



LUNDS UNIVERSITET

Ekonomihögskolan

Institutionen för informatik

Bias inom AI-baserad rekrytering

Vad företag som tillhandahåller automatiserade rekryteringstjänster med inslag av AI gör för att minska bias i rekryteringsprocessen

Kandidatuppsats 15 hp, kurs SYSK16 i Informatik

Författare: Johanna Alm
Amanda Löwgren
Elin Melkersson

Handledare: Benjamin Weaver

Rättande lärare: Björn Svensson
Christina Keller

Bias inom AI-baserad rekrytering: Vad företag som tillhandahåller automatiserade rekryteringstjänster med inslag av AI gör för att minska bias i rekryteringsprocessen

ENGELSK TITEL: Bias in AI-based recruiting: Actions taken to reduce bias by companies that provide automated recruiting services with AI

FÖRFATTARE: Johanna Alm, Amanda Löwgren och Elin Melkersson

UTGIVARE: Institutionen för informatik, Ekonomihögskolan, Lunds universitet

EXAMINATOR: Christina Keller, Professor

FRAMLAGD: maj, 2021

DOKUMENTTYP: Kandidatuppsats

ANTAL SIDOR: 91

NYCKELORD: bias inom AI, AI i rekrytering, artificiell intelligens, etik och AI

SAMMANFATTNING (MAX. 200 ORD):

Teknikens framsteg har förändrat stora delar av rekryteringsprocessen. Sedan millennieskiftet till år 2015 har erorna Digital Recruiting 1.0 och 2.0 transformerat rekryteringsbranschen tack vare internets etablering. Digital Recruiting 3.0 påbörjades då behovet av att inkludera tekniker med artificiell intelligens uppkom på grund av internet, då enorma mängder ansökningar behövde bearbetas. Dock har dessa tekniker presenterat nya utmaningar; att de enkelt riskerar att innehålla bias och därmed inte kan leverera etiska resultat, vilket leder till diskriminerande anställningsprocesser. Uppsatsen syftar till att undersöka hur bias i AI-företags rekryteringstjänster uppkommer samt hur de arbetar med att minska dessa bias. I litteraturen bearbetas begreppet AI och tillhörande tekniker, biaser samt Europeiska kommissionens riktlinjer för tillförlitlig AI. För att besvara forskningsfrågan genomfördes en kvalitativ studie där fyra AI-företag och ett rekryteringsbolag intervjuades om bias och hur de arbetar med bias. En diskussion förs kring vilka biaser som uppkommer samt hur detta kan leda till mindre bias. Studien visar hur olika biaser kan uppkomma på

grund av olika komponenter och processer i liknande tjänster. Studien visar även hur företag minskar bias i sina tjänster utifrån Europeiska kommissionens riktlinjer.

Innehåll

1	Introduktion.....	1
1.1	Bakgrund	1
1.2	Problemformulering.....	2
1.3	Forskningsfråga	4
1.4	Syfte.....	4
1.5	Avgränsningar	4
2	Litteraturgenomgång	5
2.1	Artificiell intelligens och maskininläring	5
2.1.1	Artificiell intelligens	5
2.1.2	Maskininläring	5
2.2	Bias inom AI.....	6
2.3	Riktlinjer för tillförlitlig AI	8
2.3.1	Bakgrunden till riktlinjerna för tillförlitlig AI.....	8
2.3.2	Mänskligt agentskap och mänsklig tillsyn	8
2.3.3	Integritet och dataförvaltning	9
2.3.4	Transparens	10
2.3.5	Mångfald, icke-diskriminering och rättvisa	10
2.3.6	Samhällets och miljöns välbefinnande	11
2.3.7	Ansvarsskyldighet	11
2.4	Sammanfattning av ramverk med koppling till litteratur	12
3	Metod	13
3.1	Forskningsstrategi och forskningsansats	13
3.2	Litteraturstudie.....	13
3.3	Urval	14
3.4	Empirisk datainsamling	15
3.4.1	Etiska överväganden	17
3.5	Dataanalys	18
3.6	Kvalitetsmått.....	19
3.6.1	Validitet.....	19
3.6.2	Reliabilitet	19
3.7	Metodreflektion	20
4	Empiri	21
4.1	Expertintervju med Wise Consulting AB	21

4.2	Företagens tjänster	21
4.2.1	Tengais tjänst.....	22
4.2.2	Exparangs tjänst	22
4.2.3	Huberts tjänst.....	23
4.2.4	Devtens tjänst	23
4.3	Artificiell intelligens.....	23
4.3.1	AI-tekniker	24
4.3.2	Bias inom AI	24
4.4	Europeiska kommissionens riktlinjer för tillförlitlig AI.....	27
4.4.1	Mänskligt agenskap och mänsklig tillsyn	28
4.4.2	Integritet och dataförvaltning	28
4.4.3	Transparens	29
4.4.4	Mångfald, icke-diskriminering och rättvisa	30
4.4.5	Samhället och miljöns välbefinnande	31
4.4.6	Ansvarsskyldighet	32
5	Diskussion.....	33
5.1	Insikter i rekryteringsbranschens syn på AI	33
5.2	AI-tekniker.....	34
5.3	Bias inom AI.....	35
5.4	Europeiska kommissionens riktlinjer för tillförlitlig AI.....	37
5.4.1	Mänskligt agenskap och mänsklig tillsyn	37
5.4.2	Integritet och dataförvaltning	37
5.4.3	Transparens	38
5.4.4	Mångfald, icke-diskriminering och rättvisa	40
5.4.5	Samhället och miljöns välbefinnande	41
5.4.6	Ansvarsskyldighet	41
6	Slutsats	44
6.1	Vidare forskning	45

Tabeller

Tabell 1: Ramverk med koppling till litteratur.....	12
Tabell 2: Översikt över intervjuobjekt samt relaterad information.....	15
Tabell 3: Intervjuguide	16
Tabell 4: Översikt för färgkodning och ID-taggar	18
Tabell 5: Utdrag från den transkriberade intervjun med Tengai, se appendix B	19
Tabell 6: Översikt för hämtad data av företagens tjänster	22
Tabell 7: Översikt för hämtad data av artificiell intelligens.....	23
Tabell 8: Översikt av företagens AI-tekniker.....	24
Tabell 9: Identifierade biaser i tjänsterna	27
Tabell 10: Översikt för hämtad data av Europeiska kommissionens riktlinjer	27

1 Introduktion

1.1 Bakgrund

Likt resten av samhället, lämnas inte rekryteringsprocessen oberörd av teknikens framsteg och har sedan mitten på nittiotalet gått från att vara en fullständigt analog process till att idag oftast genomförs helt, eller delvis digitalt. Black och van Esch (2019) förklarar att förr publicerades ofta tjänster i tryckt media. Som arbetssökande var man då tvungen att besöka varje enskilt företag för att fylla i en arbetsansökan på plats som sedan bearbetades av företagets HR-avdelning (Black & van Esch, 2019). De anställda på HR-avdelningen var därmed ansvariga för att attrahera kandidater, sortera och granska ansökningar samt avgöra vilka kandidater som skulle gå vidare i urvalsprocessen (Black & van Esch, 2019). Black och van Esch (2019) skriver att tydliga begränsningar uppstod av mängden information som var fysiskt möjligt för en människa att ta in och orka bearbeta. På grund av den omständliga processen fick företag för att tillsätta en tjänst göra en avvägning mellan att nå ut till så många relevanta kandidater som möjligt, gentemot en rimlig kostnad för att publicera jobbannonser eller mot att anlita företag som genomförde sökningen åt dem (Black & van Esch, 2019). Utöver detta menar Black och van Esch (2019) att urvalsprocessen präglades av rekryterarnas bias i olika former. Black och van Esch (2019) beskriver hur en rekryterare kunde förlita sig för mycket på den första informationen som delgavs om en kandidat, att rekryteraren endast tog till sig information som vidare bekräftade dennes uppfattning om kandidater eller att rekryteraren omedvetet favoriserade kandidater som var lik rekryteraren själv.

Black och van Esch (2019) beskriver hur eran *Digital recruiting 1.0* inleddes när stora delar av rekryteringsprocessen digitaliserades genom att jobbannonser lades ut på internet. Detta ledde till att information om tjänster nådde den stora massan för nästan ingen kostnad (Black & van Esch, 2019). Eran *Digital recruiting 2.0* triggades av två faktorer; dels jobbsajter som aggregerade annonser från flera företags olika plattformar och dels genom introduktionen av digitala, professionella sociala nätverk, såsom LinkedIn (Black & van Esch, 2019). Black och van Esch (2019) menar att de arbetssökande behövde endast besöka en hemsida. De sociala plattformarna influerade rekryteringsprocessen genom att delningen av information ökade och möjliggjorde även att företag kunde rikta vissa jobbannonser till specifika, mer intressanta användare (Black & van Esch, 2019).

Företagsvärlden närmade sig vad som kallas *Digital Recruiting 3.0* år 2015, en förändring som grundas i nyttjandet av artificiell intelligens, vidare benämnt AI, under rekryteringsaktiviteter (Black & van Esch, 2019). De två tidigare digitala transformationerna var essentiella för framskridandet av *Digital Recruiting 3.0* (Black & van Esch, 2019). De genererade en ofantlig mängd ansökningar till en viss annonserad position, samt att humankapitalet på företagen värderades högre, då man insåg fördelarna med att anställa rätt person för rätt position, vilket satte större press på rekryteringsprocessen (Black & van Esch, 2019). Black och van Esch (2019) menar att företag tvingades antingen spendera mer tid eller

arbetskraft på att granska alla arbetsansökningar samtidigt som de var på jakt efter den perfekta kandidaten för ett visst arbete. I denna kontext växte behovet av AI-verktyg fram för att effektivisera rekryteringsprocessen (Black & van Esch, 2019).

En rapport från Oracle (2019) visar att AI-verktyg kan utöver att vara behjälpliga i att tackla de utmaningar *Digital Recruiting 3.0* innefattar, även kan optimera rekryteringsprocesser till en ny nivå. Då talangförvärv kan vara tidskrävande, ha inbäddad bias och ofta är en reaktion på tidigare händelser snarare än en proaktiv investering, kan en rekryteringsprocess optimeras och effektiviseras på ett flertal olika sätt genom att inkorporera AI (Bogen & Rieke, 2018). Sourcing definieras som steget där arbetsgivare efterlyser interna och externa kandidater till de nya arbetstillfällena som publicerats (Bogen & Rieke, 2018). Black och van Esch (2019) menar att vid detta steg kan AI vara behjälplig genom att upptäcka de främsta kandidaterna för en tjänst och annonsera tjänster till kandidater via riktade kampanjer. I dagsläget tar företag normalt emot mellan 20 till 200 ansökningar för var enskild tillgänglig tjänst, vilket sätter press på rekryteringsprocessen och ökar behovet av AI-tjänster (Black & van Esch, 2019).

Användningen av AI inom rekrytering har en märkbar effekt, något som Black och van Esch (2019) ger flera exempel på. Exempelvis nyttjade L'Oréal AI genom annonsering av sina 5000 öppna tjänster till både aktivt sökande och passiva kandidater; något som resulterade i att företaget tog emot två miljoner Cv:n under år 2017 när AI:n implementerades (Black & van Esch, 2019). Under intervjuprocessen kan AI bistå som en digital assistent som kan schemalägga och interagera med kandidaten (Oracle, 2019). Exempelvis nyttjade L'Oréal en AI-chattbot som interaktivt frågade kandidater tre frågor och sedan jämförde deras svar med svar från högpresterande anställda på företaget (Black & van Esch, 2019). AI-botten vägde även in meningsstruktur och ordval i sin analys och genererade sedan ett testresultat för varje kandidat, som användes som underlag för rekryterare när de skulle besluta om vilka kandidater som skulle intervjuas av dem (Black & van Esch, 2019). En AI kan även effektivisera granskningsprocessen av Cv:n som skickas in avsevärt (Black & van Esch, 2019). Hilton Hotels & Resorts minskade *time-to-hire* från 42 till fem dagar, med andra ord hela 90 procent, genom att nyttja AI (Black & van Esch, 2019).

1.2 Problemformulering

Bakgrunden klagör hur rekryteringsbranschen har genomgått stora teknologiska förändringar, från helt analoga processer till AI-verktyg som filtrerar hundratals jobbansökningar och vilka effekter det har haft på branschen i sin helhet. Dattner et. al (2019) påpekar hur dessa teknologier även presenterar helt nya utmaningar för branschen ur legala och etiska aspekter. Silberg och Manyika (2019) menar att dessa förändringar har triggat en helt ny debatt om rättvisa och etiska frågor. I vilka kontexter kan en AI minska eller öka risken för bias? Vilka processer krävs för att testa samt minimera bias? Hur påverkas branschen idag av underliggande biaser? Hur arbetar människa och AI-system bäst för att undvika bias? Det är många frågor som kräver svar, samtidigt som teknikens frammarsch bevisligen går snabbt (Silberg & Manyika, 2019).

Europeiska kommissionen (2021) har påbörjat arbetet med en ny förordning relaterad till AI-system. Vidare poängterar kommissionen (2021) att de tekniker som kan vara gynnsamma för samhället även skapar nya risker och negativa konsekvenser för individer eller samhället i sin

helhet. Europeiska kommissionen (2021) har utformat en kategorisering av områden som AI nyttjas inom för att se till att systemen är etiska, transparenta, objektiva och säkra. De har bedömt AI inom anställning och personalförvaltning som ett högriskområde; vilket är en nivå från att vara en oacceptabel risk (Europeiska kommissionen, 2021). Den bias som genomsyrar den traditionella rekryteringsprocessen återfinns även i dagens AI-verktyg vilket kan upprätthålla diskriminerande anställningsprocesser (Capgemini, 2020).

En rapport från Capgemini (2020) visar att nästan 40 procent av de tillfrågade kunderna uppger att de hade återgått till mänsklig interaktion efter en negativ upplevelse med ett AI-system. Detta innebär att en AI kan bli väldigt kostsam för de företag som levererar produkten eller tjänsten (Capgemini, 2020). Kundernas tillit till interaktion med AI ökat, från 30 procent år 2018 till 49 procent år 2020 (Capgemini, 2020). 66 procent förväntar sig att AI:n är helt rättvis och fördomsfri (Capgemini, 2020). Capgemini (2020) menar att dessa förväntningar är svåra att leva upp till då AI-system har orsakat etiska problem för nio av tio organisationer. Exempelvis har mörkhyade personer tilldelats oförtjänt hårda straff i amerikanska rättsalor i jämförelse med vita personer samt att Ubers ansiktsgenkänning misslyckades med att registrera sina förare som var transpersoner (West, Whittaker & Crawford, 2019). Ett omtalat exempel relaterat till rekrytering beskrivs i en artikel av Dastin (2018) där Amazon utvecklade ett internt HR-system som de tillslut fick överge. Detta berodde på att kvinnliga sökanden underminerades då systemets förutsägelser baserades på data från främst manliga ingenjörers Cv:n (Dastin, 2018). Iansiti & Karim (2020) beskriver hur Cv:n som innehöll ordet "kvinnlig", som i "kvinnlig schackklubb", eller innefattade examen från colleges med enbart kvinnliga studenter sållades bort. Ytterligare ett exempel är ett företag som introducerade ett CV-granskande verktyg och upptäckte att modellen identifierade namnet "Jared" och att ha spelat lacrosse i *high school* som stora framgångsfaktorer kopplat till arbetsprestation (Bogen & Rieke, 2018).

Capgemini (2020) visar på hur företag rör sig mot att bli mer medvetna om etik kopplat till AI-system. 2019 hade endast 5 procent av organisationer tagit fram etiska riktlinjer för utveckling av AI, något som till år 2020 ökat med 40 procentenheter (Capgemini, 2020). Vidare så var endast 35 procent av ledare inom organisationer medvetna om diskriminerande bias i AI-system, vilket har ökat till 65 procent år 2020 (Capgemini, 2020). Organ såsom Europeiska kommissionen och amerikanska FTC har tagit fram riktlinjer som företag kan rätta sig efter för att uppnå en mer etisk AI (Europeiska kommissionen, 2019a; Federal Trade Commission, 2016). Riktlinjerna som den Europeiska kommissionen (2020) publicerat, syftar till att europeiska länder ska samlas runt en samordnad AI-plan för att gynna den politiska utvecklingen och de nationella strategierna gällande artificiell intelligens. Trots detta verkar företag göra blygsamma framsteg inom områdena transparens, förklarbarhet, rättvisa och granskning av AI-system (Capgemini, 2020). Benbya, Pachidi och Jarvenpaa (2021) menar att chefer förväntar sig att mänsklig bias kommer elimineras som följd av automatiserade beslut. Vidare menar Benbya, Pachidi och Jarvenpaa (2021) att det har konstaterats att AI-system som använts i bland annat rekryteringssammanhang har bidragit till nya former av bias vilket kan leda till en ökning av systematisk diskriminering.

Benbya, Pachidi och Jarvenpaa (2021) menar att det finns ett stort behov av att undersöka vilka typer av bias som uppkommer och hur. Vidare menar Benbya, Pachidi och Jarvenpaa (2021) de att det finns ytterligare ett behov av att undersöka hur risken för bias kan minimeras. Behovet av att ha en insikt i hur bias uppkommer i AI-system samt hur företag ska arbeta med detta har utmynnat till vår forskningsfråga.

1.3 Forskningsfråga

Hur uppkommer bias inom artificiell intelligens i rekryteringstjänster? Hur arbetar företag idag med att minska bias inom artificiell intelligens i deras automatiserade rekryteringstjänster som har inslag av artificiell intelligens?

1.4 Syfte

Syftet är att genom en kvalitativ intervjustudie undersöka hur företag, som tillhandahåller automatiserade rekryteringstjänster med inslag av artificiell intelligens, arbetar för att minska bias inom artificiell intelligens. Studien syftar även till att klargöra hur dessa biaser uppkommer. Studien ska belysa de åtgärder och arbete som görs för att i sin tur minska risken för att bias inom automatiserade rekryteringsprocesser uppstår.

1.5 Avgränsningar

Vi har avgränsat oss till att endast undersöka företag som tillhandahåller tjänster med inslag av artificiell intelligens som stödjer rekryteringsprocesser. De företag som valts ut är verksamma på den svenska marknaden.

2 Litteraturgenomgång

Under kapitlet litteraturgenomgång kommer relevant fakta och teorier tas upp och förklaras. Litteraturgenomgången är uppdelad i följande delkapitel; artificiell intelligens och maskininlärning, bias inom AI och riktlinjer för tillförlitlig AI. I första delkapitlet presenteras AI och olika AI-tekniker. I andra delkapitlet presenteras de biaser som riskerar att uppkomma inom AI och till sist presenteras Europeiska kommissionens riktlinjer för tillförlitlig AI.

2.1 Artificiell intelligens och maskininlärning

2.1.1 Artificiell intelligens

AI är ett begrepp som är svårt att definiera och avgränsa till en definition (Dignum, 2020). I denna uppsats definieras AI enligt nedan:

Artificiell intelligens avser system som uppvisar intelligent beteende genom att analysera sin miljö och vidta åtgärder – med viss grad av självständighet – för att uppnå särskilda mål. AI-baserade system kan vara helt programvarubaserade och fungera i den virtuella världen (t.ex. röstassistenter, bildanalysprogram, sökmotorer, tal- och ansiktsgenkänningssystem), eller inbäddas i hårdvaruenheter (t.ex. avancerade robotar, självkörande bilar, drönare eller tillämpningar för sakernas internet) (Europeiska kommissionen, 2019b, s.1).

Syftet med AI är bland annat till att stödja repetitiva uppgifter samt uppgifter som kräver fysisk ansträngning eller kognitiva egenskaper; något en robot kan göra med hjälp av AI, då den känner av sin omgivning och kan assistera människor i olika uppgifter (Benbya, Pachidi & Jarvenpaa, 2021). Detta är möjligt med hjälp av bland annat maskininlärning, vidare benämnt ML, som diskuteras ytterligare i nedanstående kapitel (Benbya, Pachidi & Jarvenpaa, 2021). När ett AI-system tar ett beslut behöver det inte betyda att systemet är självständigt, utan att en människa kan krävas för det slutgiltiga beslutet (Europeiska kommissionen, 2018). För en organisation är AI användbart och bidrar till mycket nytta, men det kan även resultera i oavsiktliga eller avsiktliga konsekvenser som påverkar individer och även organisationer i sin helhet (Benbya, Pachidi & Jarvenpaa, 2021).

2.1.2 Maskininlärning

ML bidrar till att göra datorsystem smarta och intelligenta och möjliggör att systemen ständigt kan anpassa sig efter nya omständigheter (Jordan & Mitchell, 2015; Alpaydin, 2010). Genom att lära systemen den önskade in- och outputen kan systemet uppnå självständighet, istället för att programmera varje tänkbar utfall som kan ske när systemet används (Jordan & Mitchell, 2015; Alpaydin, 2010). Det är sedan datorns uppgift att ständigt utvecklas och förbättras utifrån den data som tillförs (Alpaydin, 2010).

Jordan och Mitchell (2015) menar att det finns olika ML-algoritmer för olika användningsområden. De menar vidare att flesta algoritmer fokuserar på att precisera sina resultat och att göra förutsägelser som överensstämmer med verkligheten. Inlärningsprocessen

går ut på att mata systemet med välkända in- och output-par, för att på så sätt generera noggrannhet för algoritmen (Jordan & Mitchell, 2015).

ML innehåller även icke-vägled inläring, vilket betyder att det endast är input-data som används och ingen bestämd output-data som används (Alpaydin, 2010). Tanken med icke-vägled inläring är att identifiera regelbundenheter i input-data (Alpaydin, 2010). Icke-vägled inläring utgörs bland annat av *klustering*, där målet är att hitta grupperingar bland inputen (Jordan & Mitchell, 2015). Metoden används för att skapa en regel för att förutspå data (Jordan & Mitchell, 2015; Rai & Sing, 2010).

Natural language processing, vidare benämnt NLP, handlar om hur ett datorsystem kan förstå och producera ett eller flera mänskliga språk och används inom rekrytering (Bogen & Rieke, 2018). Bogen och Rieke (2018) menar att NLP består ofta av fyra olika delar; generering av input, analys av input, generering av output samt förstärkningsinläring. Generering av input handlar om att ge algoritmen data genom antingen text eller röst (Bogen & Rieke, 2018). Input-analysen hanteras olika, beroende på om inputen är textbaserad eller röstbaserad (Bogen & Rieke, 2018). Om data är textbaserad används *natural language understanding*, NLU, för att förstå innebörden av data, medan data som är röstbaserad analyseras med hjälp av en kombination av *automatic speech recognition*, ASR, och NLU för att förstå datas innebörd (Bogen & Rieke, 2018). Google har ett eget ASR-verktyg som är baserad på *cloud-computing*, som transformerar tal till text genom Googles API-tjänst och AI-tekniker (Google Cloud, n.d). Vidare innebär output-generering att *natural language generation*, NLG används för att formulera ett svar till personen som interagerar med AI:n (Bogen & Rieke, 2018). Till sist handlar förstärkningsinläring om att ML-algoritmen förfinar outputen över tid för att göra den så relevant och korrekt som möjligt (IBM Cloud Education, 2020).

Conversational AI innebär teknologi som en människa kan integrera med genom skrift eller tal och möjliggörs av NLP och ML (IBM Cloud Education, 2020). Idag används tekniken inom flera olika områden såsom Amazons smarta högtalare Alexa, Google translate, stavningskontroll, textförslag, vid intervjuer och *scrapping* (IBM Cloud Education, 2020).

Googles BERT är ett av flera AI-system som lär sig hur människor talar och skriver utifrån mängder av information som finns tillgänglig online; såsom gamla böcker, wikipedia-sidor och nyhetsartiklar (Metz, 2019). Metz (2019) beskriver hur denna inlärningsmetod kan innebära att flera decennier, och till och med århundraden, av bias därmed kan bakas in i algoritmerna. Metz (2019) menar att när dataforskaren Robert Munro testade vilket kön BERT associerade vissa ord till, kunde det konstateras att orden "jewelry", "baby", "horses", "money" och "action" i 99 fall av 100 associerades med män. Vidare menar Metz (2019) att ordet "mom" var undantaget som associerades med kvinnor. Vidare menar Metz (2019) att forskare fortfarande inte vet varför teknologi likt BERT beter sig som den gör och kan bli överraskade över algoritmens misstag. Google är dock medvetna om detta deras talesperson att sagt att minska bias i deras system är en topprioritet (Metz, 2019).

2.2 Bias inom AI

Bias är inte ett nytt problem som har uppkommit med AI, men AI kan ha bidragit till en ökning av redan existerande bias och kan även leda till nya klassificeringar och kriterier vid val som i sin tur kan skapa nya typer av bias (Ntoutsis, Fafalios, Gadiraju, Iosifidis, Nejdli, Vidal, Ruggieri, Turini, Papadopoulos, Krasanakis, Kompatsiaris, Kinder-Kurlanda, Wagner,

Karimi, Fernandez, Alani, Berendt, Kruegel, Heinze, & Broelemann, 2020) Hellström, Dignum och Bensch (2020) menar att den allmänna juridiska definitionen av bias är beslut som baseras på förutfattade meningar eller fördomar, istället för en opartisk utvärdering utifrån fakta. I och med att ML till stor del bygger på observationer från verkliga världen, som delar föregående definition av bias, menar författarna (2020) att definitionen även kan användas när man talar om bias i ML. Även Ntoutsis et al. (2020) menar att bias inom AI innebär ett AI-system som tar fördomsfulla beslut mot eller för en person eller grupp och som antingen är gynnsamt eller ofördelaktigt, speciellt på ett sätt som anses orättvist. Ntoutsis et al. (2020) fortsätter med att beskriva hur viktigt det är att skaparna och utvecklarna är medvetna om bias samt diskriminering för att då kunna säkerhetsställa att teknologin kan nyttjas ansvarsfullt.

Enligt Hellström, Dignum och Bensch (2020) finns det flera olika typer av bias som kan påverkar algoritmers utfall. Den första typen av bias som författarna (2020) beskriver är *inductive bias*, även benämnt *learning bias*, som innebär att en algoritm tränas med indata som kopplas till en viss utdata för att algoritmen på så sätt ska förstå relationen mellan de olika parametrarna. Algoritmen ska i sin tur kunna göra antaganden även utifrån data den inte tränats med (Hellström, Dignum & Bensch, 2020). Alla typer av induktiv inlärning kräver någon form av *inductive bias* för att fungera, där en kritisk del av inlärningsprocessen är valet av hur stort själva urvalet ska vara (Hellström, Dignum & Bensch, 2020).

Historical bias är en annan typ av bias som innebär att ord som algoritmen tränas med, kan ha en viss typ av betydelse eller innebörd som är fördomsfull och som i sin tur kan leda till att algoritmen kan innehålla bias (Hellström, Dignum och Bensch, 2020). Då modellen utformas och tränas baserat på hur omvärlden just nu uppfattar, eller har uppfattat, ord vid tillfället då algoritmen utformas kommer utfallet från algoritmen antagligen innehålla bias trots att data är kontrollerad (Hellström, Dignum och Bensch, 2020). ML bygger ofta på korrelationer och Hellström, Dignum och Bensch (2020) menar då att ord kan korreleras med andra, korrelationen skapar i sin tur bias. Vidare tar författarna (2020) upp ordet matlagning som exempel, som har kunnat korreleras med det kvinnliga könet.

Specification bias innebär den bias som kan uppstå när utvecklarna väljer vad som ska vara in- och ut-parametrar till algoritmen (Hellström, Dignum och Bensch, 2020). Valet av data, vilket görs av systemtillverkarna, kräver hög förståelse av det problem som algoritmen syftar till att lösa menar Hellström, Dignum och Bensch (2020). Genom att avsiktligt eller oavsiktligt välja data som är fördomsfull, kan det leda till ett sämre resultat men också till att vissa minoritetsgrupper systematiskt missgynnas (Hellström, Dignum & Bensch, 2020).

Hellström, Dignum och Bensch (2020; Dattner et al., 2019) fortsätter beskriva problematiken med bias och hur det kan uppstå genom *proxys*. Författarna menar (2020) För att göra systemet mindre fördomsfullt kan utvecklarna välja att helt ta bort begrepp som in- och ut-parametrar under algoritmens lärandeprocess som anses vara fördomsfull, exempelvis etnicitet (Hellström, Dignum & Bensch, 2020). Problemet är att andra attribut, som inte aktivt tas bort från algoritmen, kan leda fram till samma bias som uppkommer av det attribut som tas bort och på så sätt agera som *proxys* för det ursprungliga ordet (Hellström, Dignum & Bensch, 2020). Författarna (2020) tar upp exemplet att attributet postnummer kan vara en *proxy* till etnicitet vilket innebär att trots att utvecklarna aktivt har valt att ta bort exempelvis etnicitet som in-parameter, kan samma fördomar uppkomma genom att veta var en person bor.

Hellström, Dignum och Bensch (2020) menar att *measurement bias* beskriver de oavsiktliga fel som görs vid mätningar av data. Författarna (2020) menar att det kan innebära många saker, bland annat att felaktiga mätredskap används vilket leder till systematiska fel. Vidare innebär *observer bias* den skillnaden som är mellan det som utspelar sig i verkligheten och hur observatören själv uppfattar situationen (Hellström, Dignum & Bensch, 2020).

Sampling bias uppstår vid felaktig framställning av observationer av populationer, ifall vissa grupper är över- eller underrepresenterade (Hellström, Dignum & Bensch, 2020). Hellström, Dignum och Bensch (2020) menar att detta i sin tur leder till att den presterar undermåligt hela tiden eller endast för vissa minoritetsgrupper.

Det är vanligt att vissa typer av ML-verktyg används för att generera input-data till andra algoritmer (Hellström, Dignum & Bensch, 2020). Nackdelen med detta är enligt Hellström, Dignum och Bensch (2020) att utvecklarna då har svårare att påverka biasen av input-data eftersom den genereras från en annan algoritm som de förmodligen inte själva varit med och utvecklat. Vidare menar författarna (2020) att denna typ av bias benämns *inherited bias*. Författarna menar att flera typer av ärvda ordbehandlingsprocesser ofta bidrar till *inherited bias* (Hellström, Dignum & Bensch, 2020).

2.3 Riktlinjer för tillförlitlig AI

Nedan presenteras de riktlinjer Europeiska kommissionens expertgrupp, bestående av 52 personer från akademien, näringslivet och den offentliga sektorn, har utformat i syfte av att främja tillförlitlig AI. En del av dessa riktlinjer grundas i etik, för att säkerställa att etiska principer upprätthålls i AI-system, vilket denna studie ämnar diskutera.

2.3.1 Bakgrunden till riktlinjerna för tillförlitlig AI

Europeiska kommissionen (2020) har sex prioriteringar för sitt strategiska arbete mellan åren 2019–2024; en av dessa prioriteringar är vara rustad för den digitala tidsåldern, vilket AI är inkluderat i. Europeiska kommissionens (2020) riktlinjer publicerades i syfte till att närma sig en samordnad AI-plan som sedan legat till grund för politisk utveckling och nationella strategier i EU. År 2021 publiceras även ett förslag på en ny förordning gällande AI, främst för att kontrollera att privatpersoners samt företags rättigheter och säkerhet upprätthålls samtidigt som EU:s konkurrenskraft stärks (Europeiska kommissionen, 2021). AI-system inom högriskområden kommer behöva genomgå en bedömning och uppfylla de uppsatta AI-krav som Europeiska kommissionen tagit fram (Europeiska kommissionen, 2021). I ett pressmeddelande från Europeiska kommissionen (2021) påpekas det att förtroende är ett måste gällande AI vilket ska uppnås med framtidssäkrad, etisk teknik. Trots att detta förslag på förordning inte behandlas i uppsatsen, visar Europeiska kommissionens intentioner på deras tillförlitlighet inom området, vilket är skälet till att vi nyttjar riktlinjerna i denna uppsats.

