funktionsnedsättning. Detta kan illustreras med ett exempel från USA som beskrivs i den populärvetenskapliga boken "Weapons of Math Destruction"<sup>61</sup>. Här gjorde en

---

<sup>61</sup> O'Neil, Cathy, "Weapons of math destruction", (2016) Broadway Books, s. 106.



person, som tidigare diagnostiserats med bipolär sjukdom, en webbaserad personlighetstest inom ramen för ett ansökningsförfarande till ett enkelt, lågkvalificerat arbete. Personen blev bortsorterad, ”rödflaggad”, av systemet och gick inte vidare i rekryteringsförfarandet. Han sökte flera liknande arbeten och det visade sig att alla företag där han sökte arbete använde samma test och han fick inga erbjudanden.

### **Konkreta exempel**

#### **Rekryteringsrobotar**

##### **Svenskt startup-företag**

Ett svenskt startup-företag har utvecklat en digital rekryteringsrobot som med hjälp av AI enligt företaget anses minska risken för felrekryteringar. Grunden för den automatiserade matchningen utgörs av kompetens- och personlighetsdata från hundratusentals arbetssökande och nästan 10 000 rekryteringar. Roboten gör datadrivna utvärderingar av hur arbetssökandens mjuka (personliga egenskaper) och hårda kompetenser (mätbara förmågor) matchar det sökta jobbet, vilket leder till tids- och kostnadsbesparingar för arbetsgivarna. Enligt uppgift från företaget använder flera stora företag roboten i sin rekrytering. Något som företaget nämner som en fördel med att använda en digital rekryteringsrobot är att ge alla kandidater samma förutsättningar. Vidare uppges cheferna kunna ta rekryteringsbeslut utifrån rätt kriterier, där kandidatens värdegrund och personliga egenskaper uppges vara avgörande. Robotens matchning och anonymisering uppges ge en rättvis och träffsäker rekryteringsprocess som inte diskriminerar någon kandidat. Vidare uppges kvaliteten i rekryteringen öka, samtidigt som rekryteringstiden minskar, från veckor till dagar.<sup>62</sup>

##### **AI-roboten M**

En annan robot har utvecklats på La Trobe University i Australien av professor Rajiv Khosla. Roboten användes först inom barn- och äldreomsorgen och har därefter omprogrammerats för att användas i rekryteringssammanhang. Roboten var programmerad att kunna utföra intervjuer med två sorters kandidater: säljare och kundtjänstmedarbetare. Roboten genomför en omkring halvtimme lång intervju med standardiserade frågor. Allt spelas in med robotens kameraögon, som även hjälper henne att läsa av ansiktsuttryck och gestikuleringar. Efter intervjun kan de mänskliga rekryterarna på företaget gå tillbaka och titta på den

---

<sup>62</sup> [http://www.mynewsdesk.com/se/pressreleases/digital-robot-ersaetter\\_rekryterare-artif...](http://www.mynewsdesk.com/se/pressreleases/digital-robot-ersaetter_rekryterare-artif...), hämtad 2018-08-24.

inspelade intervjun, men framförallt läsa den rapport som roboten skriver ut, där kandidatens svar och beteenden under intervjun analyserats.

Tanken med att använda en robot vid rekryteringar är att roboten kan genomföra och analysera jobbintervjuer utan att påverkas av fördomar, magkänsla och andra subjektiva faktorer, då roboten ger alla kandidater samma möjlighet att svara på frågorna. Dessutom är roboten outtröttlig och kan jobba dygnet runt vid behov. Tanken är att roboten, när den är färdigtränad, ska kunna göra det första urvalet vid till exempel större rekryteringar. Roboten ska inte fatta några beslut, utan bara ge rekryterarna ett bättre underlag.<sup>63</sup>

Roboten har testats av ett svenskt rekryteringsföretag, se också nedan under nästa rubrik.

#### **Roboten T**

Ett annat svenskt startup-företag har inlett ett samarbete med ett bemannings- och rekryteringsföretag, och de har ambitionen att ta fram vad de anser är världens första fördomsfria sociala rekryteringsrobot. Startup-företaget arbetar med konverserande AI och social robotik. Företaget, som är Stockholmsbaserat, har sitt ursprung i ett projekt från Kungliga Tekniska Högskolan och har utvecklat en avancerad plattform för sociala robotar. Bemannings- och rekryteringsföretaget specialiserar sig på fördomsfri och kompetensbaserad rekrytering.

Den färdigutvecklade roboten är en produkt av de två parternas samarbete och syftet uppges vara att den ska kunna genomföra kompetensbaserade intervjuer och assistera med objektiva urval så att inga jobbsökare väljs bort på grund av fördomar, magkänsla eller koppling till diskriminerande grunder.

Samarbetet bygger på startup-företagets kunskap om sociala robotar och bemannings- och rekryteringsföretagets erfarenhet av att utveckla och arbeta med vad de själva anser är en fördomsfri, objektiv och vetenskaplig rekryteringsprocess. Nyligen lanserades den senaste versionen av roboten. Det är på den plattformen som företagen nu gemensamt utvecklar roboten. Företagens mål med samarbetet är att under år 2019 integrera roboten i bemannings- och rekryteringsföretagets rekryteringsprocess och därefter

---

<sup>63</sup> Shortcut: "Möt Matilda – världens första rekryteringsrobot", <http://shortcut.se/artiklar/mot-matilda-varldens-forsta-rekryteringsrobot/>, hämtad 2018-08-24.

dra igång försäljningen av roboten i Sverige och på sikt även internationellt.<sup>64</sup>

#### **Virtuell assistent**

Ett svenskt rekryteringsföretag använder en virtuell rekryteringsassistent baserad på artificiell intelligens, som enligt företagets uppgifter inte låter sig påverkas av fördomar eller åsikter. Assistentens uppgift är att hantera sökningsmomentet i företagets mer standardiserade rekryteringsprocesser. Det innebär att assistenten ska söka igenom olika databaser, vilket uppges vara ett väldigt tidskrävande och tidvis mycket monotont arbete.<sup>65</sup>

#### **Roboten V**

Ett stort internationellt detaljhandelsföretag använde sig av rekryteringsroboten V vid rekrytering av personal till butiker i Ryssland. Företaget fick årligen in tusentals CV från arbetssökande. Eftersom det initiala urvalet av arbetssökande ofta var ett väldigt omfattande och tidskrävande arbete startade företaget ett pilotprojekt. Roboten användes för att kontakta arbetssökande. Enligt tidningen Veckans affärer, som citerar Washington Post, använder redan 200 bolag roboten V. Roboten hinner avverka ca 1 500 intervjuer på en arbetsdag.<sup>66</sup> Efter pilotstudien i Ryssland valde dock företaget att inte gå vidare med projektet.<sup>67</sup>