2.3.2 Mänskligt agentskap och mänsklig tillsyn

Den första riktlinje som tagits fram av Europeiska kommissionen (2019a) kretsar kring att inkludera människor som en deltagande part i ett AI-system. Europeiska kommissionen

(2019a) menar att användarens aktiva roll ska stödjas och tillsyn av AI-systemen ska enkelt tillåtas. Vidare menar författarna (2019a) att användare ska tillgodoses med information och kunskap som krävs för att förstå och interagera med AI-systemet, målet är att AI-systemet ska hjälpa användaren att göra mer välgrundade val. Europeiska kommissionen (2019a) menar att det är ett krav på att användarens autonomi är central för funktionaliteten av systemet. Detta är för att en person inte ska behöva bemöta konsekvenser av en fullt automatiserad bearbetning som utmynnat i ett beslut; något som även är förtydligt i Artikel 22; "Automatiserat individuellt beslutsfattande, inbegripen profilering" i allmänna dataskyddsförordningen (Dataskyddsförordningen, 2016).

För att undvika att användarens autonomi inskränks eller andra negativa effekter krävs mänsklig tillsyn (Europeiska kommissionen, 2019a). Europeiska kommissionen (2019a) menar att *human-in-the-loop*, *human-on-the-loop* och *human-in-command* är exempel på styrmekanismer som ger mänsklig tillsyn. *Human in-the-loop* har en medverkande människa vid varje beslut i varje del av systemet (Europeiska kommissionen, 2019a). *Human-on-the-loop* ger möjligheten för människor att ingripa under systemets utformning och när det är i drift (Europeiska kommissionen, 2019a). *Human-in-command* nyttjas när människor övervakar systemets aktivitet övergripande samt beslutar om när och hur ett system ska nyttjas (Europeiska kommissionen, 2019a).

Silberg och Manyika (2019) menar att statistik och definitioner av rättvisa inte är tillräckligt för att ett AI-system ska kunna hantera den sociala omgivningen den nyttjas i, eller för att granska insamlingen av data. En algoritm kan inte själv bestämma när dess bias har minimerats tillräckligt eller i vilka situationer den kan användas helt automatiskt utan en människas ingripande; vilket skapar behovet för en människas bedömning (Silberg & Manyika, 2019). Holzinger (2016) påpekar att ML-algoritmer inte alltid presterar bättre än människor och att det därmed finns ett behov av att inkludera människor i beslutsfattandeprocessen. I resultatet av en undersökning skriven av Daumé et al. (2019) såg man att det fanns en tveksamhet till att rättvisa kunde granskas automatiskt bland anställda inom AI-branschen. Ett lovande alternativ var att vidare utforska möjligheterna med *human-in-the-loop* för att kombinera fördelarna med AI-algoritmer och mänsklig inspektion (Daumé et al, 2019).

2.3.3 Integritet och dataförvaltning

Europeiska kommissionen (2019a) menar att integritet är en rättighet som ett AI-system kan ha en stor påverkan på. Befintliga lagar såsom den allmänna dataskyddsförordningen är verkställd för att bevara denna rättighet (Europeiska kommissionen, 2019a). Europeiska kommissionen (2019) menar att för inte skada integriteten krävs det att man säkerställer en korrekt dataförvaltning, datas relevans till användningsområdet samt förmåga att behandla data och samtidigt skydda integriteten. Vidare menar europeiska kommissionen (2019) att AI-system måste garantera ett skydd av individers integritet och uppgifter under hela sin livscykel. AI-system kan dra flera slutsatser om en person utifrån den information som matas in och personer som interagerar med systemet måste kunna garanteras att data som samlas in om dem inte nyttjas för att diskriminera individen (Europeiska kommissionen, 2019a). Data riskerar att innehålla sociala konstruktioners misstag och bias vilket är en anledning till att datasets måste kvalitetssäkras innan de bearbetas av ett AI-system under träningsperioden. Sjävlärande system är särskilt känsliga mot snedvridna data (Europeiska kommissionen, 2019a).

Enligt Europeiska kommissionen (2019a) är en lösning på detta att ständigt testa de processer och datasets som används i AI-system, för varje steg som tas. Detta anser de ska ske för alla AI-system och inte bara de som utvecklats internt (Europeiska kommissionen, 2019a).

2.3.4 *Transparens*

Transparens utgörs av spårbarhet, förklarbarhet samt kommunikation i den data, system och affärsmodell som utgör ett AI-system (Europeiska kommissionen, 2019a). Spårbarhet möjliggörs genom att dokumentera datasets, processer och de beslut ett AI-system kommer fram till. Spårbarheten kan underlätta granskningen och förklarbarheten av ett system genom att det möjliggör chansen att kunna avgöra orsaken till ett felaktigt beslut (Europeiska kommissionen, 2019a).

Förklarbarhet innefattar att kunna förklara AI-systemets processer samt tillämpningsområden (Europeiska kommissionen, 2019a). Teknisk förklarbarhet innebär att personer ska ha möjlighet att förstå och spåra besluten ett AI-system har fattat (Europeiska kommissionen, 2019a). Används AI inom ett område där beslutsfattandet har en stor påverkan på en människas liv bör man få en förklaring av systemet, menar Europeiska kommissionen (2019a). För en transparent affärsmodell bör inblandade få förklarat i vilken grad ett AI-system har en inverkan på ett beslut, samt hur systemet har valts till att utformas samt varför systemet används (Europeiska kommissionen, 2019a). Utan förklarbarhet kan det vara svårt att förutse när felaktiga förutsägelser görs. Resonemanget bakom ett visst resultat riskerar att döljas i en *black-box*, på grund av dess avancerade karaktär (Knight, 2017). *Black-box* är konsekvensen av att inte veta exakt hur AI:n har kommit fram till ett beslut. Vilket vidare kan innebära att det uppkommit bias utan att utvecklarna har upptäckt detta, då beslutet inte går att härleda fram (Houser, 2019).

Enligt Adadi och Berrada (2018) finns det rent tekniskt ingen etablerad definition av förklarbar AI då det inte har funnits ett extremt behov av det förrän ML-tekniker började utvecklas. Detta behov har även ökat för att undvika att resultat med inbäddad bias ska påverka människors möjligheter eller liv (Adadi & Berrada, 2018). Adadi och Berrada (2018) menar att med hjälp av förklarbara system kan resultat rättfärdigas även om de är oväntade. Systemen blir även granskningsbara och det går att bevisa att resultaten är etiska och rättvisa (Adadi & Berrada, 2018).

2.3.5 *Mångfald, icke-diskriminering och rättvisa*

Europeiska kommissionen (2019a) menar att datasets kan försämrats av att dess historik har inbakade fördomar eller är att de är ofullständiga. Genom att inte hantera dessa så riskerar specifika grupper av människor att utsättas för diskriminering (Europeiska kommissionen, 2019a), vilket beskrivs mer ingående i föregående kapitel; AI-bias. De snedvridna och diskriminerande data eller processer som går att upptäcka, bör hanteras under insamlingen av data (Europeiska kommissionen, 2019a). Europeiska kommissionen (2019) menar att det inte bara är data som kan påverkas av bias utan även metoder för att utveckla systemen i sig. Detta kan motverkas förslagsvis genom att införa tillsynsprocesser så att systemets syfte, krav, beslut och begränsningar kan hanteras på ett transparent vis (Europeiska kommissionen, 2019a).

Enligt Europeiska kommissionen (2019a) bör företag anställa personer med olika kulturer och bakgrunder för att garantera nyanserade synvinklar. I en rapport skriven av West, Whittaker och Crawford (2019) delges en bild av dagsläget som visar på en bristande mångfald inom AI-sektorn. Statistiken från USA visar hur mer än 80% av alla professorer inom AI är män och på företagen Google och Facebook utgör kvinnor 10 % och 15 % av de anställda inom AI-forskningsavdelningen (West, Whittaker & Crawford, 2019). På Facebook, Google och Microsoft består mindre än 10% av medarbetarna av latino eller afroamerikaner (West, Whittaker & Crawford, 2019). Utöver detta menar författarna att den stora majoriteten av forskning inom AI ser kön som något binärt utifrån stereotyper vilket får historisk ojämställdhet att fortleva (West, Whittaker & Crawford, 2019). Detta överensstämmer med Farbergers uttalande i Maria Phillips (2021) artikel om näringslivet, att strukturer är skapade och anpassade för endast personer inom normen, vilket är svårt att förstå för dessa personer. Denna insikt menar Farberger att hon inte kunde få själv förrän hon genomgick en könskorrigering (Phillips, 2021). West, Whittaker, och Crawford (2019) poängterar att en bristande mångfald påverkar hur företag som dessa är utformade, vilka produkter som produceras och hur den tilltänkta målgruppen ser ut. Ett av förslagen för att hantera denna påverkan menar West, Whittaker, och Crawford (2019) är att det inte är tillräckligt att endast se över ifall människor är inkluderade i processer med AI-element, utan även vilka specifika personer det är.

Vidare menar Europeiska kommissionen (2019a) att bör AI-system utformas så de kan nyttjas universellt och personer ska inte uteslutas på grund av kön, ålder eller liknande. Dessa system ska levereras till en så stor målgrupp som möjligt och därmed inte anta att en lösning passar alla potentiella användare (Europeiska kommissionen, 2019a). Ett sätt att undvika detta är att inkludera relevanta parter och överlägga med dessa under hela systemets livscykel (Europeiska kommissionen, 2019a).

2.3.6 *Samhällets och miljöns välbefinnande*

Enligt Europeiska kommissionen (2019a) kan AI-system användas för att hantera samhällsproblem och de uppmuntras till att ha ett inkluderat hållbarhetstänk. En bedömning bör även göras över påverkan på samhället i stort menar författarna (2019a). Zhao (2018) beskriver hur termen *corporate social responsibility*, CSR, myntades vid 1950-talet och att genom att endast prioritera sina egenintressen så riskerar företag att utsättas för kritik, exempelvis som Facebook fick bemöta år 2018 efter nyheterna om hur plattformens användares personliga data utnyttjades. Genom att företag uppfyller sitt sociala ansvar förbättras ett företags rykte, vilket även kan vara fördelaktigt för företagets position på marknaden (Zhao, 2018). Gällande artificiell intelligens, som är en såpass ny teknik och utforskat område, är det sociala ansvaret stort (Zhao, 2018). Givet de områden artificiell intelligens är verksamt inom är socialt ansvar oundvikligt (Zhao, 2018).

2.3.7 *Ansvarsskyldighet*

Algoritmer, data och processen bakom utformningen av ett AI-system ska vara möjligt att granska (Europeiska kommissionen, 2019a). Om systemets tillämpningsområde kan kopplas till grundläggande rättigheter bör det vara möjligt att genomföra en oberoende granskning av systemet i fråga (Europeiska kommissionen, 2019a).

Europeiska kommissionen(2019a) menar att en åtgärd eller beslut som bidrar till ett visst resultat i systemet, måste vara möjligt att rapportera. Potentiella negativa konsekvenser av ett AI-system bör identifieras, bedömas och minimeras (Europeiska kommissionen, 2019a). Riskminimering kan ske genom konsekvensbedömning under utveckling, spridning och användning till systemet (Europeiska kommissionen, 2019a).

Europeiska kommissionen (2019a) menar vidare att utvecklingsprocessen kan leda till kompromisser. De kompromisser som uppstår bör utvärderas utifrån deras risk för etiska principer såsom grundläggande rättigheter (Europeiska kommissionen, 2019a). Fortsättningsvis menar Europeiska kommissionen (2019a) att systemet inte bör fortsätta utformas i en riktning där det inte är möjligt att göra etiskt acceptabla kompromisser. De kompromisser som görs bör vara motiverade och väl dokumenterade (Europeiska kommissionen, 2019a).

2.4 Sammanfattning av ramverk med koppling till litteratur

Tabell 1: Ramverk med koppling till litteratur

Huvudområde/tema	Aspekter/faktorer/teori	Stödande litteratur
AI	Definitioner av AI	Dignum (2020); Benbya, Pachidi & Jarvenpaa (2021); Europeiska kommissionen (2019b); Europeiska kommissionen (2018)
	Maskininlärning	Jordan & Mitchell (2015); Alpaydin (2010); Rai & Sing (2010); IBM Cloud Education (2020); Otter et al. (2021); Bogen & Rieke (2018); Google Cloud (n.d); Metz (2019)
	AI-bias	Ntoutsis et al. (2020); Hellström, Dignum & Bensch (2020); Houser (2019); Dattner et al., (2019)
AI lösningar	Mänskligt agentskap och mänsklig tillsyn	Dataskyddsförordningen (2016); Europeiska kommissionen (2019a); Silberg och Manyika (2019); Holzinger (2016); Daumé et al. (2019)
	Integritet och dataförvaltning	Europeiska kommissionen (2019a)
	Transparens	Europeiska kommissionen (2019a); Knight (2017); Adadi & Berrada (2018); Houser (2019)
	Mångfald, icke-diskriminering och rättvisa	Europeiska kommissionen (2019a); West, Whittaker, & Crawford (2019); Phillips (2021)
	Samhällets och miljöns välbefinnande	Europeiska kommissionen (2019a); Zhao, W. (2018)
	Ansvarsskyldighet	Europeiska kommissionen (2019a)

3 Metod

Under detta kapitel presenteras först den valda metoden och strategin. Därefter beskrivs datainsamlingsmetod, urvalet, etiska överväganden och dataanalys. Kapitlet avslutas med en diskussion kring studiens kvalitet i form av reliabilitet och validitet och en metodreflektion.

3.1 Forskningsstrategi och forskningsansats

Distinktionen mellan kvantitativ- och kvalitativ forskning är enligt Bryman (2018) att tonvikten i kvalitativ forskning oftast ligger på ord istället för siffror vid insamling och analys av data. Vidare föredrar kvalitativ forskningsstrategi att lägga vikten på hur individer uppfattar och tolkar sin sociala verklighet (Bryman, 2018). Med hänsyn till studiens syfte, att undersöka hur företag som tillhandahåller automatiserade rekryteringstjänster med inslag av AI, arbetar för att minska bias inom AI samt belysa de åtgärder och arbete som görs för att i sin tur minska risken för att bias uppstår i tjänsterna, kommer fokus i studien att vara diskussioner baserat på respondenternas egna ord, därav valdes den kvalitativa forskningsstrategin. Vidare existerar inte många företag i Sverige som tillhandahåller automatiserade rekryteringstjänster med inslag av AI, vilket i sin tur medför att det hade varit svårt att göra en kvantitativ studie då det ofta krävs många respondenter vid kvantitativ studie för att få en trovärdig slutsats. Graziano och Raulin (2013) menar vidare att det finns en skillnad vid planeringen mellan kvalitativ- och kvantitativ forskningsstrategi då planeringen är mer flytande och flexibel i kvalitativ forskningsstrategi och således kan processen ständigt ändras eller modifieras. Då vårt forskningsområde ständigt förändras och också är förhållandevis nytt har det varit en viktig aspekt vid val av forskningsstrategi att ha möjlighet kunna modifiera såväl frågeställning som innehåll för att anpassa till nuvarande samhällsutveckling (Bryman, 2028).

Bryman (2018) menar att med kvalitativa metoder kan studien utgå från en deduktiv eller induktiv forskningsansats. Deduktiv forskningsansats innebär att studien utgår ifrån en teori för att i sin tur skapa en hypotes som sedan undersöks empiriskt om den stämmer överens med verkligheten (Bryman, 2018). Utifrån studiens resultat formas en teori som antingen förblir som den tidigare har varit, alternativt att den förändras utifrån resultatet (Bryman, 2018). Denna studie utgår från en deduktiv forskningsansats eftersom studien jämför det empiriska materialet med den teoretiska referensramen.

3.2 Litteraturstudie

En litteraturstudie gjordes inledningsvis för att få en förståelse för vilka teorier och modeller som är relevanta för uppsatsen. Detta utgjorde i sin tur grunden för själva teoridelen vilket utmynnade i den intervjuguide som utformats. Litteraturstudien inleddes med att leta efter relevanta teorier och information kring ämnet "bias i AI" och "AI-bias inom rekrytering". Då en av studiens skribenter hade läst kursen *Digitalisering och AI ur ett organisation och samhällsperspektiv* på Ekonomihögskolan vid Lunds Universitet och fann kursens innehåll relevant och lärorik, påbörjades sökandet efter information utifrån kursens litteraturlista. Vidare identifierades flertalet relevanta artiklar genom diskussion och hjälp av uppsatsens handledare, Benjamin Weaver.

De tillhandahållna artiklarnas relevanta referenser söktes sedan upp via Google Scholar och Lunds universitetbibliotekstjänst LUBsearch. Dessa artiklar kompletterades i sin tur med annan litteratur som hittades genom AIS, LUBsearch och Google Scholar där vi filtrerade resultatet utifrån olika avgränsningar. Vi försökte i största möjliga mån avgränsa oss till artiklar som var peer-reviewed och publicerade inom ett tidsspänn där informationen om vårt ämne fortfarande är relevant. För att hitta artiklarna som kunde besvara studiens frågeställning användes bland annat sökord såsom “AI-bias”, “automated recruitment process”, “bias in artificial intelligence”, “human-in-the-loop”, “artificial intelligence in hiring process”, “ethical AI”, “natural language processing”, “trustworthy artificial intelligence”, “big data” och “discriminating system”. Utefter sökresultatet valdes artiklarna, dels utifrån antalet citeringar på Google Scholar, dels efter att ha läst sammanfattningarna av artiklarna då flertalet av dessa publicerades nyligen och därmed inte hunnit bli citerade flertalet gånger ännu. Ännu en följd av att forskningsområdet är nytt är att det inte finns stora mängder publicerade artiklar om ämnet. Därav har vi även använt oss av rapporter, uttalande från experter inom området och nyhetsartiklar i bakgrunden och problemformuleringen. Litteraturen har i sin tur legat till grund för utformningen av frågorna som ställdes vid intervju tillfällena.

3.3 Urval

För att besvara studiens frågeställning har vi använt oss av ett *målstyrt urval* som är den grundläggande principen i kvalitativa studier (Bryman, 2018). *Målstyrt urval* innebär att forskaren strategiskt försöker välja ut deltagare så att de är relevanta för de forskningsfrågor som formulerats, i detta fall företag som tillhandahåller automatiserade rekryteringstjänster med inslag av AI (Bryman, 2018). För att få ett bredare perspektiv på frågan har vi även ett företag från kundsidan intervjuats. Tillvägagångssättet vid val av urval var en blandning mellan urval av *typiskt fall*, som innebär att vi valt företag som kan exemplifiera den kategori företag vi undersöker, samt *kriteriestyrt urval* där de företag vi valt uppfyller kraven att tillhandahålla tjänster med inslag av AI som stödjer rekryteringsprocesser, är verksamma i Sverige och marknadsför sig med fördomsfri AI (Bryman, 2018). Utöver målstyrt urval, har vi även använt oss av *bekvämlighetsurval* som innebär att respondenterna till viss del består av de personer som fanns tillgängliga (Bryman, 2018). Denscombe (2016) menar att bekvämlighetsurval kan ses som fördelaktigt vid studier där man har begränsad tids- och kostnadsbudget; precis som i denna studies fall där vi endast haft två månader till förfogande och inget utrymme för kostnader. Detta ledde till att ett av de företag som användes som pilotintervju och respondenten från kundföretaget har en anknytning till en i forskningsgruppen. Vidare har vi kontaktat de respondenter på företagen som vi ansett bäst kunnat besvara vår forskningsfråga.

Urvalsprocessen inleddes med att vi sökte efter företag som arbetar med AI och rekrytering genom Googles sökmotor. Sökord som “AI-rekrytering” i kombination med “företag” och “Sverige”, samt “startups” i kombination med “AI-rekrytering” användes för att hitta företagen. Vidare sökte vi även igenom Swedish AI-startup landscapes hemsida, som är en plattform som visar AI-startups i Sverige, för att hitta fler relevanta företag. Även vår handledare kom med förslag på relevanta företag att undersöka. De företag vi ansåg vara relevanta och som även tackade ja var: Devten, Exparang, Hubert och Tengai, vi kontaktade även två företag till som valde att tacka nej. För att få en förståelse för hur rekryteringsbranschen ser ut idag och hur den ställer sig till de automatiserade tjänsterna med

inslag av AI, tog vi även kontakt med Wise Consulting som är ett konsultföretag inom HR. Vi kontaktade alla företag via mail förutom Devten och Wise Consulting där vi hade kontakter inom dessa företagen och därför nådde ut till dem via Facebooks Messenger respektive sms. Alla företag var intresserade av att ställa upp i intervjuer via Google Meet.

Tabell 2: Översikt över intervjuobjekt samt relaterad information

Företag	Roll och relevant erfarenhet	Typ av intervju	Längd på intervju	Datum	Inspelad
Devten	Grundare, datavetenskapsstudent	Google Meet	33:30	19 april 2021	Ja
Exparang	Medgrundare och 12 års erfarenhet av chefsrekrytering	Google Meet	43:38	21 april 2021	Ja
Tengai	Produktägare, 10 års erfarenhet inom rekrytering	Google Meet	37:20	21 april 2021	Ja
Hubert	Grundare, marknadsansvarig	Google Meet	26:12 + 8:00*	22 april 2021 samt 7 maj 2021	Ja
Wise Consulting	Regionchef, Stockholm	Google Meet	22:21	3 maj 2021	Ja

* = Kompletterande intervju med Hubert

3.4 Empirisk datainsamling

För att kunna besvara vår forskningsfråga valdes som tidigare nämnt en kvalitativ intervjustudie för att uppnå ett djup och en god förståelse för ämnet. Bryman (2018) menar att intervjuer oftast är den vanligaste insamlingsmetoden vid kvalitativa studier. Denna studie tar sin utgångspunkt i semistrukturerade intervjuer då vi vill ha svar på frågor inom följande specifika teman; AI-tekniker, bias inom AI, rekryteringsmetoder och EU riktlinjer. I sista delen av intervjun, där vi berör EU-riktlinjerna, finns inslag av en strukturerad intervju utformning, för att på så sätt säkerställa hur företagen arbetar jämfört med EU-riktlinjerna. För att kunna genomföra de semistrukturerade intervjuerna utformade vi en intervjuguide med ett antal frågor som vi skulle förhålla oss till, se Tabell 3. Denna utformades utifrån ramverket för litteraturkopplingar som visas i tabell 1. Bryman (2018) menar att intervjuguiden fungerar som en minneslista över de områden som ska beröras i en intervju.

Vidare menar Bryman (2018) att formuleringen av frågorna inte ska vara för specifika, utan mer öppna, för att inte hindra alternativa idéer eller synsätt från respondenter under intervjuernas gång. De teman som berördes i intervjuguiden kompletterades under intervjutillfället med följdfrågor som anpassades utefter intervjuens utveckling, något som också är ett kännetecken vid en semistrukturerad intervju (Alvehus, 2013). Bryman (2018) menar att ordningsföljden av frågorna i intervjun kan varieras från intervjuguiden vilket det även gjordes under samtliga intervjutillfällen.

Tabell 3: Intervjuguide

Tema/perspektiv	Teori/kategori	Faktorer/variabler	Litteratur	Frågor till tjänsteföretag
Rekryteringsmetoder	<ul style="list-style-type: none"> - Beskrivning av företag - Utformning av AI-tjänst - Tjänstens funktion 	<ul style="list-style-type: none"> - Om företaget 	Black & van Esch (2019)	<ul style="list-style-type: none"> - Vilka är ni och vilken affärsidé har ni? - Vilken roll har du? - Hur definierar ni det ni erbjuder/hur ser er affärsmodell ut? Som en produkt, tjänst, verktyg eller liknande?
Artificiell intelligens	<ul style="list-style-type: none"> - Inlärningsmetoder - Maskininläring - Bias i AI 	<ul style="list-style-type: none"> - Natural Language Processing 	Alpaydin (2010); Jordan & Mitchell (2015); Rai & Sing (2010); Ntoutsis et al., (2020); Hellström, Dignum och Bensch (2020); Houser (2019)	<ul style="list-style-type: none"> - Vad använder ni för AI tekniker för er tjänst? - Vilken typ av inlärningsmetod använder ni? - Hur gör ni för att säkerställa minimalt med bias? - Har ni stött på några utmaningar med att bias uppstår i er produkt? Har detta varit något ni har fått hantera? - Vad anser ni är den mest kritiska processerna/data som kan bidra till bias i ert verktyg? Varför? - Vilken roll spelar kundens tidigare data in i bedömningen? - Får kunden en chans att lära sig vad denne kan göra för att motverka bias? Hur? - Hur delaktig är kunden i algoritmens utformning? - För- och nackdelar med ert system/verktyg när det kommer till bias?
EU-kommissionens riktlinjer	<ul style="list-style-type: none"> - Lösningar för etisk AI 	<ul style="list-style-type: none"> - Mänskligt agentskap och mänsklig tillsyn - Integritet och dataförvaltning - Transparens - Mångfald, icke-diskriminering och rättvisa - Samhällets och miljöns välbefinnande - Ansvarsskyldighet 	Europeiska kommissionen (2019a); Dataskyddsförordningen (2016); Holzinger (2016); Daumé et al. (2019); Knight (2017); Adadi och Berrada (2018); West, Whittaker, &	<ul style="list-style-type: none"> - Vilken tillsyn och insyn har kunden i tjänsten? - Får kunden en chans att lära sig vad denne kan göra för att motverka bias? Hur? - Är processen helt automatiserad eller är human-in-the-loop (människa medverkar i varje beslutscykel i systemet) en del av eller liknande integrerat? - Hur kontrolleras att systemet gör korrekta förutsägelser, bedömningar? Presenteras denna statistik för kunder/kandidater? - Hur säkerställer ni att in-parametrarna eller inmatningsdata inte bidrar till bias? - Hur tillgång ni data som används vid bedömningen av en kandidat? - Hur arbetar/ser ni på transparens? - Kan systemet förklara/visa hur det kommit fram till ett beslut? - Kan man spåra systemets förlopp fram till beslutet?

			Crawford (2019); Zhao, W. (2018)	<ul style="list-style-type: none"> - Hur informeras användaren om att det är AI som gör rekommendationen? Om så är fallet? - Hur ser mångfalden ut i era utvecklingsteam/tränings-data? - Hur främjar ni mångfald i ert arbete? - Hur kompromissar ni mellan olika tekniska lösningar? Vilken resurs prioriteras? - Finns det någon utpräglad process hos er hur ni går tillväga om man misstänker problem med algoritmen? - Hur ser ni på ert framtida arbete med bias-frågan?
--	--	--	-------------------------------------	---

Frågorna skickades till respondenterna innan intervjutillfället för att de skulle vara insatta i ämnet och därmed få intervjutillfället så tidseffektivt som möjligt. Beroende på kunskap och intresse hos respondenten varierade fokuset på respektive tema mellan de olika intervjuerna. Samtliga intervjuer varade i cirka 35 minuter och genomfördes digitalt via Google Meet eller Microsoft Teams. Tre intervjuer kompletterades med följdfrågor via mail för att få svar på ytterligare frågor som uppkom under transkriberingen. Jacobsen (2002) menar att tillförlitligheten på svaren påverkas av miljön där intervjun utförs. En obekant miljö kan göra att respondenten svarar på ett annorlunda sätt än vid en naturlig miljö (Jacobsen, 2002). I och med Corona-pandemin antar vi att intervjupersonerna är vana vid digitala möten och kombinationen av detta, tillsammans med att respondenterna befinner sig i en miljö som de själva har valt, kan ha påverkat studiens resultat positivt.

Under respektive intervju hade vi olika ansvarsuppgifter. En var ansvarig för att leda intervjun genom att dels säkerställa att intervjun levde upp till de etiska principerna men också för att intervjuguidens teman besvarades av respondenten. Den andra var ansvarig för att markera vilka frågor som hade besvarats från respondenten samt spela in samtalet och den tredje var ansvarig för att föra anteckningar. Under alla intervjutillfällena tilläts alla att ställa följdfrågor för att på så sätt säkerställa så kvalitativa svar från respondenten som möjligt.

3.4.1 Etiska överväganden

Vid studier bör man enligt Bryman (2018) eftersträva följande fyra etiska principer; informationskravet, samtyckeskravet, konfidentialitetskravet och nyttjandekravet. Informationskravet innebär att forskaren ska informera deltagaren om studiens syfte, att deltagandet i studien är frivilligt samt att de moment som ingår i intervjun ska presenteras (Bryman, 2018). I studien formulerades studiens syfte i det inledande mail till deltagandet och frågorna skickades via mail till deltagarna innan intervjutillfället för att de skulle vara helt införstådda med intervjuens innehåll. Vidare presenterades även intervjuens struktur vid början av respektive intervju. Samtyckeskravet innebär att deltagaren har rätt att bestämma över sin egen medverkan, vilket i studien har beaktats genom att deltagarna har tackat ja till att delta och när som helst kan välja att inte längre medverka. Konfidentialitetskravet innebär att uppgifterna om de som deltar i intervjuerna ska behandlas med största möjliga konfidentialitet och att respondenten ska informeras vem som kommer tillgå information och hur den kommer hanteras (Bryman, 2018). Vid intervjutillfället har respondenterna tillfrågats angående om de vill vara anonyma, hur informationen de lämnar kommer användas samt vart studien kommer publiceras. Samtliga respondenter har tillåtit detta och har inte heller velat vara anonyma. Det

sista kravet är nyttjandekravet som innebär att de uppgifter som samlas in om enskilda personer endast får användas för forskningsändamålet, vilket har tagits hänsyn till under hela studiens gång (Bryman, 2018).

3.5 Dataanalys

Vid kvalitativa studier är det vanligt att stora mängder ostrukturerat textmaterial ska analyseras (Bryman, 2018). För att strukturera upp studiens analys har vi utgått från Jacobsens (2002) tre steg som kan användas vid analysprocesser. De tre stegen är; sammanställning, kategorisering av material och kombinerande (Jacobsens, 2002). För att sammanställa allt material och för att få ett så detaljerat underlag som möjligt valde vi att transkribera alla intervjuer efter respektive intervjutillfälle. Transkriberingsprotokollen lades sedan in i ett Google Sheet dokument där respektive transkribering fick ett eget blad. Det andra steget i Jacobsens teori (2002) innebär att det sammanställda materialet ska kategoriseras och på så sätt struktureras för analys. Vi valde att färgmarkera meningar och ord i de transkriberade materialet utifrån de huvudområden som vi identifierade i vår sammanfattning av ramverk med koppling till litteratur som i sin tur låg till grund för våra intervjufrågor. Vidare delade vi in huvudområdena med olika aspekter som i sin tur fick olika ID:n för att på så sätt strukturera upp materialet ännu tydligare inför analys. Nedan kodtabell visar hur kodningen är uppbyggd utefter färger och ID. Kodningen utformning sammanställs i tabell 4.

Tabell 4: Översikt för färgkodning och ID-taggar

Huvudområde	Färg	ID	Aspekt
AI	Lila	A1	AI-tekniker
		A2	Bias inom AI
Rekryteringstjänster	Blå	R1	Företagens tjänster
AI-lösningar	Grön	L1	Mänskligt agentskap och mänsklig tillsyn
		L2	Integritet och dataförvaltning
		L3	Transparens
		L4	Mångfald och icke-diskriminerande och rättvisa
		L5	Samhället och miljön välbefinnande
		L6	Ansvarsskyldighet

Nedan är ett exempel från ett transkriberingsprotokoll och hur vi har färgkodat samt ID-taggat ett svar från respondenten.

Tabell 5: Utdrag från den transkriberade intervjun med Tengai, se appendix B

23	Analyserar den på något vis de värde som finns i ord? När det kommer till manligt och kvinnligt språk eller att man kan bedöma ord på olika sätt beroende på vilken person man är och koppling.	
24	Jag skulle säga att nå inget aktivt, därför att det är väldigt simpelt på positivt och negativt. Vi har ju medvetet gjort det på det sättet för att det blir väldigt enkelt att bygga det på det sättet, för du behöver bara avgöra om det är ett positivt eller negativt svar och där blir skillnaderna mycket mindre. Jag skulle tippa på min egen fördom att manligt och kvinnligt kommer in när man börjar beskriva saker på ett annat sätt och det gör vi inte i dagsläget vi bedömer ju inte det så.	A2, L2

Efter att kategoriserat materialet utifrån de olika huvudområdena och aspekterna, påbörjade vi Jacobsens (2002) sista steg och jämförde materialet med varandra och med litteraturen. Viktigt att poängtera är de tolkningarna som gjordes i empirin gällande vilken typ av bias som förekommer i tjänsten utgår från våra egna uppfattningar av biasernas innebörd jämfört med vad som diskuterades under intervjutillfällena. Således kan fler biaser existera i företagen som inte kunde identifieras utifrån samtalsämnena under intervjun.

3.6 Kvalitetsmått

Nedan följer en reflektion kring studiens validitet och reliabilitet.

3.6.1 Validitet

Ejvegård (2003) menar att validitet innebär hur bra man mäter det som är avsett att mätas. Vidare menar Hammond (1995) att en metod måste vara reliabel för att vara valid. Ett exempel på detta kan vara om en intervjufråga missuppfattas och reliabiliteten därmed blir låg och missuppfattningen leder i sin tur till att frågan inte mäter det som den är ämnad att mäta, vilket resulterar i en försämrad validitet. Fördelen med att vi använt semistrukturerade intervjuer är möjligheten till att ställa följdfrågor för att på så sätt säkerställa att respondenten har uppfattat allt korrekt. För att undersöka om frågorna är rimliga i relation till det vi vill mäta, har vi genomfört en pilotintervju med Devten för att i sin tur kunna förfinas frågorna. Att samtliga respondenter dessutom varit relevanta för studiens syfte och haft en stor kunskap om deras produkt då två varit grundare till företagen och den tredje har rollen som produktägare och varit med sedan starten av företaget, kan också styrka validiteten.

Bryman (2018) menar att det finns en viss problematik vid kvalitativa studier att uppnå en hög extern validitet. Detta då det ofta krävs ett stort urval för att på så sätt få tillräckligt med svar och insikter som sedan kan generaliseras utanför aktuell studie (Bryman, 2018). I och med att studien endast undersöker fyra företag inom det valda området, där företagen dessutom utvecklar olika typer av rekryteringstjänster, är vi medvetna om att studiens externa validitet är låg.

3.6.2 Reliabilitet

LeCompte och Goetz (1982) skiljer på begreppen intern och extern reliabilitet. Enligt författarna innebär extern reliabilitet i vilken utsträckningen en undersökning kan replikeras (LeCompte & Goetz, 1982). Vidare menar författarna att detta ofta kan vara svårt att uppnå i kvalitativa studier då de ofta genomförs i en social miljö som är svår att exakt återskapa. Att

studien dessutom bygger på semistrukturerade intervjuer gör studien ännu svårare att återskapa då det är respondenten som till största del styr samtalet (LeCompte & Goetz, 1982). Genom att bifoga intervjuguiden, se appendix A, styrks reliabiliteten eftersom det visar en del av studiens tillvägagångssätt. Även metodkapitlet stärker reliabiliteten då det beskriver studiens tillvägagångssätt på ett transparent sätt, vilket vi har försökt att åstadkomma efter bästa förmåga (Yin, 2013). Försättningsvis menar Yin (2013) att intern reliabilitet innebär att forskarlaget har kommit överens om hur information ska tolkas. För att stärka den interna reliabiliteten kom vi därför överens om att intervjuerna ska tolkas utifrån huvudområdena och aspekterna som återfinns i tabell 4. Dock är det viktigt att ha i åtanke att vi varit tre personer som analyserat intervjuerna, vilket kan ha medfört vissa missuppfattningar då vi kan ha tolkat samma svar på olika sätt. Som tidigare nämnt finns en risk att respondenterna missuppfattar frågorna, vilket leder till att respondenterna besvarar frågorna på olika sätt vilket leder till lägre reliabilitet. För att undvika detta har vi även försökt att skapa frågor som inte är komplicerade.

3.7 Metodreflektion

Vidare har ingen av studiens skribenter genomfört en studie av denna storlek tidigare vilket bland annat kan ha påverkat genomförandet av intervjuerna. Bryman (2018) menar att ett sätt att förbättra intervjuernas kvalitet är genom att ha fler pilotintervjuer. Fler pilotintervjuer hade även kunnat ge oss fler viktiga insikter till vår intervjuguide, vilket i sin tur hade kunnat leda till ett bättre resultat. Vidare hade vi isåfall inte behövt skicka följdfrågor till företagen via mail efter intervjutillfället. När vi skulle presentera intervjuernas resultat insåg vi att temat bias i intervjuguiden hade behövt kompletteras med fler ingående frågor kopplat till litteraturgenomgången. Detta för att på ett enklare sätt kunna presentera vilka olika typer av biaser som riskerar att uppstå i tjänsterna. Istället fick vi göra egna tolkningar av vilken bias som uppstår utifrån respondenternas svar, vilket gör att punkt 4.3.2.1, *Identifierande bias*, innehåller våra konstaterande trots att det innefattas i empiri-kapitlet. Detta ansåg vi vara nödvändigt för att sedan kunna diskutera de olika formerna av bias.

I och med att vi spelade in intervjuerna, efter godkännande från respondenten, finns också en risk enligt Bryman (2018) för reaktiva effekter som kan uppstå när en person är medveten om att denne blir observerad, vilket kan leda till att den därmed beter sig mindre naturligt. Dock anser vi att fördelarna med transkribering överväger nackdelarna och dessutom upplevde vi att respondenterna inte hade några problem med att bli inspelad.

Vidare kan storleken på urvalet diskuteras då vi endast genomfört fyra intervjuer och en pilotintervju. Å andra sidan har vi kontaktat alla de företagen i Sverige som vi hittat som tillhandahåller automatiserade rekryteringstjänster med inslag av AI, varav det endast är två företag som inte svarat och därmed inte tackat ja. Därav hade vi haft svårt att få ett större urval då det inte existerar fler företag att kontakta i vår vetskap. I och med att urvalet inte är så stort valde vi att ta med vår pilotintervju för att få mer empiriskt material. Det är då viktigt att ha i åtanke att Devten är ett start-up och inte kommit så långt i sin verksamhetsutveckling, därav är inte heller materialet lika omfattande inom vissa teman som från resterande företag.

Slutligen är det viktigt att ha i beaktning att vi är en homogen grupp som genomfört studien vilket också kan medföra en risk att vår tolkning och analys av såväl teori som empiri inte inkluderar vissa infallsvinklar, som hade kunnat vara intressanta när just ämnet bias behandlas.

4 Empiri

Detta kapitel inleds med resultatet från en intervju med respondenten på ett rekryteringsbolag. Resterande empiri är strukturerad utifrån intervjuguiden och bearbetar en respondents svar inom varje område i taget. Empirin innehåller tabeller under respektive delkapitel som visar i vilken rad i transkriberingsprotokollet respondenter har besvarat området. Först presenteras företagets tjänster och vidare belyses bias inom AI. Vi har inkluderat en punkt där vi tolkar vilka biaser som uppstår i tjänsterna. Detta görs för att underbygga diskussionen om var bias inom AI uppstår. Resterande empiri består av respondenternas svar kring frågor relaterade till Europeiska kommissionens riktlinjer.

4.1 Expertintervju med Wise Consulting AB

Wise Consulting ingår i Sveriges största HR-koncern och erbjuder tjänster inom HR-konsultering, ledarutveckling och coachning (Wise Consulting:2). Enligt Wise Consultings respondent använder företaget inga automatiserade tjänster förutom systemstöd och menar att de är lite efter på den punkten (Wise Consulting:4). Om automatiserade tjänster blir relevant hade urvalsprocessen varit det rekryteringsmoment som Wise Consulting behövt hjälp med. Detta för att de söker specifika kriterier och egenskaper hos de potentiella kandidaterna, särskilt förmågan att kunna vara en bra konsult (Wise Consulting: 6). Under en rekrytering finns flera effekter som kan orsaka en fördomsfull process; egna tyckande, förutfattade meningar, vem som rekryterar och dennes magkänsla (Wise Consulting:8). Intervjun behöver vara strukturerad så den inte går över till ett vanligt samtal (Wise Consulting:20). Urvalsprocessen har därför inkluderat fler tester för att få bort fördomar som påverkar (Wise Consulting:10). På Wise Consulting nyttjas även personlighetstester i former av djupintervjuer för att förstå hur en kandidat faktiskt gör saker vilket blir ett tydligare mått på kandidatens arbetssätt (Wise Consulting:18).

Att använda AI i framtiden diskuteras och är något som några systerbolag redan börjat nyttja (Wise Consulting:22). Respondenten tror företaget kommer att nyttja detta vid själva urvalsprocessen i framtiden (Wise Consulting:24). Det är viktigt med en kombination av automatisering och handpåläggning men att en automatiserad urvalsprocess hade minimerat risken för fördomar (Wise Consulting:26). Med automatiserade processer finns även en risk för att kandidaters potential inte uppdagas. I vissa fall kan en kandidat som möjligtvis inte har exakt alla kompetenser för rollen vara den mest intressanta. Även om rekryterare ofta vill ha en kandidat som fullständigt matchar kravprofilen så är det inte givet att alla kandidater söker liknande jobb som de tidigare haft, utan vill byta riktning i sin karriär (Wise Consulting:28).

4.2 Företagets tjänster

Nedan presenteras kortare introduktioner till företagets tjänster. Detta för att ha en insikt i hur tjänsterna är utformade och fungerar, vilket vi anser kommer underlätta den vidare läsningen om bias inom AI och hur företagen förhåller sig till Europeiska kommissionens riktlinjer.

Tabell 6: Översikt för hämtade data av företagens tjänster

Företagens tjänster	Tengai	Exparang	Hubert	Devten
R1	4, 18, 20, 26	2, 4, 6, 8, 14, 24, 40, 6b	2, 6, 14, 16, 18, 52, 56	2, 8, 16, 18, 27, 43

4.2.1 Tengais tjänst

Tengais tjänst är en intervjurobot som genomför det första personliga mötet med kandidaten under rekryteringsprocessen. Intervjun består av totalt 20 grundfrågor med tillkommande följdfrågor som baseras på egenskaperna i *Big five*-modellen, som är den vanligaste modellen för att bedöma personligheter (Sjöberg & Sjöberg, 2020). Frågorna har tagits fram och bearbetats av två forskare vid Stockholms universitet, som har en stor kännedom om *Big five*-modellen, för att anpassas till roboten (Tengai:18,20). Detta har gjorts genom att utgå ifrån de egenskaper som vetenskapligt kan kopplas till arbetsprestation (Tengai:26)

Kunden kan inför intervjun vikta de relevanta egenskaperna utifrån vad de anser är viktigast för tjänsten som ska tillsättas (Tengai:18). 16 av dessa frågor är påstående och anpassade följdfrågor ställs utifrån huruvida ett svar tolkas positivt eller negativt. Till sist ställs fyra öppna, kompetensbaserade frågor som spelas in. Det sker två bedömningar av kandidaten; en automatisk som genereras av att kandidaternas svar poängsätts mellan ett till fyra av Tengai, samt en manuell då kunden får lyssna på de inspelade svaren och göra en egen bedömning (Tengai:18). Tengais automatiska poängsättning möjliggörs av forskningen om personlighetstyper i *Big five*-modellen. Den automatiserade poängsättningen lägger grund för Tengais slutgiltiga rekommendation av en kandidat (Tengai:4). Justeringar av viktningen kan även ske efter intervjutillfällena och kandidaterna tilldelas en ny score (Tengai:18).

4.2.2 Exparangs tjänst

Exparang tillhandahåller en matchningstjänst med anonyma kandidater. Exparang har nischat sig till att underlätta företags chefsrekryteringar. Syftet med anonymiseringen är att skapa bättre matchningar där kandidaterna kan vara mer ärliga (Exparang:2,4). Exparang arbetar med att bjuda in kandidater för att utöka databasen. Deras tjänst minimerar sourcing-processen från fyra veckor till 48h samt minskar kostnader (Exparang:6). Kandidater registrerar sig på hemsidan och bifogar CV samt länk till sin LinkedIn-profil, profilen skapas från den informationen och lagras i deras databas. Profilkortet skapar ett anonymiserat CV (Exparang:8,6b). Matchningen sker genom att kandidaterna fyller i ett 50-tal *items* i ett formulär och företagen skapar en annons samt svarar på frågor relaterat till det formulär kandidaterna fyllt i. Exparang gör en viktad poängsättning av formuläret utifrån en subjektiv bedömning. Ett exempel på denna viktning är att en viss kompetens kan vara viktigare för en viss tjänst än att kandidaten kan tala ett språk flytande (Exparang:24,14). Överstiger kandidatens poäng en viss procent sker en matchning automatiskt, där sedan ett första möte kan ske och kandidaten kan visa sin identitet (Exparang:40).

4.2.3 Huberts tjänst

Hubert tillhandahåller en chat-bot som genomför intervjun med kandidaten. Kunderna till Hubert är större bemannings- eller rekryteringsbolag samt snabbväxande startups (Hubert:2) Hubert inleder processen med kunden genom att ta fram en testintervju baserade på existerande frågor från Hubert och kunden samt kompetenser för rollen. Utifrån dessa frågor och kompetenser framtas de anpassade frågorna till chatt-interfacet. De frågorna som ställs av chat-botten samt viktningen av kriterierna görs alltid i samförstånd med kunden. Viktningen av kriterierna baseras på en skala mellan noll och fem (Hubert: 14,16,18). Frågor som kräver längre utläggningar bearbetas genom att leta efter *star-svar*; det vill säga svar som är strukturerade utifrån *situation, task, action, result* (Hubert:56). Kunden har översikt över kandidaterna genom ett gränssnitt som Hubert tillhandahåller, där kunden kan se en sorterad lista över kandidaternas ranking under hela processen (Hubert:6). I detta interface kan kunden se kandidatens totala ranking och hur kandidaten har blivit rankad inom olika kriterier (Hubert:52).

4.2.4 Devtens tjänst

Devtens tjänst går ut på att matcha företag och anonyma utvecklare. Syftet är att ge en mer transparent arbetsbeskrivning genom att rekryteraren får mata in en bredare variation av in-parametrar (Devten:2). Fördelen för kunden är att mer relevanta kandidater dyker upp och för utvecklarens del presenteras mer relevanta erbjudanden, då en sökning görs med hela tjänsten som ett objekt och inte endast ord i en sökterm (Devten: 8). Det viktigaste för en kandidat att specificera är de programmeringsspråk och talat språk denne besitter och från rekryterarens sida är lön och plats viktigast (Devten:18). Den information användare delger hanteras av Googles NLP (Devten:16). För att matcha kandidater använder Devten sig av en enkel algoritm som är byggd på att iterera igenom flera *if-satser* med ett flertal parametrar (Devten:27). När en matchning har gjorts får både rekryteraren och utvecklaren godkänna eller avböja erbjudandet, godkänns det av båda parter presenteras utvecklarens fullständiga profil och ett första möte kan påbörjas (Devten:43).

4.3 Artificiell intelligens

Resultatet kopplat till huvudområdet artificiell intelligens är uppdelat i de två ingående aspekterna: AI-tekniker och bias inom AI. Vid AI-tekniker har vi för läsarens skull förtydligat vilka AI-tekniker företagen beskriver genom egna tolkningar utifrån litteraturen.

Tabell 7: Översikt för hämtade data av artificiell intelligens

Artificiell intelligens	Tengai	Exparang	Hubert	Devtens
A1	10,12, 2b	8, 10, 42	8, 10, 12, 56, 14b	10, 12, 16
A2	2b,10, 12, 14, 18, 20, 22, 24, 26, 36, 42, 46	6, 10, 12, 14, 16, 18, 20, 24, 26,	6, 10, 12, 18, 20, 22, 34, 48, 58, 14b, 16b	2, 14, 16, 33, 39, 51

		28, 12b, 14b		
--	--	-----------------	--	--

4.3.1 AI-tekniker

Tabell 8: Översikt av företagens AI-tekniker

	Tengai	Exparang	Hubert	Devten
AI-tekniker	NLP (Google ASR + NLU + NLG) Intents (egna träningsdata)	NLP (Googles NLU (Exparang:14b)) Klustering (klassificering)	NLP (NLU, NLG) Intents (egna träningsdata)	NLP (Google NLU)
Nyttjas till	Omvandla ljudfil till textfil som robot grundar nästa beslut i, exempelvis be kandidaten förlänga sitt svar (Tengai:10). Tengais egna NLP avgör om ett svar är positivt eller negativt (Tengai:2b) Denna ordförståelse möjliggör att roboten kan agera utifrån vad kandidaten säger (Tengai:12).	<i>Kluster-teknologi</i> i kombination med NLP för att kategorisera de olika kandidaterna utifrån deras erfarenheter och titlar som sedan lagras i databasen (Exparang:8,10,42).	Chatt-roboten är uppbyggd på ett par tusen <i>intents</i> där Hubert försöker tolka kandidatens svar eller fråga med hjälp av NLP (Hubert:8) och sedan återkoppla med ett relevant svar (Hubert:10). <i>Intents</i> baseras på svar från kandidaterna och svar som utvecklarna själva har skrivit in (Hubert:12). En upptränad modell söker star-svar. En annan modell jämför svarets relevans i förhållande till tjänsten (Hubert:56). Hubert har även en modell som kollar på kandidaternas kommunikativa förmåga (Hubert:14b).	Devten har köpt in en API-tjänst för att kunna analysera bilder och text med hjälp av NLP (Devten:16). Med denna tjänst "scrapas" de Cv:n som laddas upp av kandidater. "Scrapa" innebär att algoritmen letar efter nyckelord i ett CV för att i sin tur kunna filtrera ut relevanta kandidater (Devten:10,12).

4.3.2 Bias inom AI

Tengai anser att de mest kritiska momenten där bias kan uppstå är hur intervju-roboten tolkar det som sägs av kandidaten, men det är inget som Tengai själva upptäckt har varit ett problem ännu (Tengai:22,36,46). Vid den manuella bedömningen kan fördomar influera rekryteringsprocessen, men att själva datainsamlingen som görs av roboten sker fördomsfritt (Tengai:18). Som tidigare nämnts använder sig Tengai av *Big Five*-modellen, företaget delger att i denna modell kan bias föreligga. *Big five*-modellen har även lagt grunden för de frågor som roboten ställer, per definition är frågorna fördomsfulla då de försöker att antyda vilken typ av person du är eller inte är. Modellen är formad kring män samt att modellen testats på manliga chefer och säljare. Tengai förklarar att det mäter även en extra personlighetstyp; emotionell stabilitet, som bidrar till en mer modern syn på modellen (Tengai:20,26).

Även Exparang är medvetna om att bias kan uppstå i deras tjänst. Respondenten beskriver att de själva subjektivt poängbedömt varje moment i frågeformuläret och att det sedan manuellt "skruvar på" poängen allteftersom de uppstår fel, vilket kan bidra till bias (Exparang:14). Exparang menar också på att det kan uppstå bias i frågeformuläret då de är tre män i samma ålder som utformat formuläret (Exparang:16).

Hubert har försökt att gå ifrån de traditionella rekryteringsmetoderna, då de anser att det är där det uppstår som mest bias och fokuserar därför istället på att sätta kriterierna tillsammans med respektive kund (Hubert:18). Respondenten på Hubert förklarar att den mest kritiska processen som bias kan uppstå i är hur kriterierna sätts då de kan riskera att sättas skevt (Hubert:58). Huberts respondent säger att om bias existerar i kundens processer så är det möjligt att detta fortplantar sig till Hubert (Hubert:22). Respondenten har svårt att tro att Hubert får med sig bias in i modellen som tittar på ordval och de språk som kandidaterna väljer att använda (Hubert:14b, 16b).

Devten menar att det som kan bidra till fördomsfulla beslut är hur kandidater själv har rankat sina språkkompetenser (Devten:39). Respondenten påstår också att en kritisk process är när momentet när input-data skickas in från respondenten då de själva inte kontrollerar vad NLP algoritmen tittar på (Devten, 33).

4.3.2.1 Identifierade biaser

Som tidigare nämnts använder Tengai, Exparang och Devten Googles NLP i sin robot för att förstå vad kandidaten säger (Tengai:10, 12; Exparang;14b; Devten:16). Då Google har utvecklat algoritmen som används som input-data till företagets tjänster finns en risk för att *inherited bias* kan uppstå.

Tengai bygger själva intents för att roboten ska agera på utifrån vilka "kommandon" kandidaten säger (Tengai:10). Det kan anses att *specification bias* riskerar att uppstå. Tengai har även bestämt att robotens tolkning av kandidaternas svar ska vara positivt eller negativt (Tengai:18, 22, 24). Respondenten fortsätter med att beskriva att de inte vet kandidatens identitet, och menar att de inte vet ålder, nationalitet och kön vilket de menar ska minska risken för bias (Tengai:14). Dock kan *specification bias* ändå uppstå på grund av *proxys*. Tengai väljer att endast identifiera om ett svar från kandidaten är positivt eller negativt (Tengai:24), detta kan tolkas som att risken för *proxys* minskar och därmed även risken för bias. För att roboten ska kunna avgöra om svaret är positivt eller negativt har Tengai observerat hur människor pratar och beskriver händelser utifrån det perspektivet. De har analyserat ordvalet, följderna i meningar och även variationen av ord och meningar (Tengai:2b)

Tengais robot bygger som tidigare nämnts på *Big five*-modellen (Tengai:26) vilket i sin tur kan riskera att bidra till *specification*- och *historical bias* då tidigare forskning utgör grunden för tjänsten, som därmed baserar algoritmen på hur världen var vid forskningens framtagning vilket har en inverkan på modellens utformning. Tengai menar att *Big five*-modellen utformades genom forskning som endast genomfördes på manliga chefer vilket därmed kan leda till att modellen innehåller bias som överförs till Tengai (Tengai:26).

Exparangs databas består av 1500 kandidaters Cv:n, men menar att det krävs åtminstone 5000 Cv:n för att rimliga slutsatser kan dras (Exparang:12). Det kan anses att en risk för att *inductive bias* uppstår då det saknas tillräckligt med data. Exparang säger också att alla tänkbara kandidater måste inkluderas för en väl fungerande tjänst och att det hade extremt oansvarigt att snäva ner urvalet (Exparang:18). Det visar på en medvetenhet för *sampling bias*, då grupper inte blir under- eller överrepresenterade, samt för *inductive bias*, då det att

det är viktigt med ett stort urval för ett bra resultat. Exparangs respondent påtalar också att det blir fel om de endast väljer kandidater från vissa lärosäten på kundens förfrågan (Exparang:20). Därmed kan risken för *sampling bias* minskas genom att en viss grupp inte blir överrepresenterad.

Respondenten på Exparang berättar också att de bjuder in kandidater till sin databas (Exparang:6,12). Det kan anses att detta kan bidra till *sampling bias* då det finns en risk för felaktig framställning av populationen. Respondenten på Exparang berättar om hur snedfördelat resultat kan bli om endast är 43 åriga, vita män som läst vid MIT får alla jobb som följd av en för snäv målgrupp (Exparang:10). Detta kan anses visa på en medvetenhet för *sampling bias*.

Respondenten fortsätter att beskriva hur de själva subjektivt poängsatt varje moment i frågeformuläret som kandidaterna genomför (Exparang:14,24). Här kan det finnas risk för att *specification bias* uppstår. Risken för *specification bias* minskar när personen i fråga har stor kunskap om ämnet. Respondenten menar att han och hans kollega har arbetat med området i flera år och genomfört tusentals intervjuer (Exparang:24). Dock medger respondenten att ord kan kopplas exempelvis till olika kön (Exparang:12b). Detta kan visa på hur ord kan agera som *proxys* och därmed leda till att *specification bias* uppstår.

Exparang menar själva att det definitivt kan finnas inbäddad bias i frågeformuläret då de är en homogen grupp som utvecklat det (Exparang:16). Detta kan kopplas till en risk för *observer bias* då ens syn på omgivningen kan skilja ifrån hur den egentligen är i verkligheten. Respondenten nämner att testgruppen hos Exparang varierar kulturellt och geografiskt samt att alla anställda är involverade i beslut där något ska förändras (Exparang:26, 28). Detta kan anses minska risken för *observer bias* då världen kan ses på olika sätt av olika människor.

Huberts respondent förklarar hur de i början av processen utformar gränssnitt innehåll med kunden baserat på bland annat deras egna samt kundens existerande frågor (Hubert:6). Det kan uppstå en risk för *historical bias* då frågorna kan ha en viss fördomsfull innebörd, vilket i sin tur leder till att algoritmen i sig har bias. Då bias kan finnas i frågorna i sig, kan bias eventuellt uppstå trots att data är kontrollerad. Kunderna är även inkluderade i att sätta upp kriterier inför intervjun i gränssnittet (Hubert:20). Detta kan utgöra en risk för *historical* samt *observer bias* som vid utformningen av frågor. *Observer bias* riskerar att uppstå då vad kunden uppfattar är viktiga frågor eller kriterier inte nödvändigtvis motsvarar vad som krävs för en lyckad rekrytering i verkligheten. Hubert utgår även från kundernas existerande processer vid utformning av kriterierna (Hubert:22). Detta kan bidra till *historical bias* på samma vis som nämnts innan. Respondenten på Hubert förklarar att den mest kritiska processen som bias kan uppstå i är hur kriterierna sätts då de kan riskera att sättas skevt (Hubert:58). Det kan anses att detta utgör en risk för att *specification bias* uppstår. Hubert anser att det är långsökt att hävda att ord kan kopplas till exempelvis kön (Hubert:14b,16b). Detta kan tolkas som att risken för att det uppstår en *specification bias* där ord agerar som *proxys* är låg.

Vid upplärning av systemet nyttjas *intents* för att Hubert ska förstå vad kandidaten säger och sedan koppla detta till ett relevant svar. I början av denna process försökte de anställda på Hubert förstå vad vilka svar som ofta återkom och träna systemet att känna igen de *intents*:en (Hubert:10,12). Vid upplärningen av systemet kan *specification bias* uppstå, då in- och utdata definieras av Hubert. Vidare kan det finnas det en risk för att *observer bias* uppstår då det de anställda uppfattar som återkommande möjligtvis inte reflekterar verkligheten. Vid analys av

kandidatens svar kan det vara fördelaktigt att använda sig av nyckelord som är associerade till tjänsten eller de kompetenserna utvalda för tjänsten (Hubert:34). När detta sker kan det finnas en risk för *specification bias*, då indata i form av nyckelord väljs ut av kund och Hubert. Utdata från gränssnittet granskas av kunderna och går sedan tillbaka in i modellen, vilket gör att Hubert ser att en viss typ av kandidat blir accepterade i en högre grad än andra kandidater (Hubert:48). Här riskerar *specification bias* att uppstå.

Devten har själva valt att ta bort känsliga data. Devten nämner i intervjun att exempelvis utbildning och erfarenheter syns i kandidaternas profil kan härleda fram till bias (Devten:2). Att ta bort data hindrar inte ord att kunna agera som *proxys* vilket kan leda till *specification bias* i inmatningsdatan. Devten har även märkt att deras inköpta API inte läser in vissa typer av data. Företaget fortsätter att förklara hur inskickade CV endast bör innehålla en sida för att deras tjänst ska kunna skapa en rättvis bild av en kandidat (Devten:14,51), båda exemplen innefattar en risk för *sampling bias* kan uppstå då urvalet kan bli över- eller underrepresenterade.

I tabellen sammanfattas våra tolkningar av vilka bias som kan uppstå i respektive tjänst utifrån respondenternas svar.

Tabell 9: Identifierade bias i tjänsterna

Bias	Tengai	Exparang	Hubert	Devten
Inductive/Learning bias		X		
Historical bias	X	X	X	X
Specification bias	X	X	X	X
Measurement bias/Observer bias		X	X	
Sampling bias	X	X		X
Inherited bias	X	X		X

4.4 Europeiska kommissionens riktlinjer för tillförlitlig AI

Resultatet kopplat till huvudområdet Europeiska kommissionens riktlinjer, presenteras uppdelat i de olika riktlinjer.

Tabell 10: Översikt för hämtade data av Europeiska kommissionens riktlinjer

Europeiska kommissionens riktlinjer	Tengai	Exparang	Hubert	Devten
L1	10,18,36	8,10,14,20,30	18	-
L2	10, 14, 16, 12b, 14b	4,20, 22, 22, 24, 2b	12, 14, 16, 7b	2, 47
L3	28, 30, 46, 10b	20, 32, 38, 8b, 10b	24, 52, 54, 2b, 5b, 9b	12

L4	32, 38, 44, 46	6, 10, 12, 16, 18, 24, 26, 28, 34, 36	16, 20, 40, 42, 58, 60	2, 33
L5	36, 16	8, 10, 40	18, 22	-
L6	20, 34, 38, 42, 4b, 6b	12, 14, 26, 28, 30, 8b, 12b	22, 44, 48	33

4.4.1 Mänskligt agenskap och mänsklig tillsyn

Tengai har själva byggt *intents* vilket innebär att roboten agerar utefter detta (Tengai:10). Rekryteraren tar även vid och tar beslut utifrån inspelningen av de kompetensbaserade frågorna (Tengai: 18). De gör även iterativt ändringar och utvecklar tjänsten utifrån vad kandidaten uppfattar (Tengai:36).

Exparang anser att de behöver ha insikt och kunna ingripa i processen relativt lång tid framöver då tjänsten i nuläget inte kan vara helt automatiserad. Anledningen till detta är det komplexa i att definiera roller och titlar såsom de gör i tjänsten (Exparang:8). Exparang ser det som sitt ansvar att ingripa och säkerställa en bred målgrupp bland kandidaterna genom mänsklig handpåläggning, för att undvika ett snedvridet urval. Hur snedfördelat det kan bli exemplifieras av respondenten med ett scenario där det endast är 43-åriga, vita män som läst vid MIT får alla jobb som följd av en för snäv målgrupp (Exparang:10).

Exparang bjuder in kandidater till sin tjänst. För att kontrollera *klustringen* och definieringen av olika profiler har Exparang genomfört manuella körningar i databasen. Vidare poängsätter Exparang värdet av de olika kriterierna för tjänsten själva. Värdet av dessa poäng justeras manuellt allteftersom Exparang upptäcker att något är inkorrekt (Exparang:14). Vidare har kunden ingen möjlighet att påverka och vikta urvalskriterierna. Anledningen till att Exparang inte alltid kan godkänna den viktningen kunden vill göra, exempelvis att endast ta emot kandidater som läst vid vissa lärosäten (Exparang:20). Exparang har även en *product steering committee* som arbetar agilt med produktutvecklingen där ämnen som mångfald tas upp. Även om de i dagsläget inte har resurser för att ha etik och bias som en enskild punkt så har de ambitionen att i framtiden att strukturera kommittén på ett annat vis (Exparang:30).

Huberts viktning görs i samförstånd med kunden där de justerar poängsättningen av svaren utifrån vad respektive kund önskar. Respondenten ger ett exempel på kriteriet arbetslivserfarenhet, som inom den exakta tjänsten kan poängsättas med att över tre år motsvarar fyra eller fem poäng medan ingen erfarenhet motsvarar noll poäng (Hubert:18).

4.4.2 Integritet och dataförvaltning

Ett intervju svar kan besvaras på hur många sätt som helst, därför har Tengai jobbat mycket med datainsamlingen, där de har försökt förfina processen. (Tengai:10). Som nämnt vet Tengai inte om kandidaters ålder, nationalitet och kön vilket de menar ska minska risken för bias (Tengai:14). Roboten analyserar endast det som bokstavligen har sagts och väger inte in tonläge eller ansiktsuttryck i sin data (Tengai:14). Anledningen till att de inte analyserar tonläge eller ansiktsuttryck är att det inte finns något i forskningen som styrker att exempelvis en glad person gör en bättre arbetsprestation. Tengai har därmed tagit ett aktivt val om detta och beslutat sig för att deras produkt ska vara grundad i vetenskaplig forskning

(Tengai:16). Tengai förvaltar kandidaters data genom att lagra på olika servrar och de försvårar även möjligheten till att få tag på information tillhörande specifik kandidat. Data sparas i 12 månader efter intervjun och kan väljas att raderas närsomhelst (Tengai:12b,14b).