#### **Försök att utveckla ett rekryteringsverktyg**

##### **Företag 1**

Maskinlärningsspecialister på ett stort amerikanskt internetpostorderföretag påbörjade år 2014 ett försök att utveckla ett rekryteringsverktyg för att granska jobbkandidaters meritförteckningar. I experimentet använde verktyget AI för att ge kandidater poäng på en skala från ett till fem, för att kunna rangordna dem. Efter en tid upptäckte företaget att dess nya system inte bedömde jobbkandidater som mjukvaruutvecklare och andra tekniska tjänster på ett könsneutralt sätt. Orsaken var att företagets datamodeller var tränade att granska ansökningar genom att observera mönster i meritförteckningar som hade skickats in till bolaget under en tioårsperiod. De flesta ansökningar kom från män, vilket

---

<sup>64</sup> <https://evertiq.se/news/36403>, hämtad 2018-11-22.

<sup>65</sup> Tidningen Karriär, nr 1 2019, s. 15, som finns tillgänglig i digital form på [http://pdf-flip.se/Karriar/Karriar\\_1\\_2019/](http://pdf-flip.se/Karriar/Karriar_1_2019/).

<sup>66</sup> <https://www.va.se/nyheter/2018/04/26/Mot-ikeas-nya-rekryteringsrobot-som-klarar-1-500-kandidater-pa-en-dag/>, hämtad 2018-08-24.

<sup>67</sup> <https://www.va.se/nyheter/2018/04/26/Mot-ikeas-nya-rekryteringsrobot-som-klarar-1-500-kandidater-pa-en-dag/>, hämtad 2018-08-24.

avspeglade mansdominansen i teknikbranschen. Företagets datamodeller hade alltså lärt sig att manliga kandidater var att föredra. Systemet bestraffade meritförteckningar som innehöll ord som "women's" som i "women's chess club captain". Även personer med examen från högskolor bara endast för kvinnor fick, enligt uppgift, lägre poäng. Företaget ändrade programmen för att göra dem neutrala i förhållande till dessa villkor, men det fanns ändå inte några garantier för att datorerna inte skulle komma på nya sätt att sortera jobbkandidater som skulle visa sig vara diskriminerande (enligt det aktuella landets lagstiftning), enligt uppgift från personer som arbetat med projektet. Det slutade med att företaget avslutade projektet i början av år 2017.<sup>68</sup>

## Företag 2

Ett annat exempel på utveckling av rekryteringsverktyg är ett internationellt tillverkningsföretag i kontorsteknikbranschen som utvecklat ett verktyg för sin HR-avdelning. Vid rekrytering till sin kundtjänst söker de personer som kan förväntas stanna kvar länge hos företaget och företaget har genom analys av stordata tagit fram fyra parametrar som har ett positivt samband med att stanna länge i företaget. Dessa fyra parametrar var (1) villighet att arbeta övertid en till tre timmar per vecka, (2) högskoleutbildning, särskilt teknisk högskoleutbildning, (3) tidigare erfarenhet av att arbeta i kundtjänst och få använda empati, det vill säga inte bara ta order, och (4) att bo nära arbetet eller ha tillgång till pålitlig transport. Företaget låter jobbkandidater genomgå tester för att förutsäga hur de skulle prestera i kundtjänsten, och en del av testerna handlar om attityder i förhållande till arbetet. En kandidat som får ett lågt värde på dessa parametrar riskerar alltså att sorteras bort tidigt i rekryteringsprocessen.<sup>69</sup>

## Tekniker för ansiktsigenkänning

Det har visat sig i olika tester att tekniken för ansiktsigenkänning fungerar sämre när den används för att identifiera mörkhyade personer, vilket troligen beror på bristande mångfald i träningsdata.<sup>70</sup> När det gäller muntlig och skriftlig kommunikation med robotar kan det finnas en risk att sådana system har svårare att förstå till exempel en person som har en brytning eller någon som har läs- och skrivsvårigheter. Forskarna Joy Buolamwini (MIT) och Timnit Gebru (Microsoft) har också visat att system som används

---

<sup>68</sup> <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>, hämtad 2018-10-11.

<sup>69</sup> Implicit Bias When Using Data Profiling within Recruitments and Human-Resource Management, Anders Persson, IT-department, Uppsala Universitet.

<sup>70</sup> <https://www.ajlunited.org/the-coded-gaze>, Joy Buolamwini.

kommersiellt för könsklassificering har betydligt större felmarginal vid klassificering av mörkhyade kvinnor (34,7%), än vid klassificering av ljushyade män (0,8%).<sup>71</sup>

## Krediter

Många aktörer som vänder sig till konsumenter, exempelvis detaljhandelsföretag, banker och kreditinstitut, marknadsföringsbolag och försäkringsbolag, intresserar sig för den nya tekniken och den stora mängd information om individer eller konsumenter och deras beteenden som i dag kan inhämtas i stordata.<sup>72</sup> Nedan beskrivs något om användningen den nya tekniken.

### Riskbedömning

Inom all kreditgivning är förmågan att bedöma risk, det vill säga låntagarens kreditvärdighet, central. Det är en nyckelkompetens inom finansindustrin och över tid har olika system, ofta samägda av bankerna, skapats för att standardisera förfarandet. Traditionell kreditvärdering samlar relevant finansiell information som fokuserar på data som på ett direkt sätt relaterar till en låntagares möjlighet att återbetala ett lån, genom att titta på exempelvis tillgångar och skulder från olika håll.<sup>73</sup>

Med teknikgenombrott som internet och stordata har möjligheterna att basera kreditbedömningar på andra uppgifter än tidigare ökat markant. Andra data än de som finns hos kreditupplysningsföretagen kan säga mycket, i vissa fall kanske till och med mer, om en individs kreditvärdighet. Aktörer som gör riskbedömningar har ett intresse av att få tillgång till information som kan erhållas genom den kartläggning som görs om enskilda individer, bland annat på internet. Sådan information kombinerad med verktyg för dataanalys gör det möjligt för en kreditgivare att minska sina risker. När nya analysverktyg används för att till exempel kombinera intern bankinformation om en kund med information från olika externa

---

<sup>71</sup> "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" in the Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency, tillgänglig på <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

<sup>72</sup> 2017:1 Kunskapsöversikter inom det konsumentpolitiska området, dnr 2017/34, Konsumentverket, 2017.