Exparangs system är utformat att automatiskt göra en matchning när en kund publicerar ett nytt jobb, och kunden kan därmed inte ändra i processen. Dock fyller såväl kunderna som kandidaterna i ett formulär med frågor som sedan matchningen baseras på. Där finns det en möjlighet att låta bli att svara på de frågor som de själva inte anser är relevanta (Exparang:4,20). Exparang menar att de har sett initiativ i branschen där företag försöker göra matchningar mellan kandidater och företag genom att personerna fyller i lekfulla formulär. Detta menar Exparang kan leda till matchningar som inte är kvalitativa. För att undvika detta består Exparangs matchningsformulär av en del "tråkig" information för att på så sätt säkerställa matchningens kvalitet. Exparang "scrapar" inte data från öppna källor (Exparang, 20,22,2b). Intervjufrågorna har utformats av respondenten och hans medgrundare, baserat på deras långa tidigare erfarenhet från rekryteringsbranschen, deras nätverk samt, en testgrupp bestående av deras kunder. Frågorna är därmed inte baserade på forskning utan på deras egen subjektiva bedömning (Exparang:24).

Hubert baserar de *intents* som de tränar sin chat-bot med på tidigare svar från kandidater samt med påhittade svar från anställda på Hubert (Hubert:12). De frågor som kandidaterna besvarar är framtagna i samråd med deras kunder. De använder upptränade frågor för olika kompetenser (Hubert:14,16). Hubert har ingen standardiserad process över hur de hanterar kandidaternas data efter att rekryteringen är färdigställd, dock enligt deras standardavtal med kunderna lagras data i 36 månader (Hubert:7b).

Devten tittar endast på kandidaternas kompetenser och tidigare erfarenhet i kandidatens CV utan att inkludera namn eller annan känsliga data. De menar att det finns studier som säger att genom exkludering av bild och namn, hamnar fokuset vid rekryterarens första intryck på kandidatens kompetenser (Devten:47). Med Devtens tjänst kan kandidaten välja när denne inte längre vill vara anonym. Kandidaten kan vara anonym genom hela matchningen tills personen presenteras för kund (Devten:2).

4.4.3 *Transparens*

Respondenten på Tengai säger att kandidaten får veta att intervjun bygger på *Big five*-modellen och att det är personliga egenskaper som mäts, dock får kandidaten inte reda på vilka egenskaper som mäts eller om dessa är viktade (Tengai:28). Både kunder och kandidater är helt införstådda med att rekryteringen sker med hjälp av AI, båda parterna får mycket information om hela processen (Tengai:10b). Kundsidan får en utbildning om ramverket, hur det är framtaget, hur det fungerar samt hur de själva ska nyttja det (Tengai:30). Tengai menar att de är helt transparenta om tjänsten skulle bli ifrågasatt, då de inte har något att dölja (Tengai:46).

Exparangs tjänst är begränsad på det sätt att kunden själv inte kan bestämma och vikta urvalskriterier. Respondenten på Exparang menar att kunder absolut vill påverka i vissa fall, men detta är något som inte är i linje med Exparangs tankesätt kring sin tjänst, därav begränsningen (Exparang:20). Kunden får i princip ingen information kring matchningen, vad matchningen grundas i eller viktningen av kriterierna. Anledningen till detta är att Exparang ska sträva efter att delge neutral information i största möjliga mån (Exparang:32). Företaget är transparenta med hur deras formulär fungerar och vilka matchningsvinklar som används samt

att Exparang kan redovisa för hur poängsättningen har genomförts (Exparang:38,8b). Företaget menar att detta tillvägagångssätt leder till det mest neutrala urvalet och att kunden behöver motarbetas i vissa fall och få hjälp i att bli införstådda hur fördomsfullhet kan undvikas. Exparang är tydliga med AI:ns inverkan i rekryteringsprocessen (Exparang:10b).

Hubert har inte för vana att förklara hur de kompenserar mot bias för kunderna. De menar att kunden relativt enkelt förstår att om den mänskliga kontakten mellan kandidat och rekryterare förskjuts till ett senare skede i processen så förflyttas även risken för bias. De satta kriterierna eliminerar risken för detta under den tidiga screening-processen, enligt Hubert, vilket de brukar informera kunderna om. Om kunderna själva går in på bias-frågan förklarar Hubert detta (Hubert:24).

Vid ett tillfälle under intervjun visade Hubert hur kundens gränssnitt av deras *Applicant Tracking System*, ATS, av kandidaterna såg ut och fungerade. Kunden ser vilka kandidater som uppfyller de formella kraven och vilken ranking en viss kandidat har. Exempelvis kunde kunden se hur en kandidat hade en score på 90 av 100 gällande tidigare erfarenhet som krävs för en viss tjänst samt att denne är bekväm med att arbeta i minusgrader. Alla parametrar leder till ett summerat resultat för kandidaten (Hubert:52). Varför en person får en viss poäng beror på vad Hubert och kunden kommit överens om för ett visst kriterium (Hubert:54). Exempelvis har de kommit fram till att en viss mängd erfarenhet ger ett visst antal poäng (Hubert:2b). Hubert kan då ta fram en kandidats svar och hänvisa hur poängsättningen har genomförts och varför en kandidat inte gått vidare. Företaget försöker vara så transparenta som möjligt (Hubert:5b,9b).

Devten menar att problem kan uppstå om kandidater inte utformar sitt CV läsvänligt för en algoritm (Devten:12)

4.4.4 *Mångfald, icke-diskriminering och rättvisa*

De anställda på Tengai arbetar i Sverige där de har kontor i Stockholm och Malmö. Det finns en jämn spridning avseende kön och i utvecklingsgruppen finns en mångfald av nationaliteter. Detta är inget som de aktivt har eftersträvat utan det har skett av en slump (Tengai:32). Tengai går regelbundet igenom buggar och fel som kan uppstå. De arbetar även ständigt med om vad och hur roboten talar till kandidaten. Det viktigaste för Tengai är kandidatupplevelsen under intervjun, att personen har fått berätta om sin kompetens som tolkas på ett fördomsfritt sätt. Detta uppnås genom att roboten aldrig programmeras om för enskilda rekryteringar. Det enda Tengai tar med i beräkningen är vad som sagts på intervjun och den data tvättas från allt det personliga. Det är en ren ljudfil som omvandlas till text där analysen om vilken typ av person kandidaten är grundas på (Tengai:38,44,46).

Exparang möjliggör för oväntade kandidater att träffa arbetsgivare utan att behöva ha kontakter eller ett nätverk, då det är matchningsdatan som bestämmer (Exparang:6,12). Företaget är ödmjuka med att endast använda AI baserad på äldre data vid matchningen, då detta lätt blir snedvriden data och rekryteringen sker fördomsfullt istället för tvärtom. Exparang poängterar att det är deras ansvar som företag att ha en bredare målgrupp och förhindra snedvridna data (Exparang:10).

Exparang medger att de som skapat tjänsten är tre män födda på 70-talet som har utformat grunden till frågeformuläret, dock har ytterligare en medgrundare tillkommit som är kvinna som har bidragit med andra perspektiv samt infallsvinklar. Exparang säger att de troligtvis

aldrig kommer att kunna garantera att frågorna blir helt fördomsfria (Exparang:16). Deras roll är att se till att det hamnar bra kandidater på chefsbefattningar, de har inget köns- eller mångfaldsfokus i dagsläget. Dock anser Exparang att det är oansvarigt att minska urvalet då alla typer av människor behövs (Exparang:18).

Exparangs frågeformulär är inte baserat på forskning utan subjektivt utformat av grundarna, dessa har även testats i samråd med en testgrupp och kunder i deras egna nätverk (Exparang:24). Utformningen av testgrupperna har sett olika ut, både arbetsmässigt men även en mångfald av nationaliteter och kön. Bland Exparangs egna kunder har det inte funnits någon strategi, utan de är homogena. På företaget har alla en chans att tycka till om hur produkten ska vara utformad, Exparang kallar det *product steering committee* (Exparang:26,28). Kunden får endast tillgång till de matchningar som passar den eftersökta tjänsten, då Exparang vill hjälpa kunderna förstå att det är lätt att hamna i omedveten fördomsfullhet. De trycker på att kandidater och arbetsplatser ska tänka utanför boxen och se saker från nya perspektiv (Exparang:34,36).

Hubert utformar sina frågor i samråd med deras kunder inför varje rekrytering. Det är dessa frågor som bedömer kandidaternas kompetenser (Hubert:16,20). Hubert har en mångfald i deras utvecklingsteam från Sverige och Asien, där det finns en variation av kön och åldrar (Hubert:40,42). Hubert anser att den mest kritiska delen i deras tjänst är att kriterierna sätts på ett snedvridet sätt. Risken för detta är mycket liten, då Hubert tillsammans med sina kunder har stor erfarenhet av detta sedan tidigare och lärt sig vad som är rätt och fel (Hubert:58,60).

Devten har valt att anonymisera kandidater, men det kan finnas information angående tidigare arbetslivserfarenhet eller annan känslig information som kan koppla kandidaten till en person trots anonymiteten (Devten:2). De poängterar vikten av att vara noga med vad för typ av indata som används, Devten gav exemplet med om endast svenska namn användes för att träna algoritmen, är det kandidater med svenska namn som kommer att väljas ut (Devten:33).

4.4.5 Samhället och miljöns välbefinnande

Tengai arbetar med att ständigt utveckla sin produkt utifrån hur kandidaterna uppfattar och vrider på svaren från roboten. Respondenten poängterar hur de fått kritik för att roboten efterliknar en vit person och att de i framtiden, när tekniken tillåter, ska låta kandidaten själv få välja utseendet på roboten inför en intervju (Tengai:36). Tengais intervjurobot analyserar inte tonläge eller ansiktsuttryck då det inte finns forskning som stödjer att vara extra glad eller nervös inför en arbetsintervju korrelerar till arbetsprestation (Tengai:16).

Exparang ser det som sitt ansvar att inte automatisera sin tjänst helt då den är komplex (Exparang:8). De är även ödmjuka inför att använda AI för matchnings- och rekommendationsfunktionerna i sin tjänst. Anledningen är att det funnits tidigare fall där företag har misslyckats med att göra detta på ett fungerande sätt vilket resulterat i snedfördelade resultat (Exparang:10). Respondenten menar att de är försiktiga och ödmjuka inför att det lätt kan bli fel med användning av AI inom nya användningsområden (Exparang:40).

Huberts respondent menar att de tar ett steg ifrån att försöka imitera den manuella processen, då AI kan ha svårt att jämföra två helt olika svar på samma fråga, vilket utgör grunden för punkten där mest bias kan uppstå (Hubert:18). Hubert utgår från kunders existerande processer när kriterier inför en rekryteringsprocess sätts upp, men menar att risken för inbakad

bias i dessa kriterier är låga och att deras kundkrets utgörs i princip helt av etablerade rekryteringsbolag. Huberts respondent säger att om bias existerar i kunders processer så är det möjligt att detta fortplantar sig till Hubert (Hubert:22).

4.4.6 Ansvarsskyldighet

Tengai nyttjar som nämnts *Big five*-modellen som underlag till frågorna intervjuroboten ställer. Frågorna har sedan bearbetats och validerats av forskarna Anders och Sofia Sjöberg vid Stockholms universitet, dessutom har de låtit testgrupper utföra ett vanligt personlighetstest och sedan fått träffa Tengais robot för att kontrollera att samma resultat har uppnåtts - vilket har konstaterats att det har gjort (Tengai:20). Vidare ser respondenten från Tengai inget annat sätt att genomföra sitt arbete än att ha det fullständigt baserat på forskning (Tengai:42). Tengai nyttjar tjänsten Jira för att registrera buggar och liknande för att rapportera, prioritera och åtgärda de problem som uppstår. Vid sprintmöten tas upptäckta problem upp och diskuteras. Testningen av eventuella fel och lösningar sker kontinuerligt, det sker även automatiska kontroller för att justera om roboten svarat fel (Tengai:34,4b). För att kontrollera att deras ASR tjänst inte innehåller bias har de kontinuerliga kontroller med Google (Tengai:6b). Tengai prioriterar främst kandidatupplevelsen och att kandidaterna ska känna att de blir behandlade på ett bra sätt (Tengai:38).

Respondenten från Exparang förklarar hur de gör flera körningar i databasen för att förstå relaterade uppgifter och analyserar varför dessa relaterar till varandra baserat på den data som hämtats (Exparang:12). När ett fel upptäcks undersöker Exparang var det härstammar från och analyserar vad som kan göras, exempelvis manuellt ändra faktorer som blivit poängsatta för högt (Exparang:14,8b). Utformningen av frågeformuläret kandidaterna fyller i granskas flera gånger av en testgrupp för att säkerställa frågornas relevans och skapa ett så brett perspektiv som möjligt (Exparang:26,12b). Alla i teamet involveras i beslut som påverkar hur produkten ska förändras. Åsikter om frågeformulering, dess relevans och vilka alternativ som ska finnas får en chans att yttras. En diskussion kan lösas genom att grundarna får ta de slutgiltiga besluten (Exparang:28). Vid dessa sprint-möten tas ämnen såsom etik och bias inte upp på ett strukturerat vis, då en avvägning gjorts mellan strukturen och tiden samt pengarna det hade krävts för att genomföra uppstyrningen (Exparang:30).

En granskning av kandidater sker då kunderna själva ser över samtliga och sedan accepterar de mest intressanta kandidaterna (Hubert:48). Hubert kompromissar med sin granskning av eventuell bias i sin tjänst genom att förlita sig på att deras kunder har ordentligt utformade, existerande processer (Hubert:22). Huberts respondent beskriver hur kompromisser sker konstant mellan tid och budget, men att bias ändå prioriteras högt av både Hubert och kunderna. Dock ser prioriteringen annorlunda ut beroende på vilken kund de arbetar med och att frågan om bias är mer aktuellt hos de företagen som inte är rekryteringsbolag, där tid och pengar prioriteras mer. Hubert omprioriterar internt utifrån kundens prioriteringar (Hubert:44).

Devten granskar inte den indata som en användare skickar in och det sker inte en kontroll av vad algoritmen ska undersöka (Devten:33).

5 Diskussion

Kapitlet inleds med en diskussion kring hur rekryteringsbranschen ser på AI-tekniker samt hur detta eventuellt kan påverka AI-företag. I diskussionen behandlas definitionen av AI och klargör varför tjänsterna vi undersökt, kan definieras som AI-system. Vidare diskuterar vi de bias som uppstår i företagens tjänster samt undersöker huruvida företagen efterlever Europeiska kommissionens riktlinjer, som ett sätt att mäta hur de arbetar för att minska bias i sina tjänster. För enkelhetens skull görs kortare sammanfattningar i slutet av varje diskussionspunkt som ett sätt att summera nyckelpunkterna av varje punkt.

5.1 Insikter i rekryteringsbranschens syn på AI

Tengai och Hubert gör bedömningar grundade i de mer traditionella måtten, såsom *Big five*-modellen, som dock är medkompletterad med en extra personlighetstyp som bidrar till en modernare syn på modellen, och *star-svar*. Vi tycker det är intressant att se de traditionella personlighetstesterna integreras i de moderna tekniska verktygen. Detta visar även på att rekryteringsbranschens traditionella tillvägagångssätt är möjliga att cementera i moderna verktyg. Detta tror vi kan innebära en lättare övergång till en mer modern rekrytering; dels för att företag fortfarande har de traditionella metoderna att förlita sig på, dels för att AI-företagen anammar de traditionella metoderna som har granskats och prövats under en längre tid, vilket vi tror kan resultera i mindre fördomsfulla verktyg. Dock påpekade Hubert att direktöversätta traditionella metoder ökar bias på grund av AI:s förmåga att tolka olika svar på samma fråga; då AI ibland kan ha det svårt att se olika svar som lika bra. De traditionella metoderna går alltså igenom en ny digitaliserad granskningsprocess, vilket vi ser som ett sätt att kunna minimera bias då metoderna revideras på nytt. Dock tror vi att vissa tester kan vara svåra att automatisera, exempelvis som Wise Consultings djupintervjuer, vilket även Hubert påtalar som tidigare nämnt i stycket. Vi tror den komplexa automatiseringen av mer avancerade frågor är anledningen till att AI-företagen låter en mänsklig agent hantera svaren på de frågor som kräver längre utläggningar av kandidaten.

Wise Consulting betonade även vikten av att ha en mänsklig handpåläggning i rekryteringsprocessen. De tror att en av de stora utmaningarna med automatiserade tjänster inom rekrytering är att kunna identifiera de mindre självklara valen av kandidater som besitter potential för en viss roll. Wise Consulting menade att en kandidat som har de exakta kompetenserna som krävs för en viss roll vill möjligtvis inte söka exakt samma tjänst igen eller alltid arbeta inom samma bransch. De kandidaterna tror vi att de automatiserade tjänsterna idag har svårt att hitta. Jordan & Mitchell (2015) menar att de flesta algoritmer fokuserar på att precisera resultat och göra förutsägelser som överensstämmer med verkligheten. Vi kan se hur algoritmens utformning för att hitta den optimala matchningen, skapar ett hinder för vissa beslut rekryterare hade velat ta under en rekryteringsprocess. Detta tror vi kan hindra kandidater från att byta eller klättra i karriären, vilket vi misstänker kan minska intresset för de automatiserade tjänsterna. Enligt Europeiska kommissionens (2019a) ska systemen som utvecklas levereras till en så stor målgrupp som möjligt och därmed inte anta att en lösning passar alla potentiella användare. En följd av dagens utformning är att systemen inte efterföljer riktlinjen till fullo, vilket vi ser som en framtida utmaning för AI-företagen. Företagen som intervjuades, med undantag av Exparang, nämnde inte denna utmaning.

Wise Consulting anser att behovet av automatiserade tjänster är stort vid urvalsprocessen och att implementeringen av sådana tekniker kan minska fördomar under rekryteringsprocessen. Wise Consulting menade att det finns flera effekter som kan orsaka en fördomsfull process; såsom eget tyckande, förutfattade meningar och vem som sköter rekryteringen. Urvalet är även stadiet som samtliga intervjuade AI-företag utvecklat sina tjänster för och därmed möter det behovet Wise Consulting uttryckt. Att AI-företagen kan erbjuda fördomsfri rekrytering är något som även uppmärksammas från kundsidan även om branschen i helhet fortfarande kan anses som traditionell. Vi tror på grund av att kundsidan ser fördelarna med automatiserade tjänster så visar det att rekryteringsprocessen överlag kan bli mindre fördomsfull allteftersom dessa tjänster, där fördomsfrihet är centralt, blir mer etablerade.

Sammantaget kan man se hur rekryteringsbranschen närmar sig ett förändringsskede men att det är en väldigt stor omställning i jämförelse med de traditionella tillvägagångssätten som är etablerade idag. Vi kan även se hur Wise Consulting som potentiell kund till AI-företagen lyfter frågor och krav som de inte tillhandahåller idag, exempelvis hur hanteringen av kandidater som inte är optimala i teorin men ändå intressanta för kunden ska ske.

5.2 AI-tekniker

Vad som kan inkluderas i definitionen av AI är diffust menar Hellström, Dignum och Bensch (2020). Europeiska kommissionen (2019b) har valt att definiera AI som ett system som anses vara intelligent, vilket det anses vara när det kan ta beslut eller assistera en människa. Vidare menar Benbya, Pachidi och Jarvenpaa (2021) att definitionen AI även kan appliceras på system som kan beslutfatta, problemlösa, innovera, uppfatta situationer och genomföra repetitiva uppgifter där kognitiva egenskaper krävs. Vi instämmer med Hellström, Dignum och Bensch (2020) att begreppet AI kan vara svårt att definiera, vilket uppdragas inte minst när de intervjuade företagens tjänster benas ut. Företagen marknadsför sig med att tillhandahålla AI-tjänster, men de delar i tjänsterna som kan betraktas som AI i bemärkelsen av avancerad ML är begränsade i sin storlek. Att varken företagen själva, Europeiska kommissionen eller forskare inom området kan göra en konkret gränsdragning gällande definitionen av AI visar verkligen på områdets komplexitet. Dock kan tjänsterna exempelvis samla in information om kandidater, analysera kandidaternas svar på intervjufrågor och automatiskt matcha den mest optimala kandidaten till en tjänst, vilket vi anser återspeglar Benbya, Pachidi och Jarvenpaas (2021) definition av AI. Enligt Europeiska kommissionen (2018) behöver inte AI-systemet vara självständigt, utan det kan även betyda att en person måste vara med i ett beslut. I tjänsterna som vi har undersökt gör AI:n en rekommendation och i de flesta fallen krävs en mänsklig handpåläggning. Här ser vi hur Europeiska kommissionens (2019b) definition av AI överensstämmer med det företagen vi intervjuat levererar. Europeiska kommissionen (2019b) och Benbya, Pachidi och Jarvenpaas (2020) definitioner av AI anser vi därför är applicerbara på företagens tjänster i sin helhet, vilket är anledningen till att vi kallar dem för AI-system. Vi applicerar därför Europeiska kommissionens riktlinjer för att kunna diskutera hur tillförlitlig företagens AI-system är och därmed nyttja det som utgångspunkt för att se hur företagen arbetar med att minimera bias.

Sammantaget kan vi konstatera att det inte finns mycket ren, avancerad AI i företagens tjänster, men att utifrån definitionerna av Europeiska kommissionen (2019b) och Benbya, Pachidi och Jarvenpaas (2020) kan deras tjänster innefattas i begreppet AI-system. Vidare applicerar företagen AI på olika delar av sina tjänster, men samtliga företag använder sig

endast av de AI-tekniker de själva uppfattar att de behärskar och därmed placeras tjänsten inte i en *black-box*, vilket vi ser som ett välgrundat val för att minska risken för bias.

5.3 Bias inom AI

Ntoutsis et al. (2020) menar att bias innebär ett AI-system som tar fördomsfulla beslut mot eller för en person eller grupp och som antingen är gynnsamt eller ofördelaktigt, speciellt på ett sätt som anses orättvist. Vidare fortsätter Ntoutsis et al. (2020) med att beskriva hur viktigt det är att skaparna eller utvecklarna är medvetna om bias för att då kunna säkerhetsställa att teknologin kan nyttjas ansvarsfullt. Att bias uppkommer i de intervjuade företagens tjänster, att företagen är medvetna om detta samt att de redan idag försöker motarbeta fördomsfullhet, kan konstateras. Vi anser att medvetenheten och agerandet från företagens sida talar för att Ntoutsis et al. (2020) uttalande om ansvarsfullt nyttjande efterföljs.

Gemensamt för företagen är risken för att *specification bias* uppstår. Detta då företag tagit beslut gällande in- och utparametrar för algoritmerna baserat på erfarenhet och eget tyckande, eller utifrån forskning, vilket Hellström, Dignum och Bensch (2020) menar kan vara en risk för att bias uppstår. Hellström, Dignum och Bensch (2020) menar att det krävs en hög förståelse av det problem som algoritmen syftar till att lösa för att valet ska bidra till minskad bias i algoritmen. Detta är något vi anser att Tengai, Exparang och Hubert har då de besitter erfarenhet från rekryteringsbranschen, samt att det i Tengais fall även vägs in forskning inom området, vilket minskar risken för *specification bias*. Devten är ett start-up där de anställda fortfarande är studenter, vilket medför en högre risk för att *specification bias* uppstår, då kunskap om rekrytering saknas. Således anser vi att en viktig del för att minimera bias är att inkludera erfarenhet av rekrytering vid algoritmens utformning.

Majoriteten av företagen har valt att ta bort vissa in-parametrar som de anser kan bidra till bias, exempelvis namn och kön. Dignum, Hellström och Bensch (2020) menar att *specification bias* ändå kan uppstå genom *proxys* av orden. Vid flera intervjuer kunde vi identifiera bristande kunskap gällande *proxys* och anser att risken för *specification bias* är närvarande. Vidare anser vi att det kan vara intressant att titta på hur ut-parametrarnas risk för bias påverkas av in-parametrarna och hur det kan motarbetas, såsom Tengai gör genom att klassificera ut-parametrarna som endast positiv eller negativ. Tengai menade att detta kan minska risken för biasen som kan uppstå vid *proxys*. Även om detta inte är en lösning som diskuterats i litteraturgenomgången tycker vi att detta är något som bör lyftas inom de andra företagen för att minska risk för bias.

En intressant skillnad mellan företagen är hur de bestämmer bedömningen av kandidaten, där Tengais robot baseras på forskning, Exparang sätter kriterierna själva och Hubert bestämmer hur bedömningen ska ske tillsammans med kunden. Vi anser att samtliga företags beslut av tillvägagångssätt riskerar att leda till bias, men på olika sätt. Att Tengai baserar bedömningen på forskning kan vid första anblick inge en uppfattning av att den är rättvis. Dock är forskningen baserad på ett urval där kvinnor inte existerar och är dessutom endast genomförd på chefer, något som gör oss kritiska till att modellen kan appliceras på andra kön eller roller än just dessa; något som resulterar i en risk för bias. Att Exparang förlitar sig på sin egen erfarenhet och Hubert på sin och kundens erfarenhet kan leda till *observer bias*. Att företagen arbetar på olika sätt anser vi är fördelaktigt för branschen för att utröna det optimala tillvägagångssättet för att minimera bias. Men idag anser vi att det är svårt att ge ett definitivt

svar på hur vilket tillvägagångssätt som är att föredra då det finns risk att bias uppkommer i alla fallen.

Samtliga företag använder sig av någon form av NLP för att tolka ord eller tal. Att Tengai, Exparang och Devten har valt att använda sig av Googles NLP medan Hubert har utformat algoritmen själv är intressant, då vi anser det kan påverka risken för bias på olika sätt. Nackdelen med att inte använda sin egen algoritm är att utvecklarna har svårare att påverka biasen som kan uppstå när utfallet genereras från en annan algoritm som de förmodligen själva inte varit med och utvecklat. Således finns en risk för att *inherited bias* uppstår (Hellström, Dignum & Bensch, 2020). Vi menar att risken med att använda andras algoritmer påvisas genom Googles BERT, då forskare konstaterat problem med bias i systemet samt att utvecklarna själva inte förstår hur systemet beter sig. Att denna problematik uppstår hos en så etablerad leverantör som Google, gör oss tveksamma till möjligheten att lösa problemen. För att minimera risken för *inherited bias*, ser vi det som fördelaktigt att utveckla algoritmen *inhouse* för att ständigt kunna kontrollera algoritmens process. Å andra sidan menar Hellström, Dignum och Bensch (2020) att en viktig aspekt för att minimera risken för *sampling bias* är en korrekt framställning av populationen, vilket kräver en stor mängd träningsdata från olika människor. Även om vi inte vet om Google har en korrekt framställning av populationen tror vi att väldigt få har en större datamängd som kan nyttjas som träningsdata. En följd av att samtliga företag är förhållandevis nya och ännu inte speciellt etablerade, är att algoritmens träningsdata inte är representativ av populationen. Att då överväga att använda en annan algoritm än sin egen, likt Tengai, Exparang och Devten, kan vara en lösning för att minska risken för *sampling bias*, men kan också öka risken för *inherited bias*.

Vi ser också en risk för att *inductive bias* kan uppkomma hos de företag som inte har kunskap om hur stort urvalet till algoritmen bör vara. Ett av företagen valde att inte använda sin algoritm då de identifierat risken för denna bias. Denna insikt hos företaget ser vi som mycket positiv då vi tror att det kan vara en kritisk process att avgöra när en algoritm är tillräckligt tränad för att användas i kommersiellt syfte.

Ännu en utmaning som kan konstateras hos alla företagen är risken för *historical bias*. Detta i och med att biasen riskerar att uppkomma när algoritmen utformas och baseras på hur världen är just nu eller tidigare har varit vid tillfället då den utformas (Hellström, Dignum & Bensch, 2020). Vi anser att detta är något som kommer ta en lång tid att lösa, men kan minimeras i takt med att samhällets syn på jämställdhet och diskriminering utvecklas. Att företagen ständigt granskar sin tjänst ser vi som ett bra sätt för att minimera risken för denna bias, då det kan vara ett sätt att identifiera gamla data som är fördomsfull som i sin tur kan raderas.

Sammantaget kan vi konstatera att samtliga företags tjänster riskerar att innehålla bias, men att de är medvetna om detta och kontinuerligt arbetar med detta. Generellt kan konstateras att företagen inte kommit så långt i sitt arbete med *proxys*. Vi vill påpeka att företagens utformning av sina tjänster kan ge upphov till bias på olika sätt; exempelvis genom att basera sin tjänst på forskning eller egna erfarenheter, eller nyttja sin egen eller någon annans algoritmer. Att inget tillvägagångssätt verkar kunna minimera bias fullständigt utan att bjuda in till en annan sorts bias visar på den komplexa utmaning företagen har framför sig. Slutligen såg vi att det fanns en hög risk för *historical bias*, vilket vi tror företagen kommer få tampas med en lång tid framöver.

5.4 Europeiska kommissionens riktlinjer för tillförlitlig AI

5.4.1 Mänskligt agentskap och mänsklig tillsyn

Silberg och Manyika (2019) menar att AI-system idag inte är avancerade nog för att självständigt kunna hantera den sociala omgivningen den befinner sig i och självständigt bedöma när en bias minimerats tillräckligt. Vidare menar Daumé et al. (2019) att utforska en kombination av algoritmer och mänsklig inspektion kan vara en lovande väg att gå. Vi kan konstatera att Tengai, Exparang och Hubert inkluderar en mänsklig, aktiv agent vilket säkerställer efterlevnaden av kravet på användarens autonomi. Utifrån ovanstående källor anser vi att majoriteten av företagen har tagit ett tillförlitligt beslut genom att involvera mänskliga parter i sin process. Europeiska kommissionen (2019a) beskriver hur människors aktiva roll ska stödjas och tillsyn av AI-systemet ska enkelt tillåtas. Internt har Tengai, Exparang och Hubert tydliga mänskliga roller som överser och granskar sina system. Även kunden har en roll som överser resultaten AI:n producerar som denne får en möjlighet att fatta bättre beslut utifrån, vilket Europeiska kommissionen (2019a) menar är målet med att användaren ska förstå och interagera med AI-systemet. Detta hindrar även att kandidater påverkas av konsekvenserna från ett automatiserat beslut, vilket Europeiska kommissionen (2019a) lyfter. Vi anser att företagen efterlever denna del av riktlinjen väl då de inkluderar mänskliga roller i AI-systemen. Företagen, med undantag för Devten, informerar även kunderna om processen, Tengai håller även en kurs om sin tjänst, vilket också pekar på en mer tillförlitlig AI (Europeiska kommissionen, 2019a).

Företagen arbetar utifrån *human-on-the-loop* eller *human-in-command*, vilket uppmuntras av Silberg och Manyika (2019), Daumé et al. (2019), Holzinger (2016) och Europeiska kommissionen (2019a). Den aktiva, mänskliga rollen är alltså ett medel för att minimera bias som alla företagen har anammat. Vi tycker det är intressant att se hur kundens handlingsförmåga begränsas av Exparang, då det vid en första anblick verkar trotsa riktlinjen till viss del, vilket uppmanar till en mindre tillförlitlig AI. Exparang anser det motsatta, att ifall kunden får för mycket bestämmanderätt över vilka kriterier som sätts kan detta bjuda in till mer bias, vilket i sin tur innebär en mindre tillförlitlig AI. Vi tycker det är ett intressant ställningstagande som visar att följa riktlinjen inte garanterar minskad bias då den kan uppstå på andra sätt, utifrån Exparangs synsätt. Hubert som genomför förberedelserna inför en intervju helt i samråd med kunden, riskerar att öppna upp för bias genom att kundens processer potentiellt innehåller bias, vilket isåfall kan fortplanta sig i Hubert menar respondenten. För minska risken för att bias uppstår måste därmed Hubert ställa höga krav på kunders redan existerande processer, för att syftet med riktlinjen ska uppnås. Att människans aktiva roll kan bidra till bias bekräftas rent historiskt, medan en brist på det kan leda till samma resultat, tycker vi är väldigt intressant.