<sup>73</sup> Benjamin Ståhl, FinTech – Den digitala transformeringen av finansmarknaden, Blue Institute, Stockholm, 2017, s. 34f; Kevin Petrasic, Benjamin Saul, James Greig, Matthew Bornfreund och Katherine Lamberth, Algorithms and bias: What lenders need to know, White & Case LLP, 2017, s. 3.

aktörer så kan man få detaljerad information om den enskilda kunden, vilket leder till flera risker för den personliga integriteten.<sup>74</sup>

Argument har framförts om att traditionell kreditvärdering utesluter mycket relevant data om låntagaren. Det gör det svårt för många kundsegment, till exempel unga människor, migranter och för små företag, att få en rimlig kreditbedömning eftersom de ofta inte har någon kredithistoria. Det har därför utvecklats modeller och algoritmer som utgår från andra källor än de traditionella för kreditvärdering, för att identifiera kreditvärdiga människor och företag där bankerna inte gör det. Det öppnar för nya kundgrupper och gör produkterna mer konkurrenskraftiga. Alternativa datakällor kan vara mobilanvändning, hur ofta man skickar pengar till andra, punktlighet med att betala skolavgifter, vad man köper på nätet och i handeln eller om man pendlar regelbundet. En persons sociala nätverk och beteende kan också vara relevant, och enkla psykometriska<sup>75</sup> frågeformulär används i vissa fall för att bedöma betalningsbenägenhet. Vilken data som samlas in beror förstås på vad som är lagligt, lämpligt och kulturellt accepterat. Användning av alternativa datakällor är enligt dess förespråkare en förutsättning för att fler ska få tillgång till finansiella tjänster, men det är också en balansakt där personlig integritet och diskriminering är risker som måste hanteras.<sup>76</sup>

### **Handel med personuppgifter**

Det finns en tydlig trend som visar på ökad digitalisering inom finansmarknaden, inte minst genom utvecklingen av och de ökade investeringarna i så kallade FinTech-bolag (Fintech är en förkortning av engelskans Financial Technology). Begreppet FinTech används ofta som ett begrepp för all teknologi som används inom bland annat verksamheterna bank, kapitalmarknad och personlig finansiering.<sup>77</sup> Utvecklingen har drivits på i takt med den ökade digitaliseringen i samhället, vilket bland annat påverkat konkurrensen om betalningsströmmar och associerade intäkter. I

---

<sup>74</sup> Hur står det till med den personliga integriteten? – en kartläggning av Integritetskommittén (SOU 2016:41), Stockholm, 2016, s. 419ff.

<sup>75</sup> Begreppet psykometri används om teori om och metoder för mätning av psykiska egenskaper, attityder, intressen, kunskaper och färdigheter, <https://www.psykologiguiden.se/psykologilexikon/?Lookup=psykometrik%2C+psykometri>, hämtad 2019-04-01.

<sup>76</sup> Benjamin Ståhl, FinTech – Den digitala transformeringen av finansmarknaden, Blue Institute, Stockholm, 2017, s. 34f; Kevin Petrasic, Benjamin Saul, James Greig, Matthew Bornfreund och Katherine Lamberth, Algorithms and bias: What lenders need to know, White & Case LLP, 2017, s. 3.

<sup>77</sup> Consumer protection on the financial market, Finansinspektionen, maj 2016, s. 29.

Sverige finns flera exempel på FinTech-bolag som verkar inom bland annat banksektorn.<sup>78</sup>

Teknikutvecklingen har medfört att aktörer med det specifika intresset att handla med personuppgifter vuxit i betydelse och omfattning. Dessa aktörer kallas ofta informations- eller datamäklare. Det är marknadsaktörer, ofta med en mindre publik framtoning, som är inriktade på att samla in information om konsumenter från många källor och sedan sälja profiler till andra näringsidkare. Konsumentverket har i en rapport från 2017<sup>79</sup> påtalat att det i mycket saknas forskning om de datamäklare aktörerna och vad deras affärsverksamhet betyder för konsumenter samt hur verksamheten utvecklas och påverkar andra branscher när de utnyttjar datamäklarnas konsumentprofilering.<sup>80</sup> Det finns i dag några företag i Sverige som säljer tjänster i form av modeller för kreditvärdering (eng: credit scoring) av privatpersoner. Genom att analysera tillgänglig information om de kredittagare som fullgör sina åtaganden genom att betala respektive de som inte betalar, hittar man samband och mönster som kan användas för att förutsäga framtida betalningsbeteenden. I vart fall ett av företagen tillhandahåller också en tjänst för automatiserade beslutsprocesser vid kreditgivning.

### **Riskbedömning i USA – exempel**

Ett tidigt exempel på användning av kreditbedömning genom prediktiv analys är det postorderföretag i USA som på 1950-talet använde prediktiv analys för att genom kreditbedömning kunna bestämma vilka som kunde få kredit av företaget. Metoden har sedan blivit vanlig i den finansiella sektorn.<sup>81</sup>

I USA ägnade sig kreditgivare tidigare åt väldigt grova generaliseringar i sin riskbedömning (så kallad redlining): till boende i vissa bostadsområden, som markerades med röda linjer på en karta, skulle inga krediter ges. Med dagens stordataverktyg har det öppnats möjligheter att uppnå samma effekt som med hjälp av redlining, som i dag är förbjuden. Begreppet weblining

---

<sup>78</sup> Benjamin Ståhl, FinTech – Den digitala transformeringen av finansmarknaden, Blue Institute, Stockholm, 2017, s. 12.

<sup>79</sup> 2017:4 Personuppgifter som betalningsmedel, Konsumentverket, 2017.

<sup>80</sup> 2017:4 Personuppgifter som betalningsmedel, Konsumentverket, 2017, s. 22 och 41.