Sammantaget ser vi att företagen har ett svårt beslut att ta gällande i vilken grad människan bör vara aktiv i systemet; särskilt när riktlinjer och erfarenhet i branschen pekar på olika saker och målet är att vara mindre fördomsfull.

5.4.2 Integritet och dataförvaltning

AI-system ska kunna säkerställa att relevant data nyttjas för att skydda kandidaters integritet. Den data som företagen väljer att använda ska inte kunna diskriminera individen, vilket dock kan vara svårt då data riskerar att innehålla sociala konstruktioners misstag. Detta innebär att

det är viktigt att ständigt testa och utvärdera företagens val av data (Europeiska kommissionen, 2019a). Majoriteten av företagen vi undersökt har valt att anonymisera kandidaters data under processens gång som ett sätt att undvika risken för bias, vilket vi tycker är ett ansvarstagande första initiativ. Dock påpekar Europeiska kommissionen (2019a) att det finns en risk för att sociala konstruktioner är inbakade i all data, vilket är svårt att undkomma. Vi ser positivt på företagens försök till att skydda individen genom att anonymisera känsliga data, men på grund av att data har sociala konstruktioner inbäddat är målet ännu svårt att uppnå.

Flera av företagen har en väl genomtänkt process för hur kandidaternas data ska förvaltas, vilket Europeiska kommissionen (2019a) menar krävs för att inte kandidatens integritet ska skadas. Detta tror vi kan vara en följd av den nuvarande lagstiftning, GDPR. Deras genomtänkta arbetssätt för att hantera av kandidaternas data, menar vi minskar risken för bias.

Exparang låter inte kunden vara delaktig i valet av data genom att kunden inte får vara delaktig i beslutet om vilka parametrar som inkluderas i rekryteringsprocessen. På detta vis anser Exparang att de säkerställer en tillförlitlig användning av relevant data för rekrytering, vilket överensstämmer med riktlinjen. Vi tror att respondenternas tidigare erfarenheter bidrar till en tydlig bild av vilken data som är nödvändig eller inte, vilket talar för att relevant data väljs ut. Även genom forskning kan relevant data urskiljas då Tengai exempelvis väljer att inte använda sig av tonläge eller ansiktsuttryck i bedömningen av en kandidat; då det inte finns forskning som visar på en korrelation mellan detta och arbetsprestation. Att företagen underbygger valet av data, dels genom erfarenhet, dels genom forskning, ser vi som ett bra sätt att minska risk för bias. Hubert väljer ut sin data baserat på vad kunden anser är relevant. Risken med detta är att det är svårt för AI-företagen att verkligen säkerställa att data är relevant och således säkerställa kandidaternas integritet. Här ser vi att vad som kan uppfattas som relevant data skiljer sig åt mellan företagen, vilket visar på svårigheten i att tolka riktlinjerna och således hur minimering av bias kan ske. Ens egna uppfattning av vad som är relevant data kan alltså vara fördomsfull, vilket är anledningen till att en tydligare riktlinje kring detta hade behövts.

Sammantaget försöker samtliga företag aktivt efterleva denna riktlinje genom att de exkluderar data som är irrelevanta utifrån mätbarhet av arbetsprestation. De exkluderar data såsom namn och etnicitet, då alla företagen anser att detta är en stor del för att minska risken för att bias uppstår. Till sist underbygger företagen sina val av data vilket vi anser minskar risken för bias.

5.4.3 *Transparens*

Europeiska kommissionen (2019a) menar att transparens grundas i förklarbarhet, spårbarhet samt kommunikation; av data, system och affärsmodell som utgör ett AI-system. Gällande teknisk förklarbarhet är det inget vi djupgående tagit upp med företagen på en sådan nivå att de förklarat detaljerat vad i tekniken som leder till ett visst beslut. Vi har därför svårt att uttala oss om hur väl de efterlever denna tekniska del av riktlinjen. Dock kan samtliga hänvisa till specifika viktningar, kriterier och liknande för att förklara en viss poängsättning. Hubert menar exempelvis att de kan visa för kandidaten exakt vilka kriterier som orsakade en bortsällning.

Det vi har frågat och fått svar på av företagen kretsar mer kring förklarbarhet och hur transparenta de är gentemot kunden samt kandidaterna. Som en påminnelse av Europeiska

kommissionens (2019a) definition av förklarbarhet, är det förmågan att kunna förklara AI-systemets processer samt tillämpningsområden. Vidare är teknisk förklarbarhet möjligheten att förstå och spåra de beslut AI-systemet har fattat. Vår uppfattning är att samtliga företag kunde förklara hur kriterier och viktningen genomförs, då detta görs av mänskliga, aktiva roller i systemet. Det vi hade kunnat argumentera för är att människors beslut i detta stadie kan befinna sig i en *black-box*, då anledningen till att en rekryterare vill vikta ett visst kriterium tyngre inte klargörs. Om människans aktiva roll är en del av AI-systemet innebär detta en risk för att bjuda in till bias, vilket var anledningen till att Exparang begränsade rekryterarens roll, vilket diskuteras i punkt 5.4.1. Att Exparang är mer återhållsamma med information till kunden under processens gång är för att skydda kandidaten från att kunden ska kunna vara fördomsfull. Vi tycker detta är intressant att se då Exparang egentligen delvis går emot riktlinjen om transparens i syfte av att vara mindre fördomsfulla.

Tengai är inte fullt transparenta gentemot kandidaterna genom att inte berätta vilka egenskaper i *Big five*-modellen som mäts, vilket innebär att de inte delger alla delar av processen i AI-systemet. Vi tycker det är intressant att se hur transparens generellt inte är verkar vara en viktig del vid traditionella rekryteringsprocesser men att det ställs ett högre krav på det i ett AI-system, särskilt då rekrytering är ett område som absolut har en stor påverkan på en persons liv. Att vara införstådd i ett AI-system är såklart relevant inom alla områden, men konflikten som uppstår just inom rekrytering är intressant att se. Då Tengai menar att det kan bli problematiskt att delge för mycket till kandidaterna inför en intervju för att det kan påverka deras svar.

Samtliga företag anser vi har en positiv attityd gällande att dela med sig av information och öka kunders och kandidaters förståelse när det efterfrågas. Både Exparang och Hubert uttrycker att kunder ofta inte är intresserade av processen bakom utan mer av resultatet. Samtliga menar även att kunder och kandidater är införstådda i att AI nyttjas i tjänsten, vilket följer Europeiska kommissionens (2019a) riktlinje. Som Knight (2017) menar riskerar avancerade algoritmers resonemang att hamna i en *black-box*. Vi anser att företagen fortfarande överläter de mest avancerade delarna av systemet till mänskligt beslutsfattande, såsom viktning av olika kriterier, vilket resulterar i att risken för en *black-boxad* algoritm minskar. Allteftersom deras tjänster utvecklas lär de även bli mer avancerade, vilket ökar risken för oförklarbara resultat; något som i sin tur kan riskera en ökning av bias i tjänsterna. Dock är det vår uppfattning att företagen är ödmjuka och ser över sin teknik på ett sådant vis att de är införstådda i vikten av ett förklarbart system och kommer fortsätta arbeta för det i framtiden.

Just att det rent tekniskt inte finns en definition på förklarbar AI, som Adadi och Berrada (2018) påpekar, kan även försvåra för företagen att kalla sitt system förklarbart. Adadi och Berrada (2018) menar även att ett förklarbart systems beslut går att rättfärdiga, vilket företag kan ha svårt för på grund av förklarbar AI:s bristande definition. Utan färdiga definitioner kan det vara svårt för oss att utgöra huruvida denna riktlinje följs till fullo eller inte. Dock tycker vi det är intressant att se att behovet av förklarbar AI ökar i takt med att tekniken blir mer komplex och har en större inverkan på människors liv, vilket företagen vi intervjuat är ett bevis på. Vi ser förklarbarhet som en viktig aspekt, inte bara för att säkerställa att resultat inte grundas i bias, men även för att kandidater ska kunna få en bra behandling under processens gång och veta varför de eventuellt inte går vidare i rekryteringen; vilket även påtalas i Europeiska kommissionens (2019a) riktlinje.

Sammantaget blir förklarbarhet svårt att klargöra om det efterlevs; dels för att det inte finns en

tydlig definition av förklarbarhet gällande AI, dels för att vi inte gått in på resonemanget bakom ett visst resultat tjänsterna levererat på en detaljerad, teknisk nivå. Dock är företagen väldigt öppna och inkluderar sina kunder i de flesta fall. Vi anmärker hur en process som annars inte har satta krav på transparens plötsligt får det bara för att det hanteras av ett AI-system. Vi tyckte det var intressant att se hur två riktlinjer kan tala emot varandra; då en mänsklig handpåläggning kan bidra till ökad transparens då resonemang inte hamnar i en algoritmisk *black-box*, samtidigt som det även kan minska transparensen då människans resonemang i många fall inte förklaras.

5.4.4 Mångfald, icke-diskriminering och rättvisa

Europeiska kommissionen (2019a) menar att företag bör anställa personer med olika kulturer och bakgrunder för att öka chansen till nyanserade synvinklar som en del i att uppnå tillförlitlig AI. Vi kan konstatera att företagen uppfattar att de har mångfald bland sina anställda vilket bidrar till att de till viss grad har nyanserade synvinklar. Dock vill vi poängtera att spektrumet av mångfald sträcker sig långt utöver män och kvinnor, exempelvis transpersoner, vilket inte diskuteras alls av företagen. Vi ser en stor risk att synvinklar från hela samhällsgrupper faller bort vilket även har en effekt på tjänsten. Detta påtalas även av West, Whittaker och Crawford (2019) som menar att en bristande mångfald påverkar utformningen av produkten, vilket vi ser ökar risken för bias. För att motarbeta detta bör företagen, som West, Whittaker och Crawford (2019) påpekar, återigen se över vilka människor som är inkluderade i processerna relaterade till AI på företaget. Dock ser vi positivt på företagets framtida arbete med detta, då de jämförelsevis med tech-företagens statistik, som West, Whittaker och Crawford (2019) presenterar, har en jämnare fördelning bland sina anställda. Att företagen har den välviljan och uppmärksammar mångfald, även om det finns arbete kvar att göra, tycker vi visar att mångfalden kommer att förbättras, vilket i sin tur kan minimera bias. Vidare tycker vi att det är mycket positivt att alla involverade i tjänsten får möjligheter att uttrycka åsikter och problem med tjänsten på samtliga företag. Vi tror detta lyfter olika synvinklar från personer mer varierade bakgrunder, vilket talar för de Europeiska kommissionen (2019a) anser bidrar till en tillförlitlig AI.

Europeiska kommissionen (2019a) menar även att datasets kan försämrats av att dess historik har inbakade fördomar och kan motverkas förslagsvis genom att införa tillsynsprocesser så systemets syfte, krav, beslut och begränsningar kan hanteras på ett transparent vis. Att det finns en mångfald i teamet är ett viktigt verktyg för att dessa tillsynsprocesser ska kunna identifiera samtliga risker. Detta eftersom att identifiera fel som inte existerar inom ens egna världsbild är svårt, till och med omöjligt, vilket överensstämmer med Caroline Farbergers uttalande (Phillips, 2021). Detta är något som vi anser att företagen är medvetna om, i en viss utsträckning, men som Caroline Farbergers uttalande visar, kan företagen inte sluta att vara självkritiska i sina tillsynsprocesser, just på grund av den bristande förmågan att identifiera samtliga fel utifrån den egna världsbilden (Phillips, 2021).

Sammantaget är mångfald en komplicerad punkt och svår att efterleva, då definitionen är så mycket bredare än vad som inkluderas i normen för manligt och kvinnligt. Vi anser att företagen därför har en viss bristande mångfald bland sina anställda och i testgrupper vilket leder till att vissa synsätt och perspektiv exkluderas från utformningen. Mångfald i team är en viktig lösning för att input från olika perspektiv ska inkluderas i systemets utformning. Att ha en homogen grupp gör det extremt svårt att inkludera tankesätt som är utanför ens egen världsbild. Dock tror vi att företagets medvetenhet om mångfald leder till att de anammar detta med tiden.

5.4.5 *Samhället och miljöns välbefinnande*

Europeiska kommissionen (2019a) menar att AI-system kan användas för att hantera samhällsproblem och bör inkludera ett hållbarhetstänk då det kan ha en påverkan på samhället i stort. Vi ser hur rekryteringsbranschen har gjort enorma omställningar de senaste decennierna, som beskrivet i bakgrunden, och teknikerna som utvecklats har varit revolutionerande i förhållande till hur rekryteringsprocesserna såg ut på 90-talet. Vi ser att företagens tjänster försöker hantera den diskriminerande rekryteringsprocess som finns i dagens samhälle. Utifrån våra intervjuer kan vi se att samtliga företag har detta i åtanke och grundar flera beslut utifrån att undvika vad som kan bedömas som fördomsfullt; något vi anser pekar på ett starkt hållbarhetstänk. Tengai, Exparang och Hubert visar alla på samhällsansvar då de motiverar val för produkten utifrån vad som leder till mindre bias. Som diskuterats tidigare nyttjas inte företagens tekniker på ett utbrett sätt och AI-verktyg är fortfarande inte normen i branschen, enligt Wise Consulting. Detta innebär att företagen har fortsatt stort ansvar när samhället närmar sig en mer digitaliserad rekryteringsprocess.

Enligt Zhao (2018) är socialt ansvar oundvikligt på grund av de områden AI är verksamt inom; vilket vi instämmer med. I synnerhet inom området rekrytering där ett bristande socialt ansvar kan ha stora negativa konsekvenser för såväl individ som samhället i stort, då snedvridna sociala strukturer riskerar att förstärkas. Som Zhao (2018) vidare menar så riskerar företag att mötas av kritik om endast egenintressen prioriteras, men genom att uppfylla ett socialt ansvar kan företagets rykte förbättras. Utifrån detta ser vi, förutom företagets välvilja att göra gott för samhället, ett incitament till att ha ett inkluderat hållbarhetstänk i sin tjänst. Tengai har exempelvis fått kritik för utseendet på deras intervjurobot, vilket de bemöter och tänker utveckla lösningar på i framtiden när tekniken tillåter; vilket visar på deras vilja att ta sitt samhällsansvar. Vi anser att företagen hamnar i en svår sits när de får hantera kritik likt detta som än så länge är tekniskt omöjligt att lösa; vilket kan utgöra ett hinder från att ta sitt fulla sociala ansvar.

För att ta detta ansvar är det mycket som talar för att arbeta på de etiska aspekterna av sin tjänst, vilket de intervjuade företagen gör utifrån vad vi diskuterat kring varje riktlinje. Detta exemplifieras genom att, som nämnt i punkt 5.4.3, att flera kunder inte är intresserade av processen bakom ett visst resultat men att de intervjuade företagen ändå tar ett socialt ansvar genom att utveckla förklarbara system. Vi tänker att ett egenintresse hade i detta fall varit att spara in på resurser för funktioner som kunderna ändå inte visar intresse för, men att det sociala ansvaret prioriteras.

Sammantaget ser vi hur företagen arbetar för samhällets välbefinnande genom att försöka leverera en tjänst som ska leda till en fördomsfri rekrytering. Vi tror även att socialt ansvar tas för att undvika kritik från samhället och för att bättra företagets rykte.

5.4.6 *Ansvarsskyldighet*

Enligt Europeiska kommissionen (2019a) bör algoritmer, data och processen bakom utformningen av ett AI-system vara möjligt att granska. Om systemets tillämpningsområde kan kopplas till grundläggande rättigheter bör det vara möjligt att en oberoende granskning genomförs. Utifrån våra intervjuer tolkar vi det som att Exparang och Tengai genomför granskningar av delar av sin tjänst; Tengai genom den valideringsrapporten två forskare producerat och Exparang genom att nyttja en testgrupp som är sammansatt av personer från

deras nätverk. Tengai genomför kontroller i samråd med Google och följer även Googles eget arbete med sin ASR som ett sätt att få en insikt i detta.

Internt genomför samtliga företag, med undantag av Devten, kontinuerliga granskningar av systemet. Vi tycker företagen verkade självkritiska och öppna för att medge att fel upptäcks, vilket pekar på att granskningen ger resultat. Detta bör leda till att bias minimeras. Dock sker granskningarna idag inte alltid oberoende, vilket vi ser som nästa steg för att minska risken ytterligare. Att Hubert utformar frågor, viktning och poängsättning helt i samråd med kunden baserat på deras existerande processer ser vi som en risk för att bias kan uppstå, då kundens processer inte granskas. Vidare konstaterade Huberts respondent att bias kan fortplanta sig från kundens processer, vilket talar ytterligare för att det som tas in till Hubert från kunden bör granskas för att säkerställa minimering av bias. Detta ser vi som en potentiell utmaning då företagen är beroende av kunderna, vilket kan resultera i att AI-företagen inte kan ställa krav på kundernas granskningsprocesser, vilket i sin tur kan leda till att bias uppstår. Som nämnts i litteraturgenomgången arbetar Europeiska kommissionen (2021) just nu med ett nytt förslag om en förordning där AI-system inom högriskområden behöver granskas av kommissionen själva. Vi ser att en oberoende granskning av företagets tjänster kan bli oundvikligt i framtiden i och med den kommande förordningen från Europeiska kommissionen, som nämnts tidigare, vilket kommer bidra till en minimering av bias.

Europeiska kommissionen (2019a) menar att beslut och åtgärder som påverkar systemet måste rapporteras, samt att potentiella negativa konsekvenser bör identifieras, bedömas och minimeras. Denna riskminimering kan ske under utveckling, spridning och användning av systemet. Samtliga företag, med undantag av Devten, har implementerade rutiner för hur felrapportering ska hanteras. Anledningen till att Devten inte har utarbetade rutiner tror vi kan bero på att de dels köper in API-tjänsten, dels för att företaget består av två studenter, vilket gör att de i princip rapporterar till varandra; något vi kan tänka oss sker naturligt. Gällande rapportering av beslut som har en påverkan på systemet uppfattade vi inte från intervjuerna huruvida strukturerat detta gjordes, utan det lät mer som att personer gick in och finjusterade systemet allteftersom. Om detta är fallet, efterlevs denna del av riktlinjen inte så noggrant. Något vi ser som väldigt positivt är att företagen gör en konsekvensbedömning under olika delar av AI-systemets livscykel, vilket Europeiska kommissionen (2019a) poängterar att man bör göra.

Europeiska kommissionen (2019a) menar att utvecklingsprocessen kan leda till kompromisser och att de kompromisser som uppstår bör utvärderas utifrån deras risk för etiska principer. Idag kompromissar majoriteten av företagen med etik, med undantag från Devten, men till vilken grad och tillvägagångssätt varierar. Något som vi ser som en intressant skillnad är att ett av företagen sätter kandidatens upplevelse i fokus vid kompromisser, medan de andra företagen oftast prioriterar utifrån kunden. Att prioritera utifrån kandidaten tror vi kan vara ett effektivt sätt för att minimera risken för bias i och med att det just är individen tjänsten syftar till att skydda. Att kompromissa utifrån kunden kan vara fördelaktigt för att få nöjda kunder vilket kan leda till fler kunder och därmed större databas, vilket kan minska risken för *sampling bias och inductive bias*. Men å andra sidan kan detta vara ett riskabelt tillvägagångssätt för att minimera risken för biaserna, då kontrollen hamnar utanför företaget. I och med att företagen kan ha svårt att veta hur kunderna arbetar mot diskriminering, anser vi att detta kan vara ett mer ineffektivt sätt för att säkerställa att risken för bias minimeras. Europeiska kommissionen (2019a) menar att de kompromisser som görs bör vara motiverade och väl dokumenterade. Att valen är väl motiverade hos samtliga företag är något vi anser att de förmodligen är då flera människor involveras i beslutsprocesserna. Men huruvida

motiveringarna är dokumenterade har vi ingen vetskap om. Vi ser att detta kanske kan bli en viktigare del av processerna i takt med att företagen växer vilket i sin tur kan minska risken för bias.

Samtliga företag genomför en intern granskning vilket är positivt och följer riktlinjen i denna bemärkelse. Dock genomför inte alla externa granskningar, vilket är negativt i förhållande till riktlinjen. Vi tar i beaktning att detta lär ske i framtiden ifall Europeiska kommissionens nya förordning verkställs kommer extern granskning bli en norm för samtliga företag. Företagen rapporterar fel allteftersom de dyker upp, vilket är positivt; dock är vi osäkra på huruvida justeringar rapporteras vilket kan påverka hur väl riktlinjen efterlevs. Idag kompromissar företagen på olika sätt och tar etiska principer i beaktning vid resonemang, dock i varierande grad.

6 Slutsats

Det går att konstatera att bias inom artificiell intelligens uppkommer i rekryteringstjänster på ett flertal sätt. De företagen själva menar kan bidra till uppkomst av bias är bland annat bedömningsmetodens framtagning och när analys av kandidaternas svar genomförs av andra algoritmer än företagens egna. Vi kunde se att företagen har bristande kunskap om *proxys* vilket vi kan konstatera som en risk för uppkomst av bias. Vi har även uppfattat risker för att samtliga bias som beskrivits i litteraturgenomgången kan uppkomma av olika anledningar inom olika AI-system. Risk för *specification bias* och *historical bias* har kunnat identifieras som mest förekommande, och är också något som vi ser som en utmaning att minimera. Detta då den mänskliga rollen i systemet, samt att algoritmer påverkas av den sociala kontext de utformas under, alltid kommer innebära en risk för bias.

Det är intressant att företagen i princip behöver välja vilka bias som bör minimeras, då alla tillvägagångssätt verkar riskera någon form av bias. Exempelvis uppkommer risken för *inherited bias* hos samtliga företag som använder sig av andra företags algoritmer. Fördelen med detta är tillgången till en större databas och därmed en minskad risk för *sampling bias*. Risken för *sampling bias* är högre för de mindre företag som använder sin egen algoritm, då vi tror att deras databaser riskerar att vara förhållandevis små. Som ett ytterligare exempel har företag valt att basera sin tjänst på egna erfarenheter, medan ett annat företag på forskning. Båda tillvägagångssätten har i diskussionen konstaterats leda till olika bias. Utifrån detta kan vi se att bias uppkommer hos företagen och att inget tillvägagångssätt verkar kunna minimera bias fullständigt utan att bjuda in till en annan sorts bias. Detta visar den komplexa utmaningen företagen har framför sig.

Trots att utmaningen är stor, kan vi konstatera att företag idag arbetar på en mängd olika sätt för att försöka minimera bias. Företag arbetar väl med mänskligt agentskap och tillsyn, genom att en mänsklig handpåläggning genomsyrar stora delar av tjänsternas processer; vilket leder till en minimering av bias. Dock är minimeringen inte garanterad, då mänskligt inflytande även kan bidra till bias. Här ser vi en intressant skillnad mellan hur riktlinjen menar att en mänsklig roll minskar bias och hur företags erfarenhet inom branschen visar att mänsklig roll kan öka bias. Vi tror därför det är viktigt att involvera rätt personer som ett sätt att minimera bias, vilket West, Whittaker och Crawford (2019) styrker. Just vilka personer som bör utses för detta inser vi är otroligt svårt att avgöra, sett till riktlinjen om mångfald, icke-diskriminering och rättvisa. Vi ser hur det är väldigt svårt att anamma tankesätt utanför ens egen världsbild. Detta, i kombination med att termen mångfald har en konstant föränderlig innebörd, bidrar till att en hög grad av mångfald är otroligt svårt för företag att uppnå; något som i sin tur kan hämma minskningen av bias. Även om mänskligt deltagande ämnar till att minimera bias, kan människans oförklarliga resonemang riskera att resultera i en *black-box* i AI-systemet. Detta resulterar i en bristande förklarbarhet som ökar risken för bias, vilket riktlinjen om transparens styrker. Vi kan konstatera att avvägningen mellan olika riktlinjer, och därmed tillvägagångssätt för att minimera bias, är komplex.

Utöver den mänskliga *black-boxen* är företag måna om sin öppenhet och efterlever därmed riktlinjen gällande transparens till viss del. Detta är ytterligare ett sätt som företag arbetar med att minimera bias. Företagens transparens gentemot användaren varierar beroende på kundens intresse, samt hur mycket information som kan delges kandidaten utan att rekryteringsprocessens syfte äventyras.

Vi kan konstatera att företag aktivt arbetar med att ta sitt samhällsansvar och strävar efter att vara socialt hållbara genom att erbjuda fördomsfri rekrytering; vilket ämnar till att motverka ett utbrett samhällsproblem. Att företag inte bara är medvetna om problemet utan aktivt arbetar för att hantera det, visar på hur de minskar risken för bias och efterlever riktlinjen utifrån vad som idag är tekniskt möjligt.

Bias minimeras till viss del genom studiens sistnämnda riktlinje; ansvarsskyldighet. Idag efterlever företag denna riktlinje delvis då granskning, rapportering och kompromisser görs. Detta görs inte i den utsträckning riktlinjen egentligen förespråkar, men det minimerar bias delvis.

Under uppsatsens gång har vi kommit till ett flertal generella insikter om området samt hur forskningsfrågan ska behandlas och slutligen besvaras. Utifrån studien kan vi konstatera att AI-företagen generellt anammar de traditionella rekryteringsmetoderna som har granskats och prövats under en längre tid. Att metoderna dessutom genomgår en ny, digitaliserad granskningsprocess och därmed revideras på nytt, anser vi kunna resultera i mindre fördomsfulla AI-system. En nackdel vi har identifierat i de nya, automatiserade tjänsterna är att personer som vill byta eller klättra i karriären kan exkluderas från urvalsprocessen, vilket det intervjuade kundföretaget poängterade som ett problem. Detta anser vi pekar på att utformningen av systemen inte anpassas till alla potentiella användare, vilket Europeiska kommissionen (2019a) menar att systemen bör göra.

Företagen använder sig av olika processer och komponenter för att nå liknande slutresultat; att rekommendera lämpliga kandidater till kund. Något som var en intressant insikt var hur lite avancerad AI-teknik företagens tjänster innehöll. Då företagen marknadsför sig som AI-företag, hade vi en förväntning på mer avancerad AI i tjänsterna. Dock, vilket Europeiska kommissionens (2018) definition av AI styrker, så kan AI appliceras på mer än den generella uppfattningen av begreppet. Rekryteringsbranschens användning av AI är idag i ett tidigt stadi. Vi tror användningen stegvis kommer att öka i takt med att dessa tjänster blir mer tillförlitliga, då branschen fortfarande är traditionell men intresserad av AI-teknikens möjligheter.

Avslutningsvis vill vi poängtera att det är svårt att besvara frågor om vad exempelvis mångfald och fördomsfrihet är, då dessa begrepp är abstrakta och svårdefinierade. Begreppens komplexitet ser vi som ett hinder som försvårar processen med att minska risken för bias i tjänsterna.

6.1 Vidare forskning

Då studiens forskningsämne är nytt och blir viktigare i takt med att samhället automatiseras, ser vi ett stort behov av vidare forskning inom området. Med utgångspunkt i vår studie anser vi att en mer ingående analys bör göras där följderna av olika biaser tydligare presenteras för att underlätta tillfällena där företag måste göra kompromisser. Vidare tycker vi att det hade varit intressant att genomföra vår studie med deltagande observation under en längre tid för att få en djupare förståelse för hur minimering av bias präglar företagens arbetssätt. Slutligen anser vi att vår studie kan användas som utgångspunkt för att analysera hur bias uppkommer samt vilket arbete som görs för att minimera bias i övriga högriskområden som den Europeiska kommissionen (2021) identifierat.

Appendix A

Intervjuguide

Intro:

- Vilka är vi?
 - Studiens syfte är att förklara hur företag som tillhandahåller AI-rekryteringstjänster arbetar för att motverka bias. Då AI-bias påverkar många idag, ska studien belysa de åtgärder och arbete som kan göras för att i sin tur minska bias i tjänsterna.
- Etiska överväganden
 - Rätt att spela in, transkribering kommer ligga som underlag i vår uppsats
 - Samtyckeskravet om ni inte vill medverka så tar vi självklart bort materialet
 - Vill du vara anonym i uppsatsen, ska företaget vara anonymt.
 - Kommer publiceras på lunds uppsatshemsida
- Lite om er och företaget
- Sen mer tekniskt med ML och bias
- EU har riktlinjer
 - Syftet med riktlinjerna är att främja tillförlitlig AI.

Tema Om företaget

- Vilka är ni och vilken affärsidé har ni?
- Hur ser er affärsmodell ut?

Tema ML

- Vad använder ni för AI tekniker för er tjänst?
 - Vilken typ av inlärningsmetod använder ni?
- Använder ni personlighetstester, hur fungerar det isåfall?

Tema Bias inom rekrytering

- Hur gör ni för att säkerställa minimalt med bias?
- Vad anser ni är den mest kritiska processerna/datan som kan bidra till bias i ert verktyg?
 - Varför?
 - Vilken roll spelar kundens tidigare data in i bedömningen?
- Hur delaktig är kunden i algoritmens utformning?
- För- och nackdelar med ert system/verktyg när det kommer till bias?

Tema EU-riktlinjer

Mänskligt agentskap och mänsklig tillsyn

- Vilken tillsyn och insyn har kunden i tjänsten?
- Får kunden en chans att lära sig vad denne kan göra för att motverka bias? Hur?
- Är processen helt automatiserad eller är human-in-the-loop (människa medverkar i varje beslutscykel i systemet) en del av eller liknande integrerat?

Integritet och dataförvaltning

- Hur säkerställer ni att inparametrarna eller inmatningsdata inte bidrar till bias?
- Hur tillgår ni datan som används vid bedömningen av en kandidat?

Transparens

- Hur arbetar/ser ni på transparens?
 - Kan systemet förklara/visa hur det kommit fram till ett beslut?
 - Kan man spåra systemets förlopp fram till beslutet?

Mångfald, icke-diskriminering och rättvisa

- Hur ser mångfalden ut i era utvecklingsteam/träningsdatan?

Ansvarsskyldighet

- Hur kompromissar ni mellan olika tekniska lösningar?
- Finns det någon utpräglad process hos er hur ni går tillväga om man misstänker problem med algoritmen?

Tema Framtid

- Hur ser ni på ert framtida arbete med bias-frågan?

Appendix B

Transkribering av intervju med Tengai

Rad	Information	Aspekt
1	Vem är du och vad gör Tengai?	
2	Namnet är Sinisia och jobbar som produktägare för Tengai och har varit med från start. Det var ett projekt från början på ett rekryteringsföretag som heter TNG, som ni säkerligen känner till, jobbade där som rekryterare och jobbat med rekrytering sedan 2007 på lite olika roller. Men sedan 2018, typ juni, som jag har jobbat med Tengai och sedan så blev det projektet ett bolag som heter Tengai AB i dagsläget och idag så jobbar jag bara med utveckling med olika funktioner så som roboten ska prata bete sig och också hur den bedömer kandidater och hur de scorar och också de ramverket som finns bakom.	R1
3	Vad är det egentligen ni säljer till kunden, vad är er tjänst?	
4	Tengai är ju en intervjurobot, folk brukar säga att det är en rekryterings robot men det är ju egentligen inte det utan den gör ju intervjuer , på så sätt tar den egentligen bara ett moment i rekryteringsprocessen. Rekryteringsprocessen är ju ganska bred, det finns många moment men vi har valt att bara fokusera på det personliga mötet . Fram tills fem -sex veckor sen var man tvungen att träffa roboten fysisk, alltså tengai, att sitta med den fysiskt i ett rum helt och hållet själv, den sköter intervjun själv , men var tvungen att gå till roboten. Men fem-sex veckor tillbaka har vi fixat videointervjuer så att du istället använder teams eller zoom eller google meet så du gör helt enkelt intervju med roboten i din telefon . Det vi levererar till kund är egentligen det vi pratar om fördomsfria intervjuer, fördomsfri bedömning av kandidater, utifrån de frågor och ramverket som roboten jobbar efter . Vi försöker att effektivisera det första urvalet, för vi rekommenderar att man använder roboten tidigt i processen så att istället för att intervjua fem-sex personer så kan man intervjua alla istället . Så att alla kandidater får komma till tals. resultatet efter en intervju är ett score som utfaller , det finns ett automatiskt och ett manuellt det kan jag återkomma i detalj längre fram hur det funkar. Sen är de upp till rekryterare att ta det vidare, men vi har en rekommendation på resultatet, som kandidat behöver du komma upp i en viss siffra för att anses vara "bra", motsvarar de förväntningarna som finns på dig.	R1
5	Videointervjuer? är det ett svar på pandemin eller är det något ni tänkt på länge?	
6	Det har funnits absolut tankar, på att göra en digital app det har bara påskyndats av pandemin. Ja är svaret.	
7	Är det då några funktioner som försvinner när det sker via Zoom eller presterar den lika väl som i verkligheten?	

8	Vi var lite rädda för att det skulle vara sämre, men det har visat sig inte alls stämna. Jag ska inte säga att det har blivit bättre, utan det är exakt lika bra. Nog bättre på grund av att kandidaten inte behöver gå till ett kontor och kan boka in sig själv, även boka om sig själv. Det uppstår ett nytt kandidat beteende, man är ju lite mer “herre över sin egen tid” och göra intervjun när man själv vill.	
9	Vad bygger roboten/produkten på för AI tekniker?	
10	Bygger på NLP, man kan säga så att roboten inte är AI egentligen, utan det som är AI är ju förståelsen för det som kandidaten säger. Att helt enkelt omvandla den ljudfilen som uppstår till text och bygga sitt beslut/nästa steg på det. Till det använder vi Googles ASR, även tittat på Microsoft men Google funkar så pass väl på både svenska och engelska, de språk vi har i dagsläget och vi har inte haft behov av andra leverantörer eller tankar på att byta ut detta. Så att ju bättre den blir på att förstå det svenska språket desto bättre kommer roboten bli på att förstå det svenska språket. i dagsläget så är vi väldigt nöjda. Sen har vi själva byggt intents där vi säger att om roboten hör en viss kommando eller ex antal ord som har ett visst kontextuellt förståelse så agerar den på lite olika sätt. Till exempel att den förstår att detta är ett ja svar eller ett nej svar att här är kandidaten osäker. Det kan också vara att kandidaten ber om mer tid och liknanden, där kan människan, som är kandidaten, uttrycka sig på tio olika sätt, där har vi jobbat ganska mycket med datainsamlign och försökt att förfina det på många olika sätt för att roboten ska uppstå som att den lever och agera och att kandidaten ska få rätt respons.	L1, L2, A1, A2
11	Där är det ni som satt input datan och att den ska generera en viss output?	
12	Ja exakt! Bygger helt och hållet på ord så roboten behöver förstå det ordet och agera därefter. Det är ingen maskininlärning än, när det gäller de mer kontextuella förståelserna eller kanske semantiska förståelse att kandidaten antyder det här i vad det nu kan vara, det har vi inte börjat med men insett att det finns ett önskemål att börja med det, men där är vi inte än. Utan det är enklare saker som roboten behöver förstå utifrån ett svar. Det är väl typ det som är AI i roboten. Resten är väl mer eller mindre hårdkodat så att den betar ju sig på samma sätt alltid mot alla kandidater. Det är väl det som är det som är fördomsfritt att alla får exakt samma upplevelse, exakt samma behandling av roboten.	A1,A2, L2
13	Om ni utvecklar till ML finns det en risk att det hade varit svårt att få den fördomsfri?	
14	Bra fråga, det som vi inte heller har i den datan är vem som har sagt vem personen bakom är, vem kandidaten är, ålder, nationalitet, man eller kvinna osv. det finns ju inte i datan på så sätt tror vi att vi kommer längre ifrån från att vara fördomsfull för vi har bara det som bokstavligen sagts. Vi har inte heller tonläge eller ansiktsuttryck men vi är inte där än, svårt att säga. men jag skulle tro att så länge man tvättar bort den datan och bara bokstavligen har det som har sagts i form av ord så kommer man ganska långt eller rättare sagt behåller det här unfördomsfull perspektivet.	A2, L2
15	Är det ett aktivt val att inte har tonläge/ansiktsigenkännings algoritmer för att inte skapa bias?	

16	<p>Det är ett aktivt val för att det inte finns forskning som säger att/pekar på att om du är glad eller förvånad eller vad det nu kan vara för känsla eller om du är lite smånervös eller att du är extra glad, det finns ingen forskning som säger att om du är på det sättet så kommer du prestera bättre på ett jobb. Så att det gör att det inte finns i roboten heller, det är aktivt val för att vi inte kan validera det på något sätt. Kanske längre fram när det finns studier, för teknolgin finns ju men den är ju inte hundra procentig. Det är jättesvårt för robotar eller för vilken AI som helst att som jag känner till i världen som kan avgöra om jag är förvånad eller inte och att vara hundra procentig. För det är typ det den behöver vara annars så faller ju hela om den är det till 98% så det fortfarande så att 2% kommer att bli bedömda på fel sätt.</p>	L2, L5
17	<p>Den produkten ni har idag har ni upptäckt bias i den, med tanke på ordval och att man pratar på olika sätt från olika delar?</p>	
18	<p>Jag kan förklara lite hur det går till för att det är en kontext: Roboten ställer 20 grundfrågor, sen kommer följdfrågor på det. 16 av de frågorna är påståenden, skulle du säga att du är på det här sättet? Då ska kandidaten uppge antingen ja eller nej, rakt av ja eller nej, eller att den helt enkelt berättar hur vad den tycker och det är här den här NLP/NLU grejen som kommer in i bilden. Du kan prata, eller vi har mätt det att upp till 250 tecken av ett svar och roboten kan avgöra om det är ett positivt eller negativt svar. Det är en ganska hög träffsäkerhet på det, vi är upp i 97,3 eller nåt sånt där. Vilket gör att vi egentligen bara tar hänsyn till det som har sagts och kan avgöra om det är positivt eller negativt i förhållande till påståendet. och därefter kommer då ett följd påstående utifrån ditt ja eller nej. Det tolkas också på det sättet. Sen så har vi 4 öppna frågor så kallad e kompetensbaserade, där roboten helt enkelt samlar in information och spelar in svaret, ljudfil då, den ljudfilen spelas upp sen för en rekryterare i vår mjukvara som heter Tengai recruit. Påståenden är ju knutna till personlighet, till Big Five och där har vi byggt ett automatiskt scoring. För att vi “vet” genom Big Five att om du svara på frågorna på det här sättet så har du en viss personlighet och det kan vi vara ganska säkra på att så ligger det till enligt den forskning som finns om Big Five. Där har roboten redan ett score två sekunder efter att intervjun är klar, så utfaller ett visst score utifrån din personlighet. Sen finns det också dom här 4 frågorna som är öppna frågor, där kan kandidaterna svara och babbla hur länge du vill där kommer roboten lyssna och humma och bete sig som att den lyssnar och låta kandidaten prata. Är kandidaten för kort kommer roboten säga att det här inte är tillräckligt och ge mig lite mer information för jag behöver det för vidare bedömning. Den bedömningen sker av en rekryterare. Där skulle man kunna säga att rekryteraren gör den “andra” bedömningen av svaren att då kommer fördomarna in, för rekryteraren hör då kanske att det här är en gammal människa eller nån som inte kan prata svenska bla bla eller vad de nu kan vara, så är de ju. Men vi komme inte ifrån det därför att det är så det funkar det är där den kontextuella förståelsen kommer in i bilden att den är gnaksa svår. men då kan vi med gptt samvete säga att datainsamlingen har skett på ett fördomsfritt sätt. Ingen kan påverka den, kandidaterna har fått samma chans till att besvara frågorna.</p>	A1,A2, R1, L1
19	<p>Utformningen av de 20 frågorna, hur har ni gått tillväga för att säkerhetsställa att de inte har någon inbakad bias?</p>	

20	<p>Intressant fråga, därför att de har per definition bias i sig därför att de försöker att avgöra/antyder någonting alltså i ett påstående skulle du säga att du oftast tar den ledande positionen i en grupp det är ju en liten antydning på att antingen är du det eller så är du inte det. Per definition är de ju lite fördomsfull, så det kommer vi inte ifrån ska jag säga. hur vi har tagit fram de är utifrån de big five testerna som finns ute i världen och tagit hjälp av två forskare på stockholms universitet Anders och Sofia Sjöberg som då har själva fått ta fram de här frågorna utifrån deras kunskap om Big five. Anders och Sofia Sjöberg är gurus när det gäller psykometri. och då har vi låtit dom validera dessa påståenden helt enkelt, två ggr egentligen, två olika grupper där kandidaterna har fått göra personlighetstest och sen har de fått träffa Tengai och kan Tengai göra samma sak som ett personlighetstest och svaret är ja på den frågan till och med lite bättre till och lite högre validitet än det testet som vi... Testet får NEO-PI (som det heter), så att vi vet att de är validerade på så sätt men det är ju per definition fördomsfull. men det andra fyra frågorna, dom här öppna, är inte det, men där är de ju också så att de inte hellet är prediktiva. För påståenden är prediktiva, svarar du på ett visst sätt är du på ett visst sätt som person enligt Big Five och så har vi också byggt det. Men enligt dom här öppna frågorna kan man säga att där är det deskriptiv analys, dvs där kandidaten kan egentligen bara beskriva sin kompetens, antingen upplevd eller hypotetiskt så vet du lite grann. För en av frågorna är till exempel "ge mig en situation där du löst ett problem" alla har ju löst ett problem, alla kan beskriva typ iallafall hur dom gör då är de upp till rekryteraren i ett senare skede att avgöra hur väl passar du för just den här tjänsten.</p>	A2, L4, R1, L6
21	<p>Vad skulle du annars säga är de mest kritiska när det kommer till bias med roboten?</p>	
22	<p>Hmm du är nog lite inne på det här hur den pratar hur den tolkar vad som sägs, vi har ju inte märkt det för att vara helt ärliga vi har ju haft roboten nere i skåne, också de som kanske pratar mindre bra svenska har ju också. Våra kunder har gjort x antal rekryteringar där det inte heller har varit ett problem med deras uttal och inte heller vad de säger för återigen så handlar det om positivt eller negativt för att roboten ska förstå, och det är ganska enkelt. Men vi människor tenderar att börja våra svar på liknande sätt sen finns det dom där dubbla negationerna där kandidaten börjar så: "Jo men fast jag skulle inte tro det men å andra sidan, fast kanske!" då kommer roboten reagera på det och säga så här att det inte korrekt och rakt svar, kan du antingen förtydliga, summera det eller ge mig ett ja eller nej. Jag skulle säga tal, men vi har inte sett något bevis på det än men och sen är de väl i rekryterarens bedömning om det men i övrigt kan jag inte säga vad det skulle kunna vara helt ärligt.</p>	A1,A, L4
23	<p>Analyserar den på något vis de värde som finns i ord? När det kommer till manligt och kvinnligt språk eller att man kan bedöma ord på olika sätt beroende på vilken person man är och koppling.</p>	
24	<p>Jag skulle säga att nä inget aktivt, därför att det är väldigt simpelt på positivt och negativt. Vi har ju medvetet gjort det på det sättet för att det blir väldigt enkelt att bygga det på det sättet, för du behöver bara avgöra om det är ett positivt eller negativt svar och där blir skillnaderna mycket mindre. Jag skulle tippa på min egen fördom att manligt och kvinnligt kommer in när man börjar</p>	A2, L2

	beskriva saker på ett annat sätt och det gör vi inte i dagsläget vi bedömer ju inte det så.	
25	Hur sker matchningen av score:et gentemot kunden? Hur matchas scoret mot kunden? Hur går det till helt enkelt?	
26	Dom här 16 påståendena är knutna till 6 olika egenskaper som vi har byggt på big five och dom är ganska hårt knutna till samvetsgrannhet, extraversion, emotionell stabilitet. Varför dom tre och inte fem är för att dom tre enligt den forskning som finns korrelerar med arbetsprestation, med generell arbetsprestation. Extraversion och samvetsgrannhet det är amerikanska studien och emotionell stabilitet tillkommer i engelska studier det har man sett ganska stora skillnader. Här kommer vi in med fördomar: byggt på män, manliga chefer i och med att det oftast varit manliga chefer som säljare och liknande som testats och byggt hela big five på det. Men som sagt så finns det mer och mer studier som är mer moderna och då kommer emotionell stabilitet in. Dom sex egenskaperna där kan rekryteraren vikta, vilka egenskaper är viktiga, som har en hög viktning medel och låg, utifrån då den rollen som de rekryterar till. Då läggs helt enkelt kandidatens resultat på den här viktningen och då utfaller ett score om hur nära kandidaten är den här profilen. Och rekryteraren kan också vikta om när den vill om den kanske upptäck att en viss typ av egenskap är viktigare än någon annan och då räknar systemet om de redan gjorde intervjuerna till en ny score, och scoren är mellan 1-4 och som kandidat behöver du komma så nära fyran som möjligt men det finns liksom värden på när du blir grön, halvgrön eller röd.	L4, A1, A2, R1
27	Informeras kandidaten om att företaget gör en viktning om vilken egenskap som är viktig?	
28	Dom får ganska mycket information men dom får inte veta exakt vilka egenskaper och vad dom egenskaperna heter däremot får det veta det är personliga egenskaper och det är från big five så. De får inte vet att de egenskaperna kan viktas, bra grej! men det har vi nog inte gjort faktiskt utan vi har valt att berätta så mycket som möjligt om intervjun men inte att det viktas eller vilka egenskaper det är, nej.	L3
29	Men ni berättar för kunden om hur hela processen går till?	
30	ja de får en utbildning i hur ramverket och hur det är framtaget och hur de själva ska använda det, yes.	L3
31	När ni sitter och utvecklar hur ser mångfalden ut internt?	
32	Vi sitter bara i Sverige, sitter i stockholm och malmö, jag är ensam i stockholm och resten sitter i malmö. ålder på oss; jag är äldst och jag är 42 och yngst. Blandat manligt och kvinnligt, fyra fyra faktiskt bra där det har jag inte tänkt på. Så att det är jag som är på produkt och sen har jag en till som är på produkt i Malmö. Sen är det Elin som är vd och grundare tillsammans med mig och sen är resten utvecklare, vissa på frontend och vissa på backend och	L4

	olika nationaliteter. Alla pratar inte svenska och två pratar engelska. Det har nog löst sig av sig självt.	
33	Hur gör ni när ni upptäcker att det uppstått fel med algoritmen?	
34	När det blir fel, det är väl i så fall när det blir buggar som kan uppstå. Hur det går till är väl att vi använder oss av Jira som vi använder dagligen både om att lägga in tasks, buggar, stories skapa nya epics så det är väl så som vi jobbar med att rapportera in om vad som ska åtgärdas och prioriteras. Vi jobbar i sprintar som ska vara klara på två veckor. Så att vi hade faktiskt igår ett sprintmöte där i började en ny sprint där vi körde lite retro. lite demo och planering för de närmaste två veckorna. Det Är oftast jag som ställer till problem för det är oftast jag som upptäcker något problem som ska ändras och ja då blir det lite diskussioner om vi ska prioritera om eller om vi ska fortsätta så som det är. Det är ongoing testning som sker självklart.	L6, L4
35	Är ni medvetna om någon bias och hur arbetar ni iterativt för att förebygga dessa?	
36	Absolut och främst skulle jag nog vilja säga i vad roboten säger till kandidaten och på vilket sätt. Det är en ständig utveckling, därför att jag/vi märker att i kandidaten hur de uppfattar och vrider på svaret. Så det är definitivt ett ongoing arbete. Vi tänker på det ständigt men det syns inte hela tiden, som till exempel hur ansiktet ser ut i dagsläget så finns det en sån önskan om att kandidaten själv ska kunna önska om vilket ansikte som den blir intervjuad av. Det är en sån där längre fram plan som vi tänker göra. I dagsläget går inte det då är en slags projektor som gör så att det blir ljus, eller det går att göra ett mörkt ansikte men skulle fortfarande vara ljust på grund av projektorn exempelvis. Den kritik vi har fått har kommit från att varför det är en vit person och varför det inte är en vad de nu kan vara. Det är helt enkelt teknik.	A2, L1, L4, L5
37	Hur gör ni annars när ni kompromisser mellan olika lösningar och problem? Vilket är prio ett, viktigaste med produkten?	
38	Viktigaste är kandidatupplevelsen. Det har det varit och kommer alltid vara det viktigaste för alla. Att kandidaten upplever att den blivit behandlad på ett bra sätt. Vad är ett bra sätt egentligen? Joo på ett professionellt, vänligt och ett sätt där kandidaten har fått berätta om sin kompetens på ett fördomsfritt sätt. sen behöver inte det fördomsfria sättet uppfattas positivt av den kandidaten, de ska jag ju säga, man är ju van vid att få hjälp på vägen när man blir intervjuad, styrd osv. Man är inte medveten om att det är de rekryteraren gör, den styr ju dig för att den tror och antar saker om dig. Det kan uppfattas som ganska kallt för att roboten hjälper inte och bekräftar en på det sättet osv. Det finns ju aspekter av det att för att det är en fördomsfri robot, den säger aldrig tack det är bra. Den säger tack. medan en rekryterare skulle säga tack bra svar! Doesn't happen!	L6, L4
39	Är det som ni fått som feedback av er produkt, att det är obehagligt?	

40	Inte obehagligt tycker dom, de som tycker det tycker de dom första minutrarna och sen så har de glömt bort att det är en robot efter 1-3 frågor. Det är nog två anledningar till det: att den beter sig tillräckligt som en människa, vi lurar hjärnan att det är en människa. Och nummer två är det är ett starkare driv att du fått en fråga och du vill besvara den frågan, vare sig det är positivt eller negativt vill du besvara den frågan. Det är mänskligt på nåt sätt, får du en fråga vill du besvara den. Så att vi spelar väldigt mycket på biologi i produkten.	
41	Det känns som att er allt er arbete baseras på forskning? Stämmer det?	
42	Japp så är det. För det finns inget annat sätt att göra det på för ingen annan har gjort något sådant innan. Det är väldigt enkel anledning till varför det är som det är. När vi började hade vi ingen aning om hur vi skulle göra detta så vi tittade på vad som fanns för forskning.	L6, R1
43	Ni tar inte in någon data? CV eller något de matar in själva?	
44	Nej utan det är bara intervjun. vi har gjort det väldigt enkelt för oss själva, utan vi har byggt ett ramverk och det är det vi gör. Vi ändrar inte på någonting inför en enskilt rekrytering eller konsult eller kandidat utan det är vad det är.	L4
45	När roboten får lära sig, vad är det för data som används? Hur säkerhetsställer ni att bias inte råkar bakas in i språkinläringen?	
46	Hmm ja egentligen tillbaka till det svaret innan, den datan som tolkas och skickas till NLP/NLU är helt och hållet tvättad från det som har en mer personlig information att göra. Man vet inte ålder, nationalitet, namn på kandidaten ingenting sånt, det är en ren ljudfil som tolkas till text och sen används texten till att bygga förståelsen. Ganska tror jag och hoppas jag att vi kommer ifrån men definitivt kan det finnas risk beroende på för vilka ord man använder. Den risken finns ju, inget snack om saken. jag tror väl att vi kommer ifrån den, även om det finns kandidater som talar på ett annorlunda sätt är de inte representativa på gruppnivå. Vi kommer ju fånga in dom absolut menade, kommer kanske var 1-5 procent, men det är inte fler så på sätt så så länge vi håller så kommer vi en bra bit på vägen eller åtminstone minska risken. Vi är jätttransparenta om vi blir ifrågasatta om vår produkt och AI, vi har liksom ingenting att dölja.	A1, A2, L3, L4, R1

Appendix C

Transkribering av intervju med Exparang

Rad	Information	Aspekt
1	Vem är du och berätta lite mer om Exparang	
2	Jag är Filip Strömbäck och jag är medgrundare till Exparang som kom ur idén av att egentligen automatisera och förbättra sourcing av chefskandidater eller nyckelpositioner i näringslivet. Jag har jobbat med headhunting eller chefsrekrytering i tolv-tretton år innan detta startades tillsammans med två medgrundare, den andra har samma bakgrund som jag och den tredje är CTO och har egentligen ingen erfarenhet av rekrytering på det sättet utan bara egenskaper av att ha många år erfarenhet av att vara chef. Men han har byggt produkten och vi har liksom bett om innehållet i produkten utan att ha någon som helst erfarenhet av IT sådär sen innan. Tanken var att bygga ett verktyg som anonymiserar kandidat-data för sin matchning och det är det främsta argumentet för att vi vet att om man som kandidat får lov att vara anonym i första steget i en kontakt så är det mycket högre sannolikhet att man är ärlig om vad man vill tjäna vad, vad man vill bo, vilken bransch man vill jobba inom, vilka jobb man vill ha och sådär. Så vi får fram ärligare matchnings-data genom att låta vara folk anonyma, sido-effekten blir då att vi också får fram mer mångfald i processen då första steget är anonym. Så det är ett högre syfte vi har.	R1
3	Vad är det ni säljer egentligen till företagen? Är det just den här matchningstjänsten?	
4	Ja exakt, de som betalar för vår tjänst är företag. Vi har två sidor kan man säga, vi är en sådan double-sided plattform. Vi bygger en databas av kandidater och kandidater registrerar sig då på ena sidan och sen så har vi då kunder i form av företag på andra sidan som betalar för att använda plattformen till sina rekryteringar. de laddar upp ett jobb och får automatiska matchningar genom den delen av databasen som passar just för det här jobbet. Och dem betalar på lite olika sätt men den långsiktiga planen är att kunna ha företag som på abonnemangs-basis använder detta verktyg hur mycket de vill. Men i dagsläget gör vi på väldigt olika sätt, det kan successfeedbaserat (?) så att du gör en grej och är i behov av en rekrytering och då hjälper vi till med bara den. Och i vissa andra fall som med rekryteringsföretag som gör flera hundratal rekryteringar på ett år så har vi mer av ett abonnemang där man får nyttja verktyget för ett fast pris.	A1, R1
5	Vad är liksom ert främsta säljargument, just när ni säljer in detta verktyget till företagen?	

6	<p>Främsta säljarargumentet är faktiskt att sourcing av kandidater går väldigt fort och det är extremt billigt på kort sikt i och med att vi har automatiserat den delen av processen. Detta är ju en del av en rekryteringsprocessen och man går till traditionell executive search firma så tar detta ungefär fyra veckor. Arvode till en rekryteringsfirma ligger idag på någonstans mellan 200 000 till en miljon för att rekrytera en chef, och vi löser ofta sourcingbiten på mindre än 48 h och kostar några tusenlappar om du bryter ned det per rekrytering. Så det är tid och pengar framförallt som är den stora grejen. Och än sålänge har vi en såpass liten databas, så vår stora utmaning är att fylla databasen med tillräckligt många kandidater så vi kan vara det enda verktyget som våra kunder använder. Men idag är vi oftast ett komplement till en annan process, att man sköter rekryteringen internt eller att man outsourcar det till en firma. För nu har vi bara ca 1500 kandidater i databasen och vi ska bli 10 000 innan året är slut, men det är fortfarande inte tillräckligt mycket för att vara det enda verktyget de kan använda. Men ett annat säljargument är just det här att vi lägger oerhört mycket tid på att bjuda in kandidater och aktivera kandidater som vanligtvis inte finns i en executive search firmas databas. Syftet blir på så sätt att få upp mer oväntade namn i rekryteringsprocesser eftersom matchningsdatan får tala sitt egna språk. Det har inget att göra med vilket nätverk man har, vem man känner osv för våra kandidater. Då kan vi lika gärna fylla på med hur många kandidater som helst. Sålänge matchnings-datan talar sitt egna språk så tror vi det kommer bli lättare för våra kunder att hitta kandidater som de vanligtvis inte kommer i kontakt med.</p>	A1, A2, R1, L2, L4
7	<p>Och hur går själva den grejen till? Genomför ni exempelvis personlighetstester eller är det mer CV-screening. Hur får man fram kandidaternas data?</p>	
8	<p>Kandidaternas data idag är än sålänge helt och hållet byggt på kandidaten själv. Inledningsvis när vi startade det här började vi inledningsvis undersöka hur data från LinkedIn, som är den mest naturliga plattformen att hitta data om jobb, men det är ganska utmanande att bygga något som integrerar information från LinkedIn, för de släpper inte med så mycket information. Än sålänge är vi rätt beroende av att kandidaterna själva tar sig tid och fyller i sin information men vi jobbar, på tal om AI, med kluster-teknologi så vi får hjälp av NLP så försöker vi att klustra ihop, förstå vad olika titlar betyder, att en economicchef är det samma som CEO, med det kommer vi kunna ta data. Vi behöver något form av CV från kandidaterna då, de behöver länka till sin LinkedIn, och de behöver ladda upp CV. Med hjälp av AI kan vi ta den datan, tolka den och skapa deras profil. Så det är ambitionen, men det är ganska dyrt och en ganska komplex process - just rekrytering av chefer eller generalister på en högre nivå. För en marknadschef i ett litet svenskt bolag på en viss nivå är inte alls samma sak som en marknadschef på ett amerikanskt, börsnoterat bolag. det är helt olika roller men exakt samma titel. det kommer krävas oerhört iterering och testning för att komma i närheten av detta, så vi behöver våra ögon och screening och mänsklig touch i processen en ganska lång tid framöver. Det kan inte vara helt automatiserat.</p>	A1, R1, L1, L5
9	<p>Är det något ni har påbörjat nu, eller är det idag helt utan AI i er matchning?</p>	

10	<p>Vi har det här med klustertänket och det med att få in data på det sättet, det har vi haft sedan början. Så det har vi redan, men vi kommer ta mer och mer utrymme i det att skapa kandidatens historiska data-profil. Sen så är vi fortfarande lite ödmjuka inför att använda AI för att matcha och rekommendera baserat på annan marknadsdata. Där har vi sett så många rekryterings-exempel som har gått snett. Där man har använt sig av marknadsdata för att matcha folk till jobb. Det finns en oerhörd stor risk för att det blir tvärtom och snedfördelat istället för fördomsfritt. Det blir väldigt snävt och plötsligt är det bara vita män som är 43 år gamla och har pluggat MIT som får alla jobb. Och det är varför det också behövs mänsklig touch för att bygga databasen och hur kandidat-urvalet blir. Det blir väldigt mångfacetterat. Det är vårt ansvar att se till att urvalet som sker inte bara är dessa 43 åriga männen från MIT utan att det är mycket bredare målgrupp.</p>	A1, A2, R1, L1, L4, L5
11	<p>Hur har ni jobbat nu med den AI ni har? Är det NLP och hur har ni jobbat för att det inte ska uppkomma bias?</p>	
12	<p>I det fallet så med tanke på att vi är ganska små och vi har en sluten databas så kandidat tillflödet rasar inte in, vi bjuder in. Detta betyder att i princip varje person som kliver på i databasen hade vi själva kunnat screena och kurera. Vi kan själva försöka testa oss fram. Vi har gjort som exempel körningar i databasen för att förstå alla som är relaterade till försäljning baserat på vad de har gett oss för data i sitt CV eller LinkedIn-profil. Så vi har tittat på databasen ur det perspektivet, hur många av våra kandidater jobbar med försäljning. Det blir ganska träffsäkert men det är långt ifrån bra skulle jag säga änsälänge. Dels behöver vi fler kandidater för att det ska bli vettigare, vi behöver över 5000 Cv:n att processa och träna på för att det ska bli någon vettig slutsats. Som jag var inne på tidigare med roller och titlar, key account manager är exempelvis en roll som är i allra högsta grad försäljning men det står det inget om i titeln. Så de hamnar på olika ställen och det är något vi får förebygga.</p>	R1, L1, L4, L6, A2
13	<p>Okej, så ni går alltså in idag och kollar själva - är det så ni kontrollerar att det inte ska uppstå bias?</p>	

14	<p>Att det inte ska uppkomma bias i kund-urvalet när de träffar kandidater är två olika delar kan man säga. Den ena är att vi påverkar inte själva matchningen, om ett företag laddar upp en rekrytering så sker matchningen helt automatiskt mot kandidat-urvalet. Där har vi ingen handpåläggning då vi byggt det på ett sätt med rutinmatris som utifrån ett femtiotal items som våra kandidater svarar på i ett formulär, som en survey. Och företagen som lägger upp en annons svarar på frågor som är direkt kopplade till de frågor som kandidaterna svarat på. Vi har helt subjektivt poängbedömt varje moment där. På språkfärdigheter, om det är en-till-en-match att företaget vill ha engelska som språkkrav och en kandidat kan engelska är det 100 poäng säger vi. Vid funktionell erfarenhet säger företaget att de vill ha erfarenhet från financial control vilket en kandidat har, så kanske det är fem gånger så viktigt som språket. I rekryteringssammanhang är det väldigt viktigt att den funktionella erfarenheten är hundraprocentig. vi kan inte matcha en art director med en finans-roll, det blir helt fel. Vi skulle kunna matcha någon som har lite knackig engelska där engelska är ett krav. Det är inte super men inte helt fel heller. Det är rent subjektiv som vi satt dessa poäng. Så alla kandidaters formulär summeras till en match score. Om denna score överstiger en viss procent mot ett matchad uppdrag så blir man automatiskt matchad. Så där har vi ingen handpåläggning, det enda är att vi går in i vårt system och skruvar på poängen allteftersom. Det gör vi hela tiden när vi märker att något slår fel, om vi får för många marknadschefer matchade till ekonomichefs-roller - varför? Och då kan vi se att det är på grund av en faktor vi satt för hög poäng på. Så det gör vi manuellt.</p>	A2, R1, L1, L3, L6
15	<p>Kan ni se att det finns en inbäddad bias i detta frågeformuläret?</p>	
16	<p>Självklart, det gör det definitivt. Vi är tre stycken gubbar, män födda på 70-talet som byggt formuläret. Nu har vi en fjärde medgrundare Valerie som är tjej och tolv år yngre än vad vi är och tekniker. Produktperspektiv och så sedan hon har kommit in så har vi fått lite annat perspektiv på hur frågor ställs. Men det är något jag tror att vi aldrig kommer komma ifrån det, att beroende på hur frågor ställs så finns det risk för fördomar.</p>	A2, L4
17	<p>Låter ändå som att ni är väldigt medvetna om det då, jag tänker om ni håller på att utveckla handlingsplaner för att hantera det?</p>	
18	<p>Vi försöker ju se till att så att vi adresserar. Sådär, för oss är det viktigaste att se till att lyckas med att matcha till dessa nyckelpositionerna, vårt uppdrag är att se till att det blir bra folk på chefsbefattningar. Vi skiter lite i vem de är. Vi har egentligen inget köns-fokus eller mångfalds-fokus per se, men det vore riktigt idiotisk att snäva ner vårt urval. Vi vill att det ska bli så bra som möjligt och då måste vi inkludera alla tänkbara kandidater och det är lika viktigt med hudfärg eller kön som ålder. En oerhört vanlig omedveten eller medveten fördom i vårt segment är att en högutbildad chef tenderar att rekrytera andra som kommer från samma universitet. Det är en fördom som vi försöker bekämpa genom att vi outar inte vilket universitet man har examen från för det ska inte betyda någonting, om man varit på Chalmers eller i Lund.</p>	A2, L4
19	<p>Kan kunden gå in och påverka?</p>	