<sup>81</sup> Stanley Greenstein, Our Humanity Exposed – Predictive Modelling in a Legal Context, Stockholm 2017, s. 100–102.

har börjat användas som term när en individs digitala fotspår används för att dra slutsatser om individens kreditvärdighet.<sup>82</sup>

### Riskbedömning under utveckling

Som framgår ovan har det framförts argument om att traditionell kreditvärdering utesluter mycket relevant data om låntagaren, och att det gör det svårt för vissa kundsegment att få en rimlig kreditbedömning. I avsaknad av kreditvärdering för vissa konsumenter förekommer att FinTech-företag och finansiella institutioner vid långivning utnyttjar stordata för att bygga poängsystem för kreditvärdering. Förespråkare menar att detta ger stora möjligheter för nya företag och kreditgivande banker och för individer som inte har någon ekonomisk uppbackning eller tidigare kredithistorik.<sup>83</sup>

Det finns några exempel på utländska företag som använder alternativa datakällor för kreditvärdering. Företagen uppger att de bygger kreditvärdering genom att värdera kreditsökandens aktivitet på sociala medier, surfhistorik på internet, geositionering och annan data från smarta telefoner.<sup>84</sup> Ett företag uppger att deras algoritm analyserar 12 000 variabler, som efter tre minuter resulterar i ett kreditpoäng eller kreditvärdering (eng: credit score). Ett annat företag uppger sig använda enbart Facebook API för att poängsätta användare, och vem du är vän med, vad du är intresserad av och andra vanor på internet kan bidra till bättre kreditpoäng.<sup>85</sup>

Det finns vissa invändningar mot att använda sociala data som grund för kreditvärdighet. Exempelvis kan gamla data och feltolkningar komma att utgöra hinder för individer som i andra sammanhang skulle visa låg risk för underlåtenhet att betala.<sup>86</sup> Det finns även integritetsfrågor som aktualiseras. Det är därför viktigt att beakta och förstå samt ta hänsyn till bland annat

---

<sup>82</sup> Hur står det till med den personliga integriteten? – en kartläggning av Integritetskommittén (SOU 2016:41), Stockholm, 2016, s. 421.

<sup>83</sup> <https://www.fernsoftware.com/blog/fintech-offers-credit-scores-without-fico/>, hämtad 2019-03-22.

<sup>84</sup> <https://www.fernsoftware.com/blog/fintech-offers-credit-scores-without-fico/>, hämtad 2019-03-22.

<sup>85</sup> <https://www.fernsoftware.com/blog/fintech-offers-credit-scores-without-fico/>, hämtad 2019-03-22.

<sup>86</sup> <https://www.fernsoftware.com/blog/fintech-offers-credit-scores-without-fico/>, hämtad 2019-03-22, som hänvisar till en artikel i ämnet som finns tillgänglig på <https://gomedici.com/can-web-search-history-become-a-holy-grail-of-credit-scoring/>.



konsumenternas perspektiv på användning av deras data och dataskyddsreglering.<sup>87</sup>

### **Kreditgivares skyldighet att efterfråga information enligt lagen (2004:297) om bank- och finansieringsrörelse**

I sammanhanget ska nämnas att kreditgivare och banker enligt svensk lag måste ta hänsyn till de regelverk som gäller på den marknad där de verkar. Detta kan alltså påverka vilken information en kreditgivare måste fråga en individ om, innan ett beslut om kreditgivning fattas. Följande är några exempel.

Innan en bank beviljar en kredit ska banken enligt 8 kap. 1–4 §§ lagen (2004:297) om bank- och finansieringsrörelse pröva risken för att de förpliktelser som följer av kreditavtalet inte kan fullgöras. Banken får bevilja en kredit bara om förpliktelserna på goda grunder kan förväntas bli fullgjorda. En banks kreditprovning ska vara organiserad så att den som fattar beslut i ärendet har tillräckligt beslutsunderlag för att bedöma risken med att bevilja krediten. Vid krediter till konsumenter är, enligt 12–13 §§ konsumentkreditlagen (2010:1846), kreditgivaren skyldig att pröva om konsumenten har ekonomiska förutsättningar att fullgöra vad han eller hon åtar sig enligt kreditavtalet. Om krediten inte beviljas så ska konsumenten underrättas om orsaken till detta, och om avslaget beror på uppgifter som härrör från en extern databas så ska konsumenten underrättas om uppgifterna och vilken databas som använts. Om ett beslut att inte bevilja en bostadskredit helt grundar sig på en automatisk behandling av uppgifter så ska konsumenten upplysas om detta. Av lagen (2003:862) om finansiell rådgivning till konsumenter följer att banken även vid finansiell rådgivning är skyldig att inhämta viss information om den enskilde.<sup>88</sup>

Sammanfattningsvis kan det alltså finnas skyldigheter enligt lag för vissa marknadsaktörer att efterfråga information om en kredit- eller låntagare.

Finansinspektionen ansvarar för bland annat tillsynen och tillståndsprovningen som rör finansiella marknader och finansiella företag samt för samordningsorganet för tillsyn enligt förordningen (2009:92) om åtgärder mot penningtvätt och finansiering av terrorism. Även länsstyrelserna i Skånes, Stockholms och Västra Götalands län utövar viss

---

<sup>87</sup> <https://www.fernsoftware.com/blog/simple-messages-help-consumers-understand-big-data/>, hämtad 2019-03-22.

<sup>88</sup> Hur står det till med den personliga integriteten? – en kartläggning av Integritetskommittén (SOU 2016:41), Stockholm, 2016, s. 417–418.

tillsyn över att bland annat banker och kreditmarknadsbolag följer lagen om åtgärder mot penningtvätt och finansiering av terrorism.<sup>89</sup>

### **DO:s tillsynsbeslut**

Inom ramen för DO:s verksamhet har DO gjort två tillsyner där automatiserad databehandling varit involverad.

#### **SBAB Bank AB (publ)**

DO har genomfört en tillsyn av den statliga banken SBAB Bank AB (publ) och en situation där en bankkund hade nekats ett utökad bolån med anledning av hans ålder. Hans aktuella inkomst hade av bankens datasystem skrivits ned med 30 procent eftersom man beräknade att hans inkomst inom en viss framtid skulle minska vid pensionering. I beslutet (GRA 2017/80) kom DO fram till att den schabloniserade kalkylmetod som tillämpades av banken vid bedömning 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.

#### **Klarna Bank AB (publ)**

DO har genomfört en tillsyn av Klarna Bank AB (publ) (Klarna) och en situation där en bankkund inte tilläts vara kund i banken grund av sin ålder. Enligt Klarna kunde bankens system inte särskilja om personer var födda på 1900-talet eller 2000-talet, och systemet kunde därmed inte avgöra om åldern var 1 eller 101 år. Av säkerhetsskäl hade bankens system därför begränsat de tillgängliga betalsätten till banköverföring (kredit- eller debetkort i dessa fall) eftersom banken inte vill riskera att acceptera kreditköp för personer under 18 år. Kunden hade därför på grund av de tekniska begränsningarna inte kunnat handla via faktura. DO konstaterade i sitt beslut (TIL 2018/22) att bankens förfaringssätt att neka en kund att bland annat betala med faktura med hänvisning till att kunden uppnått åldern 100 år utgjorde en överträdelse av diskrimineringsförbudet som hade samband med ålder.