20	<p>Det är en jättebra fråga, jag tror det är en väg för oss att ta. Våra kunder vill såklart påverka; vi har en dialog med ett konsultföretag just nu som rekryterar management konsulter - de har en önskan om att bara vilja ha kandidater från ett visst antal lärosäten och de vill inte ha kandidater som jobbat på x antal konsultfirmor för de känns inte lika heta. Där blir det fel för oss, vi kan inte tänka på det sättet utan vi måste än sålänge känna att vi äger matchningsbiten av vår produkt, så kunden kan inte göra någonting egentligen. De kan fylla i formuläret, sen kan de välja att skita i att fylla i vissa delar om inte det är viktigt för dem, men de kan inte själva bestämma och vikta om vad som är urvalskriterier. Sen finns det tanke på att i förlängningen sälja vårt verktyg till företag och då kan de företag använda det i sin egen regi men där är vi inte än. Så kunden kan inte påverka idag.</p>	A2, L1, L2, L3, L4
21	<p>Bara för att förtydliga, det är alltså dessa frågor ni ställer som avgör själva matchningen?</p>	
22	<p>Ja exakt. Det är ett ganska bökitigt formulär, som kandidat är det många frågor, mycket som har med historisk data att göra i och med att vi vill att det ska bli rätt. Vi har sett andra initiativ i vår bransch som man sett misslyckas på grund av att man har tagit för lätt på hur svårt det är att göra matchningar och göra ifyllningen av data lekfull så man ska bli sugen på att göra det. Då ställs frågor som bara är härliga, "vad man drömmer om fem år" och "vad tycker du är kul" men det räcker inte för att göra en kvalitativ matchning. Vi behöver tråkig data som hur många år du har varit, antalet år i sig är inte det viktiga men vi behöver data för att skapa en vettig profil. Men precis så, kandidaten fyller i en massa items och företagen fyller i lika många. Företagen är ju rätt vana vid detta, de ska lägga ut en annons på linkedin eller liknande, så måste de göra detta så för deras del är det nog ok.</p>	R1, L2, L6
23	<p>Du säger att ni har valt ut frågorna, hur har ni egentligen kommit fram till dessa?</p>	
24	<p>De bygger på att potpori av grejer. Från början är det jag och Carl som jobbat med executive search och intervjuer i så många år och gjort tusentals intervjuer och reflekterat över vad som är viktigt att ha med sig in i den första intervjun med en kandidat. Vi har gjort en produkt som bara sourcar fram relevanta och intresserade kandidater. Så det är en första screening, sen börjar processen med intervjuer och sådär. Men vad vill man veta om en människa när man drar igång en intervju? Man behöver ha koll på vissa dealbreakers, lönenivå ska vara rimlig med jobbet, erfarenheten ska vara ungefär vad som förväntas och att kandidaten är intresserad. Att plocka fram de på någon form av lista av kandidater. Med det som utgångspunkt har vi byggt formuläret utifrån tre delar: aspiration: vad vill du göra, var vill du bo, vilken typ av roll vill du ha. experience: vad har du med dig in? values/kultur: vill du vara i ett stort bolag, tillväxtorienterat, börsnoterat, statligt? Några särskilda värderingar med att arbeta med hållbarhetsfokus. Med det som grund har vi försökt plocka in relevanta saker, EU:s miljöpolicy med 17 globala mål är en fråga, Man tvingas som kandidat besvara vad som är intressant för den själv. Man gör en självskattning, om man är nätverkande eller analytiker. Det är inte baserat på någon vetenskaplig grund utan subjektivt byggt upp från vår sida. Det är i övrigt frågor som vi själva</p>	A2,R1, L2, L4, L6

	kommit på tillsammans med en testgrupp av kunder i vårt nätverk som fått testa detta och se om det är rimligt	
25	Hur har den testgruppen varit utformad? har det funnits en mångfald i den?	
26	Vi är i ett tidigt stadi, vi släppte en nyversion i oktober som vi är igång med nu och så utvecklar vi hela tiden. Vår testgrupp förändras därmed hela tiden. Dels har vi utvecklare i Polen, både kvinnor och män. Kulturell och geografiskt en grupp människor som är annorlunda från mig och Carl. Sen i Valeries nätverk har vi en del produktägare som har jobbat med andra typer av lanseringar som kommer från olika platser med andra nationaliteter. Våra kunder är ganska homogena, vi har inte haft en strategi där. Uppdragsgivarna är å andra sidan män och kvinnor i olika åldrar. Det är ingen medveten strategi men inputen är ganska spridd i åldrar och kön iallafall.	A2, L4, L6
27	Om man upptäcker att produkten är fördomsfull på något sätt, finns det en process för hur man går tillväga isåfall? har utvecklarna en chans att komma med input exempelvis?	
28	Ja det gör de. Nu är de ingenjörer och har sitt sätt att se på världen och pratar man med en gäng marknads-personer så har de ett helt annat sätt att se på världen men vi har varit väldigt öppna för att involvera alla i vårt team, oavsett om det är fast anställda eller konsulter. Alla är involverade i ett beslut som har att göra med hur vår produkt ska förändras. Alla får tycka till hur frågor ska formuleras, om de behövs, vilka alternativ som ska finnas osv. Vi kallar det för product steering committee. Det är egentligen vi grundare, men vi tar in data från alla. Ibland blir det ett väldigt enkelt röstningsförfarande och ibland lite mer diskussion kring om vi vet att det är viktigt för affären med någon annan även om någon inte håller med så kan vi ändå tvinga igenom det för att våra kunder vill ha det.	A2, L4, L6
29	Denna product steering committee, har ni en punkt som är fokuserat på etik och bias eller är det något som ni tar upp mer allteftersom det uppkommer?	

30	Den gruppen arbetar väldigt agilt, så vi arbetar i sprintar med produktutvecklingen och varje sprint innehåller en massa saker vi måste ta hänsyn till, såsom kundvärde, mångfald, kandidattillflöde, kostnad. Det finns inte en specifik punkt skulle jag säga för något, allt blandas ihop. I och med att vi växer så kommer vi nog behöva strukturera upp det där så att det blir på ett annat sätt men det har vi varken råd med eller tid till änsålänge.	L1, L6
31	När ni väl kommit fram till en matchning, visar ni scoren för varje område till kunden och visar hur det nått en viss matchning?	
32	Nej, kunden ser ingenting i princip. kunden får reda på att det finns ett antal matchningar till jobbet, men de vet inte vilka av matchningsfaktorerna som är starka eller svaga - de vet bara att den totala bilden av denna kandidat matchar. Kunden får se en anonym profil på kandidaten som en avpersonifierat CV och efter detta inleder kunden och kandidaten en dialog som i slutändan leder fram till att kandidaten kan välja att blotta sin identitet för att gå vidare till en intervju. Kunden ser ingenting. Det har vi pratat mycket om, just vad man ska dela med sig om. I dagsläget i produkten, detta kommer förändras snabbt, redan innan ni skrivit klart, för att vi hittar fel. Vi har en procentsats på varje kandidats profilkort som kan vara väldigt missvisande. Internt för oss kan det vara relevant och syftar till total matchscore och den bygger på hur mycket du som kandidat har fyllt i. Låt oss säga att du har fått femtio frågor att fylla i, och du bara fyller i 25 av dessa men dessa svar är spot on, så får du fortfarande inte en jättehög matchscore för du har fortfarande inte fyllt i hela formuläret. Sen kan en person som fyllt i hela och fått matchningspoäng på var enskild grej, men var enskild grej kan vara en sämre matchning än den här andra personen som bara fyllt i hälften så blir den förstas matchscore ändå högre. Så den procenten är jättemissvisande och ska ryka. Vi diskuterar om detta ska ersättas med någon form av ranking men kommit fram till att vi bara ska dela med oss av information som är så neutral som möjligt även om det inte finns något som är helt neutralt. Det var ett långt svar på frågan men de får inte se det nedbrutet.	R1, L3
33	Och poängen med detta är att det ska vara så neutralt som möjligt?	
34	Huvudsyftet är exakt det, ett exempel är att en kund rekryterar en CFO och bland matchningar finns det person som idag jobbar som marknadschef så är det väldigt lätt att denna kund säger att det inte är relevant utan att tänka längre än så, men i det verkliga fallet kan det finnas en massa erfarenheter i bagaget som är tillräckligt starka rent skill-set-mässigt i rollen och det kan finnas jättestarka motivationskrafter; som att gå till ett visst företag, jobba på en viss plats som alltihop är mycket starkare än att personen idag jobbar som marknadschef. Man får försöka kämpa mot kunden där och hjälpa kunder att förstå hur lätt det är att falla dit i att vara omedvetet fördomsfull.	L2, L3, L4
35	Så detta är något som ni säger under er säljprocess, för att de inte ska vara så fördomsfull då?	
36	Ja, verkligen. Vi pushar våra kunder att connecta med kandidater som de spontant känner att det inte är rätt. Vi pushar för att man ska connecta, testa, connecta, testa. Man vet inte.	L4, L5

37	Vad har kunderna för insyn i själva processen, om man ska skilja på information de får om kandidaten och information om hur matchningsprocessen går till?	
38	De är medvetna i olika utsträckning, vissa kunder bryr sig inte om det utan vill bara ha ett antal namn snabbt på bordet och skiter lite i på vilket sätt. Den allra största andelen av företag vill bara anställa någon, att det ska gå fort och vara billigt. Vi är fullt transparenta med hur formuläret fungerar och vilka matchningsvinklar är och sådär. Så det är egentligen bara en fråga om tid och intresse från användaren på kundsidan.	L3
39	Lite mer ingående om denna AI ni har, hur är denna matchningen AI?	
40	I själva matchningen så är det ingen AI idag, utan det är bara i datainsamlandet på kandidaten som är AI. Det är bara i datainsamlanden som i att bygga kandidatens profil. Men själva matchnings-motorn är en programvara som vi byggt, som bygger på den poängscoring så där finns ingen "intelligent förutom våra inputs" haha. Och där kommer punkten som är jätteintressant med matchning, att så fort man börjar bygga in någonting smart i matchning så blir det jäkligt obehagligt och svårt. Där har vi ett kundintresse, såklart, och det finns mängder av vägar att gå. Vi kan till exempel guida våra kunder i lönenivåer baserat på lönedata-uppgifter; så när de registrerar ett uppdrag om en ny CFO-tjänst kan vi rekommendera en lönenivå baserat på AI och hur många användningsområden som helst. Men jag tror att vi behöver vara försiktiga och väldigt ödmjuka inför att det kan bli fel. Men i datainsamlandet tror jag man kan jobba så mycket man vill med automatisering för att skapa profiler, men matchningen får vi se hur det går.	A1, R1, L5
41	Och då är det att de förstår att flera ord kan betyda en roll, det är det man använder AI till för att skapa själva profilen?	
42	Idag är det det, idag är det det ja. Vi kommer kunna bygga ut detta ganska mycket, med klusterteknologi som generellt är väldigt intressant, som netflix kategoriserar filmer - på precis samma sätt kan vi kategorisera kandidater på deras erfarenhets-paket till exempel.	A1

Appendix D

Transkribering av intervju med Hubert

Rad	Information	Aspekt
1	Berätta lite mer om Hubert och vilken roll du har?	
2	Vi har hållit på nu i fyra års tid, 5 års tid faktiskt. Det började faktiskt som ett verktyg för att samla in medarbetarfeedback, genom de här automatiserade intervjuer. Det gjorde vi de första 1,5-2 åren någor sånt, under tidens gång under tiden vi sålde in det här verktyget fick vi mer och mer request på att använda det inom just rekrytering, så de implementerade vi till det för ungefär 2,5 år sedan. Vi har haft en fungerande produkt i 2 år ungefär. Jobbar dels med lite större bemannings eller rekryteringsbolag och även lite snabbväxande startups som voi, foodora, manpower, academic work osv. Jag är då grundare och marknadsansvarig.	
3	Hur ser er affärsmodell ut?	
4	Det ser ut som mer delar än vad det faktiskt är på hemsidan, affärsmodellen är faktiskt ganska enkel. Man betalar per volymbaserad affärsmodell. Minska paket ligger strax över 1000 intervjuer i månad, och kostar 8000kr. Sen ända vägen upp till vårt största paket, upp till 500.000 intervjuer på ett år.	
5	Men om vi säger att en kund till ska anställa folk, hur ser den processen ut med ert verktyg?	
6	Brukar alltid börja med en pilot för det första, där sätter vi upp en snabb testintervju där vi utgår dels för rekrytering för jobbet, existerande frågebatteri och ja vi kollar lite på vilken kompetenser som krävs för det här jobbet och så försöker vi omsätta kundens existerande frågor för att passa i det här chat interfacet. I vissa fall kan vi integrera med ATSerna och då gör vi såklart det, i andra fall så får man då fylla i sitt namn och e-mail när man kommer in i chatten för första gången så vi kan koppla chatten till rätt kandidat. Sen så är det egentligen bara så att kunden bjuder in kandidaterna om det inte sker direkt via ATS, sedan rullar det på, kunde håller koll på resultaten under tiden och där kommer då alla kandidater in löpande och blir rankade och sortera efter högsta till lägsta score.	A2, R1
7	Arbetar ni med AI och isållfall vilka tekniker använder ni?	
8	Vi använder ett par olika AI-tekniker men den största delen använder vi i rankingen då, i själva chat-interfacet så är det en annan stort AI mycket mer NLP-baserat. Men i rankingen så använder vi ett gäng olika tekniker, vi har precis börjat experimentera SEO-shot delar med för att komma framåt i rankingen, men det krävs nästa olika tekniker för varje månad, så vi testar lite olika för att se vad som funkar bäst.	A1,

9	Vilken träningsdata ni använder för att träna NLP algoritmen exempelvis?	
10	Systemet är upplagt på olika intensions så att Hubert försöker förstå vad kandidaten svarar eller frågar och sedan koppla det till ett relevant svar. Ett intension är till exempel om man frågar, har du varit i en stressad situation någon gång, kan du berätta lite om det och utkomsten? Och så säger kandidaten, nej jag har aldrig varit i en stressad situation. Då fattar Hubert att detta är ett nej och att det behövs en uppföljningsfråga, så ställer Hubert frågan men någon gång måste du varit i en stressad situation, kan du inte berätta om det? Det finns då ett par tusen intensions som Hubert är tränad på.	A1, A2
11	Är det ni(era utvecklare) som valt in- och outputen då?	
12	Det kan man säga, från första början hade vi inga intensions alls, sen försökte vi förstå vad som återkom ofta och träna upp Hubert att känna igen de här olika intensionerna. Det sker dels genom de svar som kommer in från kandidaterna, men också att vi själva sitter och skriver in påhittade svar.	A1, A2, L1, L2, L3
13	Har ni kontrollerat de frågor som ställs i chattet och hur har ni kommit fram till frågorna?	
14	Genom samtal med kund och sedan har vi uppträdande frågor för olika kompetenser. Så vi har liksom tjänsten på högsta nivån, sedan kompetenser och sedan frågor som mäter dessa kompetenserna.	
15	Hur har ni kommit fram till de frågorna som mäter kompetensen?	
16	I samarbete med kunder. Vi har utgått exempelvis från academic works befintliga frågebatteri, och de använder då den här frågan för att mäta om en kandidat är strukturerad exempelvis och då har vi omsatt den i Hubert.	L2, L4,
17	I hela denna processen finns det en risk att det uppstår bias, vi har sett att ni marknadsför er med att det minskar bias i rekryteringsprocessen, hur arbetar ni för att säkerställa att det är minimalt med bias under hela processen?	
18	Det som de flesta liknande verktyg gör tror jag är att man försöker imitera den vanliga processen när man arbetar med strukturerade intervjuer eller kompetensbaserade frågor, då utgår man då från att detta är ett exempel svar för att få poängen poängen 1-5 och sedan försöker man jämföra svaret som kandidaten kommer med med de olika referenssvaren. Men vi har försökt vända på det där, för det är där den mesta biasen uppstår, för en AI är inte sådär superbra på att jämföra ett två svar som är helt olika. Vi utgår istället från de här kriterierna för de olika poängen, ex. Ett kriterium på en tidigare erfarenhets fråga, alltså om man har specifik erfarenhet från den exakta tjänsten. Kanske under och över 3 år, ger en 4 eller 5a, alltså högst eller näst högsta betyg, har man ingen erfarenhet alls så får man en 0 i score, har man liknande erfarenhet under och över 3 år kan det vara en 3a.	R1, A1, L5
19	Hur är frågorna utformade?	

20	Vi utgår alltid från de satta kriterierna som vi sätter tillsammans med kunderna, eftersom att alla blir bedömda på samma kriterier så tänker vi att det blir mindre bias till följd av det i och med att alla blivit bedömda på samma sätt och AI:n tar inte hänsyn till något annat än om de uppfyller kriterierna eller inte.	A2, R1, L4
21	Hur säkerställer ni att de inte bakas in bias i de kriterierna som sätts?	
22	Bra fråga, men vi sätter upp de tillsammans med kunder och vi utgår från att kunderna har existerande processer. Så visst, om de är bias i deras existerande processer så är det möjligt att det fortplantar sig till Hubert, men vi arbetar nästan uteslutande med ganska professionella rekryteringsbolag och rekryterare så vi tror att vi har rätt bra koll på detta.	A2, L5, L6
23	Berättar ni för kunden hur hela processen går till och exempelvis vart det kan uppstå kritiska problem med bias, hur transparenta är ni med er process mot kunden och kandidaten som genomför processen?	
24	Vi brukar inte gå in jättedjupt på hur på exakt hur bias kompensationen fungerar, det är ganska lätt för kunden att förstå att om du inte har någon mänsklig kontakt mellan kandidat och rekryterare tidigt i processen, så förflyttas hela risken för bias till ett senare skede för att du inte får det här uncunisos bias situationen när du träffar kandidaten och kanske börjar småprata med den och ni kanske har gemensamma intresse, eller något som gör att du kanske betraktar kandidaten på ett mer gynnsamt sätt. Eftersom vi utgår från de här satta kriterierna så finns det ju ingen risk för det i de tidiga screening delen. De är ungefär det vi brukar säga. Men om de går in på bias så förklarar vi det.	L3
25	Läste att ni samlar in mycket data om kandidaten och deras digitala fotavtryck, hur går den processen till och är detta något ni gör?	
26	Det var mer en tanke som vi hade för ett tag sen som vi inte har gått vidare jättemycket med. Tanken var då att dels få in data från personlighetstestning, cv parsing, ibland annars vissa kunder personlighetstester före eller efter hubert intervjun, men den datan är inget vi tar del av eller hantering.	
27	Det är inte ni som gör cv parsningen heller?	
28	Nja, vi har experimenterat lite med det men vi har inte fått sånt större intresse från våra kunder kring det. De flesta vill ju helt gå ifrån CV.	
29	Allt kretsar kring AI-chat boten då?	
30	Ja	
31	Vi läste på er hemsida att chat-boten exempelvis frågar varför kandidaten vill ha jobbet vilket kan leda till ett längre svar från kandidaten, hur går processen till när man analyserar kandidatens svar?	
32	Alltså själva motivation tänker du?	
33	Ja exakt, hur man analyserar kandidatens svar?	

34	Det är samma där, vi utgår från satta kriterier, i just motivationfrågan, kan det dels vara om man har tidigare erfarenhet av den specifika tjänsten och det kan vara att man nämner vissa nyckelord som är associerade till tjänsten och de kompetenser som vi valt ut för tjänsten.	A1, A2
35	Och då är det NLP algoritmen som avgör om svaret matchar?	
36	Då gäller det att förstå hur kandidatens svar står sig till kriterierna. Och att sätta upp kriterier på ett sånt sätt att det är antingen svart eller vitt. Men såklart, kan det vara mer gynnsamt för kandidaterna om de nämner fler nyckelord som är associerade med de här kompetenserna.	A1
37	I utvecklings teamen, hur ser mångfalden ut där? Sitter ni i Sverige eller har ni outsourcat det, vilka typer av människor är det?	
38	Som sitter och utvecklar tänker du?	
39	Ja, exakt	
40	Vi sitter allihopa i stockholm, vi har från sverige och asien.	L4
41	Hur ni många olika åldrar och kön?	
42	Åldrarna är hyfsat varierade, o könen är de som finns.	L4
43	När ni tar fram era tekniska lösningar, hur kompromissar ni mellan de tekniska lösningarna? Vad är viktigast? Om du står inför ett problem, och du har inte tid att lösa allt, prioriteras ex kundupplevelsen, att algoritmen är bias, tid.	
44	Så är det konstant, begränsat med tid och budget. Men svårt att svara på med bias ligger ju högt upp på listan för både oss och våra kunder. Lite beroende på vilka kunder det är då. Vi ser ju att bias frågan är mycket mer aktuella hos de som inte primärt sysslar med rekrytering. Inom rekryteringsbolagen är det viktigare att det sparar tid och pengar än att bias prioriteras högst, och även kandidat upplevelsen känns också som en viktigare puck för rekryteringsbolagen(internt dåligt)	L6
45	Har vi missuppfattat om ni alltså prioriterar internt då utifrån kundens prioriteringar?	
46	Det är ganska korrekt uppfattat.	
47	En fråga kring er kultur, om man misstänker ett problem med era algoritmer, har ni någon utpräglad process för hur ni löser detta?	
48	Det är ständigt pågående förbättringsarbete. Vi tar även hjälp från våra kunder iom att de får alltid granska alla kandidater och sedan acceptera de som är mest intressanta. Så vi får alltid en andra granskning som de kandidaterna som kommer genom Hubert. De går tillbaka in i modellen så att vi ser att en viss typ av kandidater, blir högre grad accepterade än andra, det är också en del av själva inläringen.	A1, A2, L6
49	Hur transparenta är ni med hur ni kommit fram score:t?	
50	Jag kanske kan visa ett exempel?	

51	Ja, jättegärna	
52	Så här ser då själva dashboarden ut för kunden, detta är då speglat från deras egna ATS. Här är då alla kandidater som är med i en process, och så kan man gå in på i den här kolumnen som uppfyller de formella kraven och så blir de sorterade från högst till lägst poäng. Här kan vi då gå in på Adam som har blivit rankad med högst poäng, och ser dels att han uppfyller det här kravet som är att man måste ha en deltidstjänst om 50 procent för att vara aktuell för tjänsten. Tidigare erfarenhet, här har han scorat 90/100. Praktiskt information, han är okej att jobba i minusgrader för det är kyllagret. Han har truckkort, han accepterar lönen för den här tjänsten. Här kan man även komma ner till de kompetensbaserade frågorna, så här är det självständigt arbete, noggrannhet, fysiskt arbete, målorienterat. Här man då dels se kandidatens svar och att han har scorat 80/100 på de svaret. Så detta är de som ligger till grund för det totala scoret. Detta är kundens vy.	R1, L3
53	Kan ni på Hubert visa varför han ex. Fått 90? Vad är det som gör att man kommit fram till scoret.	
54	Det är de här kriteriet som vi har kommit överens med kunden om, det är det som utgör varje steg i den här skalan.	
55	Så då är det som du sa innan, att han ex. Skriver att under "independent work", att han han är bra på egna initiativ, och om det då står i texten så får han ett högre score?	
56	Ofta är de här kompetensbaserade frågorna, så utgår vi från tidigare erfarenheter. Så frågan är ofta ställd, "kan du berätta om en gång när du jobbar självständigt, berätta gärna en gång vad din uppgift var och var utkomsten var? Och i de frågorna letar vi ofta efter de star-svar; situation, task, action och result. Vi har en upptränad modell som letar just efter de här 4 delarna i star, och en annan modell som tittar på relevansen för tjänsten. Så om man drar ett exempel som inte är relevant för tjänsten, kan man få ett lägre scoread. Men exakt vad de olika, så om du glömt i star så har vi kommit överens med kunden om vad de innebär rent konkret i kriterierna. På sidan kan man också se hela intervjun.	R1, A1
57	Vad anser ni som företag att den mest kritiska processen som gör att bias kan uppstå?	
58	Det skulle väl vara om vi sätter kriterierna helt skevt. Känns inte som en jätterisk att vi gör det. Men det är väl det som skulle vara värst.	A2, L4
59	Har ni en process för att säkerställa att det inte sker?	
60	Process eller process, vi sätter ju upp det ihop med kunden, och nu har vi gjort det så pass många gånger att vi börjar förstå hur man ska göra och inte göra så det är väl den processen som finns.	L4

Appendix E

Transkribering av intervju med Devten

Rad	Information	Aspekt
1	Jag tänkte om vi först kunde få höra lite mer om Devten och vad ni gör?	
2	Absolut! Jag och Robin som då startat det här kommer från utvecklar-sidan. Innan vi ens tog examen har vi tröttnat lite på rekryterare, att de ofta ger ut rätt oklara erbjudanden och inte är transparenta. Man kan inte riktigt säga vad de innebär eller vad de har för komposition. Så då fick vi idén av att ha en plattform där rekryteraren når ut till utvecklaren fast med transparenta erbjudanden där de kan berätta vad de erbjuder för techstack, vad de erbjuder för lön, om de har något bonusprogram osv osv. Alla möjliga parametrar som inte är så vanliga att man specar på Linkedin till exempel. För där kan man skriva in utbildning, erfarenhet och lite om techstack. Då skapar både rekryteraren och utvecklaren en profil så matchar vi då ihop dem. Utvecklaren är anonym tills den kan välja att inte vara det, så man kan visa sitt namn och sin profil. Och lite känslig information som kan koppla dig till vem det är, genom arbetslivserfarenhet eller liknande.	R1, L2, L4, A2
3	En fråga, kan man välja att lyfta att inte vara anonym redan från start då eller hur fungerar det?	
4	Nej, vi har ingen sån funktionalitet. Det kan vara en bra idé, absolut, att man kan få ändra det i settings.	
5	Var detta kopplat till er utbildning på något vis, om det var kopplat till KTH?	
6	Vi sökte till ett basprogram på KTH, ett inkubatorbolag som tar in start-up-bolag och hjälper dem med coaching och med husrum kanske, kontor och sådär. Då har KTH Innovation ett startprogram som vi gått med i som man ansöker till. Och jag vet att LU också har något vid Ideon eller så, På Alfa eller vad det var	
7	Ja precis. Hur skulle ni säga att ni säljer era tjänster och vilket värde erbjuder ni kunden med den tjänsten ni har?	

8	<p>För rekryteraren till att börja med får de tid. När du söker på Linkedin har du ganska få parametrar att söka med. Här söker du med hela rollen du är ute, och då får du kanske upp tre kandidater istället för 150 som när du söker på full stack engineer när du söker i Stockholm. Normalt sett har du en sökterm eller ord som på Google, men nu söker du med ett helt objekt. För utvecklaren vill vi pusha för det här med att man inte får erbjudanden som inte är så relevanta, så en senior utvecklare vill kanske inte ens ha Linkedin eller visa sin mail på något vis, då de kan bli spammade, därav också att man ska vara anonym. Det finns också företag som gör liknande saker, Honeypot som gör detta i Tyskland och Nederländerna och så har vi en konkurrent i Sverige som heter Demando</p>	R1
9	<p>Och varför ska man välja Devten över Demando?</p>	
10	<p>Det är hård konkurrens, vi hade det lite som benchmark när vi startade det här. De var före oss och sen så har vi gått igenom deras tjänst och ändrat på det vi inte gillar, exempelvis att varje gång du ska söka ett jobb så måste du fylla i din arbetslivserfarenhet och din utbildning; Det är ganska jobbigt. att fylla i formulär om och om igen. Hos oss kan du ladda upp ditt CV och scrapar det, och det är också kopplat med AI och är en rätt stor grej inom rekrytering. Att man scapar det. Jag vet att många rekryteringsfirmor gör det också. Jag vet inte om ni har skrivit om det?</p>	A1,
11	<p>Nä, vad sa du att man scrapar?</p>	
12	<p>Ni får dubbelkolla så ni inte citerar mig! Men om du är en stor HR-firma får du kanske in hundra Cv:n för en tjänst, eller vad vet jag? Men då har de ett program som scrapar ditt CV, du letar efter nyckelord. Säg att du vill ha en utvecklare till en tjänst som kan python och tenseflow, då letar den efter det ordet i ditt CV och filtrerar ut några kandidater direkt. Så att det finns också, om man googlar [ohörbart] skriver på CV då så det funkar bra för de här algoritmerna, kan vara en grej man behöver göra. Så det skapar kanske en problematik, det är kanske något ni kan undersöka. Eftersom man inte tänker på det när man skriver sitt CV, det ska vara lättläst för en algoritm och inte för en person.</p>	A1, L3, L4
13	<p>Ja men precis, som att använda trendiga ord?</p>	

14	Ja och att man har ett visst format, det har vi märkt på den vi har - går du in på Devten.io. Första steget om man trycker på sign-up, eventuellt lägga upp ditt CV, men vissa saker har vi märkt att den inte hittar saker så bra . Men den situationen uppstår säkert för en rekryterare också, att den inte plockar upp allt . Och det kan ju bero på att man har en annan struktur än andra Cv:n.	A1, A2
15	Är detta en plug-in till er algoritm eller hur har ni utformat denna scarpen?	
16	Vi har köpt in API-tjänsten . jag skulle gissa att de använder [ohörbart] för att analysera bilder och en läsare med NLP från Google .	A1, A2
17	Var behöver er tjänst för delar för att matcha rekryterare och utvecklare?	
18	Vi har inga egentliga svar utan bara antaganden. Vi har programmeringsspråk och vad för språk man talar. Sen har vi framförallt saker som kompensation. Så det viktigaste en kandidat specar är programmeringsspråk, vanligt språk och det viktigaste från andra hållet är lön och location . Men vi försöker matcha på så mycket som möjligt. Det är utveckling att få se vad som sker med tiden och vad folk tycker är viktigt.	R1
19	Går ni in på något som liknar personlighetstester?	
20	Nä, det handlar mest om att fylla i information. Vi skapar det första mötet mellan företag och utvecklare så de får se om man har de rätta kompetenserna . Sedan får de gå vidare med kompetenser, tech-intervjuer eller liknande.	R1
21	Vilka AI-tekniker använder ni?	
22	Den tjänsten vi köper in för att parse Cv:n kommer från Affinda , på deras hemsida specar de nog mer exakt. Men jag skulle gissa att det är någon form av natural language processing . Jag vet inte exakt, men ni kan säkert kontakta dem om det.	A1
23	Sen vet jag inte helt hur ni har definierat AI i er rapport, för det kan ju vara väldigt enkelt och ytligt	
24	Vi kan säga att det inte djupdyker i det tekniska, då det kanske hamnar mer inom datavetenskap än systemvetenskap kanske - så det är rätt grundläggande	

25	Man kan placera allt i ett venndiagram, så ytterst har man AI, innanför den machine learning, och innanför den deep learning. Vår matchningsalgoritm är nog utanför machine learning men inom AI. Vi använder ingen statistisk metod för att matcha ihop, för att kunna göra detta behöver man väldigt många tidigare samples, och det har inte vi. Just nu har vi två pilotkunder så vi kommer inte upp på den nivån så vi kan göra sådant just nu.	A1
26	Ni utformar alltså inte test- eller träningsdata på det sättet heller då kanske? Det blir helt enkelt bättre ju fler som signalerar upp sig	
27	Det är ju vårt mål, vi sparar alla matchningar och sådär men som sagt för att för få företag för att jobba med det nu. För vår business så är det mest en enkel algoritm som kollar igenom flera if-satser på så många parametrar som möjligt	A1, A2
28	Och hitta den optimala matchningen då så att säga	
29	Ja precis. att matcha så många parametrar som möjligt och ta fram topp tre kandidater.	
30	Hur tänker ni senare och hur vill ni arbeta längre fram?	
31	En första approach är att multilayer, ett väldigt generellt nätverk och då är input-features vad utvecklaren och företagen har specat. Så det blir input och sen går det i en loss function som kontrollerar saker binärt, jag vet inte om det säger så mycket men. Mer intressant för er är verkligen biten som granskar Cv:n	A1
32	Just med CV:granskning kanske man kan se vilka tidigare anställt så exkluderas exempelvis minoriteter eller kvinnor, så i framtiden finns en risk att dessa grupper fortsätter att exkluderas.	
33	Om man gör på det sätt jag beskrev tidigare och man råkar skicka med sitt för- eller efternamn och alla som blivit anställda har svenska namn så kommer algoritmen lära sig av det, så det beror ju helt på vad du skickar in för input-data, så man måste vara väldigt försiktig med det. Du kontrollerar inte vad algoritmen ska kolla efter	L4, L6
34	Okej, så man debuggar inte biasen på det viset?	
35	Precis. Ett intressant exempel var när en algoritm skulle kontrollera vilka vetenskapliga artiklar som skulle publiceras i en journal, så lärde sig algoritmen att kolla antalet sidor. Så den kollade bara längst ned på antalet sidor. Det är lite samma bias som kan uppstå med namn.	L6
36	Är det något ni har hunnit börja med idag?	