### **Finska Diskriminerings- och jämställdhetsnämndens beslut**

Diskrimineringsombudsmannen i Finland överlämnade ett ärende till Diskriminerings- och jämställdhetsnämnden (216/2017) gällande en man som finska diskrimineringsombudsmannen ansåg hade utsatts för diskriminering av ett kreditinstitut (Svea ekonomi) som nekat honom en

---

<sup>89</sup> Hur står det till med den personliga integriteten? – en kartläggning av Integritetskommittén (SOU 2016:41), Stockholm, 2016, s. 417.

konsumentkredit för ett köp på nätet. Mannen hade inte erhållit någon individuell bedömning av sin inkomst eller finansiella situation. Istället hade kreditinstitutet gjort en schablonmässig bedömning där den kreditsökande fick en särskild poäng (eng: credit scoring) där flera faktorer hade samband med sökandens personliga omständigheter, så som kön, hemspråk, bostadsort och ålder. Alla dessa utgör diskrimineringsgrunder enligt den finska grundlagen.

Nämnden ansåg att den sökanden hade blivit utsatt för direkt multipel diskriminering. Poängsystemets bedömning grundade sig på statistiska data och betalningsinformation som byggde på andra personers uppgifter och finansiella beteende. Sedan gjordes ett antagande om sökandens finansiella situation och i fallet bedömdes han ha en svagare finansiell situation än om den hade bedömts med andra variabler. Samtidigt ignorerades sökandens egen kreditvärdighet trots att dessa uppgifter var till hans fördel. Metoden baserades inte på någon individuell bedömning av kreditvärdighet utan på en statistisk bedömningsmetod som i huvudsak hänförde sig till lagens diskrimineringsgrunder. Finansiella skäl utgör inte ett berättigat syfte för att särbehandla en person. Genom nämndens beslut förbjöds kreditinstitutet att fortsätta tillämpa denna metod gentemot sökanden eller andra sökande. Under ärendets utredning framkom det att det troligen finns fler företag som agerar på ett liknande sätt.

## **Aspekter av betydelse för DO:s tillsyn**

När det gäller diskriminering genom ADB eller AI kan det vara svårt att upptäcka att man blivit utsatt för och göra antagligt att diskriminering föreligger. Om till exempel en arbetssökande sällas bort av en algoritm i ett rekryteringsverktyg och inte går vidare i en rekryteringsprocess, vet den arbetssökande inte varför det skett och även om hen misstänker att det har samband med exempelvis kön eller etnisk tillhörighet går det inte att veta för individen om en större andel kvinnor eller personer med utländsk bakgrund inte gått vidare i rekryteringsprocessen.

Vid en tillsyn torde det främst bli en fråga om bevisning, där DO behöver fastställa att ADB eller AI har använts och att detta haft en diskriminerande effekt. ADB och AI är att se som vilket annat verktyg som helst som används inom ramen för verksamheten hos en arbetsgivare eller kreditinstitut. Utgångspunkten är att DO vid en tillsyn kan kräva att ett tillsynsobjekt ger in dataprogram och liknande för att möjliggöra en granskning av programmets uppbyggnad. Ett tillsynsobjekt är enligt 4 kap. 3 § 1 DL skyldig att lämna de uppgifter om förhållandena i verksamheten som kan ha

betydelse för tillsynen. I bestämmelsens andra stycke finns dock ett undantag: Skyldigheten gäller inte om det i ett enskilt fall finns särskilda skäl som talar däremot. Enligt förarbetena<sup>90</sup> är undantaget tillämpligt på uppgifter som gäller utrikestjänsten, försvaret, affärs- eller driftsförhållanden eller andra uppgifter där motstående intressen är så starka att uppgiften inte bör lämnas ut. Undantaget ska tillämpas restriktivt.

Mot bakgrund av vad som framgår i förarbetena om undantagsbestämmelsen så kan ett tillsynsobjekt göra gällande att det exempelvis finns starka affärsintressen som gör att uppgifterna inte kan lämnas ut, och på det sättet förhindra att DO får del av uppgifterna. Det finns ännu inte någon rättspraxis avseende denna paragraf och det är därför svårt att generellt säga vad som skulle kunna utgöra så starka motstående intressen att uppgifterna inte kan lämnas ut. En sådan bedömning kommer också vara beroende av omständigheterna i det enskilda fallet.

Även om DO får tillgång till efterfrågad programmeringskod eller liknande, så kan det vara svårt att bedöma koden och programmet i arbetet med att förstå algoritmens eller programmets uppbyggnad och om det bidragit till, orsakat eller medför risk för diskriminering. Den tekniska utvecklingen kan komma att ställa nya krav på teknisk kompetens hos jurister.<sup>91</sup>

## Slutsatser

### DL:s tillämplighet

DL är tillämplig om en arbetssökande diskrimineras av en presumtiv arbetsgivare inom ramen för ett rekryteringsförfarande. Det följer av att diskrimineringsförbudet i lagen anger en personkrets som skyddas, och inte specifika situationer. Vidare anges vem som förbudet att diskriminera riktar sig mot och vem som ansvarar – nämligen arbetsgivaren (eller den som är att likställa med arbetsgivaren). Diskrimineringsförbudet på arbetslivsområdet är avsett att vara heltäckande, vilket framgår av lagens förarbeten. Om arbetsgivaren använder sig av 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.

---

<sup>90</sup> Prop. 2007/08:95, s. 546.

<sup>91</sup> Stanley Greenstein, *Our Humanity Exposed – Predictive Modelling in a Legal Context*, Stockholm 2017, s. 448.

Diskrimineringsförbudet i DL vid tillhandahållande av tjänster, exempelvis kreditgivning, till allmänheten träffar den fysiska eller juridiska person som, i enlighet med de rekvisit som uppställs i bestämmelsen, tillhandahåller en tjänst. Utifrån ordalydelsen i och syftet med bestämmelsen i 2 kap. 12 § DL kan man dra slutsatsen att ansvaret för diskriminering som sker vid tillhandahållande av en tjänst ligger hos den fysiska eller juridiska person som tillhandahåller tjänsten. Detta oavsett vilken insyn tillhandahållaren av tjänsten har i hur den algoritmen eller det program som algoritmen förekommer i fungerar.