37	Nej inte riktigt, vi kollar bara på kompetensen.	
38	Ja och det är något som är rätt unfördomsfull som det är eller?	
39	Ja det som kan dyka upp är när man specar nivån av språk man talar så är det basic, intermediate, fluent och native och då kan swedish native fångas upp.	L2, L4
40	Ja och då kan det kanske också vara så att man missar native och bara skriver swedish så kan man rankas på ett annat sätt, men bias kanske kan uppstå vid inmatning om vilket universitet man läst vid?	
41	Jo, men det kan också vara ett värde. Jag tror inte det är så vanligt att företag använder en ML-algoritm hela vägen till anställning. Det är kanske mer att man scrapar nyckelord i Cv:n	
42	Du nämnde att det inte är en helt automatiserad process, vilken roll har personer i denna rekryteringsprocessen?	
43	Efter matchningen egentligen och att man börjar prata med kandidater. Så först presenteras företaget för kandidater, och de kan välja proceed eller avböj. Väljer man proceed skickas det vidare till utvecklaren och den får veta att det har kommit ett nytt erbjudande. Utvecklaren får sedan också välja ifall denne vill fortsätta eller avböja. Har man gjort detta visas den fullständiga profilen och ni kan chatta.	R1
44	Och visas hela CV och liknande då?	
45	Ditt namn, CV, länkar till Github och linkedin. Portfolio. Vad kan du för programmeringsspråk och verktyg. Språk som vi pratade om vidare. Vad man har för utbildning och tidigare erfarenhet, det vanliga egentligen. Med mer fokus på text och den erfarenhet du har av vissa programmeringsspråk.	R1
46	Ser du ert upplägg av vilken input data som behövs för er tjänst som rent av en fördel för att minimera bias?	
47	Ja, det tror jag. Det finns ju studier om hur avgörande första intrycket och liknande är vid rekrytering som ni kan kolla på, om första intrycket inte inkluderar en bild eller namn så handlar det bara om dina kompetenser	L2,L4
48	Ja det har vi sett, att man har svårt att förbise det första intrycket och att man tar med det i avvägningar rent av omedvetet.	
49	Istället för att ha namn och bild på CV, vilket är rätt unikt för Skandinavien.	

50	Just när man scrapar, så är det väl egentligen ett ganska fult CV som algoritmen svarar bäst på? Just med att ingen bild ska vara med, det ska vara intabbat	
51	Du ska gärna ha ett one page resume som är packat med information, något som egentligen är rätt svårt att läsa.	A1, A2
52	Är användarna medvetna om att AI används? Att det står skrivet någonstans?	
53	Nej det står inte, bara att den fyller i saker åt dig. Det som är positivt är just att man undviker mänsklig bias, vilket är något ni kan skriva om. Så dessa input-feautres blir väldigt viktiga på det sättet.	L4

Appendix F

Transkribering av intervju med Wise Consulting

Rad	Information
1	Då ska vi se här, vem är du och vilken roll har du?
2	Jag heter Ann-Sofie Mattson och jobbar på Wise Consulting som ingår i Sveriges största HR-koncern. Vi är ett av bolagen där vi erbjuder tjänster inom HR-consulting, ledarutveckling och coachning. Min roll är att jag är ansvarig för våra Stockholmskontor kan man säga som regionchef.
3	Sspännande! Använder ni några automatiserade tjänster idag vid kandidat-urval?
4	Vi är lite efter där faktiskt ska jag säga. Vi jobbar inte så mycket mer rekrytering, vi rekryterar till oss själva. Men nej, inget automatiserat förutom systemstöd. Annars är det mycket handpåläggning fortfarande.
5	Om ni skulle börja med det, vad tycker ni är viktigast för att byta ut handpåläggning mot det automatiserade?
6	Då skulle vi nog behöva hjälp vid själva urvalsprocessen där man hade kunnat få slänga in kriterier, för vi anställer ju mest HR-konsulter och då handlar det om en viss kompetens man måste ha - en HR-kompetens. Men sedan är det även konsultmässigt, att du behöver ha konsult-hatten på dig. Det är inte alla som förstår det eller vet vad det innebär. Där hade det varit fint ifall man hade någon form av urvalsprocess i båda delarna, både kompetens men även mindset. Just konsultmässigheten där, när det gäller egenskaper kan man säga.
7	Jag förstår! Vad tycker du är den mest fördomsfulla informationen under en rekrytering?
8	Det finns en massa effekter egentligen, det är ens egna tyckanden och sina egna förutfattade meningar som kan spela in. Då handlar det mycket om vem som rekryterar, ens egna värderingar som styr. Det blir lite magkänsla.
9	Det har vi sett mycket om. Vi har hittat dessa företagen som jobba för att man inte ska gå på magkänsla utan att man ser mer på den hårda faktan.
10	Det man har gjort mycket, som vi också gjort en del, så har vi börjat med urvalsprocessen med mer hantering av tester. Man har testat för att få bort fördomar som påverkar. Man gör tester och sen tar man urvalet utifrån de kriterier man satt innan och vad de testerna ger. Det är ett sätt man kan göra det på.
11	Vi har förstått att det är mycket tester, personlighetstester, vissa har spelbaserade tester.
12	Ja det kan verkligen skilja sig åt, vilken process man har.
13	Okej, vad är viktigt för en rekryterare att tänka på vid en rekrytering?

14	Hos oss är det mest våra teamchefer som rekryterar våra konsulter och det viktigaste är att man klarar av hantverket, att man har kompetensen helt enkelt och att den verifieras men även i vårt fall att man behöver ha den här konsultmässigheten, som kan vara svår att mäta. Så där blir det ganska mycket utifrån deras erfarenhet. Sen ser man på hur man mäter affärsmässighet och så, men det får vi ut ganska mycket av våra personlighetstester. Vi gör ju djupintervjuer för att trycka mer på det och en mer kompetensbaserad intervjuteknik. Då handlar det egentligen mer om att man går djupare in på frågorna, exempelvis kan en person beskriva ett projekt den genomfört, då stannar man inte upp vid projektet utan tar det vidare och fråga mer om vilket sätt man gjort det på, vad blev utfallet, vilka var inkopplade, vad var din roll? Man får gräva sig in i hur man faktiskt gör saker och att man kan mäta arbets sättet mycket tydligare.
15	Är det så att ni mäter nyckelord som man använder eller mer hur man själv beskriver saker?
16	Jag skulle säga att det handlar mer om beskrivandet. Sen är nyckelord i den mån om vilket vokabulär man vill använda, hos oss behöver man ha lite mer säljhatten på. Konsultmässigt vill vi gärna att man pratar mer affär och nyckeltal, särskilt inom HR som lätt kan bli luddigt och mjukt ibland - och då behöver man kunna prata om hårda fakta. Vad ger det för effekt? vad ger det för resultat?
17	Sker bedömningen av detta i samråd med flera av era rekryterare exempelvis? Eller är det en person som bedömer svaret?
18	Det är egentligen två personer skulle jag säga som är involverade i vår rekryteringsprocess. För att också kunna ta lite olika roller. det kan vara bra att en driver samtalen och den andra observerar och ha lite olika roller. Så då kan man fånga upp saker där som används vid nästa intervju. Då brukar det vara att om vi väljer att gå vidare med en kandidat så har vi ofta gör denna NPA, ett djuppersonlighetstest, och då kan de frågetecken som dykt upp innan djupas lite i det.
19	Är det standardiserade frågor ni har vid varje intervjudel?
20	Jo, strukturerat på det vis att vi vill hålla det kompetens-fokuserat, då behöver det vara strukturerat i intervju-förfarandet så man inte hamnar i ett samtal. Oj vad trevligt vi har haft det en timme liksom utan att man är tydlig med att det är en intervjusituation
21	Hur är er generella inställning mot att ha AI som stöd vid rekrytering i framtiden? Är det något som diskuteras?
22	Ja, framförallt våra systerbolag - våra bolag som jobbar med våra kunder och deras rekryteringsprocesser. De använder faktiskt en AI som heter GLEM på vårt Göteborgskontor och de har testat detta i urvalsprocessen. Jag är lite dålig på hur det urvalet har gått till och hur utfallet av det har varit. Det vet jag inte riktigt men de har testat det ganska länge.
23	Ja vi har intervjuat företag som arbetar med en massa spännande tjänster, så det är väldigt spännande att se.
24	Verkligen, och vi kommer säkert dit mer och mer i själva urvalsprocessen. Men det finns ju även flera aspekter i det där. Det finns andra aspekter som är viktiga, som att en viss person är spännande eller har en viktig nyanserad.
25	Det har många sagt, att man inte automatiserar hela vägen utan bara under ett visst moment under rekryteringsprocessen. Många har ändå stor respekt för AI och vill inte tappa hela handpåläggningen.

26	Det tror jag är viktigt, att det finns en kombination. Men jag tror också det kan vara bra i själva urvalet för då tappar man detta, som ni var inne på, med denna fördomsfråga. Att man går mycket på egen känsla och att man gillar en person. Det blir att man istället aidentifierar sig lite.
27	Vad tror du är de stora riskerna ifall man väljer att ha automatiserade processer, rent ur ett bias-perspektiv?
28	Ja, den största risken är att man tappar potentialen i folk. Man kan hitta det när man sitter i intervjuer där man ser andra infallsvinklar och kan läsa av potentialen i en kandidat, där det är ganska kul att slänga in en sådan person i en rekryteringsprocess som är lite mer av ett wildcard eller att det inte är någon som stämmer överens med den exakta kravprofilen. Det är ändå viktigt att man vågar, att man kan lära sig mycket på jobbet och att man ser potentialen i en person. Att det finns något som man kan driva vidare på i vår verksamhet och att man inte tappar chansen att göra det. Den kan jag känna är, när man rekryterar kan jag känna att, utifrån min erfarenhet, att man tittar på kravprofilerna så vill man ha personer som gjort exakt vad kravprofilen säger. Men ofta vill den personen göra något nytt eller ta en svängning i sin karriär. Det är kanske sällan en person vill gå från ett jobb till ett exakt likadant. Man vill kanske ha någon form av utveckling och då måste man kunna läsa av potential.
29	Angående jobbannonser och kravprofil - vad vill man annonsera? Vad är viktigt att få med?
30	Den svängningen har ju gjorts, att det varit kravställande mot kandidaten men nu är lite tvärtom. Kandidater idag väljer sin arbetsplats mycket mer idag än vad man kanske gjorde förr. Nu är det mer ett tvåvägsval som ska göras. Som bolag behöver du tänka till lite - hur vill du presentera dig? Vilken arbetsgivare vill du vara och vilket värde har du? Tänka på medarbetarlöften och tänka på den djupare idéen. Vad får du om du kommer till oss? Vad kan vi erbjuda dig? Det tror jag man behöver tänka till på för att göra sig attraktiv. Och om man har ett större syfte, inte bara inom verksamheten utan även jobbar med CSR-frågor och hållbarhetsfrågor kanske, det säger man också blir mer och mer intressant för kandidater - speciellt de yngre som då vill ha ett högre syfte. Att man vill göra något mer och stå för något mer och det gör att företagets värderingar behöver beabeta ordentligt. Så det är större krav egentligen på arbetsgivare och företag idag, vilket är bra.
31	Så det är alltså något man vill trycka på i annonseringar?
32	Ja, så jobbar vi iallafall. Det är något vi vill erbjuda och det vi brukar vara väldigt autentiska i det. Man ska liksom inte sminka grisen som man brukar säga och tro att man är någon man inte är. Det är ett medarbetarlöfte som ska hålla hela vägen även när man är anställd. Så man måste ha gjort det jobbet. Men det är bolagen som får mest intresse.
33	Brukar ni försöka målgruppsanpassa annonserna? Om man vill rikta annonser mot en viss typ av person? Vi ser vissa fall där AI används för att skraddarsy annonser till en viss målgrupp så jag tänkte ifall det är något ni gör idag rent manuellt?

34	<p>Ja men det tror jag - det är också väldigt viktigt att inte skriva rätt ut att man söker en kvinna på en ålder mellan si och så. För det är diskrimineringsskyldigheter i det där. Sen hos oss är vi väldigt många tjejer, vi är en minoritet av män. Man hade velat skriva att vi gärna ser manliga sökanden. Men det ska man vara försiktig med. Det handlar också om att det finns detta förklädda som du säger, att man använder ord som appellerar en viss målgrupp. Att det är lite förtäckt kommunikation. Jag kan ändå tycka, i vissa fall, att det görs för ett gott syfte. Exempelvis inom IT-branschen så vill man locka tjejer på ett annat sätt, och då kan det vara en smart och klokt sätt att få in tjejer. Det behövs även i vår bransch, att det finns en dynamik mellan killar tjejer och etniska bakgrunder också. Ofta pratas det om killar och tjejer men man behöver också ta det ett steg längre. Det vore häftigt att ha denna mångfalden så gott det går.</p>
35	<p>Hur uppfattar det den tekniska utvecklingen idag inom branschen? Är det något som lyfts i kommunikationen, den strategiska planeringen och liknande? Kommer det ta lång tid innan detta kommer in?</p>
36	<p>Jag tycker att det pratas alldeles för lite, det är många bolag som inte har kommit så långt. Eller några bolag har kommit en bit, jag vet inte vilka bolag ni hade intervjuat nu men jag vet något rekryteringsbolag som legat i framkant. Några ligger i framkant, flera som verkligen har kommit en bit men jag tycker det är långt ifrån att man pratar om det och veta hur man ska använda sig av det på rätt sätt. Jag tycker fortfarande att det är ganska traditionellt. Det finns mycket att göra där.</p>
37	<p>Ja EU la ju precis fram en ny förordning för AI, där de lyfter just rekryteringssammanhang så det är väldigt nytt på det sättet.</p>
38	<p>Ja så har man GDPR på det så man har en process på det med mera. Så det blir svårt att hämta fakta när det är nytt liksom.</p>

Appendix G

Mailtråd med kompletterande frågor till Tengai

Rad	Information	Aspekt
1	Hur har ni kommit fram till vilket svar som ska anses positivt resp. negativt i er matching?	
2	Vi har, tillsammans med Furhat och forskare på Stockholms Universitet, tittat på människor pratar och beskriver "saker" utifrån positivt och negativt. Vilka ord man använder, i vilken följd, hur variationen ser ut osv. Det är det som utgör huvudmotorn i NLPn	A1, L2, L6
3	Om beslutet tas med hjälp av NLP, ifrågasätter ni någon gång hur NLP kommit fram till beslutet?	
4	Det sker automatiska kontroller, då roboten kan uppfatta svaret fel och det får vi veta, och kan justera därefter. Dvs. "roboten" försöker avgöra vad kandidaten har sagt genom olika beräkningar och sannolikhet. Exempelvis så gjorde vi precis en justering där roboten uppfattade att kandidaten har bett om mer tid, mitt i ett svar, och så var det inte. Kandidaten använde sig av ord såsom "behövde mer tid just då" och då tolkade Tengai det som att kandidaten ber om mer tid.	L6
5	Hur kontrollerar ni att de algoritmer som är skapade av andra företag såsom Googles ASR inte innehåller någon form av inbäddad bias?	
6	Kontinuerliga kontroller i samråd med just Google. De driver en hel del arbete själva, och vi kontrollerar på vår sida hur utfallet blir.	L6
7	Hur förvaltas data från personer under processens gång? Samt även efter en person har gått igenom processen?	
8	Osäker på vad ni exakt menar här?	
9	Är kunder och kandidater införstådda i hur AI har haft en inverkan på beslut relaterade till rekryteringsprocessen?	
10	Jag skulle svara JA på den. Kandidaterna och rekryterarna får väldigt mycket information om hur allt går till. Men såsom det alltid när något nytt lanserats så finns det säkerligen områden/information som skulle kunna utökas/breddas.	L3

Appendix H

Transkribering av kompletterande intervju med Hubert

1	Kan ni förklara både internt inom företaget samt för kunder hur algoritmen kommer fram till en viss poängsumma utifrån de svaret som kandidaten lämnat? Alltså kan algoritmens beslut är förklaras?	
2	Som jag sa förra gången utgår vi alltid ifrån bedömningar på olika kriterier, så vi sätter upp så vad är kriteriet för att få 1,2,3,4,5 poäng på denna frågan. det är det hubert försöker komma fram till. jag kan ta fram ett exempel.*exempel visas* En annan modell kollar hur lik denna erfarenhet är som kompetensen är för det jobbet.	R1, L3
3	Okej, detta var mycket tydligare! Då förstår jag hur ni har kommit fram till poängsättningen, tack för det!	
4	Om kund hade frågat varför man blivit bedömd på detta sättet, hade ni kunnat plocka fram då vad en kandidat sagt som gör att den inte uppfyller ett krav?	
5	ja exakt. vi har försökt göra detta så transparent som möjligt. Man får poäng på dels specifika kompetenser, och man ser även då exakt vad man blivit bortsållad på.	L3
6	Hur förvaltas data från personer under processens gång? Samt även efter en person har gått igenom processen?	
7	Hanteras inte på något särskilt vis eller processas på något särskilt vis efter rekryteringen är klar. Det ligger kvar i databasen i 36 månader enligt vårt standardavtal med kunderna. Vi har idéer av vad vi ska göra med det här men det är inget vi gör just nu.	L2
8	Är kunder och kandidater införstådda i hur AI har haft en inverkan på beslut relaterade till rekryteringsprocessen?	
9	Ja det skulle jag säga, det är vi rätt öppna och transparenta med. Att förklara detta både för kandidaternas skull innan man kommer in i intervjun och det tar vi också upp med kunderna innan. Sen är det inte alla som läser den.	L3
10	Nä det är inte självklart, men det kommunicerar ni ändå!	
11	När en person säger ett visst ord så kan det vara laddat med olika värden och innebörder. Bedömer er NLP värdet av olika ord? På vilket sätt isåfall? Och ser ni om detta kan bli ett problem?	
12	Hur tänkte du med värde på ord?	
13	Exempelvis kan man se om man nämner att man lagar mat eller något så kan det ordet kopplas till kvinnor, så vissa ord kan ha vissa associationer. Men kan det påverka er algoritm på något vis?	

14	Ja det har jag väl svårt att se att det kan påverka på något sätt. Vi har en modell som kollar på kommunikativ förmåga; alltså hur väl man kan uttrycka sig, hur snabbt man skriver osv. Det är en modell i sig som kollar på hur avancerade ordval man gör och vilket språk man för sig med. Men jag har svårt att se hur man kan koppla detta till kön. jag antar att det är det ni vill kolla på?	A1,A2
15	Ja eller egentligen allt som på något vis kan bli diskriminerande, exempelvis uttrycker sig en äldre kanske annorlunda än en yngre - värdesätts då en äldre persons språk mer, som att det är bättre än en yngres. om det kan ha en viss påverkan?	
16	Det skulle kunnat ha en viss påverkan om man kollar just på ordval, men samtidigt kan en ung person prata lika väl som en äldre person också. Kanske, men det känns långsökt.	A2

Appendix I

Mailtråd med kompletterande frågor till Exparang

Rad	Information	Aspekt
1	Vi vill gärna höra lite mer om hur er insamling av data och skapandet av kandidaters profiler fungerar. Vilka tekniker använder ni för att göra detta?	
2	Ang insamling av data är våra kandidater helt ansvariga för att själva ge oss sin data. Vi skrapar den inte från öppna källor över huvudtaget. Både pga tekniska utmaningar att få relevant och bra data och pga att den data vi främst är intresserad av är inget som öppna källor kan ge oss.	L2
3	Hur avgör ni vilken data som är relevant att samla in om kandidaterna? Hur tolkas och bearbetas denna data sedan?	
4	Datan kategoriseras i tre grupperingar: Experience/Skills, Values och Aspirations. Vi ”poängsätter” respektive datapunkt i relation till de jobben som laddas upp på plattformen, och utefter det sker matchning.	
5	Du nämnde sist att kandidaten har ett profilkort - är detta något som arbetas fram av AI?	
6	Profilkortet är en anonymiserad CV, där vi tar en rad förutbestämda datapunkter som kandidaten har gett oss. Kandidaten kan själv se sitt profilkort samt göra justeringar i sin profil.	R1
7	Kan ni förklara internt hur algoritmen kommer fram till en viss poängsumma utifrån de 50 frågorna? Alltså varför ett svar från en kandidat genererar en viss poäng?	
8	Yes - vi kan förklara poängsumman. Däremot är den helt och hållet subjektiv utifrån vår erfarenhet från rekrytering. Vi justerar den över tid för att få en så träffsäker och inkluderande matchning som möjligt.	L3, L6
9	Är kunder och kandidater införstådda i hur AI har haft en inverkan på beslut relaterade till rekryteringsprocessen?	
10	Vi informerar kunder och kandidater ang AI:s inverkan ja, men i vilken utsträckning de är medvetna är svårt att säga.	L3

11	Ord kan vara laddade med värden och innebörder. Exempelvis kan gender bias uppkomma genom andra ord än just från input data som kön och namn. Tidigare erfarenheter eller intressen såsom matlagning har bevisats oavsiktligt kunnat kopplas till kvinnor. Hur ser och arbetar ni med detta?	
12	Ang ord kopplade till gender bias håller jag helt med. Otroligt svårt! Vi låter regelbundet testpersoner från olika grupper gå igenom vårt frågeformulär så att vi får ett så brett perspektiv som möjligt på frågorna. Men ändå, supersvårt.	L5
13	Ville bara höra om ni har tränat upp er NLP själva eller använder exempelvis Googles NLP?	
14	Vi använder Googles NLP för tillfället!	A1

Referenser

- Adadi, A., & Berrada, M. (2018) Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) [pdf] Tillgänglig: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8466590&tag=1> [Hämtad 15 april 2021]
- Aichholzer, C. (2018). What Are Game-Based Assessments?, HireVue. Tillgänglig: <https://www.hirevue.com/blog/hiring/what-are-game-based-assessments> [Hämtad 26 april 2021]
- Alpaydin, E. (2010). Introduction to Machine Learning. 2. uppl., Cambridge, Massachusetts: The MIT Press
- Alvehus, J. (2013). Skriva uppsats med kvalitativ metod: en handbok, Stockholm: Liber.
- Bajec, M., & Krisper, M. (2005). A methodology and tool support for managing business rules in organisations. *Information Systems*, vol. 30, no. 6, s. 423-443, Tillgänglig: https://www.sciencedirect.com/science/article/pii/S0306437904000481?casa_token=JtqnRQYXXr4AAAAA:To7X7eJWqQI8hiE0scUpP_11f1wc8SgIb5VXsT5w8TKEptngKrOI8txLpII08PSvEk6O9pGgOh4 [Hämtad: 16 april 2021]
- Black, J.S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it?, *Business Horizons*, vol. 63, iss. 2 s.215-226, Tillgänglig: <https://www.sciencedirect.com/science/article/pii/S0007681319301612> [Hämtad 23 mars 2021]
- Benbya, H., Pachidi, S., & Jarvenpaa, S.L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research, *Journal of the Association for Information Systems*, vol. 22, iss. 2, s.281-303. Tillgänglig: <https://aisel.aisnet.org/jais/vol22/iss2/10/> [Hämtad 3 april 2021]
- Bogen, M. and Rieke, A. (2018). Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias. Upturn. [pdf] Tillgänglig: <https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20--%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf> [Hämtad 27 april 2021].
- Breakwell, G.M. (1995). Interviewing, in Breakwell G.M., Hammond S. & Fife-Schaw C. (eds), *In Research Methods in Psychology*, New York: Sage Publications Ltd.
- Bryman, A. (2018). Samhällsvetenskapliga metoder. 3. uppl., Stockholm: Liber AB
- Capgemini. (2020). AI and the Ethical Conundrum - How organizations can build ethically robust AI systems and gain trust [pdf], s. 7-10, Tillgänglig: <https://www.capgemini.com/wp-content/uploads/2020/10/AI-and-the-Ethical-Conundrum-Report-1.pdf> [Hämtad: 2 april 2021]
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women, Reuters, 11 oktober, Tillgänglig: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/> [Hämtad 4 april 2021]
- Dataskyddsförordningen (GDPR). (2016). Europeiska Unionen
- Dattner, B., Chamorro-Premuzic, T., Buchband, R., & Schettler, L. (2019). The Legal and Ethical Implications of Using AI in Hiring. Harvard Business Review. [pdf],

- Tillgänglig: <https://egn.com/dk/wp-content/uploads/sites/3/2020/06/The-legal-and-ethical-implications-of-using-AI-in-Hiring.pdf> [Hämtad: 3 april 2021]
- Daumé, H., Dudík, M., Holstein, K., Wallach, H., & Wortman Vaughan, J. (2019). Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? [pdf] Tillgänglig: <https://arxiv.org/pdf/1812.05239.pdf> [Hämtad: 15 april]
- Denscombe, M. (2016). *Forskningshandboken: för småskaliga forskningsprojekt inom samhällsvetenskaperna*. 3., rev. och uppdaterade uppl., Lund: Studentlitteratur AB
- Dignum, V. (2020). Responsibility and Artificial Intelligence, in M.D. Dubber, F. Pasquale and S. Das (eds), *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press.
- Ejvegård, R. (2003). *Vetenskaplig metod*. Lund: Studentlitteratur AB
- Europeiska kommissionen. (2019a). Etiska riktlinjer för tillförlitlig AI. Tillgänglig: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> [Hämtad 9 april 2021]
- Europeiska kommissionen. (2019b). En definition av AI: Viktigaste förmågor och discipliner, Tillgänglig: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> [Hämtad 9 april 2021]
- Europeiska kommissionen. (2020). Spetskompetens inom och förtroende för AI, Tillgänglig: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_sv [Hämtad 8 maj 2021]
- Europeiska kommissionen. (2021). Europa rustat för den digitala tidsåldern EU-kommissionen föreslår nya regler och åtgärder för spetskompetens och förtroende för artificiell intelligens, Tillgänglig: https://ec.europa.eu/commission/presscorner/detail/sv/IP_21_1682 [Hämtad: 8 maj 2021]
- Federal Trade Commission. (2016). Big Data - A Tool for Inclusion or Exclusion? [pdf] Tillgänglig: <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> [Hämtad: 6 april 2021]
- Graziano, A.M., & Raulin, M.L. (2013). *Research methods: a process of inquiry*, 8. uppl. Boston: Pearson.
- Google Cloud. (n.d.). Cloud Speech-to-Text - Speech Recognition, Tillgänglig: <https://cloud.google.com/speech-to-text>. [Hämtad 26 april 2021]
- Hammond, S. (1995). Using psychometric tests, in G.M. Breakwell, S. Hammond, & C. Fife-Schaw (eds), *In Research Methods in Psychology*, New York: Sage Publications Ltd.
- Hellström, T., Dignum, V., Bensch, S. (2020). Bias in machine learning - what is it good for?, in A. Saffiotti, L. Serafini, & P. Lukowicz (eds.), *NeHuAI 2020 : First International Workshop on New Foundations for Human-Centered AI: Proceedings of the First International Workshop on New Foundations for Human-Centered AI (NeHuAI) co-located with 24th European Conference on Artificial Intelligence (ECAI 2020)* s.3-10
- Holzinger, A. (2016) Interactive machine learning for health informatics: when do we need the human-in-the-loop?, *Brain Informatics*, vol. 3, no. 2, s.119-131 Tillgänglig: <https://link.springer.com/content/pdf/10.1007/s40708-016-0042-6.pdf> [Hämtad 14 april 2021]
- Houser, K.A. (2019). Can AI Solve the Diversity Problem in the Tech Industry: Mitigating Noise and Bias in Employment Decision-Making, *Stanford Technology Law Review*, [e-journal] vol. 22, s. 290. Tillgänglig: via LUSEM Library website <https://lusem.lu.se/library> [Hämtad 26 mars 2021]

- IBM Cloud Education. (2020). What is Conversational AI?, IBM, 31 augusti, Tillgänglig: <https://www.ibm.com/cloud/learn/conversational-ai> [Hämtad 15 april 2021]
- Iansiti, M., & Karim, R.L. (2020). *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World*, Boston: Harvard Business Review Press.
- Jacobsen, D. (2002). Vad, hur och varför? Om metodval i företagsekonomi och andrasamhällsvetenskapliga ämnen, översatt av G. Sandin, Lund: Studentlitteratur
- Jordan, M.I. and Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects, *American Association for the Advancement of Science*, vol. 349, iss. 6245, s.255–260, Tillgänglig: <https://science.sciencemag.org/content/349/6245/255/tab-pdf> [Hämtad 15 mars 2021]
- Knight, W. (2017). The Dark Secret at the Heart of AI, MIT Technology Review, 11 april, Tillgänglig: <https://www.technologyreview.com/2017/04/11/51113/the-dark-secret-at-the-heart-of-ai/> [Hämtad 20 april 2021]
- LeCompte, M.D., & Goetz, J.P. (1982). Problems of reliability and validity in ethnographic research, *Review of Educational Research*, [e-journal] vol. 52, iss. 1, s.31-60 Tillgänglig: via LUSEM Library website <https://lusem.lu.se/library> [Hämtad 26 mars 2021]
- Metz, C. (2019). We Teach A.I. Systems Everything, Including Our Biases, The New York Times, 11 november, Tillgänglig: <https://www.nytimes.com/2019/11/11/technology/artificial-intelligence-bias.html> [Hämtad 11 maj 2021]
- Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdli, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C. & Broelemann, K. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, vol. 10, iss. 3, s.1-14. Tillgänglig: <https://arxiv.org/pdf/2001.09762.pdf> [Hämtad 12 april 2021]
- Oracle. (2019). AI in Human Resources - The Time is Now [pdf] Tillgänglig: <https://www.oracle.com/a/ocom/docs/applications/hcm/oracle-ai-in-hr-wp.pdf> [Hämtad: 20 februari 2021]
- Otter, D.W., Medina, J.R., & Kalita, J.K. (2021), A Survey of the Usages of Deep Learning for Natural Language Processing, *IEEE Transaction on Neural Networks and Learning Systems*, vol. 32, no. 2, s.604–624 Tillgänglig: <https://ieeexplore.ieee.org/abstract/document/9075398> [Hämtad 27 april 2021]
- Phillips, M. (2021). Erik Selins uttalande om jämställdhet väcker skarp kritik: ”Cementerar en bild”, Dagens Industri, 8 maj, Tillgänglig: <https://www.di.se/nyheter/erik-selins-uttalande-om-jamstalldhet-vacker-skarp-kritik-cementerar-en-bild/> [Hämtad 12 maj 2021]
- Rai, P., & Singh, S. (2010). A Survey of Clustering Techniques, *International Journal of Computer Applications*, vol. 7, no.12, s.1-5 Tillgänglig: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.206.5219&rep=rep1&type=pdf> [Hämtad 26 april 2021].
- Silberg, J., & Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans), *McKinsey Global Institute*, Tillgänglig: <https://www.mckinsey.com/~media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/mgi-tackling-bias-in-ai-june-2019.ashx> [Hämtad 21 april 2021]
- Sjöberg, A., & Sjöberg, S. (2020). Tengai och Personlighet. En konvergerande och diskriminerande valideringsstudie av extraversion och samvetsgrannhet [pdf]

- Tillgänglig: <https://www.tengai-unbiased.com/validation-study-robot-interview/>
[Hämtad 24 april 2021]
- West, S.M., Whittaker, M., & Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI, AI Now Institute. [pdf] Tillgänglig:
<https://ainowinstitute.org/discriminatingystems.html> [Hämtad: 2 april 2021]
- Yin, R.K. (2013). Kvalitativ forskning från start till mål, översatt av J. Retzlaff, Lund: Studentlitteratur
- Zhao, W. (2018). How to improve corporate social responsibility in the era of artificial intelligence?, *IOP Conference Series: Earth and Environmental Science*, vol. 186, iss. 6, Tillgänglig: <https://iopscience.iop.org/article/10.1088/1755-1315/186/6/012036/pdf>
[Hämtad 27 april 2021]