### Aspekter av betydelse för DO:s tillsyn

I DO:s tillsyn kommer det vara viktigt att få tillgång till information om hur en algoritmen har fattat ett beslut eller på annat sätt utgjort en del i ett beslutsunderlag, samt vilka data som använts. Med stöd av 4 kap. 3 § andra stycket DL så kan dock ett tillsynsobjekt göra gällande att det exempelvis finns starka affärsintressen som gör att uppgifterna inte kan lämnas ut, och på det sättet förhindra att DO får del av uppgifterna. Det finns ännu inte någon rättspraxis avseende denna paragraf och det är därför svårt att generellt i vilka situationer som undantagsbestämmelsen kan åberopas med framgång, och de av DO efterfrågade uppgifterna således inte behöver lämnas ut. Det är också för myndigheten viktigt att ha kunskap om hur tekniken fungerar, i vilka sammanhang tekniken används samt vilka konsekvenser användningen kan få.

### Utvecklingen och användningen av ADB och AI

Utvecklingen av ADB och AI pågår ständigt och användningsområdena kommer sannolikt att förändras och utvecklas.

Inom området **rekrytering i arbetslivet** kan sammanfattningsvis sägas att produktutvecklare anser att de digitala rekryteringsverktyg (inbegripet robotar) som utvecklas ska bidra till att bland annat effektivisera rekryteringsprocesser, genomföra och analysera jobbintervjuer utan att påverkas av fördomar, magkänsla och andra subjektiva faktorer, samt vara tidsbesparande. Det pågår projekt i Sverige och i andra länder för att utveckla AI-baserade rekryteringsverktyg och robotar.

ADB eller AI används, eller kan förväntas komma att användas, för bland annat följande ändamål:

- Bakgrundskontroller via sökningar eller sökrobotar på internet.
- Ansökningar via webbförmulär som sedan behandlas av robotar.

- Granskning av CV och personliga brev i en första sällning av kandidater innan en människa tar vid i nästa skede i rekryteringsprocessen.
- Digitala eller webbaserade tester av exempelvis intelligenskvot.
- Numerisk analys, läsförståelse och personlighet.
- Intervjuer med robotar.
- Teknik för analys av ansiktsuttryck och ansiktsrörelser på videofilmade intervjuer.
- Så kallade chatbotar på rekryteringsföretags webbplatser.
- Automatiska sökningar i kandidatdatabaser med efterföljande automat-genererade mail till potentiella kandidater.

Inom området **krediter** har banker och kreditinstitut börjat intressera sig för den nya tekniken och den stora mängd information om individer som i dag kan inhämtas i så kallad stordata, som genereras av bland annat användning av smarta telefoner och internet. Den information man kan hämta i stordata kan bland annat ge underlag för att göra modeller och analyser som används i syfte att försöka förutspå framtida scenarier samt beteenden hos individer. Teknikutvecklingen har också medfört att en marknad för informations- eller datamäklare har växt fram. Dessa aktörer samlar in information om konsumenter från många källor och säljer profiler till näringsidkare.

I dag finns modeller och algoritmer som utgår från andra källor än de traditionella för kreditvärdering, för att identifiera kreditvärdiga personer och företag där bankerna inte gör det. Det öppnar för nya kundgrupper och anses göra produkterna mer konkurrenskraftiga. Information hämtas från alternativa datakällor, som kan vara mobilanvändning, hur ofta man skickar pengar till andra, vad man köper på nätet och i handeln eller om man pendlar regelbundet. En persons sociala nätverk och beteende kan också vara relevant. Användning av alternativa datakällor är enligt dess förespråkare en förutsättning för att fler ska få tillgång till finansiella tjänster, men det är också en balansakt där personlig integritet och diskriminering är risker som måste hanteras.<sup>92</sup> Förekomsten av

---

<sup>92</sup> Benjamin Ståhl, FinTech – Den digitala transformeringen av finansmarknaden, Blue Institute, Stockholm, 2017, s. 34f; Kevin Petrasic, Benjamin Saul, James Greig, Matthew Bornfreund och Katherine Lamberth, Algorithms and bias: What lenders need to know, White & Case LLP, 2017, s. 3.

kreditgivning från svenska kreditgivare med stöd av information från alternativa datakällor är okänd.

## Risker för diskriminering

Användningen av ADB och AI kan komma att medföra en risk för diskriminering, beroende på hur tekniken och verktygen eller systemen programmeras, tränas, utvecklas och används samt vilken möjlighet till insyn i programvara och algoritmer som kommer att finnas. Därtill kommer eventuell framtida reglering på området i form av till exempel lagar, riktlinjer, rekommendationer, branschetiska regler eller vad som kan komma att anses vara lämpligt.

Vid utvecklingen och användningen av AI är det viktigt dels att beakta de etiska aspekterna i form av att respektera grundläggande rättigheter, tillämplig lagstiftning samt grundläggande principer och värderingar, dels göra tekniken robust och tillförlitlig (eng: trustworthy) för att undvika avsiktlig och oavsiktlig skada. Möjligheten att kontrollera, följa upp och utvärdera uppbyggnaden, utvecklingen och användningen av algoritmer är också betydelsefullt – det som inbegrips i det som kallas transparens (eng: transparency). Detta för att kunna identifiera exempelvis felkällor i de resultat en algoritm producerar eller felaktigheter (eng: bias) i den data som en algoritm tränats i eller programmerats med.<sup>93</sup>

Sammanfattningsvis kan konstateras att det är viktigt med kunskap om hur tekniken fungerar, i vilka sammanhang tekniken används samt vilka konsekvenser användningen kan få. I DO:s tillsyn kommer det vara viktigt att få tillgång till information om hur en algoritm har fattat ett beslut eller på annat sätt utgjort en del i ett beslutsunderlag, samt vilka data som använts. Det kan också vara viktigt att göra allmänheten uppmärksam på problematiken.

## Avslutande diskussion

Användningen av ADB och AI kan medföra risk för diskriminering, men det är helt beroende på hur verktygen används. Som redogjorts för i denna promemoria finns flera orsaker till att användningen av ADB och AI kan ge, eller avsiktligt användas för att ge, ett diskriminerande resultat.

---

<sup>93</sup> Se till exempel EU-kommissionens High-Level Expert Group on Artificial Intelligence, Draft Ethics Guidelines for Trustworthy AI, Working document for stakeholders' consultation, Bryssel, 18 december 2018, s. i; Nationell inriktning för artificiell intelligens, artikelnummer N2018.14, Stockholm 2018.

Ett förslag på hur man ska kunna **öka kunskapen** om risker för diskriminering vid användning av ADB och AI framförs av Frederik Zuiderveen Borgesius i Europarådets studie.<sup>94</sup> I studien föreslås att likabehandlingsorgan bör överväga att anordna informationskampanjer för att medvetandegöra frågan inom organisationer i både offentlig och privat sektor, eftersom organisationer i många fall omedvetet använder diskriminerande ADB och AI, se ovan Omedveten diskriminering, sidan 29, och att det därför kan finnas ett värde i att medvetandegöra det hela. Det är också viktigt att informera allmänheten om riskerna med diskriminerande AI.

Professor Frederik Zuiderveen Borgesius framhåller vidare att likabehandlingsorgan (eng: equality bodies) och övervakningsorgan för mänskliga rättigheter (eng: Human rights monitoring bodies) bör se till att skaffa **teknisk expertis** om AI genom att involvera personer med programmeringskompetens. Om det finns problem med ett AI-system så kan det ofta upptäckas genom en enkel observation av ett systems indata (eng: inputs) och utdata (eng: outputs). Personer med programmeringskompetens kan bättre känna igen och förstå risker än till exempel jurister som saknar programmeringskompetens, och de kan genomföra utredningar av AI-driven diskriminering. Professor Frederik Zuiderveen Borgesius föreslår också att **likabehandlingsorgan bör samarbeta med dataskyddsmyndigheter**, som i Sverige är Datainspektionen, eftersom det kan vara värdefullt att utbyta information och lära av varandras erfarenheter. Andra möjliga organ som enligt studien kan vara värdefulla att samarbeta med är konsumentskyddsmyndigheter (Konsumentverket) och konkurrensrättsmyndigheter (Konkurrensverket).<sup>95</sup>

I en rapport från Konsumentverket reflekterar forskaren Stefan Larsson, i fråga om konsumentfrågor, över om det hos tillsynsmyndigheter behöver utvecklas vad som i rapporten benämns ”datordrivna labb”, för att möjliggöra **tillsyn** över dagens och framtidens allt kraftfullare och individualiserade konsumentprofileringsmetoder, i syfte att kunna motverka något av den växande informationsasymmetri<sup>96</sup> som i och med

---

<sup>94</sup> Discrimination, artificial intelligence and algorithmic decision-making, prof. Frederik Zuiderveen Borgesius, s. 30.

<sup>95</sup> Discrimination, artificial intelligence and algorithmic decision-making, prof. Frederik Zuiderveen Borgesius, s. 31.

<sup>96</sup> Begreppet ”informationsasymmetri” kan användas för att utifrån ett konsumentperspektiv beskriva att en konsument ofta är den svaga parten i förhållande till marknadsaktörer, på så sätt att individer i och med digitaliseringens utveckling kan få sämre inblick i hur marknadsaktörer använder sig av individernas data eller hur vissa grupper tenderar att vara mindre



den digitala utvecklingen finns mellan marknads aktörer och konsumenter.<sup>97</sup>

Konsumentverket konstaterar också i en rapport<sup>98</sup> att datainsamling och konsumentprofilering är relevant för en rad nya marknader såväl som för traditionella marknadens omvälvning, och att både forskning och policydiskussioner behöver stimuleras utifrån ett väldigt brett konsumentperspektiv. Det konstateras att det är **svårt att få insyn** i och bedöma hur långt olika marknadsaktörer kommit i sin digitala konsumentprofilering och datadrivna analys av sina kunder. Mycket av den litteratur som Konsumentverkets rapport bygger på beskriver eller är baserad på studier av en amerikansk kontext, vilket för de mest gränsöverskridande marknaderna också inkluderar Sverige och svenska konsumenter. För många marknader är det dock ännu oklart hur långt konsumentprofileringen kommit att bli en utbredd praktik för svenskt vidkommande. I rapporten konstateras att konsumentinformation, det vill säga data om konsumenters beteende, köpvanor, demografiska och andra tillhörigheter, utgör grunden för både analys och handel på och mellan olika marknader, och som en marknad i sig, på ett sätt som gör att det utifrån ett konsumentperspektiv fortsatt väcker en rad behov av insyn, fortsatta studier och frågor kring reglering, tillsyn, ansvar, integritet och nytta.

Sammanfattningsvis kan följande konstateras. Det krävs en ökad medvetenhet hos allmänheten om AI. Det krävs en också ökad kunskap om AI och den bakomliggande tekniken hos de myndigheter och organisationer inom vilkas verksamhetsområden AI kan ha betydelse. I princip omfattas alla samhällsområden. Vidare kommer samarbete och erfarenhetsutbyte mellan myndigheter att bli viktigt.

---

skickade att medvetet kunna värja sig från fallgropar, se 2016:12 Digitalisering och konsumentintresset – En litteraturöversikt, Konsumentverket 2016, s. 41ff.

<sup>97</sup> 2017:1 Kunskapsöversikter inom det konsumentpolitiska området, dnr 2017/34, Konsumentverket, 2017, s. 44.

<sup>98</sup> 2017:1 Kunskapsöversikter inom det konsumentpolitiska området, dnr 2017/34, Konsumentverket, 2017, s. 44f.

## Ordlista för vissa centrala begrepp för kartläggningen

- **ADB:** Automatisk databehandling.
- **Algoritm:** En systematisk procedur eller formel för hur man genom ett begränsat antal steg löser ett problem, kommer fram till beslut och utesluter andra svar. I många fall vill man kunna använda detta för prediktiva analyser. Att exempelvis förutsäga om en sökande kommer att bli 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 maskiner som uppvisar intelligent beteende.
- **Datautvinning** (data mining, databrytning): Innebär att gå igenom stora mängder data för att finna mönster, samband eller trender.
- **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.
- **Maskinläring:** Maskinlärande står för ett självlärande dataprogram. Efter att det en gång programmerats så "tränas" det till att själv hitta olika mönster. Dessa mönster behöver inte nödvändigtvis motsvara verkliga orsakssamband. Att klargöra varför ett självlärande program fattat ett visst beslut, exempelvis varför någon uteslutits från en ansökningsprocess, är ofta mycket svårt.
- **Prediktiv analys:** Databaserad analys för att förutsäga individers och andra aktörers beteende och framgång, exempelvis vid nyanställning eller vid studier.
- **Proxygrupper** eller **jämförelsegrupper:** Används allt mer för så kallade prediktiva analyser av arbetssökande, låntagare, konsumenter, försäkringstagare, bidragstagare och kriminella. Vid analyser med proxygrupper jämförs en individ med det kartlagda beteendet hos olika grupper utifrån antagandet att sannolikheten är stor att denne ska bete sig på ett liknande sätt.
- **RPA** eller **Robotstyrd processautomation:** Att genom datorprogram automatisera olika processer, exempelvis ansökningar. Berör framförallt regelstyrda processer som handlägger ärenden i stor omfattning.

- **Stordata** (eng: big data): När ett stort antal källor används som underlag för automatiserat beslutsstöd. Stordata betyder just mängder av data från olika datakällor. Det är när algoritmer och maskinlärande kombineras och stordata läggs till som de riktigt stora effekterna blir möjliga. Det kan handla om att bearbeta miljoner eller fler informationsbitar från en rad källor och i form av såväl ord, siffror som ljud och bilder. Även klick på internet, sökningar, vad du skriver på sociala media, vem du känner och umgås med eller vilka låtar du väljer på Spotify kan användas för olika analyser.

## Referenser

### DO-dokument

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

Underlag för prioritering av undersökningsområden för AI-projektet, LED 2018/387, handling 7

Beslut om fördjupad kartläggning inom ramen för uppdraget AI och diskriminering, LED 2018/387, handling 8

### Offentligt tryck

Proposition med förslag till lag om jämställdhet mellan kvinnor och män i arbetslivet m. m. (prop. 1978/79:175)

Proposition om en ny jämställdhetslag m.m. (prop. 1990/91:113)

Ett starkare skydd mot diskriminering (prop. 2007/08:95)

Hur står det till med den personliga integriteten? – en kartläggning av Integritetskommittén (SOU 2016:41), Stockholm, 2016

### EU-rättsakter

Fördraget om Europeiska unionen, FEU, (konsoliderad version 2016), EGT C 202, 7.6.2016

Fördraget om Europeiska unionens funktionssätt, FEUF, (konsoliderad version 2016), EGT C 202, 7.6.2016

Europeiska Unionens stadga om de grundläggande rättigheterna (2010/C 83/02), C 83/392

Europaparlamentets och Rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning)

Europaparlamentets och rådets direktiv 2006/54/EG av den 5 juli 2006 om genomförandet av principen om lika möjligheter och likabehandling av kvinnor och män i arbetslivet

## EU-publikationer

Independent High-Level Expert Group on Artificial Intelligence, set up by the European Commission: Ethics Guidelines for trustworthy AI, publicerade den 8 april 2019, <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

Yttrande från Europeiska ekonomiska och sociala kommittén om Artificiell intelligens, yttrande på eget initiativ (2018/C 440/01), vid dess 537:e plenarsession den 19–20 september 2018, publicerat i Official Journal den 6 december 2018

## Vetenskapliga artiklar

Barocas S. och Selbst A.D., ”Big Data’s disparate impact (2016) Calif Law Review, vol. 104, s. 671–732

Buolamwini, Joy och Gebru, Timnit “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification” in the Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency, tillgänglig på <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>

Minds and Machines, 28(4): 689–707

Naarttijärvi, Markus, Rättsstatlighet och algoritmiska svarta lådor, Jubileumsskrift till Juridiska Institutionen 40 år, Umeå Universitet, Juridiska institutionen, Umeå 2017, s. 250–253

Persson, Anders, Implicit Bias When Using Data Profiling within Recruitments and Human-Resource Management, IT-department, Uppsala Universitet

## Övriga artiklar

Petrasic, Kevin, Saul, Benjamin, Greig, James, Bornfreund, Matthew och Lamberth, Katherine, Algorithms and bias: What lenders need to know, White & Case LLP, 2017

Ståhl, Benjamin, FinTech – Den digitala transformeringen av finansmarknaden, Blue Institute, Stockholm, 2017

Tidningen Karriär, nr 1 2019, s. 15, som finns tillgänglig i digital form på [http://pdf.flip.se/Karriar/Karriar\\_1\\_2019/](http://pdf.flip.se/Karriar/Karriar_1_2019/)

## Europarådet

Europarådet: <https://www.coe.int/en/web/cepej>.

<https://rm.coe.int/presentation-note-en-for-publication-4-december-2018/16808f699d>, hämtad 2019-04-02

Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. Strasbourg: Europarådet, Directorate General of Democracy

## Litteratur

Fransson/Stüber, Diskrimineringslagen – en kommentar, 2 uppl., Norstedts, 2015

Greenstein, Stanley, Our Humanity Exposed – Predictive Modelling in a legal context (akademisk avhandling, Juridiska institutionen, Stockholms Universitet) 2017

Lunning/Toijer, Anställningsskydd, Wolters Kluwer, 11 uppl. 2016

O’Neil, Cathy, ”Weapons of math destruction”, (2016) Broadway Books

## Myndighetspublikationer

Consumer protection on the financial market, Finansinspektionen, 2016

2016:12 Digitalisering och konsumentintresset – En litteraturöversikt, Konsumentverket 2016

2017:1 Kunskapsöversikter inom det konsumentpolitiska området, dnr 2017/34, Konsumentverket, 2017

2017:4 Personuppgifter som betalningsmedel, Konsumentverket, 2017

Nationell inriktning för artificiell intelligens, artikelnummer N2018.14, Regeringskansliet, Stockholm 2018

## Websidor

<https://www.fernsoftware.com/blog/fintech-offers-credit-scores-without-fico/>, hämtad 2019-03-22

<https://gomedici.com/can-web-search-history-become-a-holy-grail-of-credit-scoring/>

<https://www.fernsoftware.com/blog/simple-messages-help-consumers-understand-big-data/>, hämtad 2019-03-22

<https://www.va.se/nyheter/2018/04/26/Mot-ikeas-nya-rekryteringsrobot-som-klarar-1-500-kandidater-pa-en-dag/>, hämtad 2018-08-24

<https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>, hämtad 2018-10-11

<https://www.ajlunited.org/the-coded-gaze>, Joy Buolamwini

[http://www.mynewsdesk.com/se/pressreleases/digital-robotersaetter\\_rekryterare-artif...](http://www.mynewsdesk.com/se/pressreleases/digital-robotersaetter_rekryterare-artif...), hämtad 2018-08-24

<https://www.psykologiguiden.se/psykologilexikon/?Lookup=psykometrik%2C+psykometri>, hämtad 2019-04-01

<https://evertiq.se/news/36403>, hämtad 2018-11-22

Shortcut: “Möt Matlda – världens första rekryteringsrobot”,  
<http://shortcut.se/artiklar/mot-matlda-varldens-forsta-rekryteringsrobot/>,  
hämtad 2018-08-24